Learning and Decision-Making in Competitive and Uncertain Systems

Tanner Fiez

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Lillian Ratliff, Chair

Sam Burden

Lalit Jain

Kevin Jamieson

Program Authorized to Offer Degree:
Electrical and Computer Engineering

University of Washington

Abstract

Learning and Decision-Making in Competitive and Uncertain Systems

Tanner Fiez

Chair of the Supervisory Committee:
Lillian Ratliff
Department of Electrical and Computer Engineering

As a result of the demonstrated potential for impact in traditional use cases, progressively more is being asked of machine learning methods. This evolution has lead to a renewed focus on learning and decision-making systems. In this domain, theoretical challenges relating to competition and uncertainty are emerging from the practical considerations that have motivated this paradigm shift.

There is an increasing awareness that learning and decision-making algorithms will eventually need to be or already are being embedded into complex systems where game-theoretic considerations naturally arise owing to the presence of competing, self-interested entities. Moreover, it has become clear that the artificial introduction of competition in game-theoretic abstractions of machine learning problems can often be a convenient and effective modeling technique for many problems of interest. Consequently, tools from game theory are now critically needed to analyze coupled learning and decision-making algorithms for the purposes of characterizing the outcomes that can be expected from competitive interactions and computing meaningful solutions such as equilibria in machine learning problems. Meanwhile, the demands of learning and decision-making algorithms operating under uncertainty are both changing and becoming more challenging. This transformation includes a movement towards more general, yet structured feedback models and objectives that reflect the desire to enable downstream tasks and future inferences. To this end, important problems remain to be solved pertaining to designing theoretically sound sequential decision-making algorithms tailored to such tasks.

This discussion motivates the research on learning and decision-making in competitive and

uncertain systems presented in this thesis. Together, the contents of this thesis can be summarized by a pair of themes that form Parts I and II: game-theoretic methods for analyzing decision-making algorithms and solving machine learning problems, and machine learning methods for designing and analyzing sequential decision-making algorithms under uncertainty.

The former theme is approached from a top-down perspective: general formulations of games and gradient-based learning algorithms are studied, theoretical characterizations are developed, and then the results are connected to specific problems of interest. In contrast, the latter theme is approached from a bottom-up perspective: models of practical sequential decision-making tasks are developed and then theoretically justified algorithms and solutions are constructed.

While learning and optimization in games is a well-studied topic, the majority of past research has focused on highly structured settings. Part I of this thesis moves away from this practice and presents studies of nonconvex games on continuous strategy spaces and gradient-based learning algorithms within them. The intent of this research is to develop appropriate notions of game-theoretic equilibria, characterize and understand the behaviors of so-called 'natural' learning dynamics, and establish methods for computing equilibria to solve machine learning problems formulated as games.

Chapter 2 lays the foundation for Part I and is built upon thereafter. Based upon the idea of viewing the underlying interaction structure as a Stackelberg game, both a local Stackelberg equilibrium concept and a corresponding characterization in terms of gradient-based sufficient conditions called a differential Stackelberg equilibrium are presented. Learning dynamics emulating the natural game structure are then constructed and convergence guarantees to differential Stackelberg equilibrium are proven. Chapter 3 follows along this path to study the role of timescale separation on the convergence of the canonical gradient descent-ascent learning dynamics in the subclass of nonconvex-nonconcave zero-sum games. The results characterize the timescales for which the dynamics both locally converge to differential Stackelberg equilibrium and locally avoid points lacking game-theoretic meaning. Finally, Chapter 4 considers zero-sum games in which the minimizing player faces a nonconvex objective and the maximizing player optimizes a Polyak-Łojasiewicz or strongly-concave objective. For this class of games, global convergence guarantees for gradient descent-ascent with timescale separation to only differential Stackelberg equilibrium are proven.

Throughout Part I, the implications of the theoretical results for both competitive decision-making and methods for solving machine learning problems are discussed.

Traditionally, the study of sequential decision-making under uncertainty in machine learning has focused on problems in which the evaluation criterion is directly linked to the immediate feedback. However, it has become clear that decision-making under uncertainty is often also pertinent to problems where the goal of the learner is instead to acquire information for the purpose of drawing inferences or fulfilling targets only partially linked to the immediate feedback. Part II of this thesis presents a pair of studies on well-motivated sequential decision-making problems with structured feedback models that fall under this theme. The intent of this research is to design sequential decision-making algorithms for solving practical problems that emerge in the real-world with desirable theoretical guarantees by exploiting structured feedback models.

Chapter 5 commences Part II by formulating the task of ranking papers to reviewers in peer review bidding systems as a sequential decision-making problem. A model of this problem is developed that identifies a pair of misaligned objectives: ensuring that each paper obtains a sufficient number of bids to be matched adequately with qualified reviewers, and respecting the preferences of reviewers by showing them relevant papers early in the list. To balance the competing objectives, a sequential decision-making algorithm is constructed that exploits the objective structure and it is shown both theoretically and empirically to have a number of advantages over baselines currently used in practice. Chapter 6 then concludes Part II with an analysis of pure exploration transductive linear bandits, a problem that arises naturally in experimental design settings. A decision-maker in this problem sequentially samples measurement vectors from a given set and observes a noisy linear response with an unknown parameter vector. The goal is to infer with high confidence the item from a separate set of vectors that has the maximum inner product with the unknown parameter vector while taking a minimal number of measurements. The optimal achievable sample complexity for this problem is characterized and a near-optimal algorithm that exploits the information structure of the feedback model to enhance the sample efficiency is developed.

Together, the contributions of this thesis take steps towards developing important theoretical foundations for learning and decision-making with competition and uncertainty.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I have often heard the saying "it takes a village to raise a PhD" as a play on the proverbial phrase. In my experience, this saying could not be more accurate. I have been fortunate to be supported by many people along my academic journey and it is with great gratitude that I now take the opportunity to acknowledge and thank my "village".

My path toward toward electrical engineering and computer science began in high school. I was fortunate to have an amazing math teacher, Paul Buchanan, with a sense of humor and teaching ability that could make math fun for anyone. To this day, I often find a sign that was hung on his wall incredibly useful for my research, it read "when math gets hard, make it easy."

My undergraduate experience at Oregon State University was instrumental in leading me to where I am today. It is where I learned the foundations of electrical engineering and computer science, found a passion for learning and working with others, and developed self-confidence in my ability as a student. This amazing environment was fostered by many wonderful professors that provided advice, support, and time along the way. The trio of Un-Ku Moon, Karti Mayaram, and Bella Bose made me feel at home in the Kelley Engineering Center. Each of them gave me significant guidance throughout my undergraduate degree and in navigating my future goals and ambitions. Huaping Liu and Raviv Raich taught core signal processing classes that helped guide me toward the direction of research I pursued for graduate school. I remember asking Raviv how he chose a graduate school and he told me that his choice was based on the school where he would get the maximum amount of mentorship. I now know how good of advice of this was and feel fortunate that the University of Washington has turned out to be that school for me. Lastly, without Matt Johnston, it is questionable if I would have pursued a PhD at all. I distinctly remember a conversation we had in the Summer of 2015 following the 2nd year of my undergraduate degree and shortly after Matt had generously accepted to mentor me in a research project. I was interested in graduate school, but hesitant about pursuing a PhD since I lacked research experience and had what I thought would need to be over a years worth of credits left to take. Matt laid out reasons for doing a PhD and encouraged me to think about applying to graduate school in the Fall and then pushing to finish my undergraduate degree early. By the next day, I knew that I would end up doing exactly that. In the final year of my undergraduate degree, Matt taught me how to do research, advised my senior design project, and also met with me often to discuss anything and everything. Perhaps the most influential thing Matt taught me was indirect; he was the consummate role model of how to work with and treat others. Matt went above and beyond with his time to help me and I am incredibly thankful for that.

When I was looking at graduate schools, I was drawn to the University of Washington by the northwest connection and strong electrical engineering and computer science programs. However, I was concerned that there was not an ideal research fit for me at the time. Then, in January of 2016, I received an email from Lillian Ratliff telling me that she would be starting as an assistant professor in Fall 2016 at the University of Washington and that she wanted to talk about the

opportunities for me. That day we had a video call and by the conclusion of it I was confident that I had stumbled upon an amazing opportunity. This has turned out to be an understatement. I simply cannot say enough about how great of an adviser Lily has been. Being the 1st PhD student for Lily has always been a point of both motivation and pride for me. I knew from the beginning of my PhD that progress for each of us would be tied together. Given the time and effort Lily devoted to developing me as a researcher along with the opportunities she brought to me, I always wanted to do everything in my power to hold up my end of the bargain. As she goes on to an amazing career, I will always have satisfaction knowing I played a part in kicking this off with her together. Her expertise in dynamical systems theory and game theory is amazing, and she has personally taught me almost everything I know about these topics and invigorated my passions for the subjects. I will always admire her work ethic, selflessness, and passion and I want to thank her for being such a great teammate and friend over the years. While we may not work as closely together in the future, I look forward to more times catching up and shooting the breeze over some drinks, they can be on me from now on!

I have been extremely lucky to have several 'unofficial' advisers along my PhD journey. Prior to the 2nd year of my PhD, Shreyas Sekar joined Lily's group as a postdoctoral researcher. It was a pleasure to work closely together for a year with such a kind and patient person. During that time, I learned a lot from Shreyas about writing papers and working through technical proofs as I moved towards more theoretical research. Kevin Jamieson has had a substantial role during my PhD. When Kevin joined the University of Washington I was just beginning to become interested in multi-armed bandits in the Summer of 2017. After realizing he was an expert on the topic, I spent at least a week that summer reading everyone one of his papers and my excitement built. We began collaborating as I entered my 3rd year, and since then Kevin has graciously included me in his group. I have learned so much about multi-armed bandits, designing algorithms, and developing intution for solving a problem from him. He is also an amazing teacher and a fun, upbeat person to be around, which I have always appreciated. During my 3rd year, Kevin introduced me to Lalit Jain, who was his postdoctoral researcher at the time. Lalit and I then spent practically every day for the next 3 months together working on pure exploration for linear bandits. This effort resulted in the contents of Chapter 6 together with Kevin and Lily. Lalit pushed me as hard as anyone did during my PhD during that time, and I am better off for it. Since then, Lalit has been a great mentor and friend, along with a source of expertise that I know I can always bounce ideas back and forth with. Kevin also connected me with Houssam Nassif at Amazon. My internship experiences at Amazon have been excellent, and it has been so interesting learning about the practical considerations of multi-armed bandits as I pursued theoretical aspects in my research. Houssam is undoubtedly one of the most considerate people I have ever met, and I am thrilled to be joining Amazon to continue to work with him. I began working with Nihar Shah during my 3rd year after he visited the University of Washington to give a seminar and connected with Lily. It was then that Nihar introduced me to his research program on developing principled methods for improving peer review. From the across the country, Nihar has been a mentor to me ever since and the work presented in Chapter 5 is a result of this collaboration. I am extremely thankful for his dedication and patience with me and the attention to detail and preciseness that he instilled in me as a researcher. While I was presenting an initial version of the work in Chapter 2

of this thesis at the smooth games workshop held at NeurIPS in 2019 during the 3rd year of my PhD, I met Georgios Piliouras after he asked me several questions at the conclusion of my talk. When Georgios, Lily, and I continued to chat after, it was clear that we spoke the same language of game theory and dynamical systems. This spur of the moment meeting has resulted in a fun and productive collaboration together along with Ryann Sim and Stratis Skoulakis for the past several years. Georgios taught me a totally different viewpoint on learning and games than I was familiar with and significantly expanded my horizons in this regard. His amazing creativity also has helped me in becoming a more outside the box thinker. I would also like to acknowledge Chi Jin and Praneeth Netrapalli for the insightful discussions and exciting collaboration together with Lily in this past year.

I want to thank Maryam Fazel and Sasha Aravkin for serving on my qualifying exam committee, Mehran Mesbahi for serving as my GSR for my general and final exams, and Baosen Zhang for research collaborations early in my PhD. I would also like to thank Sam Burden for serving on my exam committees and being a relentlessly positive presence in the department. I took courses with Sam on linear systems and control in my first year. At a time where I was adjusting to graduate school life, his attitude was an amazing asset that helped me get excited about topics that have become so fundamental in my research on game theory.

I am thankful to the many fellow students that have helped make my graduate school experience enjoyable. I sat between Yuanyuan Shi and Yize Chen for the duration of my time in the office. I enjoyed our many great conversations over the years and the company when we found ourselves in the office together on weekends. Each of them were invaluable resources for me, who I could always turn to and ask questions that they would happily discuss with me. I owe Chase Dowling gratitude for working with me in the 1st year of my PhD and being an overall fun person to have debates with to kill some time in the office. Ben Chasnov has been a great collaborator for projects on learning dynamics in games and also taught me a lot about coding and machine learning software frameworks. The work presented in Chapter 2 of this thesis would not be possible without him. Liyuan Zheng has been a great collaborator and is so sharp. We took many classes together and I always enjoyed our times spending hours talking over problem sets. I cannot count how many times I asked Dan Calderone a technical question and he would literally drop what he was doing and spend hours teaching me topics and working with me. His mathematical intuition and ability to condense anything into a picture is absolutely amazing and his random musings always kept me entertained. Mitas Ray and Jimin Kim have been great office mates over the years. Omid Sadeghi helped me numerous times with optimization related topics over the years. Eric Mazumdar has been a terrific collaborator and I learned a lot from his research. It has been great getting to know and working with Evan Faulker and Adhyyan Narang in this past year and I am excited to see their achievements during graduate school. I also want to thank everyone in Kevin Jamieson's and Jamie Morgenstern's group meetings, especially for the weekly zooms during COVID-19 that kept some semblance of normal communication with peers.

Finally, I want to thank my family and friends for the support during my PhD. I am lucky to have a group of friends that gave me important reprieve from graduate school life and research. I have convinced myself that the time I wasted each day debating sports and fantasy football over group messages has been a net-positive for my overall productivity. My roommate for 4 years and

fellow ECE PhD student, Yu-Chia Chen, has been a great friend and our chats at the end of most days were always a highlight. Having a friend in a similar position in life was immensely valuable. My parents and brother have always been there for me, and in the last 5 years the phone calls each week were so important to me. My dad always had the ability to give me perspective and was an unwavering calming presence that I often needed. My mom, and academic herself, has been the ultimate confidence builder for me. She had the ability in a 30 minute call to convince me I was the top student in the country at times that I doubted if I would ever be able to finish a PhD. My brother kept me relaxed by talking with me about anything but research and always making me laugh by making fun of me. It goes without saying that I am forever indebted to my wife Lydia. Lydia has been with me since the freshman year of my undergraduate degree and seen it all. She supported me when I decided that I was going to push to finish undergraduate a year before her and then likely leave to another state for 5 years. Then, through 4 years of living in separate states, she was the person that stuck by my side, talked daily with me, and brought me happiness. This milestone could not be possible without her patience and understanding of my late nights and weekends spent working on papers. I look forward to this transition in life and spending as much time as I can with you in the future Lydia, I owe you that and much more. Lydia's family also played an instrumental role in the last portion of my PhD. When COVID-19 hit, working each day in my bedroom and hardly having any interaction with people was a real struggle for me. Luckily, Lydia and I were able to move in with her family in Portland for several months. The daily home-cooked meals and time together reinvigorated me and gave me the motivation that I needed to finish my PhD. I am also thankful they didn't make too much fun of me the several times they found me early in the morning asleep on the couch of the living room with papers and a computer sprawled around after long nights preparing for the NeurIPS deadline. I have to end my acknowledgements by mentioning Willy and Luna(tic), the kittens Lydia and I adopted during COVID-19 that became my de facto desk mates and gave me so much entertainment and company during the last year and a half spent working at home alone.

DEDICATION

To Lydia, this milestone would not be possible without you

Chapter 1

Introduction

As a result of the demonstrated potential for impact in traditional use cases, progressively more is being asked of machine learning methods. This evolution has lead to a renewed focus on learning and decision-making systems. In this domain, theoretical challenges relating to competition and uncertainty are emerging from the practical considerations that have motivated this paradigm shift. There is an increasing awareness that learning and decision-making algorithms will eventually need to be or already are being embedded into complex systems where game-theoretic considerations naturally arise owing to the presence of competing, self-interested entities. Moreover, it has become clear that the artificial introduction of competition in game-theoretic abstractions of machine learning problems can often be a convenient and effective modeling technique for many problems of interest. Consequently, tools from game theory are now critically needed to analyze coupled learning and decision-making algorithms for the purposes of characterizing the outcomes that can be expected from competitive interactions and computing meaningful solutions such as equilibrium in machine learning problems. Meanwhile, the demands of learning and decision-making algorithms operating under uncertainty are both changing and becoming more challenging. This transformation includes a movement towards more general, yet structured feedback models and objectives that reflect the desire to enable downstream tasks and future inferences. To this end, important problems remain to be solved pertaining to designing theoretically sound sequential decision-making algorithms tailored to such tasks.

This discussion motivates the research on learning and decision-making in competitive and uncertain systems presented in this thesis. Together, the contents of this thesis can be summarized by a pair of themes that form Parts I and II: game-theoretic methods for analyzing decision-making algorithms and solving machine learning problems, and machine learning methods for designing and analyzing sequential decision-making algorithms under uncertainty. The former theme is approached from a top-down perspective: general formulations of games and gradient-based learning algorithms are studied, theoretical characterizations are developed, and then the results are connected to specific problems of interest. In contrast, the latter theme is approached from a bottom-up perspective: models of practical sequential decision-making tasks are developed and then theoretically justified algorithms and solutions are constructed.

While learning and optimization in games is a well-studied topic, the majority of past research has focused on highly structured settings. Part I of this thesis moves away from this practice and presents studies of nonconvex games on continuous strategy spaces and gradient-based learning algorithms within them. The intent of this research is to develop appropriate notions of game-theoretic equilibrium, characterize and understand the behaviors of so-called 'natural' learning dynamics, and establish methods for computing equilibrium to solve machine learning problems formulated as games. On the other hand, the traditional study of sequential decision-making under uncertainty in machine learning has focused on problems in which the evaluation criterion is

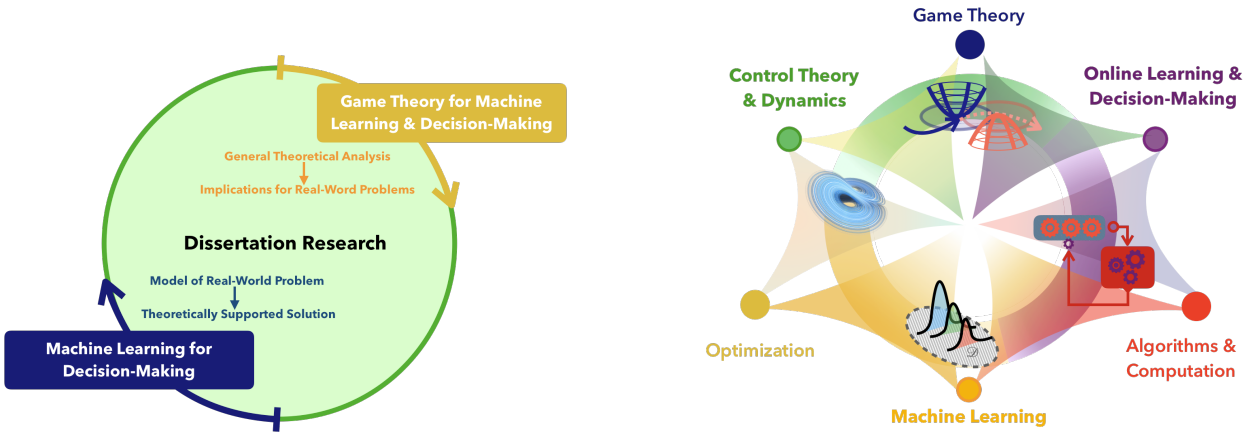


Figure 1.1: Graphical overview of the approach and methods that form the focus of this thesis. Game theory is used for analyzing decision-making algorithms and solving machine learning problems and machine learning is used for designing effective decision-making algorithms. This research spans several fields and the results rely on using a diverse set of tools.

directly linked to the immediate feedback. However, it has become clear that decision-making under uncertainty is often also pertinent to problems where the goal of the learner is instead to acquire information for the purpose of drawing inferences or fulfilling targets only partially linked to the immediate feedback. Part II of this thesis presents a pair of studies on well-motivated sequential decision-making problems with structured feedback models that fall under this theme. The intent of this research is to design sequential decision-making algorithms for solving practical problems that emerge in the real-world with desirable theoretical guarantees by exploiting structured feedback.

Together, the contributions of this thesis take steps towards developing important theoretical foundations for learning and decision-making with competition and uncertainty. This research is timely owing to the emergence of high-impact, real-world problems where theoretically guarantees are essential to the design of algorithms and for developing understanding of methods already commonly deployed in practice. A distinguishing feature of this research is the diversity of tools and methodologies that are needed to achieve the breadth and depth of results. Dynamical systems theory will provide a unified framework to analyze coupled learning algorithms, game theory will be used to assess and characterize strategic behaviors, machine learning and optimization techniques will be adopted to analyze and synthesize learning algorithms with provable guarantees, and experimental evaluations will connect theory and practice.

Organization. In terms of organization, in the remainder of this introduction, we provide further motivation for Part I of this thesis and then describe the contents of each chapter, and after which, we motivate further Part II of this thesis and describe the contents of each chapter. At the end of this introduction, we include bibliographic remarks regarding the contents of this thesis. We point out that Part I and Part II of this thesis are generally distinct, thus they can be read separately. Moreover, the chapters of Part II are also distinct so they can be read individually and in any order. However, the contents of Part I follows a fairly direct line of work and are best read in order. That being said, if they are read out of order, we remark that the first chapter of Part I contains a significant amount of background information that motivates the latter chapters, and also more

details on the overall analysis methods and approaches.

1.1 Part I: Learning and Optimization in Games

Learning algorithms are now often embedded in real-world systems and tasked with acquiring information to enable effective decision-making. However, learning agents are rarely acting in isolation within complex systems; instead they are typically in the presence of multiple autonomous agents who may be optimizing conflicting objectives and consequently acting as competitors in the environment. Simply put, game-theoretic constraints are a naturally occurring phenomena in the real-world. This fundamentally changes what outcomes can be expected of algorithms designed for static environments and the types of algorithms that can achieve desired objectives.

The emergence of competition introduces a number of distinct questions and challenges in the study of learning and decision-making. To begin, it requires rethinking what it even means to act optimally or solve a problem. Typically, doing so necessitates considering an equilibrium notion as a solution concept. However, even narrowing down a proper notion of equilibrium is a nuanced discussion without a unifying resolution. Then, given a solution concept, effective decision-making demands accounting for the inherently dynamic and non-stationary environment. This means it is imperative to not only consider the stationary objective being optimized, but also how interactions with competing agents and the environment affect the pursuit of that objective while learning.

In commonly arising game-theoretic settings, the goal is often to characterize the expected outcomes of interactions between rational agents implementing some 'natural' learning rule that reflects this characteristic. In fact, a typical game-theoretic point of view is that learning in games provides axiomatic backing for equilibrium in the sense that they arise and can be explained through the outcomes of iterative competitions for optimality (Fudenberg et al., 1998). This perspective motivates the analysis of learning algorithms which reflect any natural structure present in a problem and doing so often suggests refinements of equilibrium notions and methods for computing equilibrium.

While the aforementioned viewpoint is certainly important, it alone cannot harness all that game-theoretic abstractions can provide as a modeling framework. Indeed, artificial formulations of games can often be extremely useful. This has become readily apparent as a number of machine learning problems are now being successfully formulated as games. Prominent examples of this paradigm include generative adversarial network formulations (Goodfellow et al., 2014) and adversarial training objectives (Madry et al., 2018). Solving problems formulated in this manner is clearly dependent on the ability to efficiently compute a meaningful solution to a game. In this regard, respecting rational behavior is no longer important. Instead, there is room for the design of learning algorithms that can guarantee convergence to a suitable equilibrium notion.

In this part of the thesis, we present work which falls into the category of characterizing the expected outcomes of natural learning dynamics in games and that which is primarily focused on computing equilibrium in problems that allow for algorithm design. While there is no shortage of existing research that can be placed among the viewpoints described on learning in games, the vast majority of past work concerns highly restrictive classes of games.

We deviate from existing work by studying much more general classes of games in which minimal assumptions are made on the objectives being optimized by the players in the game. A significant motivation for the study of such relaxations is the increasing practical relevance with regard to applications in machine learning and the deployment of learning algorithms in competitive environ-

ments. Moreover, from a theoretical perspective, interesting questions arise since solution concepts, algorithms, and analysis methods suitable for restricted classes of games often do not easily transfer to or are unfit for more general classes of games. In other words, there is the opportunity to build from the ground up and this thesis presents several steps in this direction, while also outlining important questions that are actively being pursued within learning in games.

1.1.1 Contributions of Part I: Learning and Optimization in Games

The analysis of gradient-based learning in games has traditionally concerned restricted classes of games such as those with potential, bilinear, and convex cost structures. While each of the restricted types of games has merits and a place of relevance, many of the problem formulations appearing in machine learning do not fall into any of them. Instead, nonconvex games on continuous action spaces are the foundation of machine learning problems including generative adversarial networks, adversarial learning, robust supervised learning, and multi-agent learning, among others. The generality of this class of games poses challenges as equilibrium are not unique and global solution concepts are generally unattainable. That being said, this class of games forms the focus of the work in this thesis, in which we make a number of advances in terms of developing equilibrium characterizations and understanding the convergence of gradient-based learning algorithms. We now give an overview of the contents of each chapter in Part I of this thesis.

1.1.1.1 Chapter 2: Nonconvex Games: A Stackelberg Viewpoint

In Chapter 2, we begin our study of gradient-based learning dynamics in nonconvex games on continuous strategy spaces. The notion of solving a game belonging to this class is quite intricate. In past work, the standard global Nash equilibrium concept for continuous strategy spaces has been refined to a local definition and characterized in terms of gradient-based sufficient conditions (Ratliff et al., 2013; Ratliff et al., 2016). Strategies satisfying the sufficient conditions for a local Nash equilibrium are known as differential Nash equilibrium and finding them has frequently been taken as a goal of a learning algorithm in the emerging line of work studying this class of games. However, there has been empirical evaluations showing that various algorithms which do not guarantee convergence to only local Nash equilibrium reach acceptable solutions in machine learning problems (Berard et al., 2020). Moreover, there are classical equilibrium notions in game theory beyond the Nash equilibrium concept that have not yet been considered in this context. To reconcile these disconnects, we explore complimentary equilibrium notions for the class of nonconvex games along with gradient-based algorithms to compute them.

The Stackelberg equilibrium concept is typical in games with either an implicit or explicit order of play and it generalizes the notion of a minmax equilibrium from zero-sum to general-sum games. Yet, it has not been studied in the modern literature on nonconvex games. This chapter begins by presenting a refinement of the standard game-theoretic formulation of a Stackelberg equilibrium to a local definition along with a characterization in terms of gradient-based sufficient conditions. We term points in the strategy space which satisfy the sufficient conditions differential Stackelberg equilibrium. Given the equilibrium notions, a number of characteristics and connections are presented. In particular, it is shown that local Nash equilibrium form a subset of local Stackelberg equilibrium in zero-sum games, and also that in general-sum games one player always prefers a local Stackelberg equilibrium over any local Nash equilibrium. Moreover, the class of differential

Stackelberg equilibrium are proven to be both generic amongst local Stackelberg equilibrium and also structurally stable within the class of nonconvex-nonconcave zero-sum games.

Given the equilibrium foundations, this chapter then shifts to focus on the behaviors of gradient-based learning algorithms in relationship to the solution concepts. This includes a study of the prototypical simultaneous gradient dynamics (gradient descent-ascent in zero-sum games) and an analysis of a novel set of Stackelberg gradient dynamics that are formulated using the implicit function theorem to reflect the hierarchical decision-making structure of Stackelberg games. Based upon stability analysis using the local linearization of the dynamics, conditions and classes of games are presented under which the simultaneous gradient dynamics are locally stable around differential Stackelberg equilibrium in nonconvex-nonconcave zero-sum games. In contrast, it is shown that the formulated Stackelberg gradient dynamics are only locally stable around Stackelberg equilibrium in nonconvex-nonconcave zero-sum games. Building upon the desirable local stability characterization, extensive convergence analysis is presented for the Stackelberg gradient dynamics in both nonconvex-nonconcave zero-sum and nonconvex general-sum games with both deterministic and stochastic gradient information. In the class of nonconvex-nonconcave zero-sum games, it is notable that local convergence to only game-theoretically meaningful differential Stackelberg equilibrium is proven with a rate that depends on natural properties of the equilibrium.

An extensive empirical study is presented to complement the theoretical results. In particular, illustrative examples are given to highlight the differences between local Nash and Stackelberg equilibrium both in zero-sum and general-sum games. Moreover, the experiments show that the Stackelberg gradient dynamics result in stable learning trajectories compared to the simultaneous gradient dynamics. This behavior can be theoretically expected in zero-sum games by inspecting the local linearizations. Finally, experiments with generative adversarial networks highlight that the gradient-based learning dynamics are often converging to neighborhoods of differential Stackelberg equilibrium and reaching satisying solutions, which underscores the importance of this equilibrium concept in the optimization landscape of machine learning problems.

Together, this work has lead to expanded viewpoint of what it means to solve a nonconvex game and has been a subject of several follow-up works that build on the algorithm and equilibrium notion. In fact, the following chapters in this part of the thesis build closely on this work.

1.1.1.2 Chapter 3: Nonconvex Zero-Sum Games: Gradient Descent-Ascent with Timescale Separation

In Chapter 3, we continue our study of gradient-based learning dynamics in nonconvex continuous strategy space games, but the focus shifts to revisiting the behavior of the gradient descent-ascent dynamics in nonconvex-nonconcave zero-sum games. Understanding the local stability and convergence properties of this particular set of dynamics is important since it is an algorithm that is deployed ubiquitously in practice for machine learning problems owing to the simplicity and computational efficiency of the method. Chapter 2 began taking steps in this direction theoretically and empirically by providing some conditions and classes of games for which the gradient descent-ascent dynamics are locally stable around differential Stackelberg equilibrium and illustrating the dynamics often converge to differential Stackelberg equilibrium that are not differential Nash equilibrium in numerical experiments. However, these results fall short of a complete theoretical characterization of the types of critical points the dynamics locally converge around and the precise

connections with the set of differential Stackelberg equilibrium in nonconvex-nonconcave zero-sum games. Moreover, only the setting when players employ equal learning rates is considered.

Concurrent with the work from Chapter 2, Jin et al. (2020) provide a study of gradient descent-ascent in nonconvex-nonconcave zero-sum games and explore the local stability around critical points. The key result presented in relation to this thesis is that gradient descent-ascent with a ratio of learning rates (timescale separation) between the players approaching infinity is only locally stable around differential Stackelberg equilibrium across the spectrum of zero-sum games. This result gives an equivalent guarantee as is provided in Chapter 2 for the Stackelberg gradient dynamics (with any learning rate ratio) in nonconvex-nonconcave zero-sum games. From a practical perspective, a weakness of the aforementioned continuous-time stability result for gradient descent-ascent with timescale separation is that ensuring an equivalent local stability and convergence result in discrete-time requires the learning rate to tend toward zero. Moreover, the continuous-time local stability result fails to provide insights into the empirical success and observed equilibrium convergence of the gradient descent-ascent learning dynamics with a finite timescale separation.

This observation motivates the development of a more fine-grained analysis of the local stability of the gradient descent-ascent dynamics around critical points as a function of the ratio of learning rates. The research presented in Chapter 3 follows this viewpoint and provides comprehensive local stability, instability, and convergence characterizations. Given any critical point of the gradient descent-ascent dynamics satisfying benign non-degeneracy assumptions, we show there exists a range of finite learning rate ratios that can be explicitly characterized for which the point is locally stable if and only if it is a differential Stackelberg equilibrium. The stability result is complimented by an instability result: we show that the gradient descent-ascent learning dynamics avoid critical points which are not differential Stackelberg equilibrium for a range of finite learning ratios that can also be constructed. The aforementioned results are extended to obtain local convergence guarantees to differential Stackelberg equilibrium including in a formulation tailored to generative adversarial network training. Together, these results provide a near-complete characterization of gradient descent-ascent with finite timescale separation in nonconvex-nonconcave zero-sum games and give theoretical backing for the common usage of a finite timescale separation in practice along with the efficacy of the differential Stackelberg equilibrium concept in machine learning problems.

Chapter 3 also contains extensive experimental evaluations. The numerical simulations highlight the manner in which timescale separation in the gradient descent-ascent dynamics warps the vector field around critical points to influence local stability, the regions of attraction, and also the rate of convergence. The experiments with generative adversarial network training demonstrates that often a reasonable finite timescale separation results in the fastest convergence to a desirable solution and also that when introducing a gradient penalty for the discriminator there is an important interplay between the choice of timescale separation and regularization. It is worth noting that such observations are commensurate with that observed for simpler problems as well.

1.1.1.3 Chapter 4: Nonconvex Zero-Sum Games: Global Convergence Guarantees

This is the final chapter of Part I of this thesis and it concludes the study of gradient-based learning dynamics in nonconvex continuous strategy space games. Chapter 4 continues the study of gradient descent-ascent learning dynamics with timescale separation, but with a restricted focus on semi-structured classes of zero-sum games. Specifically, this chapter considers zero-sum games

where the minimizing player faces a nonconvex optimization problem and the maximizing player optimizes a Polyak-Łojasiewicz (PŁ) or strongly-concave (SC) objective. These classes of games have been subject to a significant amount of study in the last several years owing to the relevance in regards to certain machine learning formulations. However, the results on gradient-based learning dynamics in these classes of games have had a distinct style compared to those for the more general setting of nonconvex-nonconcave zero-sum games. Specifically, while a primary focus in the study of nonconvex-nonconcave zero-sum games (including that of Chapters 2 and 3 in this thesis) has concerned developing local convergence guarantees for gradient-based learning dynamics to only game-theoretically meaningful equilibrium, the existing work on nonconvex-PŁ/SC zero-sum games has instead concentrated on developing global convergence guarantees for gradient-based learning dynamics to notions of stationary points. Chapter 4 is focused on bridging the gap between these types of results by building off and extending the analysis relating critical points of the gradient descent-ascent dynamics with timescale separation to differential Stackelberg equilibrium in nonconvex-nonconcave zero-sum games from Chapter 3 and methods for proving global convergence guarantees to critical points in nonconvex-PŁ/SC zero-sum games.

Concretely, the objective of Chapter 4 is to develop global convergence guarantees in nonconvex-PŁ/SC zero-sum games to only differential Stackelberg equilibrium. In pursuit of this goal, the chapter begins with a refined local stability analysis for gradient descent-ascent with timescale separation in nonconvex-PŁ/SC zero-sum games. We prove the only critical points that can be locally stable with respect to the gradient descent-ascent continuous-time system for any choice of learning rate ratio correspond to differential Stackelberg equilibrium in nonconvex-PŁ/SC zero-sum games. For the class of nonconvex-PŁ games, we exploit timescale separation to construct a potential function that when combined with the stability characterization and an asymptotic saddle avoidance result gives a global asymptotic almost-sure convergence guarantee to a differential Stackelberg equilibrium. For the class of nonconvex-SC games, we show the surprising property that the function of the game can be made a potential with timescale separation. Combining this insight with the stability characterization allows us to generalize methods for efficiently escaping saddle points in nonconvex optimization to obtain a global finite-time convergence guarantee to only differential Stackelberg equilibrium. Together, these results to the best of our knowledge provide in terms of the class of games the most general existing global convergence guarantees to only game-theoretically meaningful equilibrium. Moreover, the results highlight that by introducing some structure into the class of zero-sum games, significantly stronger theoretical guarantees are obtainable. Finally, they imply that the deployment of gradient descent-ascent with timescale separation in machine learning problems formulated as zero-sum games in these classes is ensured to reach meaningful solutions.

1.2 Part II: Sequential Decision-Making under Uncertainty

Thus far we have focused on presenting an overview of the significance and challenges of learning and decision-making in the presence of competition, along with the types of questions this thesis seeks to answer in the domain. The remainder of this introduction focuses on giving a synopsis of the importance and demands of learning and decision-making in the presence of uncertainty, along with the topics addressed in Part II of this thesis.

Learning and decision-making is often complicated by auxiliary sources of uncertainty such

as randomness, limited feedback, and dependencies between past and future choices. Typically, multiple such sources of uncertainty are present simultaneously and handling them appropriately is paramount to acting near-optimally and achieving the desired objectives. Each of the preceding forms of uncertainty poses unique challenges to learning. To deal with randomness, it may be necessary to develop estimates of unknown parameters which are robust to deviations and enable high-confidence decision-making. Then, in problems with limited feedback, it is often important to leverage any structural information of a model to infer supporting information indirectly. Finally, handling dependencies between past and future decisions requires careful consideration of the trade-offs between optimizing immediate payoffs or information gain and the final objective. Clearly, accounting for each aspect requires nuanced algorithmic design that is closely tied to the achievable analytical performance guarantees as well as the empirical effectiveness.

To give a sense of the types of uncertain sequential learning problems considered in this thesis, it is helpful to contrast with common problems studied from another perspective. In classic formulations, the decision-maker has access to a set of actions and some knowledge of how feedback is generated. The learner is then tasked with choosing actions sequentially, each of which generates some form of noisy feedback. Both the objective and feedback model are fundamental to informing how actions should be taken. Commonly, the feedback model is such that the selection of an action results in some immediate reward coupled with it and the goal of the decision-maker is to maximize the cumulative sum of rewards over a time horizon. In other words, the performance of the decision-maker is evaluated entirely while learning.

In contrast, decision-making under uncertainty is often pertinent to problems where the goal of the learner is instead to acquire information for the purpose of drawing inferences or fulfilling targets that are only partially linked to the immediate feedback obtained. This undertaking demands a distinct point of view to design effective algorithms. Rather than focusing attention on selecting actions perceived to be immediately promising, the decision-maker must consider which actions provide the maximum amount of information or progress toward an indirect goal that cannot be immediately realized. Often these paradigms are in significant conflict, which elucidates why the algorithmic methods must be tailored to each type of problem.

The research in Part II of this thesis on sequential decision-making under uncertainty is focused on designing and analyzing algorithms that facilitate downstream tasks and enable future inferences. Compared to past research on sequential learning with uncertainty, much of the novelty of the work in this thesis stems from the consideration of optimizing for multiple objectives simultaneously that cannot be immediately realized, and the design of algorithms with improved theoretical guarantees under more general feedback models. Each project in this realm is highly relevant to concrete practical applications such as facilitating interactions in markets including peer review systems and drawing inferences in web-optimization. The algorithms proposed are unique in the manner that they crucially leverage the structure of the problems to inform how to make decisions. This fact turns out to be key to obtaining desirable theoretical guarantees in the problems.

1.2.1 Contributions of Part II: Sequential Decision-Making under Uncertainty

Despite the classical nature of decision-making under uncertainty, the focus of study has typically concerned problems where decisions yield immediate feedback in the form of rewards that the learner seeks to maximize the accumulation of over time. In contrast, problems where the learner

seeks to optimize indirect objectives or acquire information using immediate feedback are now becoming commonplace. The work in this direction seeks to show that algorithms tailored to the aforementioned targets can significantly outperform methods which do not account for the differing perspective. Yet, the formulations in which results have been obtained are relatively simple and cannot fully capture a number of problems where it is necessary to model and exploit structure.

To illustrate this point, we begin by remarking that several real-world decision problems involve optimizing competing objectives whereas the majority of past research only considers a unified objective. A fitting example is an intermediary acting in a two-sided market (online labor systems, electronic commerce platforms, etc.) or abstractions thereof seeking to facilitate interactions and satisfy the preferences of each distinct party. In such settings, sequential decision-making is complicated by the obligation to balance trade-offs and understand the coupling between past and future decisions on the final objective when immediate feedback is only partially related. Furthermore, in a number of applications there is often an underlying structure that relates the values of actions. For example, the set of items available in a recommender system (movies, songs, etc.) typically have features that connect them (genre, style, etc.). To inform future inferences, it is often important to determine with high confidence which of a set of available items has the highest value while minimizing the number of requests for feedback. The presence of limited feedback magnifies the importance of accounting for structure relating actions to minimize the sample complexity needed to make decisions. The research in Part II of this thesis seeks to design and analyze sequential decision-making algorithms in uncertain, structured environments that exploit models to optimize multiple competing objectives or be sample efficient.

1.2.1.1 Chapter 5: Sequential Decision-Making in Peer Review Bidding Systems

We begin our study of sequential decision-making under uncertainty in Chapter 5. As mentioned, facilitating interactions between distinct user groups, sellers and buyers, or items and users in markets often necessitates balancing competing objectives in order to satisfy each party. A relevant market where this problem arises is that of conference peer review systems and the paper bidding systems within them. In the typical process, reviewers arrive to a management system and are shown a list of papers and asked to look through and bid on papers they would like to review. Then, using the bids and potentially supplemental information, reviewers are matched to papers and asked to review them. From the perspective of the conference, it is important that each paper obtains a sufficient number of bids so that the papers can be matched to reviewers with adequate background. In contrast, reviewers prefer to only interact with papers that are of the most interest.

Chapter 5 begins by developing a mathematical model of the bidding process in conference peer review that captures the competing objectives and the behavior of reviewers while bidding. Importantly, the order of papers shown to a reviewer is crucial to influencing the papers that they end up bidding on as a result of primacy effects. Motivated by this fact, we design a sequential decision-making algorithm that selects how to order papers to each arriving reviewer in order to effectively trade-off between optimizing for the preferences of papers and reviewers. A key challenge is that the objective being optimized cannot be fully realized until the end of the bidding process, which couples past and future decisions. Theoretically, we show a local optimality guarantee and a global optimality guarantee under a community model of reviewer preferences for the algorithm, whereas standard baselines which solely focus the preferences of the papers or the reviewers are considerably

suboptimal. Experimentally, we demonstrate on real conference data that our algorithm leads to an improved distribution of bids and adequately satisfies reviewer preferences. This work studies a problem that has not been carefully inspected before and the methods show potential for impact in real-world conference peer review systems.

1.2.1.2 Chapter 6: Sequential Experimental Design for Transductive Linear Bandits

The research presented in Chapter 5 focuses on overcoming challenges that emerge from the presence of competing objectives which cannot be immediately realized and must be carefully balanced under uncertainty. We transition in Chapter 6 to an exploration problem with a structured model. Often, exploration problems with limited feedback are effectively studied under the umbrella of multi-armed bandits. In the basic multi-armed bandit formulation, the decision-maker sequentially pulls arms (selects actions) and obtains rewards sampled from unknown stationary and stochastic distributions tied to the selections. The fixed-confidence exploration problem in multi-armed bandits describes a setting where the goal is to determine with high probability which of the arms has the maximum mean reward in a minimum number of samples. This simple, yet profound learning problem is now well-understood; algorithms have been developed for which the provable sample complexity matches information-theoretic lower bounds. However, the standard formulation fails to fully capture problems with an underlying structure that connects the rewards of arms.

We introduce and study the pure exploration transductive linear bandit problem in Chapter 6. Given a set of measurement vectors $\mathcal{X} \subset \mathbb{R}^d$, a set of items $\mathcal{Z} \subset \mathbb{R}^d$, and an unknown parameter vector $\theta^* \in \mathbb{R}^d$, the decision-maker seeks to identify the item $z \in \mathcal{Z}$ which maximizes $\langle z, \theta^* \rangle$ while minimizing the number of measurements $x \in \mathcal{X}$ that are taken which provide noisy observations of the form $\langle x, \theta^* \rangle$. This problem generalizes both the linear and combinatorial bandit pure exploration formulations. Moreover it arises in a number of applications such as recommender systems and drug discovery where there is underlying structure in the rewards of actions and where the set of available measurement vectors may be limited. A distinguishing factor of the problem is that information regarding the values of actions can be gained even without direct measurements of them. We design an algorithm based on successive phases of experimental design that exploits this feature of the problem and prove the sample complexity is optimal up to logarithmic factors. Through experimental evaluations, we demonstrate the algorithms potential to minimize the budget needed to draw inferences in web optimization applications. This work presents the first near-optimal algorithm for pure exploration linear bandits and since then has been the subject of follow-up works and improvements to remove logarithmic factors, develop complimentary asymptotically optimal algorithms, and extend to more general feedback models.

1.3 Bibliographic Remarks

For reference, we now cover some bibliographic information regarding the contents of this thesis.

The contents of Chapter 2 are primarily from a recent conference publication (Fiez et al., 2020a). This work was also previously presented in a workshop (Fiez et al., 2020a) and the technical report version (Fiez et al., 2019b) has some results not included in the conference version for the sake of brevity. However, we remark that perhaps more so that the other chapters, the presentation of this work has been revised and extended. Moreover, we note that Chapter 2 has additional

background information on dynamical systems theory and convergence analysis methods that is useful in understanding the chapters that follow in Part I of the thesis.

> Tanner Fiez, Benjamin Chasnov, and Lillian Ratliff. Implicit learning dynamics in stackelberg games: equilibrium characterization, convergence analysis, and empirical study. In International Conference on Machine Learning, pages 3133–3144, 2020.

The contents of Chapter 3 have been appeared as a conference publication as given below. The work also appeared in a workshop (Fiez and Ratliff, 2021) and a somewhat extended version is available as a technical report (Fiez and Ratliff, 2020). The presentation in this thesis is a bit of a mix between these versions.

> Tanner Fiez and Lillian Ratliff. Local Convergence Analysis of Gradient Descent Ascent with Timescale Separation. In International Conference on Learning Representations, 2021.

The contents of Chapter 4 is the most recent work contained in this thesis (Fiez et al., 2021b). Specifically, the paper it corresponds to is currently under review and given below.

> Tanner Fiez, Lillian Ratliff, Eric Mazumdar, Evan Faulkner, Adhyyan Narang. Global Convergence to Local Minmax Equilibrium in Classes of Nonconvex Zero-Sum Games. Under Review, 2021.

The contents of Chapter 5 have appeared as a conference publication (Fiez et al., 2020b) as given below. We remark that this work initially appeared in a workshop (Fiez et al., 2019d) and the technical report version (Fiez et al., 2020c) is extended from the conference publication and closely mirrors the presentation in this thesis.

> Tanner Fiez, Nihar Shah, and Lillian Ratliff. A super* algorithm to optimize paper bidding in peer review. In Conference on Uncertainty in Artificial Intelligence, pages 580–589, 2020.

The contents of Chapter 6 have appeared as a conference publication (Fiez et al., 2019c).

> Tanner Fiez, Lalit Jain, Kevin G Jamieson, and Lillian Ratliff. Sequential experimental design for transductive linear bandits. In Advances in Neural Information Processing Systems, pages 10667–10677, 2019.

The final chapter of this thesis (Chapter 7) includes discussion of the work presented along with directions of future work. Moreover, it will connect to several recent works that are not included in this thesis.

Part I

Learning and Optimization in Games

Chapter 2

Nonconvex Games: A Stackelberg Game Viewpoint

This chapter begins the study of learning and optimization in games that is presented in Part I of this thesis. While game theory is a classical field, the vast majority of study historically has focused on highly structured classes of games and the behaviors of learning algorithms within them. In the past decade, and particularly the last five years, renewed attention has been given to the topic of learning in general classes of games with minimal assumptions on the cost functions that players are optimizing. This direction of research has primarily been motivated by the emergence of machine learning problems formulated as games and the desire to understand the behaviors of learning algorithms in competitive and complex environments. Specifically, nonconvex games on continuous strategy spaces and gradient-based learning algorithms within them have come into focus as a result of these motivations. In this class of games, important problems include developing appropriate notions of game-theoretic equilibria, characterizing the behaviors of rational learning dynamics, and establishing methods for computing equilibrium.

A dominant perspective of this class of games has emerged in the early works seeking to provide answers to these questions. That is, treating the underlying game between players as a simultaneous play game. Consequently, much of the existing research focus has been on refining the Nash equilibrium concept to a suitable local characterization for this class of games, analyzing the relationship between the stable critical points of the simultaneous gradient dynamics and local Nash equilibrium, and designing gradient-based learning dynamics for finding only local Nash equilibrium.

This chapter deviates from this perspective and presents a novel study of nonconvex games that arises from treating the underlying game between players as a Stackelberg (hierarchical play) game. As is detailed in the chapter, this form of interaction structure between players often implicitly emerges in machine learning problems and is explicitly present in a number of traditional game-theoretic settings. Together, this chapter contributes equilibrium characterizations, analysis of the local stability and convergence of gradient-based learning algorithms, and empirical insights.

Specifically, a local Stackelberg equilibrium concept is presented and a gradient-based characterization called a differential Stackelberg equilibrium is developed. The properties (genericity and structural stability) of this equilibrium notion are analyzed and connections are drawn with the local and differential Nash equilibrium concepts. Moreover, a game-theoretic interpretation of stable critical points of the simultaneous gradient dynamics previously thought to lack meaning is given through the lens of the differential Stackelberg equilibrium concept. Furthermore, novel gradient-based learning dynamics emulating the natural structure of a Stackelberg game are derived using the implicit function theorem. Convergence analysis is provided for both deterministic and stochastic updates for zero-sum and general-sum games. Notably, in zero-sum games, it is shown that the only critical points the dynamics locally converge to are differential Stackelberg equilibria. The empirical results highlight the characteristics of the Stackelberg gradient dynamics along with the role of differential Stackelberg equilibrium in the optimization landscape of nonconvex games.

2.1 Introduction

The field of game theory has a long, established history with traditional focuses including the development and characterization of equilibrium concepts along with the analysis of iterative learning algorithms in games. Classically, the foundational results of the field have concerned classes of games that admit structured cost functions. Indeed, the seminal minimax theorem of two-player zero-sum games that ensures an essentially unique solution was developed for bilinear cost structures and then generalized to convex-concave cost structures (Morgenstern and Von Neumann, 1953; Neumann, 1928; Sion, 1958). Similarly, fundamental equilibrium concepts and characterizations in multi-player general-sum games were initially developed for bilinear cost structures (Nash, 1951, 1950) and then generalized to convex cost structures (Rosen, 1965), and also cost structures admitting a joint potential function (Monderer and Shapley, 1996). In structured classes of games, there has been extensive study of the behaviors of rational or 'natural' learning algorithms as well as algorithms for computing equilibrium solutions. This line of work dates back to the study of zero-sum games with bilinear cost structures (Brown, 1949; Robinson, 1951), and has since been extended to convex-concave zero-sum games (Golshtein, 1974; Korpelevich, 1976; Nedić and Ozdaglar, 2009; Nemirovski, 2004; Nemirovski and Yudin, 1978; Uzawa, 1958), and more generally multiplayer convex games (Rosen, 1965).

The past decade has witnessed an emerging coupling between game theory and machine learning. This can be credited to the formulation of learning problems as interactions between competing objectives and strategic agents. Specifically, generative adversarial network (Goodfellow et al., 2014), robust supervised learning (Madry et al., 2018; Sinha et al., 2018), reinforcement and multi-agent reinforcement learning (Dai et al., 2018; Zhang et al., 2019), and hyperparameter optimization (Lorraine et al., 2020; MacKay et al., 2019) problems have all been cast as zero-sum or general-sum continuous strategy space games. A common theme in these problem formulations is the presence of nonconvexity in the cost functions. However, this is a property that is not as well-studied from a game-theoretic perspective given the traditional focus on structured classes of games. Thus, it has not been entirely clear the proper game-theoretic notions of meaningful solutions and if gradient-based learning algorithms converge to them. Consequently, significant attention as of late has given to both developing suitable equilibrium concepts and characterizations for classes of nonconvex games on continuous strategy spaces and analyzing the convergence of gradient-based learning algorithms within such games toward computing meaningful solutions.

In terms of equilibrium notions, early works in this direction have developed local refinements of the Nash equilibrium concept along with gradient-based sufficient conditions relevant to analyzing the convergence of gradient-based learning algorithms (Adolphs et al., 2019; Daskalakis and Panageas, 2018; Jin et al., 2020; Mazumdar and Ratliff, 2019; Mazumdar et al., 2019; Mescheder et al., 2017, 2018; Nagarajan and Kolter, 2017; Ratliff et al., 2013; Ratliff et al., 2014, 2016). From a game-theoretic perspective, the Nash equilibrium concept is the typical solution notion when viewing the underlying interaction structure between players as a simultaneous move game. This is reflected in the definition which, informally, states that no player can benefit from unilaterally deviating (locally) when holding the rest of the players fixed at the joint Nash equilibrium strategy. Taking the simultaneous play perspective, natural gradient-based learning dynamics reflecting the structure of the game correspond to each player descending the gradient of their cost function with respect to their own choice variable. In general-sum games, this set of gradient dynamics is often

referred to as the simultaneous gradient dynamics, while in zero-sum games, this set of gradient dynamics is often referred to as gradient descent-ascent. Several recent works have studied the simultaneous gradient dynamics in both nonconvex general-sum games, and the important subclass of nonconvex-nonconcave zero-sum games. The studies and results have highlighted the inherent challenges of gradient-based learning in games. In particular, as a result of non-symmetric nature of the Jacobian of the dynamics, cyclic behavior and rotational forces can plague convergence. Moreover, it has been shown that the simultaneous gradient-dynamics can converge to stationary points that do not correspond to local Nash equilibrium, even in nonconvex-nonconcave zero-sum games (Daskalakis and Panageas, 2018; Mazumdar et al., 2020).

The perceived shortcomings of the simultaneous gradient dynamics have motivated the development and analysis of more sophisticated gradient-based learning dynamics, primarily focused on nonconvex-nonconcave zero-sum games. A number of techniques have been proposed including optimistic and extra-gradient algorithms (Daskalakis and Panageas, 2018; Daskalakis et al., 2018; Mertikopoulos et al., 2019), gradient adjustments (Balduzzi et al., 2018; Mescheder et al., 2017), and opponent modeling methods (Foerster et al., 2018; Letcher et al., 2019; Schäfer and Anandkumar, 2019; Zhang and Lesser, 2010), among others. While each of the the aforementioned gradient-based learning algorithms have merits compared to the simultaneous gradient dynamics, and in some cases useful game-theoretic interpretations, none of them can guarantee local convergence to only game-theoretically meaningful points of the optimization landscape. That being said, a pair of gradient-based learning algorithms exploiting higher order gradient-information that only locally converge to local Nash equilibrium have now been developed (Adolphs et al., 2019; Mazumdar et al., 2019).

Given this background, it is worth pointing out that the existing work on nonconvex games has generally focused on the simultaneous play perspective of the game and the corresponding Nash equilibrium solution concept. However, in game theory, a common interaction structure that is studied beyond simultaneous play is a hierarchical order of play. This type of game is known as a Stackelberg game (Von Stackelberg, 1934, 2010). In the simplest formulation of a Stackelberg game, one player acts as the leader who is endowed with the power to select an action knowing the other player (follower) plays a best-response. Consequently, the leader uses this knowledge to its advantage when selecting a strategy. The hierarchical decision-making structure of a Stackelberg game emerges naturally in economic competitions, control problems, and mechanism design. Moreover, given the non-symmetric nature of the formulation, it also a plausible model for problems with an implicit order of play such as nonconvex-nonconcave zero-sum games. The Stackelberg equilibrium (Basar and Olsder, 1998; Von Stackelberg, 1934, 2010) solution concept generalizes the minmax solution to general-sum games. Informally, in a Stackelberg equilibrium, the leader cannot benefit from deviating when considering how the best-response of the follower will change, and the follower cannot benefit from unilaterally deviating given the strategy of the leader.

The hierachical play (Stackelberg) viewpoint has long been researched in structured settings from a control perspective on games (Basar and Olsder, 1980, 1998; Basar and Selbuz, 1979; Jungers et al., 2011; Leitmann, 1978; Papavassilopoulos and Cruz, 1979, 1980; Simaan and Cruz, 1973) and in bilevel optimization (Breton et al., 1988; Danskin, 1966, 1967; Dempe et al., 2007; Dempe, 2002, 2005, 2018; Dempe and Gadhi, 2010; Sinha et al., 2017; Wiesemann et al., 2013; Zaslavski, 2012). However, this viewpoint has not been thoroughly explored in nonconvex, continuous strategy space games and from the machine learning perspective with respect to gradient-based learning algo-

rithms. This chapter adopts this perspective and explores the role of local Stackelberg equilibrium in two-player nonconvex games along with the connections between this solution concept and the limit points of gradient-based learning algorithms. We now provide an overview of the contributions provided in this chapter.

2.1.1 Contributions

Motivated by the lack of understanding of the role of Stackelberg equilibrium in nonconvex games as well as the absence of gradient-based algorithms emulating a hierarchical interaction structure, this chapter provides a study of nonconvex games from a Stackelberg perspective including equilibrium characterizations, novel gradient-based learning dynamics and convergence analysis, and an illustrative empirical study. The primary benefits of this work to the community include an enlightened perspective on the consideration of equilibrium concepts reflecting the underlying optimization problems present in machine learning applications formulated as games and classical game-theoretic applications, and a set of natural learning dynamics mirroring the Stackelberg interaction structure with desirable guarantees for computing Stackelberg equilibrium.

Equilibrium Characterizations. This chapter presents a local refinement of the Stackelberg equilibrium concept suitable for nonconvex games (Definition 2.2) and a characterization via sufficient conditions on the objectives of the players that is termed a differential Stackelberg equilibrium (Definition 2.4 and Proposition 2.1). The differential Stackelberg equilibrium notion is shown to be both generic amongst local Stackelberg equilibrium (Theorem 2.1) and structurally stable (Theorem 2.2) in zero-sum games. This means except on a set of measure zero in the class of zero-sum continuous games, differential Stackelberg equilibrium and local Stackelberg equilibrium are equivalent, and lastly that differential Stackelberg equilibrium persist under smooth perturbations to the cost functions. These results mirror properties that have been shown for the corresponding local and differential Nash equilibrium concepts in zero-sum games (Mazumdar and Ratliff, 2019; Ratliff et al., 2014, 2016). Moreover, connections are drawn between the local Nash and Stackelberg equilibrium notions along with the corresponding gradient-based characterizations. Specifically, in zero-sum games, we show the sets of local and differential Nash equilibrium form subsets of the sets of local and differential Stackelberg equilibrium, respectively (Proposition 2.2). This indicates that in nonconvex-nonconcave zero-sum games, the local Stackelberg equilibrium concept is not as restrictive as the local Nash equilibrium concept, and similarly for the gradient-based characterizations. For general-sum games, we also point out a variation of a well-know fact, that the local Stackelberg equilibrium cost for the leader is lower than any local Nash equilibrium cost in the local neighborhood under certain assumptions on the best-response set of the follower (Proposition 2.3).

Local Stability of Gradient-Based Learning Dynamics in Zero-Sum Games. This chapter analyzes the well-known simultaneous gradient dynamics and a novel set of gradient-based learning dynamics emulating the natural structure of a Stackelberg game that are derived based on the implicit function theorem. For nonconvex-nonconcave zero-sum games, the local stability of each set of dynamics is analyzed around critical points using the continuous-time limiting system toward drawing connections between the limiting behaviors and the set of differential Stackelberg equilibrium. For the simultaneous gradient dynamics, we reveal that there exists locally stable critical points that are differential Stackelberg equilibrium and not differential Nash equilibrium. This insight gives meaning to a broad class of critical points previously thought to lack game-theoretic

meaning, and may give some explanation for the adequacy of solutions that are not local Nash equilibrium in machine learning problems. To characterize this phenomenon, we provide necessary and sufficient conditions for when such points exist, and also explore special cases such as scalar action games and generative adversarial networks under the realizable assumptions (Mescheder et al., 2018; Nagarajan and Kolter, 2017) (Propositions 2.4–2.8). In contrast to the simultaneous gradient dynamics, for the Stackelberg gradient dynamics developed in this chapter, we show the desirable property that the only locally stable critical points are differential Stackelberg equilibrium and also that such equilibrium must be locally stable critical points (Theorem 2.5).

Convergence Analysis of Stackelberg Gradient Dynamics. This chapter provides extensive convergence analysis for both deterministic and stochastic versions of the Stackelberg gradient dynamics in zero-sum and general-sum games. For zero-sum games with deterministic gradient information, the continuous-time local stability result is built on to show that the discrete-time system only converges to critical points that are differential Stackelberg equilibrium. This result relies on developing a discrete-time local stability result that mirrors the continuous-time characterization (Proposition 2.9) and an asymptotic saddle avoidance result (Theorem 2.6). For both zero-sum and general-sum games with deterministic gradient information, we provide local convergence rates to differential Stackelberg equilibrium (Theorems 2.7–2.8 and Corollaries 2.1–2.2), where the rate for zero-sum games decomposes into a natural quantity related to the sufficient conditions for a differential Stackelberg equilibrium. For the stochastic setting, we provide analogous asymptotic convergence guarantees for several variations of the dynamics, including when players act on a single timescale (Theorem 2.9), when the follower best-responds at each step (Proposition 2.10), and when the players use a two-timescale learning rate sequence (Proposition 2.11).

Empirical Study. This chapter concludes with an illustrative set of experiments. Empirically, we show that the Stackelberg gradient dynamics result in stable learning compared to the simultaneous gradient dynamics. A key reason for this in zero-sum games is that the Jacobian of the Stackelberg gradient dynamics does not admit complex eigenvalues as does the Jacobian for the simultaneous gradient dynamics. Moreover, several experiments highlight the differences between Nash and Stackelberg equilibrium in traditional general-sum games such as economic competitions and bimatrix games. Finally, extensive results are presented for generative adversarial network problems. To gain insights into the each equilibrium concept in the optimization landscape, we analyze the eigenvalues of relevant game objects. Notably, we observe convergence to neighborhoods of differential Stackelberg equilibrium for each set of learning dynamics.

2.1.2 Organization

The organization of this chapter is as follows. In Section 2.2, we present formulations of continuous action games (Section 2.2.1), gradient-based learning dynamics (Section 2.2.2), local Nash and Stackelberg equilibrium concepts and characterizations (Section 2.2.3), connections and relationships between the equilibrium notions (Section 2.2.4), and genericity and structural stability results for the gradient-based characterization of a local Stackelberg equilibrium (Section 2.2.5). Then, in Section 2.3, the focus shifts to study the local stability of gradient-based learning algorithms; we provide background on the analysis methods (Section 2.3.1), a study of the simultaneous gradient dynamics (Section 2.3.2), and a study of the Stackelberg gradient dynamics (Section 2.3.3). Following up on the stability analysis, Section 2.4 and Section 2.5 give convergence results for the

Stackelberg gradient dynamics in deterministic and stochastic settings, respectively. Experimental results are presented in Section 2.6. A discussion of the results in the chapter, comments on some related, concurrent, and follow-up works, and open questions is contained in Section 2.7. Following the discussion section, there is an appendix to the chapter that contains several proofs that are deferred for the sake of presentation clarity and readability.

2.2 Games, Gradient Dynamics, and Equilibria

This section formalizes the class of games studied in this chapter, formulates gradient-based learning dynamics, and presents equilibrium concepts, characterizations, and relationships. Before moving on, we set some of our notation.

Notation. We denote by $D_i f_i(x_1, x_2) \in \mathbb{R}^{d_i \times 1}$ the derivative of $f_i(x_1, x_2)$ with respect to x_i, $D_{ij} f_i(x_1, x_2) \in \mathbb{R}^{d_i \times d_j}$ as the partial derivative of $D_i f_i(x_1, x_2)$ with respect to x_j, and $D_i^2 f_i(x_1, x_2) \in \mathbb{R}^{d_i \times d_i}$ as the partial derivative of $D_i f_i(x_1, x_2)$ with respect to x_i. Moreover, given a symmetric matrix A, we indicate that it is positive definite using the notation $A \succ 0$. Finally, we denote by $\mathbb{C}_-^\circ$ and $\mathbb{C}_+^\circ$ the open left and right half complex planes (that is, the set of complex numbers for which the real parts are negative and positive respectively), $\rho(\cdot)$ as an operator returning the spectral radius (maximum modulus of the eigenvalues) of a matrix argument, $\text{spec}(\cdot)$ as an operator returning the set of eigenvalues of a matrix argument, $\det(\cdot)$ as an operator returning the determinant of a matrix argument, and $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ as operators returning the minimum and maximum eigenvalues of a matrix argument, respectively.

2.2.1 Continuous Action Game Formulations

Consider a non-cooperative game between two agents where player 1 optimizes the cost function $f_1 : \mathcal{X} \to \mathbb{R}$ and player 2 optimizes the cost function $f_2 : \mathcal{X} \to \mathbb{R}$, where $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 = \mathbb{R}^d$ with $\mathcal{X}_1 = \mathbb{R}^{d_1}$ and $\mathcal{X}_2 = \mathbb{R}^{d_2}$ denoting the action spaces of player 1 and player 2, respectively.[1] We often use the compact notation $x = (x_1, x_2) \in \mathcal{X}$ to represent the joint strategy of the players. Let us denote by $\mathcal{I} = \{1, 2\}$ the set of players in the game and following standard game-theoretic notation $-i$ denotes the set of players excluding player i. We assume throughout that each f_i is sufficiently smooth: $f_i \in C^q(\mathcal{X}, \mathbb{R})$ for some $q \geq 2$. For zero-sum games, the game is defined by costs $(f_1, f_2) \equiv (f, -f)$. Observe that no assumptions have been made on the structure of the cost functions beyond smoothness and consequently the cost function for each player may be nonconvex. In simple terms, we consider the class of two-player smooth nonconvex games on continuous, unconstrained actions spaces.

Simultaneous Play Games. The existing viewpoint that has dominated the study of this class of games has been to treat the underlying interaction structure as a *simultaneous play game*. Given this perspective, each player $i \in \mathcal{I}$ in the game is faced with the optimization problem

$$\min_{x_i \in \mathcal{X}_i} f_i(x_i, x_{-i}). \tag{2.1}$$

[1] The convergence results in this chapter hold more generally for action spaces that are precompact subsets of the Euclidean space since they are local.

Observe that under this viewpoint, each player treats the strategy of the other player in the game as fixed in the optimization problem that is to be solved.

Stackelberg Games. In this chapter, we adopt the viewpoint of treating the underlying interaction structure as a *Stackelberg game*. Let us deem player 1 the 'leader' and player 2 the 'follower'. The designation of 'leader' and 'follower' indicates the order of play between the agents, meaning the leader plays first and the follower second. Thus, a Stackelberg game is a type of hierarchical or sequential play game. In a Stackelberg game, the leader and follower aim to solve the following optimization problems, respectively:

$$\begin{aligned} \text{Leader:} \quad & \min_{x_1 \in \mathcal{X}_1} \{ f_1(x_1, x_2) \big| \; x_2 \in \arg\min_{y \in \mathcal{X}_2} f_2(x_1, y) \}, \\ \text{Follower:} \quad & \min_{x_2 \in \mathcal{X}_2} f_2(x_1, x_2). \end{aligned} \tag{2.2}$$

Observe that under this viewpoint, the leader is optimizing its cost anticipating that the follower selects a best-response and the follower optimizes its cost given the strategy of the leader. Thus, unlike in the simultaneous play perspective in which the players do not account for how the other player will act, the leader models the behavior of the follower and selects a strategy by accounting for this model. Consequently, the action of the follower is a function of the leader's action. Moreover, in this formulation, the leader has a distinct advantage in that it can enforce a strategy on the follower.

The hierarchical perspective of the interaction structure is well-motivated both from machine learning and traditional areas where game-theoretic analysis is important. In machine learning problems formulated as nonconvex-nonconcave zero-sum games, an implicit order of play naturally emerges since in general $\min_{x_1 \in \mathcal{X}_1} \max_{x_2 \in \mathcal{X}_2} f(x_1, x_2) \neq \max_{x_2 \in \mathcal{X}_2} \min_{x_1 \in \mathcal{X}_1} f(x_1, x_2)$ for this class of games. Moreover, explicit hierarchical decision-making structures often arise in market competitions in economics (Anderson and Engers, 1992), mechanism design and control problems (Ratliff and Fiez, 2020; Ratliff et al., 2019), and human-robot interaction problems (Fisac et al., 2019; Nikolaidis et al., 2017).

Depending on the viewpoint of the interaction structure between the players, gradient-based learning dynamics can be developed that reflect each players role in the game as we show in the following subsection.

2.2.2 Gradient-Based Learning Dynamics

The gradient-based learning algorithms we formulate are such that the players follow myopic update rules that take steps in the direction of steepest descent for the optimization problem that they are faced with. Accordingly, the gradient-based learning dynamics we develop can be seen as types of 'natural' learning dynamics for each perspective of the interaction structure (simultaneous or hierarchical (Stackelberg) play game).

Simultaneous Gradient Dynamics. From the simultaneous play game viewpoint, natural myopic gradient-based learning dynamics considering the optimization problem each player is faced with (recall Equation 2.1) come in the form of each player descending along their individual gradi-

ent. Thus, the vector field is given by the concatenation of each players individual gradient:

$$g(x) = (D_1 f_1(x_1, x_2), D_2 f_2(x_1, x_2)).$$

The vector field $g(x)$ forms the basis of the well-studied *simultaneous gradient dynamics*, also commonly referred to as gradient descent-ascent in the context of zero-sum games. The continuous-time simultaneous gradient dynamics are described by the system

$$\dot{x} = -g(x).$$

Moreover, the Jacobian of the continuous-time simultaneous gradient dynamics is given by

$$J(x) = \begin{bmatrix} D_1^2 f_1(x) & D_{12} f_1(x) \\ D_{21} f_2(x) & D_2^2 f_2(x) \end{bmatrix}. \tag{2.3}$$

Stackelberg Gradient Dynamics. From the Stackelberg game viewpoint, natural myopic gradient-based learning dynamics considering the optimization problem each player is faced with (recall Equation 2.2) come in the form of the leader descending along the total derivative of its cost, since the action of the follower is seen as a function of the action of the leader, and the follower descending along its individual gradient. Thus, the vector field reflecting this structure is given by

$$g_{\mathcal{S}}(x) = (Df_1(x_1, x_2), D_2 f_2(x_1, x_2)).$$

The notation $Df_1(x_1, x_2)$ is denoting the total derivative of f_1 with respect to x_1 given that x_2 is implicitly a function of x_1, capturing the fact that the leader operates under the assumption that the follower will play a (local) best response to x_1. Specifically, given a joint strategy (x_1, x_2) at which $D_2 f_2(x_1, x_2) = 0$ and $\det(D_2^2 f_2(x_1, x_2)) \neq 0$[2], the implicit function theorem (Abraham et al., 1988, Thm. 2.5.7) implies that there exists an implicit map $r : x_1 \mapsto x_2$ defined on a neighborhood $U_1 \subset \mathcal{X}_1$ such that $x_2 = r(x_1)$ and $Dr(x_1) = -(D_2^2 f_2(x_1, r(x_1)))^{-1} D_{21} f_2(x_1, r(x_1))$. Thus, under the stated assumptions, by the chain rule and the implicit function theorem,

$$\begin{aligned} Df_1(x_1, r(x_1)) &= D_1 f_1(x_1, r(x_1)) + Dr(x_1)^\top D_2 f_1(x_1, r(x_1)) \\ &= D_1 f_1(x_1, r(x_1)) - D_{21} f_2(x_1, r(x_1))^\top (D_2^2 f_2(x_1, r(x_1)))^{-1} D_2 f_1(x_1, r(x_1)). \end{aligned}$$

As a surrogate for this gradient using only local information, we define the following:

$$\begin{aligned} Df_1(x_1, x_2) &:= Df_1(x_1, r(x_1))|_{r(x_1)=x_2} \\ &= D_1 f_1(x_1, x_2) - D_{21} f_2(x_1, x_2)^\top (D_2^2 f_2(x_1, x_2))^{-1} D_2 f_1(x_1, x_2). \end{aligned} \tag{2.4}$$

The vector field $g_{\mathcal{S}}(x)$ forms the basis of the *Stackelberg gradient dynamics* studied in this chapter. The continuous-time Stackelberg gradient dynamics are described by the system

$$\dot{x} = -g_{\mathcal{S}}(x).$$

---

[2]This is a generic condition.

Moreover, the Jacobian of the continuous-time Stackelberg gradient dynamics is given by

$$J_{\mathcal{S}}(x) = \begin{bmatrix} D_1(Df_1(x)) & D_2(Df_1(x)) \\ D_{21}f_2(x) & D_2^2 f_2(x) \end{bmatrix}. \tag{2.5}$$

Unlike the canonical simultaneous gradient dynamics, the Stackelberg gradient dynamics that have been formulated are novel in terms of the construction.

The continuous-time gradient systems and the corresponding Jacobians have been defined since they play a key role in determining the local behavior of the dynamics around strategies of interest as becomes clear in Section 2.3. Before moving on to analyze the stability and convergence properties of the gradient-based dynamics that have been defined, we transition to present and define equilibrium concepts suitable for the class of nonconvex games under consideration along with gradient-based characterizations of them.

2.2.3 Local Equilibrium Concepts and Gradient-Based Characterizations

The typical equilibrium notions studied in continuous games are the pure strategy Nash equilibrium in simultaneous play games and the pure strategy Stackelberg equilibrium in hierarchical play games. We begin this subsection by introducing local refinements of the usual global notions of the equilibrium concepts in terms of the costs. Following that, we present gradient-based characterizations of each equilibrium concept relevant to analyzing gradient-based learning dynamics.

Local Equilibrium Concepts. The following Nash and Stackelberg equilibrium definitions are direct local refinements of standard global equilibrium notions (Basar and Olsder, 1998, Chapter 4). Specifically, the local definitions restrict the search space of deviations to local neighborhoods. While the former concept (local Nash) has been studied in the modern study of learning in games, the latter concept (local Stackelberg) has not been and bringing this equilibrium notion into focus is a conceptual contribution of this chapter.

A *local Nash equilibrium* is defined by a joint strategy at which no player can benefit from unilaterally deviating with a local neighborhood. The following definition formalizes this statement in terms of the costs of the players.

Definition 2.1 (Local Nash Equilibrium)**.** *Consider a game (f_1, f_2) defined by $f_i \in C^q(\mathcal{X}, \mathbb{R})$ for $i \in \mathcal{I}$ with $q \geq 2$. The joint strategy $x^* = (x_1^*, x_2^*) \in \mathcal{X}$ is a local Nash equilibrium on $U_1 \times U_2 \subset \mathcal{X}_1 \times \mathcal{X}_2$ if for each $i \in \mathcal{I}$, $f_i(x^*) \leq f_i(x_i, x_{-i}^*)$, $\forall\ x_i \in U_i \subset \mathcal{X}_i$.*

A *local Stackelberg equilibrium* is a joint strategy at which the leader cannot benefit from deviating within a local neighborhood when considering the reaction curve of the follower and the follower cannot benefit from unilaterally deviating within a local neighborhood given the strategy of the leader. The following definition formalizes this statement in terms of the costs of the players.

Definition 2.2 (Local Stackelberg Equilibrium)**.** *Consider a game (f_1, f_2) defined by $f_i \in C^q(\mathcal{X}, \mathbb{R})$ for $i \in \mathcal{I}$ with $q \geq 2$ and player 1 (without loss of generality) taken to be the leader. Consider $U_i \subset \mathcal{X}_i$ for each $i \in \mathcal{I}$. The strategy $x_1^* \in U_1$ is a local Stackelberg solution for the leader if, $\forall\ x_1 \in U_1$,*

$$\sup_{x_2 \in R_{U_2}(x_1^*)} f_1(x_1^*, x_2) \leq \sup_{x_2 \in R_{U_2}(x_1)} f_1(x_1, x_2), \tag{2.6}$$

where $R_{U_2}(x_1) := \{y \in U_2 | f_2(x_1, y) \leq f_2(x_1, x_2), \forall x_2 \in U_2\}$ *for any* $x_1 \in U_1$. *Moreover,* (x_1^*, x_2^*) *for any* $x_2^* \in R_{U_2}(x_1^*)$ *is a local Stackelberg equilibrium on* $U_1 \times U_2$.

Observe that if $R_{U_2}(x_1)$ is a singleton for all $x_1 \in U_1$, then the inequality in Equation 2.6 of Definition 2.2 is replaced by the simplified condition that $f_1(x_1^*, R_{U_2}(x_1^*)) \leq f_1(x_1, R_{U_2}(x_1))$ for all $x_1 \in U_1$. Moreover, note that the definition of a local Stackelberg equilibrium relies on the designation of leader and follower among the players. Without loss of generality, we always consider player 1 to be the leader unless otherwise noted.

While characterizing existence of equilibria is outside the scope of this work, we remark that Nash equilibria exist for convex costs on compact and convex strategy spaces and Stackelberg equilibria exist on compact strategy spaces (Basar and Olsder, 1998, Thm. 4.3, Thm. 4.8, & §4.9). This means the class of games on which Stackelberg equilibria exist is broader than on which Nash equilibria exist. Existence of local equilibria is guaranteed if the neighborhoods and cost functions restricted to those neighborhoods satisfy the assumptions of the cited results. The strategy spaces under consideration are not compact, so our focus is on interior equilibria when they exist.

Gradient-Based Equilibrium Characterizations. Often it is useful to characterize equilibria in terms of gradient-based sufficient conditions on the costs of the players. Specifically, conditions of this type are especially handy for assessing the connections between the limit points of gradient-based learning dynamics and notions of equilibrium.

We begin by stating gradient-based sufficient conditions for a local Nash equilibrium as given in Definition 2.1. Joint strategies satisfying sufficient conditions for a local Nash equilibrium have been termed *differential Nash equilibrium.* Since the local Nash equilibrium definition is described by a joint strategy at which no player can benefit from unilaterally deviating, it naturally follows that the sufficient conditions are characterized by zero individual derivatives and positive definite second-order individual derivatives to indicate each player is at a local minimum with respect to their own choice variable.

Definition 2.3 (Differential Nash Equilibrium, Ratliff et al. 2013; Ratliff et al. 2016)**.** *Consider a game* (f_1, f_2) *defined by* $f_i \in C^q(\mathcal{X}, \mathbb{R})$ *for* $i \in \mathcal{I}$ *with* $q \geq 2$. *The joint strategy* $x^* = (x_1^*, x_2^*) \in \mathcal{X}$ *is a differential Nash equilibrium if* $D_i f_i(x^*) = 0$ *and* $D_i^2 f_i(x^*) \succ 0$ *for each* $i \in \mathcal{I}$.

The differential Nash equilibrium concept has been widely adopted in the past several years as the study of gradient-based learning algorithms in nonconvex games has gained prominence. In particular, a significant research focus has been to analyze the connections between the limit points of 'natural' gradient-based learning dynamics and the set of differential Nash equilibrium (Daskalakis and Panageas, 2018; Mazumdar et al., 2020). Moreover, gradient-based learning dynamics have been crafted for finding and computing only differential Nash equilibrium in zero-sum games (Adolphs et al., 2019; Mazumdar et al., 2019). In some sense, the motivation for much of this research can be attributed to the common perspective of treating the underlying game as a simultaneous play game and then Nash being the natural solution concept from that viewpoint.

In contrast, this chapter takes the perspective of treating the underlying game as having a hierarchical interaction structure, which in turn motivates a closer look at local Stackelberg equilibrium solutions. Toward developing a characterization of local Stackelberg equilibria that is amenable to computation, we take an approach that mirrors past work studying the local Nash equilibrium concept (Ratliff et al., 2013; Ratliff et al., 2016). Specifically, we now provide gradient-based sufficient conditions for a local Stackelberg equilibrium as given in Definition 2.2 and term

joint strategies satisfying the conditions *differential Stackelberg equilibrium*. Recalling that a local Stackelberg equilibrium is described by a joint strategy at which the leader cannot benefit from deviating considering the response curve of the follower and the follower cannot unilaterally benefit from deviating given the strategy of the leader, the sufficient conditions for the leader are naturally characterized in terms of the total derivative implicitly defined by the response curve of the follower and the follower conditions are in terms of its individual derivatives.

Indeed, consider a joint strategy pair $x^* = (x_1^*, x_2^*) \in \mathcal{X}$. Sufficient conditions for the follower being at a local minimum with respect to its choice variable given the strategy of the leader are characterized by the first-order and second-order conditions $D_2 f_2(x_1^*, x_2^*) = 0$ and $D_2^2 f_2(x_1^*, x_2^*) \succ 0$ on the individual derivatives of the follower's cost function. Moreover, given that $D_2 f_2(x_1^*, x_2^*) = 0$ and $\det(D_2^2 f_2(x_1^*, x_2^*)) \neq 0$, sufficient conditions for the leader being at a local minimum along the response curve of the follower are characterized by first-order and second-order conditions $Df(x_1^*, x_2^*) = 0$ and $D^2 f(x_1^*, x_2^*) \succ 0$ on the total derivative of the leader's cost function. Specifically, recall from (2.4) that $Df(x_1^*, x_2^*) := Df(x_1^*, r(x_1^*))|_{r(x_1^*)=x_2^*}$ where the mapping $r : x_1^* \mapsto x_2^*$ is implicitly defined by $D_2^2 f_2(x_1^*, x_2^*) = 0$. Similarly, defining

$$D^2 f(x_1^*, x_2^*) := D^2 f(x_1^*, r(x_1^*))|_{r(x_1^*)=x_2^*} := D(D(f(x_1^*, r(x_1^*))))|_{r(x_1^*)=x_2^*} \tag{2.7}$$

gives rise to the second-order sufficient condition for the leader. We now formally state the conditions for a differential Stackelberg equilibrium.

Definition 2.4 (Differential Stackelberg Equilibrium)**.** *Consider a game (f_1, f_2) defined by $f_i \in C^q(\mathcal{X}, \mathbb{R})$ for $i \in \mathcal{I}$ with $q \geq 2$ and player 1 (without loss of generality) taken to be the leader. The joint strategy $x^* = (x_1^*, x_2^*) \in \mathcal{X}$ is a differential Stackelberg equilibrium if $Df_1(x^*) = 0$, $D_2 f_2(x^*) = 0$, $D^2 f_1(x^*) \succ 0$, and $D_2^2 f_2(x^*) \succ 0$.*

Noting that the Schur complement of $J_{\mathcal{S}}(x^*)$ with respect to $D_2^2 f_2(x^*)$ is identically $D^2 f(x^*)$ at points $x^* = (x_1^*, x_2^*)$ where $D_2 f_2(x^*) = 0$ and $D_2^2 f_2(x^*) \succ 0$, we give alternative but equivalent sufficient conditions as those in Definition 2.4 in terms of $J_{\mathcal{S}}(x^*)$. Moreover, for zero-sum games where $(f_1, f_2) \equiv (f, -f)$, we give another set of equivalent sufficient conditions as those in Definition 2.4 but defined in terms $D_1 f(x^*)$ and $J(x^*)$ for the leader. For a two-by-two block matrices such as $J_{\mathcal{S}}(x)$ and $J(x)$, we denote by $\mathtt{S}_1(J_{\mathcal{S}}(x^*))$ and $\mathtt{S}_1(J(x^*))$ as the Schur complement of the block matrix with respect to $D_2^2 f_2(x^*)$.

Proposition 2.1. *Consider a game (f_1, f_2) defined by $f_i \in C^q(\mathcal{X}, \mathbb{R})$ for $i \in \mathcal{I}$ with $q \geq 2$ and player 1 (without loss of generality) taken to be the leader. Let $x^* = (x_1^*, x_2^*) \in \mathcal{X}$ satisfy $D_2 f_2(x^*) = 0$ and $D_2^2 f_2(x^*) \succ 0$. Then $Df_1(x^*) = 0$ and $\mathtt{S}_1(J_{\mathcal{S}}(x^*)) \succ 0$ if and only if x^* is a differential Stackelberg equilibrium. Moreover, in zero-sum games $(f_1, f_2) = (f, -f)$ with $f \in C^q(\mathcal{X}, \mathbb{R})$ for $q \geq 2$, $D_1 f(x^*) = 0$ and $\mathtt{S}_1(J(x^*)) \succ 0$ if and only if x^* is a differential Stackelberg equilibrium.*

The proof of Proposition 2.1 can be found in Section 2.A and it follows directly from calculus and the implicit function theorem. A benefit of the sufficient conditions in Proposition 2.1 for a local Stackelberg equilibrium is that the leader higher order condition is explicitly stated in terms of quantity that depends on block components of the Jacobian matrix of the Stackelberg gradient dynamics in general-sum games and also the simultaneous gradient dynamics in zero-sum games, which are relevant objects to analyzing the local stability of the dynamics around local equilibrium.

Observe that Proposition 2.1 implies that in the special subclass of zero-sum games, the conditions for differential Nash and differential Stackelberg only differ by the second-order condition for the leader. That is, taking player 1 as the leader in a zero-sum game $(f, -f)$, the conditions are equivalent except that a differential Nash equilibrium requires $D_1^2 f(x^*)$ to be positive definite whereas a differential Stackelberg equilibrium requires the Schur complement of the simultaneous Jacobian given by

$$\mathtt{S}_1(J(x^*)) = D_1^2 f(x^*) - D_{12} f(x^*)(D_2^2 f(x^*))^{-1} D_{21} f(x^*)$$

to be positive definite. We discuss the implications of this shortly.

Before moving on, let us make a few remarks about similar, and in some cases analogous, equilibrium definitions. There is a vast historical literature on minmax (zero-sum) problems and several works develop characterizations of optimality or equilibrium in terms of first-order or second-order gradient-based conditions (see, for example, Danskin 1966, 1967; Evtushenko 1974). For zero-sum games, the sufficient conditions we present for the local Stackelberg definition that is given are equivalent to that for a strict local minmax (Stackelberg) equilibrium. The strict local minmax (Stackelberg) equilibrium concept for zero-sum games and its gradient-based characterization has appeared previously (Evtushenko, 1974). Moreover, a concurrent work (Jin et al., 2020) presents equivalent sufficient conditions for a local minmax (Stackelberg) equilibrium in zero-sum games.[3] This being said, the strict local minmax (Stackelberg) equilibrium notion and its gradient-based characterization in zero-sum games has not been given attention in the modern body of work on learning in games. Moreover, a benefit of the Stackelberg perspective in this chapter is that it extends beyond zero-sum games to general-sum settings while the minmax equilibrium notion does not. Many machine learning applications (e.g., generative adversarial networks) are in fact non-zero-sum, particularly with the introduction of regularization and other heuristics in the form of cost modifications. Similarly, traditional economic and control problems are naturally general-sum games. In the general-sum game setting, a gradient-based characterization in the form that we provide is not surprising as its derivation is simply from considering sufficient conditions for local optimality respecting the order of play. However, it is important due to the ubiquitous nature of general-sum games in settings of interest.

In the remainder of this section, we study relationships between and properties of the equilibrium notions that have been presented, and then this chapter transitions in Section 2.3 to analyze the behavior of the gradient-based learning dynamics around the equilibrium points.

2.2.4 Relationships Between Nash and Stackelberg Equilibrium

Given that we have thus far presented a pair of local equilibrium concepts and gradient-based sufficient conditions for each solution notion, it is natural to ask the relationship between the solution concepts. For zero-sum games, we provide a subset relationship, and for general-sum games, we relate the solution notions in terms of the costs at the equilibrium.

---

[3]Note that the local minmax definition presented in the work of Jin et al. (2020) is slightly different than the local Stackelberg equilibrium concept presented in this chapter when restricted to zero-sum games, and we discuss it further in a concluding discussion section.

Nash and Stackelberg Equilibrium in Zero-Sum Games. For zero-sum games, we prove that the any local Nash equilibrium is a local Stackelberg equilibrium, and analogously, any differential Nash equilibrium is a differential Stackelberg equilibrium. To provide some intuition for this statement, let us begin by considering the local definitions. Observe that the conditions for the follower (player 2) are equivalent in each definition given the action of the leader (player 1). Moreover, a local Nash equilibrium requires the leader to be at a local minimum within a neighborhood holding the follower fixed at its equilibrium strategy (which locally maximizes the function), whereas a local Stackelberg equilibrium requires the leader to be at a local minimum within a neighborhood considering how the follower deviates along its reaction curve. Observing that former requirement implies the latter leads to the relationship between the local equilibrium concepts. The relationship between the differential notions is almost immediate. In particular, Proposition 2.1 shows that the only difference in zero-sum games is that a differential Nash equilibrium requires $D_1^2 f(x^*)$ to be positive definite whereas a differential Stackelberg equilibrium requires $\mathtt{S}_1(J(x^*))$ to be positive definite. Given that $D_1^2 f(x^*)$ is positive definite, the former must imply the latter. This is formalized in the following proposition.

Proposition 2.2. *In zero-sum games $(f, -f)$ with $f \in C^q(\mathcal{X}, \mathbb{R})$ for $q \geq 2$, any local Nash equilibrium is a local Stackelberg equilibrium and any differential Nash equilibrium is a differential Stackelberg equilibrium.*

Proof. Let us prove the statement regarding the local equilibrium concepts to begin. Consider that $x^* = (x_1^*, x_2^*)$ is a local Nash equilibrium on $U_1 \times U_2 \subset \mathcal{X}_1 \times \mathcal{X}_2$ so that $f(x^*) \leq f(x_1, x_2^*)$ for all $x_1 \in U_1 \subset \mathcal{X}_1$ and $f(x^*) \geq f(x_1^*, x_2)$ for all $x_2 \in U_2 \subset \mathcal{X}_2$. For any $x_1 \in U_1 \subset \mathcal{X}_1$, denote the set of optimal responses by $R_{U_2}(x_1) := \{y \in U_2 | f(x_1, y) \geq f(x_1, x_2), \forall x_2 \in U_2 \subset \mathcal{X}_2\}$. Then, by the properties of x^* being a local Nash equilibrium, we have the following that is explained below:

$$f(x_1^*, x_2^*) = \sup_{x_2 \in R_{U_2}(x_1^*)} f(x_1^*, x_2) \leq \sup_{x_2 \in R_{U_2}(x_1^*)} f(x_1, x_2) \leq \sup_{x_2 \in R_{U_2}(x_1)} f(x_1, x_2).$$

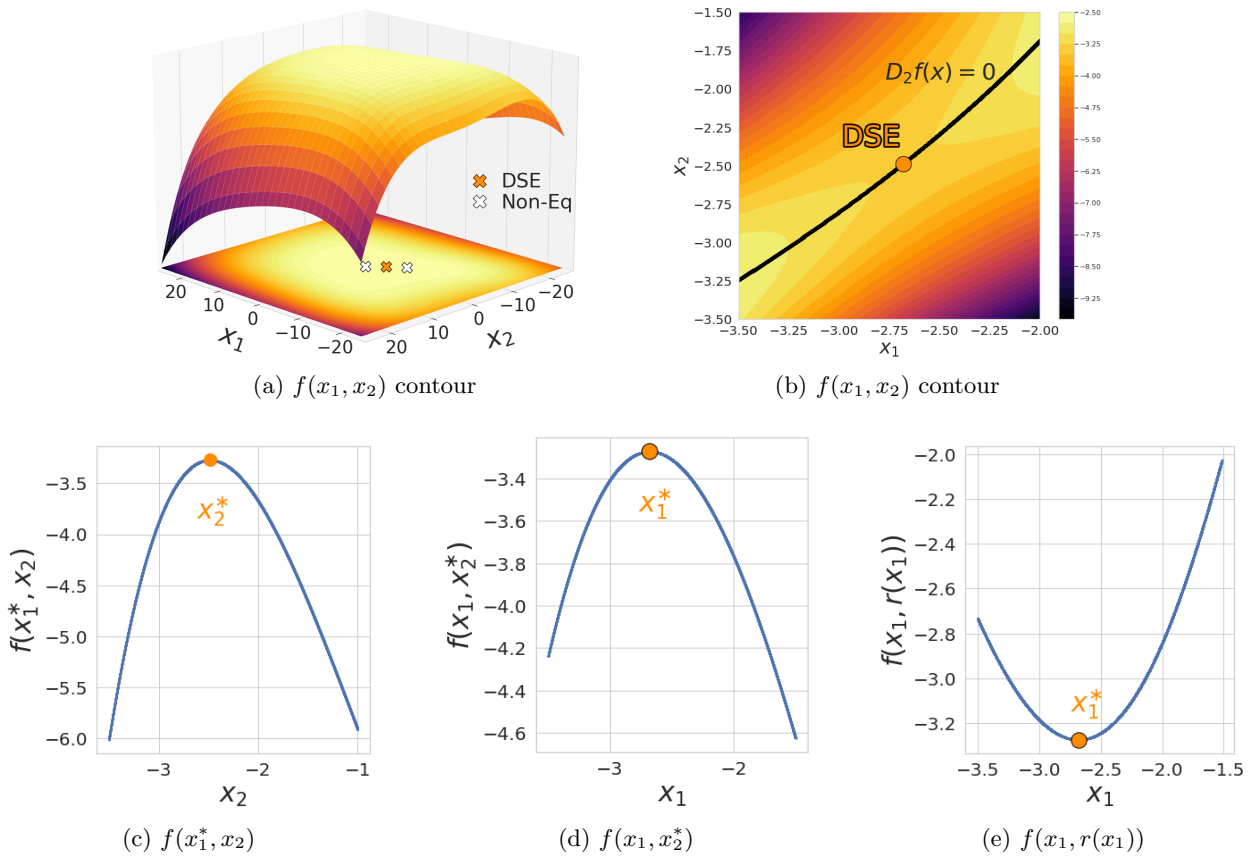
Observe that the equality follows from x_2^* belonging to $R_{U_2}(x_1^*)$, the first inequality holds since $f(x^*) \leq f(x_1, x_2^*)$ for all $x_1 \in U_1 \subset \mathcal{X}_1$, and the final inequality holds since any $x_2 \in R_{U_2}(x_1)$ locally maximizes $f(x_1, x_2)$. Hence, by definition, $x^* = (x_1^*, x_2^*)$ is a local Stackelberg equilibrium on $U_1 \times U_2 \subset \mathcal{X}_1 \times \mathcal{X}_2$.

Now suppose that $x^* = (x_1^*, x_2^*)$ is a differential Nash equilibrium. By definition, $D_1 f(x^*) = 0, D_2 f(x^*) = 0, D_1^2 f(x^*) \succ 0$, and $-D_2^2 f(x^*) \succ 0$. Proposition 2.1 then implies it only needs to be shown that $\mathtt{S}_1(J(x^*))$ is positive definite for x^* to be a differential Stackelberg equilibrium. This is immediate using that $D_1^2 f(x^*)$ and $-D_2^2 f(x^*)$ are positive definite. Indeed, from these properties:

$$\mathtt{S}_1(J(x^*)) = D_1^2 f(x^*) - D_{12} f(x^*)(D_2^2 f(x^*))^{-1} D_{12}^\top f(x^*) \succ 0.$$

Hence, by the characterization in Proposition 2.1, x^* is a differential Stackelberg equilibrium. □

It is worth making a few remarks regarding the preceding result. In general, the Nash and Stackelberg equilibrium notions coincide in zero-sum games when the minimax theorem is available (Basar and Olsder, 1998; Conitzer, 2016). As pointed out by Conitzer (2016), this should not be surprising since any solution concept that does not agree with the minimax theorem would be



(a) $f(x_1, x_2)$ contour (b) $f(x_1, x_2)$ contour

(c) $f(x_1^*, x_2)$ (d) $f(x_1, x_2^*)$ (e) $f(x_1, r(x_1))$

Figure 2.1: Illustration of Example 2.1 showing a local Stackelberg equilibrium that is not a local Nash equilibrium.

suspect. Similarly, on unconstrained strategy spaces, when the cost function satisfies the conditions of the minimax theorem the equilibrium notions generally coincide.[4] However, without structure imposed on the cost function, the equivalence is lost. Indeed, it is easy to see from the proof of Proposition 2.2 that the converse relationship does not necessarily hold. Specifically, there may exist local and differential Stackelberg equilibrium that are not local and differential Nash equilibrium, respectively in zero-sum games. This can be seen directly in the differential characterizations by observing that $\mathtt{S}_1(J(x^*))$ may be positive definite even if $D_1^2 f(x^*)$ is not positive definite. To illustrate this point, consider the following example.

Example 2.1 (Local Stackelberg equilibrium that is not a local Nash equilibrium)**.** *Consider the zero-sum game* $(f, -f)$ *defined by*

$$f(x_1, x_2) = -(0.15x_1^2 + x_2)^2 - (0.25x_2^2 + x_1)^2.$$

[4]For example, it is straightforward to see that in zero-sum strictly-convex-strictly-concave games, differential Nash and differential Stackelberg coincide since $D_1^2 f(x)$ and $-D_2^2 f(x)$ are positive definite at any point such that $D_1 f(x) = 0$ and $D_2 f(x) = 0$ so that by the proof of Proposition 2.2 the notions are equivalent in the game class.

The joint strategy $(x_1^*, x_2^*) = (-2.68, 2.49)$ is a differential Stackelberg (and hence a local Stackelberg) equilibrium but not a local Nash (and hence not a differential Nash) equilibrium of the game. The cost function $f(x_1, x_2)$ is shown in 3-dimensions in Figure 2.1a with the placement of points (x_1, x_2) where $D_1 f(x_1, x_2) = 0$ and $D_2 f(x_1, x_2) = 0$ are labeled. Only $(x_1^*, x_2^*) = (-2.68, 2.49)$ is an equilibrium. Figure 2.1b shows a contour plot of $f(x_1, x_2)$ with (x_1^*, x_2^*) highlighted and the zero-derivative line $D_2 f(x_1, x_2) = 0$ is shown that implicitly defines the reaction curve $r : x_2 \to x_1$. Figure 2.1c shows $f(x_1^*, x_2)$ as a function of x_2 and demonstrates that x_2^* is a local maximum. However, Figure 2.1d shows $f(x_1, x_2^*)$ as a function of x_1 and demonstrates that x_1^* is a local maximum (and not a local minimum) so (x_1^*, x_2^*) is not a local Nash equilibrium. Finally, Figure 2.1e shows $f(x_1, r(x_1))$ as a function of x_1 and demonstrates that x_1^* is a local minimum so that (x_1^*, x_2^*) is a local Stackelberg equilibrium.

Comparing Nash and Stackelberg Equilibrium Costs in General-Sum Games. In general-sum games, the set of Nash and Stackelberg equilibrium can be distinct. Thus, we instead connect the equilibrium notions in terms of the costs. We have alluded to the idea that the ability to act first gives the leader a distinct advantage over the follower in a hierarchical game. We now formalize this statement with a known result (that we refine in terms of local equilibrium concepts) for general-sum games that compares the cost of the leader at Nash and Stackelberg equilibrium when the follower best-response set is always unique.

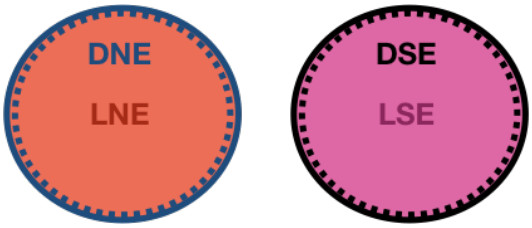
Proposition 2.3 (Basar and Olsder 1998, Proposition 4.4)**.** *Consider a game (f_1, f_2) defined by $f_i \in C^q(\mathcal{X}, \mathbb{R})$ for $i \in \mathcal{I}$ with $q \geq 2$ and player 1 (without loss of generality) taken to be the leader. Consider a local neighborhood $U_1 \times U_2 \subset \mathcal{X}_1 \times \mathcal{X}_2$ and let $f_1^{\mathcal{N}}$ denote the infimum of the set of local Nash equilibrium costs for player 1 on the neighborhood and let $f_1^{\mathcal{S}}$ denote an arbitrary local Stackelberg equilibrium cost for player 1 on the neighborhood. If $R_{U_2}(x_1) = \{y \in U_2 : f_2(x_1, y) \leq f_2(x_1, x_2) \ \forall \ x_2 \in U_2 \subset \mathcal{X}_2\}$ is a singleton for every $x_1 \in U_1 \subset \mathcal{X}_1$, then $f_1^{\mathcal{S}} \leq f_1^{\mathcal{N}}$.*

Proof. Let $x^{\mathcal{N}} = (x_1^{\mathcal{N}}, x_2^{\mathcal{N}})$ denote a local Nash equilibrium on $U_1 \times U_2 \subset \mathcal{X}_1 \times \mathcal{X}_2$ so that $f_1(x^{\mathcal{N}}) \leq f_1(x_1, x_2^{\mathcal{N}})$ for all $x_1 \in U_1 \subset \mathcal{X}_1$ and $f_2(x^{\mathcal{N}}) \leq f_2(x_1, x_2^{\mathcal{N}})$ for all $x_2 \in U_2 \subset \mathcal{X}_2$. Furthermore, suppose that $x^{\mathcal{N}}$ is the local Nash equilibrium on $U_1 \times U_2 \subset \mathcal{X}_1 \times \mathcal{X}_2$ with the minimum cost for the leader among the set of local Nash equilibrium on the neighborhood. That is, $f_1(x^{\mathcal{N}}) \leq f_1(x^{\mathcal{N}'})$ for all local Nash equilibrium $x^{\mathcal{N}'} = (x_1^{\mathcal{N}'}, x_2^{\mathcal{N}'})$ on $U_1 \times U_2 \subset \mathcal{X}_1 \times \mathcal{X}_2$. Moreover, let $x^{\mathcal{S}} = (x_1^{\mathcal{S}}, x_2^{\mathcal{S}})$ denote an arbitrary local Stackelberg equilibrium on $U_1 \times U_2 \subset \mathcal{X}_1 \times \mathcal{X}_2$ so that $f_1(x_1^{\mathcal{S}}, R_{U_2}(x_1^{\mathcal{S}})) \leq f(x_1, R_{U_2}(x_1))$ for all $x_1 \in U_1 \subset \mathcal{X}_1$ where $R_{U_2}(x_1) := \arg\min_{x_2 \in U_2 \subset \mathcal{X}_2} f_2(x_1, x_2)$ for all $x_1 \in U_1 \subset \mathcal{X}_1$ (assumed to be a singleton) and observe that $x_2^{\mathcal{S}} = R_{U_2}(x_1^{\mathcal{S}})$. These properties directly imply

$$f_1(x_1^{\mathcal{S}}, R_{U_2}(x_1^{\mathcal{S}})) \leq f_1(x_1^{\mathcal{N}}, R_{U_2}(x_1^{\mathcal{N}})) = f_1(x_1^{\mathcal{N}}, x_2^{\mathcal{N}}).$$

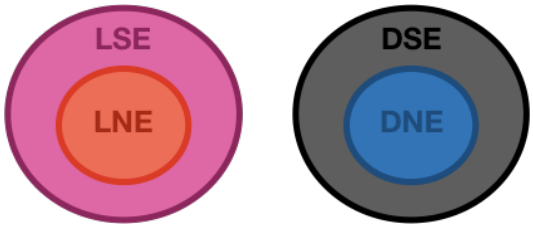
In other words, the leader can always switch to the minimum cost local Nash strategy. □

This result says that the leader never favors the simultaneous play game over the hierarchical play (Stackelberg) game in two-player general-sum games with unique follower responses. On the other hand, the follower may or may not prefer the simultaneous play game over the hierarchical play (Stackelberg) game. The fact that under certain conditions the leader can obtain lower cost in a local Stackelberg equilibrium compared to any of the local Nash equilibrium in the neighborhood may be relevant to generative adversarial networks. Specifically, the game is often general-sum



(a) (b)

Figure 2.2: Genericity results for equilibrium in zero-sum games: (a) local Nash are generically differential Nash (Mazumdar and Ratliff, 2019) and (b) local Stackelberg are generically differential Stackelberg (Theorem 2.1).



(a) (b)

Figure 2.3: Relationships between Nash and Stackelberg equilibrium in zero-sum games (Proposition 2.2): (a) local Nash are a subset of local Stackelberg and (b) differential Nash are a subset of differential Stackelberg.

Figure 2.4: Illustration highlighting some results on equilibrium characterizations in zero-sum games.

after cost modifications. Since typically the goal is for the generator to produce the desired output, intuitively it seems that equilibrium which benefit the generator (leader) may be more likely to correspond to the type of solutions that are sought. While we do not focus on this question, it is worth noting that when each player mutually benefits from the leadership of a player the Stackelberg solution is called concurrent and when each player prefers to be the leader the Stackelberg solution is called non-concurrent (Basar and Olsder, 1998). In this situation, the player that can announce a strategy before the other becomes the leader. It is worth mentioning that this scenario arises in the real-world, such in pricing competitions where the ability to set prices faster results in higher profits in market competitions (Brown and MacKay, 2021). A prototypical example of this type is presented in Section 2.6.1.

In the following subsection, we continue on the topic of equilibrium characterizations and give closer inspection to placement of differential Stackelberg equilibrium amongst local Stackelberg equilibrium. Moreover, we study the robustness properties of the solution concept.

2.2.5 Genericity and Structural Stability of Differential Stackelberg Equilibria

A natural question is how common is it for local equilibria to satisfy sufficient conditions, meaning in a formal mathematical sense, what is the gap between necessary and sufficient conditions in games. Towards addressing this, it has been shown that differential Nash equilibria are *generic* amongst local Nash equilibria and *structurally stable* in the classes of zero-sum and general-sum continuous games, respectively (Mazumdar and Ratliff, 2019; Ratliff et al., 2016). The results say that in 'almost all' games belonging to each class in a formal mathematical sense, the sets of differential Nash equilibrium and local Nash equilibrium are equivalent and also that the equilibria persist under sufficiently smooth perturbations to the cost functions.

We give analogous results for differential Stackelberg equilibria in the class of zero-sum games in this subsection. Since the proofs require a rather dense set of mathematical preliminaries, we

defer them to Section 2.B. We remark that the proofs build on the methods from Mazumdar and Ratliff (2019); Ratliff et al. (2014, 2016) and they largely follow from the fact that the class of two-player zero-sum games are defined completely in terms of a single (sufficiently) smooth function $f \in C^q(\mathcal{X}, \mathbb{R})$, so that the proofs primarily rely on lifting the properties of genericity and structural stability to the class of zero-sum games from the class of smooth functions. In particular, in the class of functions belonging to $C^q(\mathbb{R}^d, \mathbb{R})$ for $q \geq 2$, it is a generic property that all points with a zero gradient vector field are non-degenerate. Note that the usage of generic in this context means there is an open dense set of functions in the class for which the property holds. We lift this property to conclude that among zero-sum games $(f, -f)$ with $f \in C^q(\mathbb{R}^d, \mathbb{R})$ for $q \geq 2$, it is a generic property that both $\det(\mathtt{S}_1(J(x^*))) \neq 0$ and $\det(-D_2^2 f(x^*)) \neq 0$ at points $x^* = (x_1^*, x_2^*)$ such that $D_1 f(x^*) = 0$ and $D_2 f(x^*) = 0$, which then allows us to conclude that it is a generic property that local Stackelberg equilibrium are differential Stackelberg equilibrium. In the remainder of this chapter, we call a zero-sum game $(f, -f)$ with $f \in C^q(\mathbb{R}^d, \mathbb{R})$ for $q \geq 2$, a generic zero-sum game when the aforementioned property holds.

The following result based on the above reasoning allows us to conclude that for a generic zero-sum game, the set of differential Stackelberg equilibrium is equivalent to the set of local Stackelberg equilibrium.

Theorem 2.1. *For the class of two-player, zero-sum continuous games $(f, -f)$ where $f \in C^q(\mathbb{R}^d, \mathbb{R})$ with $q \geq 2$, differential Stackelberg equilibrium are generic amongst local Stackelberg equilibrium. That is, given a generic $f \in C^q(\mathbb{R}^d, \mathbb{R})$, all local Stackelberg equilibrium of the game $(f, -f)$ are differential Stackelberg equilibrium.*

As a final result in this subsection, we show that differential Stackelberg equilibria are structurally stable in the class of zero-sum games. Structural stability ensures that differential Stackelberg equilibria are robust and persist under smooth perturbations to the cost function. We remark that the perturbations considered here are those such that the game remains zero-sum.

Theorem 2.2. *For the class of two-player, zero-sum continuous games $(f, -f)$ where $f \in C^q(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}, \mathbb{R})$ with $q \geq 2$, differential Stackelberg equilibria are structurally stable: given $f \in C^q(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}, \mathbb{R})$, $\zeta \in C^q(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}, \mathbb{R})$, and a differential Stackelberg equilibrium $(x_1, x_2) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, there exists neighborhoods $U \subset \mathbb{R}$ of zero and $V \subset \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ such that $\forall\ t \in U$ there exists a unique differential Stackelberg equilibrium $(\widetilde{x}_1, \widetilde{x}_2) \in V$ for the zero-sum game $(f + t\zeta, -f - t\zeta)$.*

Before moving on, we remark that important classes of non-generic games certainly exist. In games where the cost function of the follower is bilinear, local Stackelberg equilibrium can exist that do not satisfy the sufficient conditions outlined in Definition 2.4. As a simple example, $x^* = (x_1^*, x_2^*) = (0, 0)$ is a local Stackelberg equilibrium for the zero-sum game defined by $f(x_1, x_2) = x_1 x_2$ and not a differential Stackelberg equilibrium since $D_2^2 f_2(x_1^*, x_2^*) = 0$. Since such games belong to a degenerate class in the context of the genericity result we provide, they naturally deserve special attention and algorithmic methods. While we do not focus our attention on this class of games, we do propose some remedies to allow for the gradient-based learning algorithms under consideration to successfully seek out equilibria in them. In the experiments section, we discuss a regularized version (see Section 2.6.4) of the Stackelberg gradient dynamics that injects a small perturbation to cure degeneracy problems leveraging the fact that differential Stackelberg equilibrium are structurally stable. Finally, for bimatrix games with finite actions it is common to reparameterize the problem

using a softmax function to obtain mixed policies on the simplex (Fudenberg et al., 1998). We explore this viewpoint on a parameterized bilinear game in Section 2.6.3.

Given the foundations of game perspectives and equilibrium that have been established in this section, we now move on to begin to analyze the local stability and convergence of the gradient-based learning algorithms that have been introduced in relationship to equilibrium.

2.3 Local Stability of Critical Points in Zero-Sum Games

In this section, we present a local stability analysis of the learning dynamics formulated in Section 2.2.2. Specifically, we evaluate the limiting behavior of the dynamics from a continuous-time viewpoint since the discrete-time systems closely approximate this behavior for suitably selected learning rates. This analysis allows us to draw connections between limiting behavior of each set of dynamics and Nash and Stackelberg equilibria in zero-sum games. To begin this section, we provide background on dynamical systems analysis and the terminology we adopt. This background is necessary for the remainder of Part I of this thesis. We then study the local stability in zero-sum games of the simultaneous gradient dynamics and the Stackelberg gradient dynamics in Section 2.3.2 and Section 2.3.3, respectively.

2.3.1 Background on Dynamical Systems Analysis and Terminology

The gradient-based learning dynamics we analyze in continuous-time are nonlinear dynamical systems of the form $\dot{x} = -F(x)$ where $F : \mathcal{X} \to \mathcal{X}$. Toward characterizing the local behavior of the dynamics around points of interest in the optimization landscape, we rely on methods from the nonlinear systems theory (Khalil, 2002; Sastry, 1999) for determining and characterizing stability and instability. The rest of this subsection is devoted to providing background on this topic.

Let us begin by introducing the notion of a critical (stationary) point of a dynamical system. Often in dynamical systems theory, such points are called equilibrium points, but given the use of game-theoretic terminology for equilibrium in this thesis, we do not follow this convention.

Definition 2.5 (Critical Point)**.** *We refer to stationary points of dynamical systems as critical points. That is, $x^* \in \mathcal{X}$ is a critical point of the system $\dot{x} = -F(x)$ if $F(x^*) = 0$.*

Observe that in zero-sum games $(f, -f)$, the critical points of the simultaneous gradient dynamics system $\dot{x} = -g(x)$ and the Stackelberg gradient dynamics system $\dot{x} = -g_{\mathcal{S}}(x)$ coincide. Specifically, any critical point $x^* = (x_1^*, x_2^*) \in \mathcal{X}$ of each system is such that $D_1 f(x^*) = 0$ and $D_2 f(x^*) = 0$. This also means that all critical points of the systems satisfy the first-order conditions for both a differential Nash equilibrium and a differential Stackelberg equilibrium. We state these facts now for later reference.

Lemma 2.1. *Consider a zero-sum game $(f, -f)$ defined by $f \in C^q(\mathcal{X}, \mathbb{R})$, $q \geq 2$. The critical points of $\dot{x} = -g(x)$ and $\dot{x} = -g_{\mathcal{S}}(x)$ coincide. Moreover, $x^* = (x_1^*, x_2^*) \in \mathcal{X}$ is a critical point of $\dot{x} = -g(x)$ and $\dot{x} = -g_{\mathcal{S}}(x)$ if and only if x^* satisfies the first-order conditions for both a differential Nash equilibrium and a differential Stackelberg equilibrium.*

Proof. The result holds since for any $x^* = (x_1^*, x_2^*) \in \mathcal{X}$, $D_1 f(x^*) = 0$ and $D_2 f(x^*) = 0$ if and only if $Df(x^*) = D_1 f(x^*) - D_{21} f(x^*)^\top (D_2^2 f(x^*))^{-1} D_2 f(x^*) = 0$ and $D_2 f(x^*) = 0$. The final statement is then immediate by the equilibrium characterizations in Definition 2.3 and Proposition 2.1. □

We remark that it is straightforward to see in general-sum games the critical points of $\dot{x} = -g(x)$ generally only satisfy the first-order sufficient conditions for a differential Nash equilibrium and similarly the critical points of $\dot{x} = -g_{\mathcal{S}}(x)$ generally only satisfy the first-order sufficient conditions for differential Stackelberg equilibrium.

The goal of this section is to determine the local behavior of each set of gradient-based dynamics around the set of critical points. From the point of view of seeking to compute equilibria, ideally the dynamics would be unstable locally around critical points that are not game-theoretically meaningful, and locally stable around critical points that are game-theoretically meaningful, where game-theoretically meaningful in this context refers to corresponding with the differential Nash or Stackelberg equilibrium concepts.

Notions of Stability and Instability. We begin by recalling some definitions of stability and then present methods for verifying stability pertinent to the analysis methods adopted in this chapter and those that follow. The following material can be found in standard textbooks on nonlinear systems theory (Khalil, 2002; Sastry, 1999) and is presented here for easy reference and background since it is used throughout Part I of this thesis.

Consider a time-varying nonlinear system of the form $\dot{x} = -F(x,t)$ where $F : \mathbb{R}^d \times \mathbb{R}_{\geq 0} \to \mathbb{R}^d$ and $x(t_0) = x_0$. This system is said to be time-invariant if $\dot{x} = -F(x,t)$ does not depend explicitly on t. The typical notions of stability are (uniform) Lyapunov stability, (uniform) asymptotic stability, and exponential stability. We describe each stability notion in terms of local characterizations and adopt the notation $x \in B_\delta(y)$ to indicate $\|x - y\| \leq \delta$ for any $x, y \in \mathbb{R}^d$ and $\delta > 0$.

Lyapunov Stability. A critical point x^* of $\dot{x} = -F(x,t)$, that is x^* such that $F(x^*,t) = 0$ for all $t \geq 0$, is said to be stable in the sense of Lyapunov, or simply Lyapunov stable, if for all $t_0 \geq 0$ and $\varepsilon > 0$, there exists $\delta(t_0, \varepsilon)$ such that $x_0 \in B_{\delta(t_0,\varepsilon)}(x^*)$ implies $x(t) \in B_\varepsilon(x^*)$ for all $t \geq t_0$ where $x(t)$ is the solution of the system starting from x_0 at t_0. If $\delta(t_0, \varepsilon)$ is independent of t_0 (such as in time-invariant systems), then this property is called uniform Lyapunov stability. Simply put, Lyapunov stability characterizes the phenomenon that if the system is initialized at some x_0 in a $\delta(t_0, \varepsilon)$-neighborhood around a critical point x^*, then the trajectory always remain in an ε-neighborhood around x^*. Observe that this notion of stability does not imply that the trajectory $x(t)$ tends to x^* asymptotically as $t \to \infty$.

Instability. A critical point x^* of $\dot{x} = -F(x,t)$, that is x^* such that $F(x^*,t) = 0$ for all $t \geq 0$, is said to be unstable if it is not Lyapunov stable. Considering the definition of Lyapunov stability, this means that there exists an $\varepsilon > 0$, such that for all $\delta > 0$, there exists an $x_0 \in B_\delta(x^*)$ such that there is a time $t > t_0 \geq 0$ for which $x(t) \notin B_\varepsilon(x^*)$ where $x(t)$ is the solution of the system starting from x_0 at t_0. In plain words, this means if x^* is unstable, then there is an initial condition from any arbitrarily small neighborhood from which the system leaves the neighborhood around x^* at some time. Note that instability does not preclude returning back to a neighborhood around x^*.

Asymptotic Stability. A stronger notion of stability than Lyapunov stability is asymptotic stability. Given a critical point x^* of $\dot{x} = -F(x,t)$, that is x^* such that $F(x^*,t) = 0$ for all $t \geq 0$, x^* is said to be asymptotically stable if it is Lyapunov stable and additionally it is attractive—that is, for all $t_0 \geq 0$, there exists $\delta(t_0)$ such that $x_0 \in B_\delta(x^*)$ implies $\lim_{t\to\infty} \|x(t) - x^*\| = 0$. Moreover, if x^* is uniformly Lyapunov stable and there exists a function $\gamma : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}_+$ and $\delta > 0$ such that $\lim_{t\to\infty} \gamma(t, x_0)$ for all $x_0 \in B_\delta(x^*)$ and $x_0 \in B_\delta(x^*)$ implies $x(t) \in B_{\gamma(t-t_0,x_0)}(x^*)$ for all $t \geq t_0$ where $x(t)$ is the solution of the system starting from x_0 at t_0, then x^* is said to be uniformly asymptotically stable. Observe that both Lyapunov stability and the attractive property

are needed for asymptotic stability since the attractive property does not imply Lyapunov stability. In simple terms, asymptotic stability implies that if the system is initialized at some x_0 in a $\delta(t_0, \varepsilon)$-neighborhood around a critical point x^*, then the trajectory always remain in an ε-neighborhood around x^* and it tends to x^* asymptotically as $t \to \infty$.

Exponential Stability. Exponential stability is the strongest notion of stability. A critical point x^* of $\dot{x} = -F(x,t)$, that is x^* such that $F(x^*,t) = 0$ for all $t \geq 0$, is said to be exponential stable if there exist $c, \alpha > 0$ such that $\|x(t) - x^*\| \leq ce^{-\alpha(t-t_0)}\|x_0 - x^*\|$ for all $x_0 \in B_\delta(x^*)$ with $\delta \geq 0$ and $t \geq t_0 \geq 0$ where $x(t)$ is the solution of the system starting from x_0 at t_0. The constant α is an estimate of the rate of convergence. Note that in linear systems (even time-varying) in which $F(x,t) = A(t)x$ so that $\dot{x} = -A(t)x$, uniform asymptotic stability is equivalent to exponential stability (Sastry, 1999, Theorem 5.33).

Methods for Verifying Stability and Instability. There are a pair of standard methods for determining stability and instability properties, the Lyapunov method and the indirect Lyapunov method. The Lyapunov method hinges on the ability to construct a Lyapunov function, however, for general nonlinear systems without a given structure, it is often hard to know how to go about finding a Lyapunov function. The alternative is what is often known as the indirect Lyapunov method. The indirect Lyapunov method involves studying the stability of the linearization of a nonlinear system around a critical point in order to determine the stability properties of the nonlinear system. We adopt this approach throughout Part I of this thesis and now review the conclusions that can be drawn from this method.

Consider a time-invariant nonlinear system of the form $\dot{x} = -F(x)$ where $F : \mathbb{R}^d \to \mathbb{R}^d$. Let $J_F(x^*) \in \mathbb{R}^{d\times d}$ denote the Jacobian of $F(x)$ evaluated at a critical point x^* of the nonlinear system and observe that the linearized system around x^* can be described by $\dot{x} = -J_F(x^*)x$. A critical point x^* of $\dot{x} = -F(x)$ is said to by non-degenerate if $\det(J_F(x^*)) \neq 0$ and hyperbolic if there are no eigenvalues of $J_F(x^*)$ with zero real part. All hyperbolic critical points are non-degenerate, but not all non-degenerate critical points are hyperbolic. Informally, the Hartman-Grobman theorem (Sastry, 1999, Theorem 7.3) asserts that the qualitative properties of the nonlinear system $\dot{x} = -F(x)$ in the vicinity of a hyperbolic critical point x^* are determined by its linearization $J_F(x^*)$. We remark that Hartman-Grobman can also be applied to discrete time maps with the same qualitative outcome (Sastry, 1999, Theorem 7.8). These results lead to direct methods for determining stability and instability.

The following result shows that a critical point x^* of $\dot{x} = -F(x)$ is such that $\text{spec}(-J_F(x^*)) \subset \mathbb{C}^\circ_-$, or equivalently that $\text{spec}(J_F(x^*)) \subset \mathbb{C}^\circ_+$, then x^* is a locally exponentially stable critical point of $\dot{x} = -F(x)$. In other words, ability to verify $\text{spec}(-J_F(x^*)) \subset \mathbb{C}^\circ_-$ directly implies the nonlinear system convergences at an exponential rate to x^* given an initialization in a local neighborhood. In general, we simply say that a critical point x^* of $\dot{x} = -F(x)$ is stable given $\text{spec}(-J_F(x^*)) \subset \mathbb{C}^\circ_-$.[5]

Theorem 2.3 (Stability Characterizations, Khalil 2002, Theorem. 4.6, Corollary 4.3)**.** *Consider a critical point x^* of $\dot{x} = -F(x)$. The following are equivalent: (a) x^* is a locally exponentially stable critical point of $\dot{x} = -F(x)$; (b) $\text{spec}(-J_F(x^*)) \subset \mathbb{C}^\circ_-$; (c) Given any symmetric matrix $Q \in \mathbb{R}^{d\times d}$, there exists a unique symmetric positive-definite matrix $P \in \mathbb{R}^{d\times d}$ such that $PJ_F(x^*) + J_F(x^*)^\top P = Q$.*

[5]Note that if a matrix $A \in \mathbb{R}^{d\times d}$ is such that $\text{spec}(A) \subset \mathbb{C}^\circ_-$ (that is, the real part of each eigenvalue is negative), then often A is called a Hurwitz stable, or just stable, matrix.

To contrast with the previous result, the following result shows that a critical point x^* of $\dot{x} = -F(x)$ such that $-J_F(x^*)$ has at least one eigenvalue in $\mathbb{C}_+^\circ$, or equivalently that $J_F(x^*)$ has at least one eigenvalue in $\mathbb{C}_-^\circ$, is an unstable critical point of $\dot{x} = -F(x)$. Thus, the ability to verify that $-J_F(x^*)$ has an eigenvalue in $\mathbb{C}_+^\circ$ directly implies there is an initial condition from any arbitrarily small neighborhood around x^* such that the system leaves a neighborhood around x^*. In general, we say that a critical point x^* of $\dot{x} = -F(x)$ is a strict saddle given that $-J_F(x^*)$ has an eigenvalue in $\mathbb{C}_+^\circ$ and x^* is hyperbolic.

Theorem 2.4 (Instability Characterization, Sastry 1999, Theorem 5.42)**.** *Consider a critical point x^* of a nonlinear system $\dot{x} = -F(x)$ with $J_F(x^*)$ denoting the Jacobian of $F(x)$ evaluated at x^*. If $-J_F(x^*)$ has at least one eigenvalue in $\mathbb{C}_+^\circ$, then x^* is an unstable critical point of the nonlinear system $\dot{x} = -F(x)$.*

Together, Theorem 2.3 and Theorem 2.4 show that the local stability or instability of any critical point of a time-invariant nonlinear system can be determined by assessing the eigenvalues of the Jacobian matrix, with the exception of non-hyperbolic critical points x^* for which $\text{spec}(-J_F(x^*)) \subset \mathbb{C}_-$ (marginally stable critical points). These results will be key for the remainder of the section.

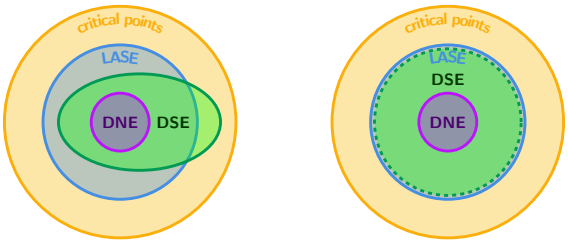
We remark that analogous characterizations hold for discrete-time systems (Ortega and Rheinboldt, 1970; Sastry, 1999). In particular, consider the discrete-time variant $x^+ = -F(x)$ of the time-invariant nonlinear system that has been discussed. Given a hyperbolic critical point x^* of $x^+ = -F(x)$, x^* is a locally exponentially stable equilibrium if and only if $\rho(J_F(x^*)) < 1$. Moreover, if $\rho(J_F(x^*)) > 1$, then x^* is an unstable critical point of $x^+ = -F(x)$. Thus, the local stability or instability of any hyperbolic critical point of a discrete-time time-invariant nonlinear system can again be determined by assessing the eigenvalues of the Jacobian matrix, with the exception of critical points x^* for which $\rho(J_F(x^*)) = 1$ (marginally stable critical points). The methods for assessing the stability and instability of critical points in discrete-time nonlinear-systems will be key to the discrete-time convergence analysis that follows this section.

2.3.2 Local Stability of Simultaneous Gradient Dynamics

In this subsection, we study the local stability of the simultaneous gradient dynamics around critical points. The goal is is to determine the relationship between the stable critical points of the system $\dot{x} = -g(x)$ and the set of differential Stackelberg equilibrium in zero-sum games. Recall that as discussed in the Section 2.3.1, any critical point of the system $\dot{x} = -g(x)$ satisfies the first-order conditions for both a differential Nash equilibrium and a differential Stackelberg equilibrium. It is worth noting that recent works (Daskalakis and Panageas, 2018; Mazumdar et al., 2020) study the relationship between stable critical points of the simultaneous gradient dynamics and the set of differential Nash equilibrium. The existing results in this direction along with the equilibrium characterizations from the previous section guide the investigation presented in this subsection.

2.3.2.1 Stability of Nash-Stackelberg Equilibrium

To begin, we remark existing results have shown that in zero-sum games any differential Nash equilibrium is a stable critical point of the simultaneous gradient dynamics (Daskalakis and Panageas, 2018; Mazumdar et al., 2020). Thus, combined with the result of Proposition 2.2 that any differential Nash equilibrium is a differential Stackelberg equilibrium in zero-sum games, this immediately



(a) $\dot{x} = -g(x)$ (b) $\dot{x} = -g_S(x)$

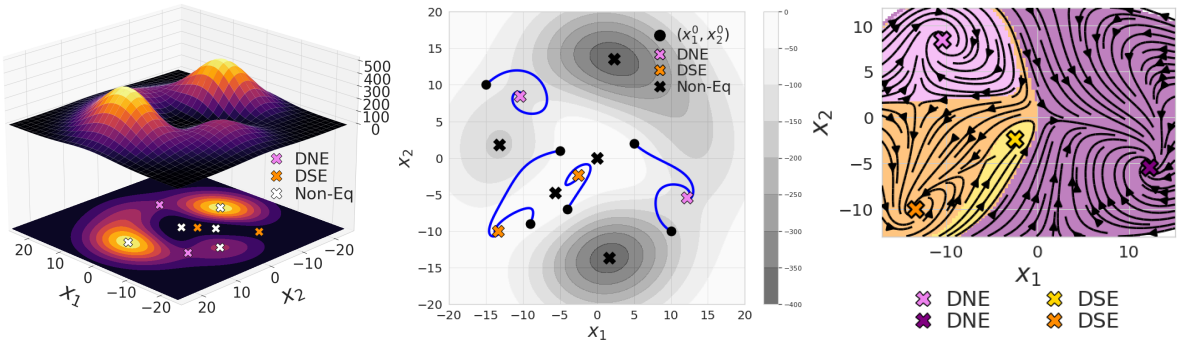
Figure 2.5: Graphical depiction of the local stability characterizations we present for critical points of the simultaneous gradient dynamics $\dot{x} = -g(x)$ (Figure 2.5a) and Stackelberg gradient dynamics $\dot{x} = -g_S(x)$ (Figure 2.5b) in zero-sum games. For the simultaneous gradient dynamics, we show that all differential Stackelberg equilibrium that are differential Nash equilibrium are locally stable (Proposition 2.4) and also that some differential Stackelberg equilibrium that are not differential Nash equilibrium are locally stable. For the Stackelberg gradient dynamics, we show that any critical point is stable if and only if it is a differential Stackelberg equilibrium (Theorem 2.5). The subset relationship between differential Nash equilibrium and differential Stackelberg equilibrium is from Proposition 2.2.

implies that any differential Stackelberg equilibrium that is a differential Nash equilibrium must be a stable critical point of the simultaneous gradient dynamics. Moreover, in zero-sum games, it has been shown that local Nash equilibria are generically differential Nash equilibria (Mazumdar and Ratliff, 2019). Thus, combined with the result of Theorem 2.1 that local Stackelberg equilibria are generically differential Stackelberg equilibria, this immediately implies that in a generic zero-sum game, any local Stackelberg equilibrium that is a local Nash equilibrium must be a stable critical point of the simultaneous gradient dynamics. We now summarize these statements in the following proposition and then discuss the implications thereafter.

Proposition 2.4. *Consider a zero-sum game $(f, -f)$ defined by $f \in C^q(\mathcal{X}, \mathbb{R})$ with $q \geq 2$. Any differential Nash equilibrium is a stable critical point of $\dot{x} = -g(x)$ and a differential Stackelberg equilibrium. For a generic zero-sum game $(f, -f)$, any local Nash equilibrium is a stable critical point of $\dot{x} = -g(x)$ and a local Stackelberg equilibrium.*

Proof. The first statement follows directly from Mazumdar et al. (2020, Proposition 8) and Proposition 2.2. Indeed, by Mazumdar et al. (2020, Proposition 8) any differential Nash equilibrium is a stable critical point of $\dot{x} = -g(x)$ and by Proposition 2.2 it must be a differential Stackelberg equilibrium. Now, suppose that $f \in C^q(\mathcal{X}, \mathbb{R})$ is a generic function. Then, by the genericity of differential Nash equilibria in zero-sum games (Mazumdar and Ratliff, 2019, Theorem 2), all local Nash equilibria of $(f, -f)$ are differential Nash equilibria. Similarly, by Theorem 2.1 (genericity of differential Stackelberg equilibria in zero-sum games), all local Stackelberg equilibria of $(f, -f)$ are differential Stackelberg equilibria so that the final statement of the result then holds by Mazumdar et al. (2020, Proposition 8) and Proposition 2.2 again. □

This result shows in zero-sum games at least some of the stable critical points of the simultaneous gradient dynamics are differential Stackelberg equilibria (and generically local Stackelberg



(a) (b) (c)

Figure 2.6: Example demonstrating existence of stable differential Stackelberg equilibria with respect to the simultaneous gradient dynamics that are not differential Nash equilibria in a zero-sum game $(f, -f)$ where f is defined in (2.8) with $a = 0.15, b = 0.25$. There are two stable points of the simultaneous gradient dynamics that are differential Stackelberg equilibrium, but not differential Nash equilibrium.

equilibria), specifically those which are also differential Nash equilibria (and generically local Nash equilibria). Moreover, it indicates learning dynamics seeking only local Nash equilibrium are also seeking a subset of the local Stackelberg equilibrium (Adolphs et al., 2019; Mazumdar et al., 2019).

2.3.2.2 Stability of Non-Nash-Stackelberg Equilibrium

To follow up on the previous result, it is then natural to ask if the simultaneous gradient dynamics are only locally stable around differential Nash equilibrium in zero-sum games. This query has been resolved negatively. In particular, past results show that there exists stable critical points of the simultaneous gradient dynamics that are not local Nash equilibria, and hence not differential Nash equilibria (Daskalakis and Panageas, 2018; Mazumdar et al., 2020, Lemma 2.5, Proposition 9). Given the focus on the Nash equilibrium solution concept, this property has commonly been interpreted as an indication that the simultaneous gradient dynamics often locally converge to strategies lacking game-theoretic meaning. However, our focus on the Stackelberg equilibrium solution concept allows us to look at stable strategies of this type through another lens.

Scalar Zero-Sum Games. The examples presented by Mazumdar et al. (2020) and Daskalakis and Panageas (2018) are constructed on scalar action spaces, that is $x_1 \in \mathbb{R}$ and $x_2 \in \mathbb{R}$, and such that at a stable critical point $x^* = (x_1^*, x_2^*)$ either $D_1^2 f(x^*) \succ 0$ or $-D_2^2 f(x^*) \succ 0$. Let us consider a zero-sum with several stable critical points having this property.

Example 2.2. *Let $a = 0.15$ and $b = 0.25$. Consider the zero-sum game defined by*

$$f(x_1, x_2) = -e^{-0.01(x_1^2 + x_2^2)}((ax_1^2 + x_2)^2 + (bx_2^2 + x_1)^2). \tag{2.8}$$

The function $-f(x_1, x_2)$ is visualized in Figure 2.6a along with the placement of equilibria in the game. Figure 2.6b shows the behavior of the simultaneous gradient dynamics from several initial-

izations on a contour plot of the function. Finally, Figure 2.6c shows the regions of attraction for critical points on top of the vector field of the dynamics. Observe that all stable critical points are differential Stackelberg equilibrium and there is a continuum of initial conditions from which the dynamics converge to differential Stackelberg equilibrium that is not a differential Nash equilibrium.

Example 2.2 serves as evidence that often stable critical points of the simultaneous gradient dynamics that are not differential Nash equilibrium are in fact game-theoretically meaningful since they can be explained through the differential Stackelberg equilibrium concept. In fact, it is straighforward to verify that in zero-sum games on $\mathbb{R}^2$, all stable critical points of the simultaneous gradient dynamics must be a differential Stackelberg equilibrium with player 1 or player 2 as the leader. In this context, it means all stable points are strict local minmax or strict local maxmin equilibrium.

Proposition 2.5. *Consider a zero-sum game $(f, -f)$ defined by $f \in C^q(\mathbb{R} \times \mathbb{R}, \mathbb{R})$ with $q \geq 2$. All stable critical points $x^* = (x_1^*, x_2^*)$ of $\dot{x} = -g(x)$ are such that $D_1^2 f(x^*) \succ 0$ or $-D_2^2 f(x^*) \succ 0$. Moreover, all stable critical points $x^* = (x_1^*, x_2^*)$ of $\dot{x} = -g(x)$ at which $-D_2^2 f(x^*) \succ 0$ that are not differential Nash equilibria are differential Stackelberg equilibria with player 1 as the leader, and all stable critical points $x^* = (x_1^*, x_2^*)$ of $\dot{x} = -g(x)$ at which $D_1^2 f(x^*) \succ 0$ that are not differential Nash equilibria are differential Stackelberg equilibria with player 2 as the leader.*

Proof. Suppose $x^* = (x_1^*, x_2^*)$ is a stable critical point of $\dot{x} = -g(x)$ so that the eigenvalues of $J(x^*)$ have positive real parts. This fact guarantees that the determinant and trace of $J(x^*)$ must be positive since the eigenvalues are either complex conjugates or both real. As a result, $(D_{12} f(x^*))^2 \succ D_1^2 f(x^*) D_2^2 f(x^*)$ and $D_1^2 f(x^*) \succ D_2^2 f(x^*)$. Observe that the latter condition is impossible if $D_1^2 f(x^*) \preceq 0$ and $-D_2^2 f(x^*) \preceq 0$ so it must be that $D_1^2 f(x^*) \succ 0$ or $-D_2^2 f(x^*) \succ 0$. This proves the first statement. Now, suppose that $D_1^2 f(x^*) \preceq 0$ and $-D_2^2 f(x^*) \succ 0$ so that x^* is not a differential Nash equilibrium. This assumption, combined with the previous conditions that follow from stability directly imply

$$\mathtt{S}_1(J(x^*)) = D_1^2 f(x^*) - D_{12} f(x)(D_2^2 f(x))^{-1} D_{21} f(x) \succ 0.$$

Thus, together with the characterization in Proposition 2.1, x^* is a differential Stackelberg equilibrium with player 1 as the leader. Following identical reasoning, but assuming that $D_1^2 f(x^*) \succ 0$ and $-D_2^2 f(x^*) \preceq 0$, it follows that

$$\mathtt{S}_2(J(x^*)) = -D_2^2 f(x^*) + D_{12} f(x)(D_1^2 f(x))^{-1} D_{21} f(x) \succ 0.$$

This proves the final statement holds. □

While it is often important to discriminate between strict local minmax equilibrium and strict local maxmin equilibrium given that they have distinct implications, the previous result at least implies that in simple scalar games all stable critical points of the simultaneous gradient dynamics are game-theoretically meaningful through the lens of the Stackelberg equilibrium concept.

Generalization to Arbitrary Zero-Sum Games. The concurrent work of Jin et al. (2020) shows that the property stated in Proposition 2.5 fails to generalize to zero-sum games of arbitrary dimension. Specifically, for a zero-sum game $f \in C^q(\mathbb{R}^2 \times \mathbb{R}^2, \mathbb{R})$ with $q \geq 2$, it is shown (Jin et al.,

2020, Proposition 27) that there exists stable critical points $x^* = (x_1^*, x_2^*)$ of $\dot{x} = -g(x)$ such that both $D_1^2 f(x^*)$ and $-D_2^2 f(x^*)$ are indefinite, implying that the joint strategy is not a local minmax (local Stackelberg with player 1 as the leader) equilibrium or a local maxmin (local Stackelberg with player 2 as the leader) equilibrium.

Thus, the focus of the rest of this subsection is on determining conditions under which a stable critical point $x^* = (x_1^*, x_2^*)$ of $\dot{x} = -g(x)$ at which $-D_2^2 f(x^*) \succ 0$ is a differential Stackelberg equilibrium. The motivation for this is to develop a deeper understanding of the role of differential Stackelberg equilibrium in the optimization landscape of zero-sum games. At the end of this subsection, we specifically discuss the implications of the results that follow.

Ideally, we would like to generalize from scalar zero-sum games to zero-sum games of arbitrary dimension the statement in Corollary 2.5 that any stable critical point $x^* = (x_1^*, x_2^*)$ of $\dot{x} = -g(x)$ at which $-D_2^2 f(x^*) \succ 0$ is a differential Stackelberg equilibrium. However, for now, we instead provide necessary and sufficient conditions for this generalization to hold. We begin by stating the results and then given an interpretation of them.

For the following pair of propositions, we need some notation that is common across the results. Let $x_1 \in \mathbb{R}^{d_1}$ and $x_2 \in \mathbb{R}^{d_2}$. For a stable critical point $x^* = (x_1^*, x_2^*)$ of $\dot{x} = -g(x)$ that is not a differential Nash equilibrium and is such that $-D_2^2 f(x^*) \succ 0$, let $\text{spec}(D_1^2 f(x^*)) = \{\mu_j,\ j \in [d_1]\}$ where $\mu_1 \leq \cdots \leq \mu_m \leq 0 < \mu_{m+1} \leq \cdots \leq \mu_{d_1}$, and let $\text{spec}(-D_2^2 f(x^*)^{-1}) = \{\lambda_j,\ j \in [d_2]\}$ where $\lambda_1 \geq \cdots \geq \lambda_{d_2} > 0$, and define $p = \dim(\ker(D_1^2 f(x^*)))$. [6] Finally, for a matrix W, let $W^\dagger$ denote the conjugate transpose.

Proposition 2.6 (Necessary Conditions)**.** *Consider a zero-sum game $(f, -f)$ defined by $f \in C^q(\mathcal{X}, \mathbb{R})$ with $q \geq 2$ and a stable critical point $x^* = (x_1^*, x_2^*)$ of the simultaneous gradient dynamics $\dot{x} = -g(x)$ that is not a differential Nash equilibrium and is such that $-D_2^2 f(x^*) \succ 0$. Given $\kappa > 0$ such that $\|D_{12} f(x^*)\| \leq \kappa$, if x^* is a differential Stackelberg equilibrium, then $m \leq d_2$ and $\kappa^2 \lambda_j + \mu_j > 0$ for all $j \in [m]$.*

Proposition 2.7 (Sufficient Conditions)**.** *Consider a zero-sum game $(f, -f)$ defined by $f \in C^q(\mathcal{X}, \mathbb{R})$ with $q \geq 2$ and a stable critical point $x^* = (x_1^*, x_2^*)$ of the simultaneous gradient dynamics $\dot{x} = -g(x)$ that is not a differential Nash equilibrium and is such that $-D_2^2 f(x^*) \succ 0$. Suppose that there exists a diagonal matrix $\Sigma \in \mathbb{C}^{d_1 \times d_2}$ with non-zero entries such that $D_{12} f(x^*) = W_1 \Sigma W_2^\dagger$ where W_1 are the orthonormal eigenvectors of $D_1^2 f(x^*)$ and W_2 are orthonormal eigenvectors of $-D_2^2 f(x^*)$. Given $\kappa > 0$ such that $\|D_{12} f(x^*)\| \leq \kappa$, if $m \leq d_2$ and $\kappa^2 \lambda_j + \mu_j > 0$ for all $j \in [m - p]$, then x^* is a differential Stackelberg equilibrium.*

The proofs of Proposition 2.6 and Proposition 2.7 are deferred to Section 2.C. The analysis primarily follows the arguments from Berger et al. (2019), but tailored to the context of this problem. It is also worth noting that the fact that $\text{spec}(-J(x^*)) \subset \mathbb{C}_-^\circ$ is not used in proving this set of result. We conjecture that using this property can lead to stronger characterizations.

We now give an interpretation of this set of results. Since it is given that $x^* = (x_1^*, x_2^*)$ is a critical point of $\dot{x} = -g(x)$ and $-D_2^2 f(x^*) \succ 0$, the only condition that remains for x^* to be a differential Stackelberg equilibrium is $\mathtt{S}_1(J(x^*)) \succ 0$. Taking intuition from the expression $\mathtt{S}_1(J(x^*)) = D_1^2 f(x^*) - D_{12} f(x^*)(D_2^2 f(x^*))^{-1} D_{12}^\top f(x^*)$, the conditions are derived from relating $\text{spec}(D_1^2 f(x^*))$ to $\text{spec}(-(D_2^2 f(x^*)^{-1}))$ via $D_{12} f(x^*)$. The necessary conditions essentially say that

[6] Observe that the set notation $[z] := \{1, 2, \ldots, z\}$ for some positive integer $z \in \mathbb{Z}^+$is being used.

when $\mathtt{S}_1(J(x^*)) \succ 0$, it must be that in the directions of $D_1^2 f(x^*)$ with negative eigenvalues the matrix $-D_{12}f(x^*)(D_2^2 f(x^*))^{-1}D_{12}^\top f(x^*)$ has sufficiently positive eigenvalues so that the sum is positive. The sufficient conditions say that if $D_1^2 f(x^*) = W_1 M W_1^\dagger$ with $W_1 W_1^\dagger = I_{d_1 \times d_1}$ and M diagonal, and $-D_2^2 f(x^*) = W_2 \Lambda W_2^\dagger$ with $W_2 W_2^\dagger = I_{d_2 \times d_2}$ and Λ diagonal, then $D_{12}f(x^*)$ can be written as $W_1 \Sigma W_2^\dagger$ for some diagonal matrix $\Sigma \in \mathbb{R}^{d_1 \times d_2}$. Note that since Σ does not necessarily have positive values, $W_1 \Sigma W_2^\dagger$ is not the singular value decomposition of $D_{12}f(x^*)$. In turn, this means that each eigenvector of $D_1^2 f(x^*)$ gets mapped onto a single eigenvector of $-D_2^2 f(x^*)$ through the transformation $D_{12}f(x^*)$ which describes how player 1's variation $D_1 f(x)$ changes as a function of player 2's choice. With this structure for $D_{12}f(x^*)$, we can show that $D_1^2 f(x^*) - D_{21}f(x^*)^\top (D_2^2 f(x^*))^{-1} D_{21} f(x^*) \succ 0$. We also note that the condition depends on conditions that are difficult to check *a priori* without knowledge of x^*. On the other hand, the results are useful for the synthesis of games, such as in mechanism design where the goal is to drive agents to particular desirable behavior.

Realizable Generative Adversarial Networks. The 'realizable' assumption in the generative adversarial networks literature says the discriminator network is zero near an equilibrium parameter configuration (Nagarajan and Kolter, 2017). The assumption implies the Jacobian of $\dot{x} = -g(x)$ at critical points $x^* = (x_1^*, x_2^*)$ is such that $D_1^2 f(x^*) = 0$ and $D_{12}f(x^*)$ is full-rank. Under this assumption, we show stable critical points that are not differential Nash equilibria must be differential Stackelberg equilibria given $-D_2^2 f(x) \succ 0$.

Proposition 2.8. *Consider a zero-sum generative adversarial network satisfying the realizable assumption. Any stable critical point of $\dot{x} = -g(x)$ at which $-D_2^2 f(x) \succ 0$ is a differential Stackelberg equilibrium.*

Proof. Consider a stable critical point $x^* = (x_1^*, x_2^*)$ of $\dot{x} = -g(x)$ such that $-D_2^2 f(x^*) \succ 0$. Note that the realizable assumption implies that $D_1^2 f(x^*) = 0$ and $D_{12}f(x)$ is full-rank. Thus, the Jacobian of $g(x^*)$ at critical points under the realizable assumption is given by

$$J(x^*) = \begin{bmatrix} 0 & D_{12}f(x^*) \\ -D_{12}^\top f(x^*) & -D_2^2 f(x^*) \end{bmatrix}.$$

Thus, along with the fact $-D_2^2 f(x) \succ 0$, it immediately follows that

$$\mathtt{S}_1(J(x^*)) = -D_{12}f(x^*)(D_2^2 f(x^*))^{-1} D_{12}^\top f(x^*) \succ 0.$$

Thus, since both $-D_2^2 f(x) \succ 0$ and $\mathtt{S}_1(J(x)) \succ 0$, by the characterization in Proposition 2.1 and the fact from Lemma 2.1 that the critical points of $\dot{x} = -g(x)$ satisfy the first-order conditions, x^* is a differential Stackelberg equilibrium. □

This result shows that under standard theoretical assumptions on generative adversarial networks, any stable critical point of simultaneous gradient dynamics where the follower is at a local optimum is in fact a differential Stackelberg equilibrium. Thus, this implies that differential Stackelberg equilibrium play a key role in the optimization landscape of generative adversarial networks and also that they may be desirable solutions of the underlying machine learning problem.

Discussion of Results. The examples and results in this subsection imply some stable critical points of $\dot{x} = -g(x)$ that are not differential Nash equilibrium are in fact differential Stackelberg equilibrium. This is a meaningful set of examples and results since recent works have proposed schemes to avoid stable critical points of the simultaneous gradient dynamics that are not differential Nash equilibrium as they have been thought to lack game-theoretic meaning (Adolphs et al., 2019; Mazumdar et al., 2019). Moreover, some recent empirical studies show a number of successful approaches to training generative adversarial networks do not converge to differential Nash equilibrium, but rather to stable critical points of the dynamics at which the follower is at a local optimum (Berard et al., 2020). This may suggest reaching differential Stackelberg equilibrium is desirable in generative adversarial networks. The study in this subsection on the limit points of simultaneous gradient descent that are not differential Nash equilibrium is some of the primary motivation for the extensive study of this learning rule and the connections to differential Stackelberg equilibrium in Chapter 3.

2.3.3 Local Stability of Stackelberg Gradient Dynamics

To conclude this section, we study the local stability of the Stackelberg gradient dynamics around critical points in zero-sum games. In fact, let us consider a generalization of the Stackelberg gradient dynamics. Thus far, the continuous-time limiting systems that have been introduced correspond to systems in which players learn on identical timescales, that is, the discrete-time systems the continuous-time dynamics correspond to are such that the learning rates for each of the players are equal. Denote the learning rate of player 1 by $\gamma_1 > 0$ and let $\tau > 0$ be the "timescale separation" parameter such that the learning rate of player 2 is given by $\gamma_2 = \tau\gamma_1$ and $\tau = \gamma_2/\gamma_1$ is the ratio of learning rates. Moreover, define the vector field

$$g_{\mathcal{S}_\tau}(x) := (Df(x), \tau D_2 f_2(x)). \tag{2.9}$$

The continuous-time limiting system for the discrete-time Stackelberg gradient dynamics when player 1 has learning rate γ_1 and player 2 has learning rate $\gamma_2 = \tau\gamma_1$ is then given by

$$\dot{x} = -g_{\mathcal{S}_\tau}(x).$$

We refer to the system $\dot{x} = -g_{\mathcal{S}_\tau}(x)$ as the τ–Stackelberg gradient dynamics, or the Stackelberg gradient dynamics with timescale separation. The Jacobian of $g_{\mathcal{S}_\tau}(x)$ is denoted $J_{\mathcal{S}_\tau}(x)$; it is equivalent to $J_{\mathcal{S}}(x)$ as defined in (2.5) with the $d_2 \times d$ block row multiplied by the timescale separation parameter τ. Specifically,

$$J_{\mathcal{S}_\tau}(x) = \begin{bmatrix} D_1(Df_1(x)) & D_2(Df_1(x)) \\ \tau D_{21} f_2(x) & \tau D_2^2 f_2(x) \end{bmatrix}. \tag{2.10}$$

For the τ–Stackelberg gradient dynamics in zero-sum games, we show that for any $\tau \in (0, \infty)$, the set of stable critical points and the set of differential Stackelberg equilibrium coincide. Moreover, for generic zero-sum games, we show an equivalent statement holds in regards to local Stackelberg equilibrium. As the proof highlights, the result follows from the structure of the Jacobian $J_{\mathcal{S}_\tau}(x^*)$ at critical points of $\dot{x} = -g_{\mathcal{S}_\tau}(x^*)$ in zero-sum games, which is lower block triangular with $\mathtt{S}_1(J(x^*))$ and $-\tau D_2^2 f(x^*)$ in the diagonal blocks. Since the spectrum of a lower block triangular matrix is the

union of the spectrum of the diagonal blocks and these blocks precisely characterize the conditions for a differential Stackelberg equilibrium, the result immediately follows.

Theorem 2.5. *Consider a zero-sum game $(f, -f)$ with $f \in C^q(\mathcal{X}, \mathbb{R})$ for $q \geq 2$. Fixing any $\tau \in (0, \infty)$, a joint strategy $x^* = (x_1^*, x_2^*) \in \mathcal{X}$ is a stable critical point of $\dot{x} = -g_{\mathcal{S}_\tau}(x)$ if and only if x^* is a differential Stackelberg equilibrium. Moreover, if f is generic, fixing any $\tau \in (0, \infty)$, a joint strategy $x^* = (x_1^*, x_2^*) \in \mathcal{X}$ is a stable critical point of $\dot{x} = -g_{\mathcal{S}_\tau}(x)$ if and only if x^* is a a local Stackelberg equilibrium.*

Proof. For a zero-sum game $(f, -f)$, the Jacobian of the τ–Stackelberg gradient vector field $g_{\mathcal{S}_\tau}(x)$ at a critical point is given by

$$J_{\mathcal{S}_\tau}(x^*) = \begin{bmatrix} \mathtt{S}_1(J(x^*)) & 0 \\ -\tau D_{21} f(x^*) & -\tau D_2^2 f(x^*) \end{bmatrix}. \tag{2.11}$$

The structure of the Jacobian $J_{\mathcal{S}_\tau}(x^*)$ follows from the fact that $D_2 f(x^*) = 0$ at any critical point as a result of Lemma 2.1 and the fact that the timescale separation does not impact the set of critical points. Indeed, observe that since $D_2 f(x^*) = 0$, we have that

$$\begin{aligned} D_1(Df(x^*)) &= D_1\big(D_1 f(x^*) - D_{12} f(x^*)(D_2^2 f(x^*))^{-1} D_2 f(x^*)\big) \\ &= D_1^2 f(x^*) - D_{12} f(x^*)(D_2^2 f(x^*)^{-1} D_{21} f(x^*) \\ &= \mathtt{S}_1(J(x^*)). \end{aligned}$$

Similarly, since $D_2 f(x^*) = 0$, we have that

$$\begin{aligned} D_2(Df(x^*)) &= D_2\big(D_1 f(x^*) - D_{12} f(x^*)(D_2^2 f(x^*))^{-1} D_2 f(x^*)\big) \\ &= D_{12} f(x^*) - D_{12} f(x^*)(D_2^2 f(x^*))^{-1} D_2^2 f(x^*) \\ &= 0. \end{aligned}$$

The eigenvalues of a lower triangular block matrix are the union of the eigenvalues in each of the block diagonal components. This implies that the eigenvalues of $J_{\mathcal{S}_\tau}(x^*)$ are purely real and also that $\mathrm{spec}(J_{\mathcal{S}}(x^*)) \subset \mathbb{C}_+^\circ$ if and only if $\mathtt{S}_1(J(x^*)) \succ 0$ and $-D_2^2 f(x^*) \succ 0$. By the characterization in Proposition 2.1 and the fact that any critical point satisfies the first-order conditions (Lemma 2.1), this means that x^* is a stable critical point of the Stackelberg gradient dynamics if and only if x^* is a differential Stackelberg equilibrium.

Now, suppose that $f \in C^q(\mathcal{X}, \mathbb{R})$ is a generic function. Then, by Theorem 2.1 (genericity of differential Stackelberg equilibrium in zero-sum games), all differential Stackelberg equilibrium of $(f, -f)$ are local Stackelberg equilibrium so that the final statement of the theorem holds. □

This result implies that with appropriate choices of learning rates, the discrete-time τ–Stackelberg gradient dynamics will only locally converge to Stackelberg equilibrium. From the perspective of seeking local Stackelberg equilibrium, this is a highly desirable property. Moreover, from the viewpoint of the τ–Stackelberg gradient dynamics being a type of 'natural' learning dynamics that emulate the interaction structure of a Stackelberg game, it implies that in zero-sum games if players myopically update considering the roles they have in the game, then reaching a Stackelberg solution can be expected in zero-sum games. It is also worth highlighting the fact

shown in the proof that the Jacobian $J_{\mathcal{S}_\tau}(x^*)$ has purely real eigenvalues at critical points of the dynamics. The significance of this is that locally around critical points in zero-sum games, the dynamics should not admit cycling behavior. In contrast, a number of works have previously highlighted that a pitfall of the simultaneous gradient dynamics is the presence of imaginary eigenvalues in the Jacobian causing rotational forces in the dynamics. Given the desirable convergence implications of this stability result, we focus on analyzing the discrete-time convergence properties of the τ–Stackelberg gradient dynamics with deterministic and stochastic gradient information for the remainder of the theoretical study presented in this chapter.

2.4 Deterministic Convergence Analysis

In this section, we provide convergence guarantees for the discrete-time deterministic Stackelberg gradient dynamics, while in the following section we provide convergence guarantees for the discrete-time stochastic Stackelberg gradient dynamics. The deterministic discrete-time τ–Stackelberg dynamics are of the form

$$x_{k+1} = x_k - \gamma_1 g_{\mathcal{S}_\tau}(x_k). \tag{2.12}$$

Recall that $g_{\mathcal{S}_\tau}(x)$ is defined in (2.9) and the Jacobian $J_{\mathcal{S}_\tau}(x)$ of $g_{\mathcal{S}_\tau}(x)$ is defined in (2.10)

2.4.1 Local Asymptotic Convergence and Avoidance

To begin the convergence analysis of the discrete-time Stackelberg gradient dynamics, we provide results that mirror those from Section 2.3.3 for the continuous-time Stackelberg gradient dynamics.

Asymptotic Convergence. Theorem 2.5 shows that the set of stable critical points with respect to the the system $\dot{x} = -g_{\mathcal{S}_\tau}(x)$ coincides with the set of differential Stackelberg equilibrium in zero-sum games. This result also implies that the rest of the critical points excluding any marginally stable critical points, the set of which only contains critical points that are not differential Stackelberg equilibrium, are unstable with respect to the the system $\dot{x} = -g_{\mathcal{S}_\tau}(x)$ in zero-sum games. We now show a discrete-time analogue to Theorem 2.5 with equivalent implications in zero-sum games.

Proposition 2.9. *Consider a zero-sum game $(f, -f)$ defined by $f \in C^q(\mathcal{X}, \mathbb{R})$ with $q \geq 2$. Fix any $\tau \in (0, \infty)$ and assume that $\sup_{x\in\mathcal{X}} \|J_{\mathcal{S}_\tau}(x)\| \leq L < \infty$. Then, any joint strategy $x^* = (x_1^*, x_2^*)$ is a stable critical point of the τ–Stackelberg gradient dynamics with $\gamma_1 < 1/L$ if and only if x^* is a differential Stackelberg equilibrium.*

Proof. The proof of Theorem 2.5 shows that at any fixed critical point $x^* = (x_1^*, x_2^*)$ of the vector field $g_{\mathcal{S}_\tau}(x)$, the spectrum of the Jacobian is given by $\text{spec}(J_{\mathcal{S}_\tau}(x^*)) = \text{spec}(\mathtt{S}_1(J(x^*))) \cup \text{spec}(-\tau D_2^2 f(x^*)) \subset \mathbb{R}$. Thus, considering any given critical point $x^* = (x_1^*, x_2^*)$, for all $\lambda \in \text{spec}(J_{\mathcal{S}_\tau}(x^*))$, we have

$$\lambda \leq \max\{\text{spec}(J_{\mathcal{S}_\tau}(x^*))\} \leq \rho(J_{\mathcal{S}_\tau}(x^*)) \leq \|J_{\mathcal{S}_\tau}(x^*)\| \leq \sup_{x\in\mathcal{X}} \|J_{\mathcal{S}_\tau}(x)\| \leq L.$$

Hence, it follows that with $\gamma_1 < 1/L$,

$$\rho(I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)) = \max_{\lambda \in \text{spec}(J_{\mathcal{S}_\tau}(x^*))} |1 - \gamma_1 \lambda| < 1$$

if and only if $\mathtt{S}_1(J(x^*)) \succ 0$ and $-D_2^2 f(x^*) \succ 0$. This implies that x^* is a stable critical point of the τ–Stackelberg gradient dynamics with $\gamma_1 < 1/L$ if and only if x^* is a differential Stackelberg equilibrium. This follows from the fact that x^* is stable with respect to the discrete-time τ–Stackelberg gradient dynamics if and only if $\rho(I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)) < 1$ and x^* is a differential Stackelberg equilibrium if and only if $\mathtt{S}_1(J(x^*)) \succ 0$ and $-D_2^2 f(x^*) \succ 0$ by Lemma 2.1 and Proposition 2.1. Since this statement holds identically for all critical points, the final result holds. □

The previous result implies that with an appropriate choice of learning rate, the discrete-time τ–Stackelberg gradient dynamics locally exponentially converge around any differential Stackelberg equilibrium in zero-sum games (Argyros, 1999; Ortega and Rheinboldt, 1970). Shortly, in Section 2.4.2, we explain this further and explicitly characterize convergence rates with deterministic gradient feedback. Before doing so, we give a complimentary asymptotic avoidance result.

Asymptotic Avoidance. Proposition 2.9 also indicates that in zero-sum games all critical points that are not differential Stackelberg equilibrium are either marginally stable or they are unstable. Let us recall the definition of instability as described in Section 2.3.1 and discuss the implications. Informally, if x^* is an unstable critical point, then there is an initial condition from any arbitrarily small neighborhood from which the trajectory of the dynamics leaves a neighborhood of the critical point. The subtle, yet important implications of this definition are that if a critical point is unstable, it does not imply that from all initial conditions the trajectory diverges away from the critical point, nor does it imply that there cannot exist an initial condition in a local neighborhood from which the dynamics converge to the critical point.

That begin said, the following result shows that the discrete-time τ–Stackelberg gradient dynamics asymptotically avoid strict saddle points of the continuous-time τ–Stackelberg gradient dynamics almost surely in general-sum games.[7] At this juncture, it is important to remark that saddle points with respect to the dynamics should not be conflated with a saddle point of the cost functions. To provide some intuition for the result, strict saddles have the property of being completely characterized by stable and unstable manifolds. Initializations on the stable manifold asymptotically converge, while initializations on the unstable manifold asymptotically diverge. Thus, by showing that the stable manifold has measure zero, the result follows.

Theorem 2.6 (Almost Sure Avoidance of Saddles)**.** *Consider a general-sum game defined by $f_i \in C^q(\mathcal{X}, \mathbb{R})$, $q \geq 2$ for $i \in \mathcal{I}$. Fix any $\tau \in (0, \infty)$ and assume that $\sup_{x \in \mathcal{X}} \|J_{\mathcal{S}_\tau}(x)\| \leq L < \infty$. Then, the discrete-time τ–Stackelberg gradient dynamics converge to strict saddle points of $\dot{x} = -g_{\mathcal{S}_\tau}(x)$ on a set of measure zero.*

The proof of Theorem 2.6 is deferred to Section 2.D.3. The technical approach mirrors closely the arguments from the proof that the simultaneous gradient dynamics avoid strict saddles of the

[7]Recall that we refer to strict saddles of the system $\dot{x} = -g_{\mathcal{S}_\tau}(x)$ as hyperbolic critical points $x^* = (x_1^*, x_2^*)$ at which $J_{\mathcal{S}_\tau}(x^*)$ has at least one eigenvalue in $\mathbb{C}_-^\circ$. In the terminology that has been used in Theorem 2.4, this means strict saddles are hyperbolic unstable critical points.

dynamics (Mazumdar et al., 2020, Theorem 12), which itself builds on similar results from single player optimization problems (Lee et al., 2016; Panageas and Piliouras, 2017). In particular, we show that $g_{\mathcal{S}_\tau}(x)$ is a diffeomorphism, and then apply the stable manifold theorem (Sastry, 1999).

Discussion of Results. We remark that differential Stackelberg equilibrium are never strict saddle points in zero-sum games as a result of Theorem 2.5. Thus together, Proposition 2.9 and Theorem 2.5 imply in zero-sum games that if all saddle points of τ–Stackelberg gradient dynamics are strict, $\sup_{x\in\mathcal{X}} \|J_{\mathcal{S}_\tau}(x)\| \leq L < \infty$, and $\gamma_1 < 1/L$, then if the τ–Stackelberg gradient dynamics converge to a critical point it must be that the critical point is a differential Stackelberg equilibrium almost surely. Observe that the if statement regarding the τ–Stackelberg gradient dynamics converging to a critical point has been made since the dynamics do not correspond to a gradient flow, so there may existing non-trivial limiting behaviors that emerge beyond convergence to a critical point.

For zero-sum games, we also remark that for a particular critical point $x^* = (x_1^*, x_2^*)$, weaker assumptions can be made than presented in Proposition 2.9 such that x^* is stable if and only if it is a differential Stackelberg equilibrium. Specifically, as long as $\gamma_1 < 1/L'$ where $L' = \max\{\text{spec}(J_{\mathcal{S}_\tau}(x^*))\} = \max\{\text{spec}(\mathtt{S}_1(J(x^*))) \cup \text{spec}(-\tau D^2 f(x^*))\}$, then x^* is stable a critical point if and only if it is a differential Stackelberg equilibrium. This property can be observed the proof of Theorem 2.5 and the proof of Proposition 2.9. Proposition 2.9 is stated in the manner it is presented so that there is a fixed learning rate which guarantees the if and only if statement for all critical points simultaneously.

It is also worth noting that in general-sum games, it may be that differential Stackelberg equilibrium are strict saddle points of the Stackelberg gradient dynamics. So the result on escaping saddles can in fact imply that the τ–Stackelberg gradient dynamics avoid some differential Stackelberg equilibrium in general-sum games.

2.4.2 Convergence Rates

To derive convergence rates for the Stackelberg gradient dynamics, we apply well-known results regarding discrete-time dynamical systems. As mentioned in Section 2.3.1, given a dynamical system of the form $x_{k+1} = F(x_k)$, when the spectral radius of the Jacobian $J_F(x^*)$ at critical point x^* is such that $\rho(J_F(x^*)) < 1$, then the operator F is a contraction at x^* so that x^* is locally asymptotically stable (Argyros, 1999; Ortega and Rheinboldt, 1970; Sastry, 1999). Moreover, an asymptotic rate of convergence can be derived to show that x^* is locally exponentially stable.[8]

Indeed, given $\rho(J_F(x^*)) < 1$, there exists $\kappa > 0$ such that $\rho(J_F(x^*)) \leq \kappa < 1$. Moreover, there exists a matrix norm $\|\cdot\|$ such that $\|J_F(x^*)\| \leq \rho(J_F(x^*)) + \varepsilon \leq \kappa + \varepsilon$ for any $\varepsilon > 0$ (Horn and Johnson, 2012; Ortega and Rheinboldt, 1970, Lemma 5.6.10, 2.2.8). It then follows from a Taylor expansion that there exists a ball $B_\delta(x^*)$ of radius $\delta > 0$ such that for any $x_0 \in B_\delta(x^*)$, and some constant $K > 0$, $\|x_k - x^*\| \leq K(\kappa + 2\varepsilon)^k \|x_0 - x^*\|$ given that $\varepsilon > 0$ is chosen such that $\kappa + 2\varepsilon < 1$. Given two real valued functions $F(k)$ and $G(k)$, we write $F(k) = \mathcal{O}(G(k))$ if there exists a positive constant $c > 0$ such that $|F(k)| \leq c|G(k)|$. For example, if $F(k) = \|x_k - x^*\| \leq M^k \|x_0 - x^*\|$, we write $F(k) = \mathcal{O}(M^k)$ where $c = \|x_0 - x^*\|$.

---

[8] We also refer the reader to Ortega and Rheinboldt (1970, Chapter 10) for background on this topic.

Given an asymptotic rate of convergence, a finite-time iteration complexity can near immediately be derived from an initial condition $x_0 \in B_\delta(x^*)$ to reaching $x_k \in B_\varepsilon(x^*)$ along with a characterization of the neighborhood size on which this holds. Note that in what follows, we call x_k an ε–differential Stackelberg equilibrium when $x_k \in B_\varepsilon(x^*)$. The analysis methods that have been described give rise to the local convergence rates for the deterministic discrete-time Stackelberg gradient dynamics presented in this section. The formal proofs of the results stated in this section can be found in Section 2.D.

Zero-Sum Convergence Rates. We begin by stating convergence rates in zero-sum games to differential Stackelberg equilibrium. As shown in the proof of Theorem 2.5, at critical points $x^* = (x_1^*, x_2^*)$ of the τ–Stackelberg gradient dynamics in zero-sum games, the Jacobian $J_{\mathcal{S}_\tau}(x^*)$ has a lower-block triangular structure so that $\text{spec}(J_{\mathcal{S}_\tau}(x^*)) = \text{spec}(\mathtt{S}_1(J(x^*))) \cup \text{spec}(-\tau D_2^2 f(x^*)) \subset \mathbb{R}$. Given that x^* is a differential Stackelberg equilibrium, this implies that $\text{spec}(J_{\mathcal{S}_\tau}(x^*)) \subset \mathbb{R}_+$. Combining this fact with the analysis methods outlined earlier in this section and optimizing the choice of learning rate to maximize the rate of convergence, we obtain the following result.

Theorem 2.7 (Zero-Sum Rate of Convergence.)**.** *Consider a zero-sum game defined by $f \in C^q(\mathcal{X}, \mathbb{R})$ with $q \geq 2$. For a differential Stackelberg equilibrium $x^* = (x_1^*, x_2^*)$, define the parameters $\alpha = \min\{\lambda_{\min}(\mathtt{S}_1(J(x^*))), \lambda_{\min}(-\tau D_2^2 f(x^*))\}$ and $\beta = \max\{\lambda_{\max}(\mathtt{S}_1(J(x^*))), \lambda_{\max}(-\tau D_2^2 f(x^*))\}$ where $\tau \in (0, \infty)$ is fixed. Then, with learning rate $\gamma_1 = 1/(2\beta)$, the τ–Stackelberg gradient dynamics locally asymptotically converge to x^* with a rate of $\mathcal{O}((1 - \frac{\alpha}{4\beta})^k)$.*

The proof of Theorem 2.7 can be found in Section 2.D.1.1. Observe that the convergence rate is defined in terms of a very intuitive and fundamental quantity. Specifically, considering $\tau = 1$, the rate depends on the ratio of the minimum and maximum eigenvalues of the union of the spectrums of $\mathtt{S}_1(J(x^*))$ and $-D_2^2 f(x^*)$. In zero-sum games, these quantities define the sufficient conditions for a differential Stackelberg equilibrium. Thus, the result is illustrating that when the equilibrium is well-conditioned in terms of the sufficient conditions, then the converge rate is faster. We also remark that the potential benefits of timescale separation via the parameter τ can be seen from this result. In particular, when $\lambda_{\min}(-D_2^2 f(x^*)) < \lambda_{\min}(\mathtt{S}_1(J(x^*)))$ and $\lambda_{\max}(-D_2^2 f(x^*)) < \lambda_{\max}(\mathtt{S}_1(J(x^*)))$, the timescale separation τ can scale $\lambda_{\max}(-D_2^2 f(x^*))$ to improve the rate.

The asymptotic convergence rate can be translated to a finite-time local convergence guarantee. Moreover, an estimate of the neighborhood on which the convergence holds can be derived. The following corollary gives these results that follow easily from Theorem 2.7 and its proof.

Corollary 2.1 (Zero-Sum Finite Time Guarantee)**.** *Given $\varepsilon > 0$, under the assumptions of Theorem 2.7, τ–Stackelberg gradient dynamics obtain an ε–differential Stackelberg equilibrium in $\lceil \frac{4\beta}{\alpha} \log(\|x_0 - x^*\|/\varepsilon) \rceil$ iterations for any $x_0 \in B_\delta(x^*)$ with $\delta = \alpha/(4L\beta)$ where L is the local Lipschitz constant of $I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)$.*

The proof of Corollary 2.1 can be found in Section 2.D.1.2. Similar conclusions can be drawn from this result as from Theorem 2.7. However, it is worth noting that this result illustrates that the region of attraction around the equilibrium depends on the timescale separation parameter τ in the Stackelberg gradient dynamics. In an analogous fashion as was described for the converge rate, scaling τ can increase the size of the region of attraction around the equilibrium.

General-Sum Convergence Rates. We now state convergence rates in general-sum games to differential Stackelberg equilibrium. Since there may be stable critical points in general-sum games that are not differential Stackelberg equilibrium in general-sum games, the guarantees are for stable differential Stackelberg equilibrium whereas the results for zero-sum games apply to any differential Stackelberg equilibrium since all differential Stackelberg equilibrium are stable with respect to the τ–Stackelberg gradient dynamics. Moreover, since the Jacobian $J_{\mathcal{S}_\tau}(x^*)$ at critical points $x^* = (x_1^*, x_2^*)$ does not decompose into a lower block triangular structure, the convergence rates depend on quantities that do not have as natural of interpretations. That being said, using the analysis methods presented at the beginning of this section, we can derive the following rates of convergence to stable differential Stackelberg equilibrium.

Theorem 2.8 (General-Sum Rate of Convergence)**.** *Consider a general sum game (f_1, f_2) with $f_i \in C^q(\mathcal{X}, \mathbb{R})$ for $q \geq 2$ and $i \in \mathcal{I}$. For a differential Stackelberg equilibrium $x^* = (x_1^*, x_2^*)$ such that $J_{\mathcal{S}_\tau}^\top(x^*) + J_{\mathcal{S}_\tau}(x^*) \succ 0$, define the parameters $\alpha = \lambda_{\min}^2(\frac{1}{2}(J_{\mathcal{S}_\tau}^\top(x^*) + J_{\mathcal{S}_\tau}(x^*)))$ and $\beta = \lambda_{\max}(J_{\mathcal{S}_\tau}(x^*)^\top J_{\mathcal{S}_\tau}(x^*))$ where $\tau \in (0, \infty)$ is fixed. Then, with learning rate $\gamma_1 = \sqrt{\alpha}/\beta$, the τ–Stackelberg gradient dynamics locally asymptotically converge to x^* with a rate of $\mathcal{O}((1 - \frac{\alpha}{2\beta})^{k/2})$.*

The proof of Theorem 2.8 can be found in Section 2.D.2.1. Analogous to the zero-sum setting, the asymptotic rate of convergence can near-directly be translated into a finite-time rate of convergence and an estimate of the region of attraction can be obtained.

Corollary 2.2 (General Sum Finite Time Guarantee)**.** *Given $\varepsilon > 0$, under the assumptions of Theorem 2.8, τ–Stackelberg learning obtains an ε–differential Stackelberg equilibrium in $\lceil \frac{4\beta}{\alpha} \log(\|x_0 - x^*\|/\varepsilon) \rceil$ iterations for any $x_0 \in B_\delta(x^*)$ with $\delta = \alpha/(2L\beta)$ where L is the local Lipschitz constant of $I - \gamma_1 J_{\mathcal{S}_\tau}(x)$.*

The proof of Corollary 2.2 is provided in Section 2.D.2.2

2.5 Stochastic Convergence Results

This section presents stochastic convergence results for the Stackelberg gradient dynamics. Specifically, in Section 2.5.1 we give results for the τ-Stackelberg gradient dynamics when players act on a single timescale, in Section 2.5.2 we give results for the scenario that the follower plays a best-response at each step, and in Section 2.5.3 we give results for the situation the players update on two-timescales. The results and proofs rely on what is known as the ordinary differential equation method, in which the flow of the limiting continuous time system starting at sample points from the stochastic updates of the players actions is compared to asymptotic pseudo-trajectories; that is, linear interpolations between sample points. By and large, we are able to apply known stochastic approximation results to obtain the results that are presented. The key distinction is the manner in which the limiting points of the stochastic dynamics are connected to the Stackelberg equilibrium concept. So far in studying deterministic dynamics, the results have relied heavily on stability analysis and the stochastic case is similar in this regard. Thus, before moving on to the results, we now present a suitable notion of stability for stochastic dynamics.

Stochastic Stability: Internally Chain Transitive Sets. To understand stability in the stochastic case, we need the notion of internally chain transitive sets. For more detail, the reader is referred to Alongi and Nelson (2007, Chapters 2–3). A closed set $U \subset \mathbb{R}^d$ is an invariant set for an ordinary differential equation $\dot{x} = -F(x)$ if any trajectory $x(t)$ with $x(0) \in U$ satisfies $x(t) \in U$ for all $t \in \mathbb{R}$. Let ϕ^t be a flow on a metric space $(\mathcal{X}, d)$. Given $\varepsilon > 0$, $T > 0$ and $x, y \in \mathcal{X}$, an (ε, T)-chain from x to y with respect to ϕ^t and d is a pair of finite sequences $x = x_0, x_1, \ldots, x_{k-1}, x_k = y$ in $\mathcal{X}$ and $t_0, \ldots, t_{k-1}$ in $[T, \infty)$, denoted together by $(x_0, x_1, \ldots, x_{k-1}, x_k; t_0, \ldots, t_{k-1})$, such that $d(\phi^{t_i}(x_i), x_{i+1}) < \varepsilon$ for $i = 0, 1, 2, \ldots, k-1$. A set $U \subseteq X$ is *(internally) chain transitive* with respect to ϕ^t if U is a non-empty closed invariant set with respect to ϕ^t such that for each $x, y \in U$, $\epsilon > 0$ and $T > 0$ there exists an (ε, T)-chain from x to y. A compact invariant set U is invariantly connected if it cannot be decomposed into two disjoint closed nonempty invariant sets. Note that every internally chain transitive set is invariantly connected.

2.5.1 Single-Timescale Analysis

In this subsection, we study the single-timescale stochastic τ–Stackelberg gradient dynamics. Specifically, the stochastic form of the update is given by

$$x_{k+1} = x_k - \gamma_k(g_{\mathcal{S}_\tau}(x_k) + w_{k+1}) \tag{2.13}$$

where $\{w_{k+1}\}$ is a stochastic noise process and $\{\gamma_k\}$ is the learning rate sequence. The limiting ordinary differential equation that (2.13) can be expected to track asymptotically is the τ–Stackelberg gradient dynamic system in continuous-time given by

$$\dot{x} = -g_{\mathcal{S}_\tau}(x).$$

For the results in this subsection, we make the following standard stochastic approximation assumptions (see, e.g., Borkar 2008).

Assumption 2.1. *The following hold:*

2.1a. The map $g_{\mathcal{S}_\tau} : \mathbb{R}^d \to \mathbb{R}^d$ is L-Lipschitz for any $\tau \in (0, \infty)$.

2.1b. The learning rate sequence satisfies $\sum_k \gamma_k = \infty$, $\sum_k \gamma_k^2 < \infty$.

2.1c. The stochastic process $\{w_k\}$ is a martingale difference sequence with respect to the increasing family of σ-fields defined by $\mathcal{F}_k = \sigma(x_\ell, w_\ell, \ell \leq k)$, $\forall\ k \geq 0$, so that $\mathbb{E}[w_{k+1} |\ \mathcal{F}_k] = 0$ almost surely for all $k \geq 0$. Moreover, for some constant $C > 0$, $\mathbb{E}[\|w_{k+1}\|^2 |\ \mathcal{F}_k] \leq C(1 + \|x_k\|^2)$ almost surely, $\forall\ k \geq 0$.

2.1d. The iterates $\{x_k\}$ of (2.13) remain bounded almost surely: that is, $\sup_k \|x_k\| < \infty$.

By viewing the combined system of the players as a single-timescale stochastic approximation update, we can near immediately invoke known results to obtain meaningful stochastic convergence and non-convergence results, specifically when put together with the local stability results for the τ–Stackelberg gradient dynamics from Section 2.3.3.

Asymptotic Convergence. We begin by presenting almost sure asymptotic convergence result for sequence $\{x_k\}$ generated by the stochastic dynamics in (2.13) under Assumption 2.1. Specifically, classical stochastic approximation results using the so-called ordinary differential equation

method immediately imply that in general-sum games, the sequence $\{x_k\}$ generated by the stochastic dynamics in (2.13) converge to a possibly sample path-dependent, compact connected internally chain transitive invariant set of $\dot{x} = -g_{\mathcal{S}_\tau}(x)$. To narrow down this set (which may include limiting behaviors beyond critical points) often requires producing a local or global Lyapunov function. In general, constructing a Lyapunov function can be challenging. However, given that a critical point $x^* = (x_1^*, x_2^*)$ is locally exponentially stable with respect to $\dot{x} = -g_{\mathcal{S}_\tau}(x)$, converse Lyapunov results imply the existence of a local Lyapunov function that can be used to conclude local asymptotic convergence of the sequence $\{x_k\}$ to x^* almost surely. Thus, for any differential Stackelberg equilibrium $x^* = (x_1^*, x_2^*)$ such that $\text{spec}(-J_{\mathcal{S}_\tau}(x^*)) \subset \mathbb{C}_-^\circ$, these techniques give rise to local asymptotic guarantees for the sequence $\{x_k\}$ converging to x^* almost surely. In zero-sum games, the result can be strengthened using the result of Theorem 2.5. Specifically, since $\text{spec}(-J_{\mathcal{S}_\tau}(x^*)) \subset \mathbb{C}_-^\circ$ holds for all differential Stackelberg equilibrium $x^* = (x_1^*, x_2^*)$ by Theorem 2.5, we can conclude local asymptotic convergence for any differential Stackelberg equilibrium almost surely, without explicitly assuming stability. This discussion leads to the following result.

Theorem 2.9. *Consider a general-sum game (f_1, f_2) such that $f_i \in C^q(\mathcal{X}, \mathbb{R})$ for some $q \geq 2$ and each $i \in \mathcal{I}$. Fix any $\tau \in (0, \infty)$ and suppose that Assumptions 2.1 holds. Then, the sequence $\{x_k\}$ generated by* (2.13) *converges to a, possibly sample path-dependent, compact connected internally chain transitive invariant set of $\dot{x} = -g_{\mathcal{S}_\tau}(x)$. Moreover, given the game is zero-sum so that $(f_1, f_2) \equiv (f, -f)$, if $x^* = (x_1^*, x_2^*)$ is a differential Stackelberg equilibrium, then $\{x_k\}$ almost surely locally asymptotically converges to x^*.*

Proof. The convergence of $\{x_k\}$ to a, possibly sample path dependent, compact connected internally chain transitive invariant set of $\dot{x} = -g_{\mathcal{S}_\tau}(x)$ is immediate given the assumptions from classical results in stochastic approximation theory (Borkar, 2008, Chapter 2, Theorem 2); (Benaim, 1996).

Suppose that $x^* = (x_1^*, x_2^*)$ is a differential Stackelberg equilibrium. By Theorem 2.5, x^* is a locally exponentially stable equilibrium of the continuous time dynamics $\dot{x} = -g_{\mathcal{S}_\tau}(x)$, that is, $\text{spec}(-J_{\mathcal{S}_\tau}(x^*)) \subset \mathbb{C}_-^\circ$. Since $\text{spec}(-J_{\mathcal{S}_\tau}(x^*)) \subset \mathbb{C}_-^\circ$, $\det(-J_\tau(x^*)) \neq 0$ so that x^* is an isolated critical point. Furthermore, exponentially stability of x^* implies that there exists a (local) Lyapunov function defined on a neighborhood of x^* by the converse Lyapunov theorem (Krasovskii, 1963; Sastry, 1999, Theorem 4.3,Theorem 5.17). Let U be the neighborhood of x^* on which the local Lyapunov function is defined, such that U contains no other critical points (which is possible since x^* is isolated). That is, let $\Phi : U \to [0, \infty)$ be the local Lyapunov function defined on U where $x^* \in U$, Φ is positive definite on U, and for all $x \in U$, $\frac{d}{dt}\Phi(x) \leq 0$ where equality holds for $z \in U$ if and only if $\Phi(z) = 0$. By Corollary 3 (Borkar, 2008, Chapter 2), $\{x_k\}$ converges to an internally chain transitive invariant set contained in U almost surely. The only internally chain transitive invariant set in U is x^*. □

The result of Theorem 2.9 for zero-sum games can be seen as a stochastic analogue of the deterministic result of Proposition 2.9. We again remark that following identical arguments, in general-sum games, given a differential Stackelberg equilibrium $x^* = (x_1^*, x_2^*)$ such that $\text{spec}(-J_{\mathcal{S}_\tau}(x^*)) \subset \mathbb{C}_-^\circ$, we can also ensure local asymptotic converge for the sequence $\{x_k\}$ to x^* almost surely. However, this requires assuming stability instead of stability being a property that is guaranteed to be satisfied. It is also worth noting that the result of Theorem 2.9 can be stated with relaxed assumptions. In particular, suppose that the set $\{\sup_k \|x_k\| < \infty\}$ has positive probability but it

does not hold almost surely, then the result can be stated without Assumption 2.1d. but amended to hold almost surely on the event $\{\sup_k \|x_k\| < \infty\}$ (Borkar, 2008, Chapter 2). This will also be the case for stochastic convergence results that appear in the following subsections. As a final note, it is possible to obtain concentration bounds and even finite time, high probability guarantees on convergence leveraging recent advances in stochastic approximation (Borkar, 2008; Borkar and Pattathil, 2018; Kamal, 2010; Thoppe and Borkar, 2019). We do not pursue such results in this chapter for the stochastic τ–Stackelberg gradient dynamics, but do give results of this nature in the Section 3.6.2 of the chapter that follows for the simultaneous gradient dynamics.

Asymptotic Avoidance. To contrast with the convergence result asymptotic convergence result, we now focus on developing a non-convergence guarantee. That is, we show that the sequence $\{x_k\}$ generated by the stochastic dynamics from (2.13) asymptotically avoid strict saddle points of the system $\dot{x} = -g_{\mathcal{S}_\tau}(x)$ almost surely. This result follows near-immediately from a classical results from Pemantle (1990). Consider a general stochastic approximation framework $x_{k+1} = x_k + \gamma_k F(x_k) + w_k$ for $F : \mathcal{X} \to T\mathcal{X}$ with $F \in C^2$ and where $\mathcal{X} \subset \mathbb{R}^d$ and where $T\mathcal{X}$ denotes the tangent space of $\mathcal{X}$.

Theorem 2.10 (Theorem 1 Pemantle 1990)**.** *Suppose γ_k is $\mathcal{F}_k$–measurable and $\mathbb{E}[w_k|\mathcal{F}_k] = 0$. Let the stochastic process $\{x_k\}_{k\geq 0}$ be defined as above for some sequence of random variables $\{w_k\}$ and $\{\gamma_k\}$. Let $x^* \in \mathcal{X}$ with $F(x^*) = 0$ and let W be a neighborhood of x^*. Assume that there are constants $\eta \in (1/2, 1]$ and $c_1, c_2, c_3, c_4 > 0$ for which the following conditions are satisfied whenever $x_k \in W$ and k sufficiently large: (i) x^* is a hyperbolic unstable critical point (strict saddle), (ii) $c_1/k^\eta \leq \gamma_k \leq c_2/k^\eta$, (iii) $\mathbb{E}[\max\{w_k \cdot v, 0\}|\mathcal{F}_k] \geq c_3/k^\eta$ for every unit vector $v \in T\mathcal{X}$, and (iv) $\|w_k\|_2 \leq c_4/k^\eta$. Then, $\mathbb{P}(x_k \to x^*) = 0$.*

The above classical result directly implies avoidance of strict saddles by the stochastic Stackelberg gradient dynamics, given the proper assumptions. The assumption that $\mathbb{E}[(w_{i,t} \cdot v)^+|\mathcal{F}_{i,t}] \geq b_i$ essentially requires the covariance of the noise to be full-rank, and is made to rule out degenerate cases where the noise forces the dynamics to stay on the stable manifold of strict saddle points.

Theorem 2.11 (Almost Sure Avoidance of Saddles.)**.** *Consider a general-sum game (f_1, f_2) with $f_i \in C^q(\mathcal{X}, \mathbb{R})$ for $q \geq 2$ and $i \in \mathcal{I}$. Given that Assumption 2.1 holds and there exists a constant $b > 0$ such that $\mathbb{E}[(w_{k+1} \cdot v)^+|\mathcal{F}_k] \geq b$ for every unit vector $v \in \mathbb{R}^d$, the sequence $\{x_k\}$ generated by (2.13) with any $\tau \in (0, \infty)$ converges to any strict saddle $x^* = (x_1^*, x_2^*)$ of the system $\dot{x} = -g_{\mathcal{S}_\tau}(x)$ on a set of measure zero.*

It is worth noting that an analogous result has been stated for the stochastic simultaneous gradient dynamics (Mazumdar et al., 2020). In fact, Mazumdar et al. (2020) invoke results of Benaim and Hirsch (1995) to state more generally almost sure avoidance of unstable cycles. With regard to this result, there are a couple of comments worth making. As pointed out by (Pemantle, 1990), this result is primarily useful when combined with a corresponding almost sure convergence result, such as Theorem 2.9. Moreover, the significance of this result hinges on the the set of strict saddles being discrete, so that the implication is that there is almost sure avoidance of all strict saddle points. Together, Theorems 2.9 and 2.11 yield similar conclusions in the single timescale stochastic setting as the results in the deterministic setting from Section 2.4.1.

2.5.2 Best-Response Analysis

We now study a variant of the τ-Stackelberg gradient dynamics in which the follower is actually playing a best-response at each step of the leader. Indeed, suppose that given the action $x_{1,k}$ of the leader, the action of the follower satisfies $x_{2,k} \in \arg\min_{x_2 \in \mathcal{X}_2} f_2(x_{1,k}, x_2)$ where the best-response is assumed to be unique for any $x_{1,k} \in \mathcal{X}_1$. Then, under the stated assumptions, there exists an implicit map $r : x_1 \mapsto x_2$ defined on a neighborhood such that $x_{2,k} = r(x_{1,k})$. Thus, the following stochastic dynamics for the leader can be defined:

$$x_{1,k+1} = x_{1,k} - \gamma_{1,k}(Df_1(x_{1,k}, x_{2,k})) + w_{1,k+1}) \tag{2.14}$$

where $x_{2,k}$ is defined via the best-reponse map $r : x_1 \mapsto x_2$ defined implicitly in a neighborhood of $(x_{1,k}, x_{2,k})$ and $\{w_{1,k+1}\}$ is a stochastic noise process and $\{\gamma_{1,k}\}$ is the learning rate sequence. The limiting ordinary differential equation that (2.14) can be expected to track asymptotically is then given by

$$\dot{x}_1 = -D(x_1, r(x_1)). \tag{2.15}$$

Observe that this is a single timescale system defined only in terms of the variable of player 1.

Let us now state the assumptions we make to analyze (2.14). To begin, we again need standard stochastic approximation assumptions (see, e.g., Borkar 2008).

Assumption 2.2. *The following hold:*

2.2a. *The map $Df_1 : \mathbb{R}^d \to \mathbb{R}^{d_1}$, $D_2 f_2 : \mathbb{R}^d \to \mathbb{R}^{d_2}$ are L_1, L_2 Lipschitz, respectively.*

2.2b. *The learning rate sequence of player 1 satisfies $\sum_k \gamma_{1,k} = \infty$, $\sum_k \gamma_{1,k}^2 < \infty$.*

2.2c. *The stochastic process $\{w_{1,k}\}$ is a martingale difference sequence with respect to the increasing family of σ-fields defined by $\mathcal{F}_k = \sigma(x_\ell, w_{i,\ell}, \ell \leq k)$, $\forall\ k \geq 0$, so that $\mathbb{E}[w_{1,k+1}|\ \mathcal{F}_k] = 0$ almost surely for all $k \geq 0$. Moreover, for some constant $C > 0$, $\mathbb{E}[\|w_{1,k+1}\|^2|\ \mathcal{F}_k] \leq C(1 + \|x_{1,k}\|^2)$ almost surely, $\forall\ k \geq 0$.*

2.2d. *The iterates $\{x_k\}$ of (2.14) remain bounded almost surely: that is, $\sup_k \|x_k\| < \infty$.*

The following assumption ensures that the follower has a unique best-response to play in reaction to any action of the leader and guarantees the implicit function mapping the leader action to the follower action is well-defined on the domain convergence is being assessed.

Assumption 2.3. *For every x_1, $\dot{x}_2 = -D_2 f_2(x_1, x_2)$ has a globally asymptotically stable critical point $r(x_1)$ uniformly in x_1 and $r : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ is L_r–Lipschitz.*

The final assumption is made so that stronger statements can be made about the internally chain transitive sets (2.14) converges toward.

Assumption 2.4. *All critical points of the system $\dot{x}_1 = -D(x_1, r(x_1))$ are isolated.*

Given Assumptions 2.2–2.4, classical stochastic approximation results imply that in general-sum games, the sequence $\{x_{1,k}\}$ generated by the stochastic dynamics in (2.13) converge to a possibly sample path-dependent, compact connected internally chain transitive invariant set of $\dot{x}_1 = -D(x_1, r(x_1))$. Since this system corresponds to a gradient-flow and all critical points are assumed to be isolated, the result implies almost sure convergence of the sequence $\{x_{1,k}\}$ to a

possibly sample path dependent-critical point contained in the internally chain transitive invariant sets. Then, by recognizing that the only isolated critical points contained in the internally chain transitive invariant sets are differential Stackelberg equilibrium solutions for the leader, we obtain the following result.

Proposition 2.10. *Consider a general-sum game (f_1, f_2) such that $f_i \in C^q(\mathcal{X}, \mathbb{R})$ for some $q \geq 2$ and each $i \in \mathcal{I}$. Suppose Assumptions 2.2–2.4 hold. Then, the sequence $\{x_k\}$ generated by (2.14) converges to a, possibly sample path-dependent, differential Stackelberg equilibrium almost surely.*

Proof. The convergence of $\{x_{1,k}\}$ to a, possibly sample path dependent, compact connected internally chain transitive invariant set of $\dot{x}_1 = -Df_1(x_1, r(x_1))$ is immediate given the assumptions from classical results in stochastic approximation theory (Borkar, 2008, Chapter 2, Theorem 2); (Benaim, 1996). Furthermore, since $\dot{x}_1 = -Df_1(x_1, r(x_1))$ corresponds to a gradient-flow and all critical points are assumed to be isolated, the only internally chain transitive invariant sets are stable critical points. That is x_1^* such that $Df_1(x_1^*, r(x_1^*)) = 0$ and $D^2 f_1(x_1^*, r(x_1^*)) \succ 0$. Then, observe that $x_{2,k} \to r(x_1^*)$ is guaranteed since r is Lipschitz and $x_{1,k} \to x_1^*$. By Assumption 2.3, it follows that $D_2 f_2(x_1, r(x_1^*)) = 0$ and $D_2^2 f_2(x_1, r(x_1^*)) \succ 0$. Thus, the pair $(x_1^*, x_2^*) = (x_1^*, r(x_1^*))$ is a differential Stackelberg equilibrium by definition. □

This result gives a strong convergence guarantee to differential Stackelberg equilibrium in general-sum games for the scenario that the follower plays a best-response at each time step. However, the assumptions that are made are quite restrictive. We remark that the global asymptotic stability assumption can be relaxed to a local asymptotic stability assumption to get a local convergence guarantee, but we do not include a statement of this result for brevity. Compared to the results from the previous subsection, the best-response analysis has value in terms of the interpretation. Specifically, the result of Theorem 2.10, characterizes what can be the expected in general-sum games when the follower actually does play a best-response versus taking (scaled) gradient steps toward reaching a best-response. This may be a more natural model in some real-world game-theoretic settings. Moreover, the result is a stronger characterization of convergence to a differential Stackelberg equilibrium in general-sum games, albeit under much stronger assumptions.

2.5.3 Two-Timescale Analysis

Now, let us consider a two-timescale approximation of the best-response dynamics from the previous section. Equivalently, the stochastic Stackelberg gradient dynamics, but with non-uniform learning rate sequences. Specifically, let $\{\gamma_{i,k}\}$ and $\{w_{i,k+1}\}$ denote the learning rate sequence and stochastic noise process for each $i \in \mathcal{I}$, respectively and consider the following set of dynamics:

$$\begin{aligned} x_{1,k+1} &= x_{1,k} - \gamma_{1,k}(Df_1(x_k) + w_{1,k+1}) \\ x_{2,k+1} &= x_{2,k} - \gamma_{2,k}(D_2 f_2(x_k) + w_{2,k+1}). \end{aligned} \tag{2.16}$$

Suppose that there is timescale separation so that $\lim_{k\to\infty} \gamma_{1,k}/\gamma_{2,k} = 0$ and equivalently $\gamma_{1,k} = o(\gamma_{2,k})$. Given this condition, x_2 evolves on a faster timescale than x_1. That is, the fast transient player is the follower (player 2) and the slow component is the leader (player 1). The system in (2.16) can be compared to the following continuous-time singularly perturbed system in the

limit as $\tau \to \infty$:

$$\begin{aligned} \dot{x}_1 &= -Df_1(x_1(t), x_2(t)) \\ \dot{x}_2 &= -\tau D_2 f_2(x_1(t), x_2(t)). \end{aligned} \tag{2.17}$$

From the perspective of the follower, x_1 appears quasi-static. Thus, it is natural to think about the behavior of x_2 via the ordinary differential equation $\dot{x}_2 = -D_2 f_2(x_1, x_2(t))$ where x_1 is fixed. Then, given that there is a solution $r(x_1)$ arising on the fast timescale, from the perspective of the leader, we can expect the stochastic dynamics to track $\dot{x}_1 = -Df(x_1(t), r(x_1(t)))$. Thus, the stochastic learning dynamics from (2.16) should be expected to approximate the best-response dynamics from the preceding subsection. We show under proper assumptions, this intuition can be formalized.

The result in this subsection again requires the following standard stochastic approximation assumptions.

Assumption 2.5. *The following hold:*

2.5a. *The maps $Df_1 : \mathbb{R}^d \to \mathbb{R}^{d_1}$, $D_2 f_2 : \mathbb{R}^d \to \mathbb{R}^{d_2}$ are L_1, L_2 Lipschitz.*

2.5b. *The learning rates satisfy $\sum_k \gamma_{i,k} = \infty$ for each $i \in \mathcal{I}$, $\sum_{i \in \mathcal{I}} \sum_k \gamma_{i,k}^2 < \infty$, $\gamma_{1,k} = o(\gamma_{2,k})$.*

2.5c. *The stochastic processes $\{w_{i,k}\}$ are martingale difference sequence with respect to the increasing family of σ-fields defined by $\mathcal{F}_k = \sigma(x_\ell, w_\ell, \ell \leq k)$, $\forall\ k \geq 0$, so that $\mathbb{E}[w_{k+1} |\ \mathcal{F}_k] = 0$ almost surely for all $k \geq 0$. Moreover, for some constant $C > 0$, $\mathbb{E}[\|w_{i,k+1}\|^2 |\ \mathcal{F}_k] \leq C(1 + \|x_k\|^2)$ almost surely, $\forall\ k \geq 0$.*

2.5d. *The iterates $\{x_k\}$ of (2.16) remain bounded almost surely: that is, $\sup_k \|x_k\| < \infty$.*

Moreover, we make the following asymptotic stability assumption that will be used along with Assumption 2.3.

Assumption 2.6. *The dynamics $\dot{x}_1 = -Df_1(x_1, r(x_1))$ have a globally asymptotically stable critical point x_1^*.*

Given Assumptions 2.3, 2.5, and 2.6 hold, then it immediately follows that $(x_{1,k}, x_{2,k}) \to (x_1^*, r(x_1^*))$ (Borkar, 2008, Chapter 6, Theorem 2). Moreover, by the global asymptotic stability assumptions, it is immediate that $D^2 f_1(x_1^*, r(x_1^*)) \succ 0$ and $D_2^2 f_2(x_1, r(x_1^*)) \succ 0$ so that $(x_1^*, r(x_1^*))$ is a differential Stackelberg equilibrium. This gives rise to the following result.

Proposition 2.11. *Consider a general-sum game (f_1, f_2) such that $f_i \in C^q(\mathcal{X}, \mathbb{R})$ for some $q \geq 2$ and each $i \in \mathcal{I}$. Suppose Assumptions 2.3, 2.5, 2.6 hold. Then, $(x_{1,k}, x_{2,k}) \to (x_1^*, r(x_1^*))$ almost surely. Moreover, $(x_1^*, r(x_1^*))$ is a differential Stackelberg equilibrium.*

Thus, this result gives an analogous convergence guarantee for the two-timescale setting as the best-response setting.

2.6 Experiments

We now present experimental results. The goal of this set of experiments is to investigate the behavior of trajectories of the gradient-based learning dynamics that have been analyzed along with the connections to the local Stackelberg equilibrium concept, and more broadly to study illustrative examples demonstrating the role of local Stackelberg equilibrium in the optimization landscape.

Section 2.6.1 begins with a canonical duopoly model that compares and contrasts the simultaneous and hierarchical play models along with the equilibrium within them. Moreover, it highlights that in general-sum games, the critical points of the simultaneous gradient only satisfy the first-order sufficient conditions for local Nash, while the critical points for the Stackelberg gradient dynamics satisfy the first-order sufficient conditions for local Stackelberg. Then, Section 2.6.2 presents a toy generative adversarial network problem for learning a covariance matrix. The simulation highlights how the Stackelberg gradient dynamics result in stable learning, without significant cyclic behavior, and consequently fast convergence. Section 2.6.3 shows how the gradient-based learning dynamics can be used in finite action games, and gives some discussion of the intricacies of mixed Stackelberg equilibrium. Finally, Section 2.6.4 presents generative adversarial network training problems in which the players are parameterized by neural networks. The results again highlight the stability of the Stackelberg gradient dynamics and demonstrate the scalability owing to modern computational tools. Moreover, the placement of differential Stackelberg equilibrium in the optimization landscape is studied by evaluating the eigenvalues of relevant game objects at convergence. The results show that the gradient-based learning algorithms are in fact converging to neighborhoods of differential Stackelberg equilibrium.

2.6.1 Duopoly Games

We begin by presenting a typical application of game-theoretic analysis: economic competitions.

Economic Competitions. Perhaps the simplest models of economic competitions come in the form of duopoly games. Consider a duopoly competition in which a homogeneous product is produced by a pair of firms. Each firm selects the quantity of product to produce. Let the production cost of firm $i \in \mathcal{I}$ producing $x_i \geq 0$ units of the product be given by $c_i x_i$ where $c_i > 0$ is the unit cost. It is common in duopoly games to model the firms as having market power, meaning the price of the product is dependent on the amount of production. Given a parameter $A \geq 3c_i$ for each $i \in \mathcal{I}$, we consider a linear price function of the form $P(x_1, x_2) = A - x_1 - x_2$. The profit (revenue minus product cost) of each firm $i \in \mathcal{I}$ is then given by $\pi_i(x_i, x_{-i}) = (P(x_i, x_{-i}) - c_i)x_i$. Thus, a general-sum game (f_1, f_2) emerges in which the cost function of each firm $i \in \mathcal{I}$ is given by $f_i(x_i, x_{-i}) = -\pi_i(x_i, x_{-i})$. That is, each firm is seeking to maximize their profit.

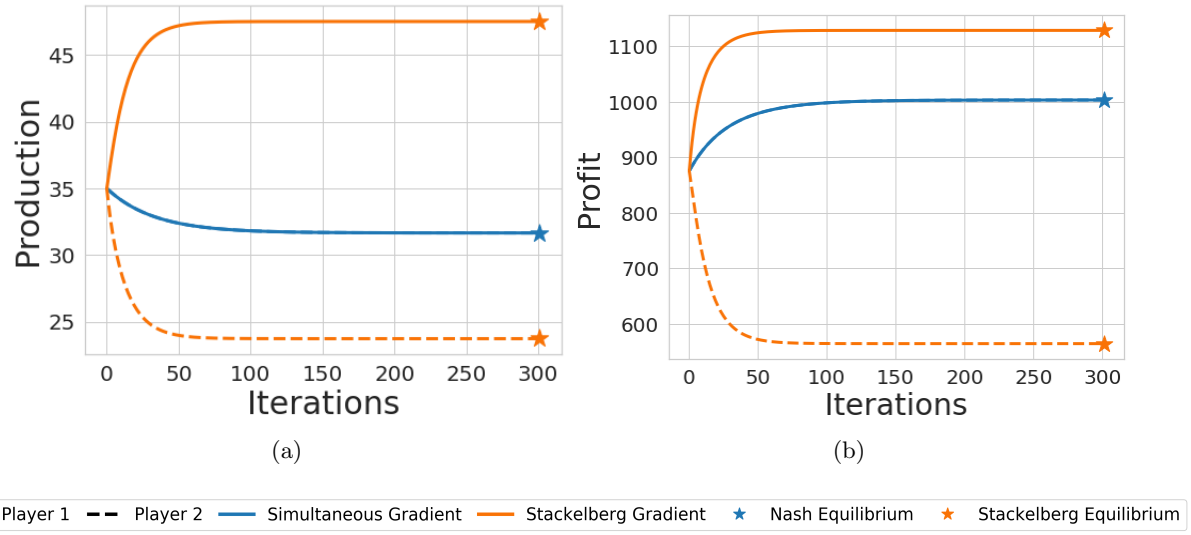
Cournot Duopoly. When the duopoly game that has been described is played simultaneously, it is known as a Cournot competition. The Nash equilibrium is the typical solution concept in Cournot competitions given the simultaneous play structure. For this game, it is relatively straightforward to solve for the unique Nash equilibrium. Recall that a Nash equilibrium in this context is a pair (x_1^*, x_2^*) such that x_1^* is a best-response to x_2^* and x_2^* is a best-response to x_1^*. Observe that for each firm $i \in \mathcal{I}$, the best-response to x_{-i} is given by $x_i^*(x_{-i}) = \frac{1}{2}(A - c_i - x_{-i})$ since $D_i f_i(x_i, x_{-i}) = c_i - A + 2x_i + x_{-i}$ and $D_i^2 f_i(x_i, x_{-i}) = 2$ so that setting $D_i f_i(x_i, x_{-i}) = 0$ and solving gives the result. Thus, plugging the best-response function of x_{-i}^* into the best-response function of x_i^* and solving, the unique Nash equilibrium is such that $x_i^* = \frac{1}{3}(A + c_{-i} - 2c_i)$ for each $i \in \mathcal{I}$. The product price at the Nash equilibrium is $P(x_1^*, x_2^*) = \frac{1}{3}(A + c_1 + c_2)$ and each firm $i \in \mathcal{I}$ obtains a profit of $\pi_i(x_1^*, x_2^*) = \frac{1}{9}(A - 2c_i + c_{-i})^2$. Observe that if $c_1 = c_2$, then at the Nash equilibrium each firm has equal profit.

Stackelberg Duopoly. When the duopoly game that has been described is played sequentially, it is known as a Stackelberg competition. The Stackelberg equilibrium is then the typical solution concept. For this game, it is also relatively straightforward to solve for the unique Stackelberg equilibrium. Recall that the static game structure is such that that the leader moves and then the follower produces a best-response to the choice of the leader. Knowing this, the leader seeks to minimize its cost function taking advantage of the power to move before the follower. Let firm 1 be the leader and firm 2 be the follower. From the previous discussion of the Cournot duopoly model, the best-response function of the follower is given by $x_2^*(x_1) = \frac{1}{2}(A - c_2 - x_1)$. This means that the leader wants to minimize $f_1(x_1, x_2^*(x_1)) = -\frac{1}{2}(A - x_1 + c_2 - 2c_1)x_1$. Observe that the optimal strategy for the leader is then $x_1^* = \frac{1}{2}(A + c_2 - 2c_1)$ since $Df_1(x_1, x_2^*(x_1)) = -\frac{1}{2}(A - 2x_1 + c_2 - 2c_1)$ and $D^2 f_1(x_1, x_2^*(x_1)) = 1$ so that setting $Df_1(x_1, x_2^*(x_1)) = 0$ and solving gives the result. Plugging x_1^* into the best-response function of x_2, the optimal strategy for the follower is $x_2^* = \frac{1}{4}(A - 3c_2 + 2c_1)$. Thus, the unique Stackelberg equilibrium in the game is given by the strategies $x_1^* = \frac{1}{2}(A + c_2 - 2c_1)$ and $x_2^* = \frac{1}{4}(A - 3c_2 + 2c_1)$. The product price at the Stackelberg equilibrium is $P(x_1^*, x_2^*) = \frac{1}{4}(A + 2c_1 + c_2)$, the profit of the leader is $\pi_1(x_1^*, x_2^*) = \frac{1}{8}(A - 2c_1 + c_2)^2$, and the profit of the follower is $\pi_2(x_1^*, x_2^*) = \frac{1}{16}(A + 2c_1 - 3c_2)^2$.

Comparing Equilibrium and Learning Dynamics. It is worth making a few remarks on the Nash and Stackelberg equilibrium in this game. Observe that the profit of the leader firm in the Stackelberg equilibrium exceeds the profit in the Nash equilibrium. This is exactly reflective of Proposition 2.3 since the best-response of the follower is unique for each strategy of the leader. Thus, the leader firm prefers the Stackelberg competition over the Cournot competition. In contrast, the follower firm always has higher profit in the Cournot competition than the Stackelberg competition under the given restriction that $A > 3c_i$ for each $i \in \mathcal{I}$.[9] Thus the follower firm prefers the Cournot competition over the Stackelberg competition. Recalling the discussion on comparing Nash and Stackelberg equilibrium costs from Section 2.2.4, this implies the Stackelberg equilibrium is nonconcurrent. Finally, observe that the product price is higher and the total production is lower in the Cournot competition than the Stackelberg competition. Consequently, the market prefers the Stackelberg competition over the Cournot competition.

This example highlights that depending on the interaction structure, the corresponding equilibrium outcomes are often disparate. Now we investigate if the 'natural' learning dynamics that have been formulated to reflect the simultaneous and hierarchical interaction structures converge to the corresponding equilibrium solutions. We simulate the deterministic simultaneous gradient dynamics and the Stackelberg gradient dynamics (firm 1 is taken to be the leader) with the parameters of the game selected to be $A = 100, c_1 = 5, c_2 = 2$ and the learning rate fixed to be $\gamma_1 = \gamma_2 = 0.01$. In Figure 2.7 we show the results of the simulation. Figure 2.7a shows the production path of each firm and Figure 2.7b shows the profit path of each firm. The simultaneous gradient dynamics converge to the unique Nash equilibrium of $x^* = (x_1^*, x_2^*) = (31.67, 31.67)$ that gives profit of $\pi(x_1^*, x_2^*) = (1002.78, 1002.78)$. The Stackelberg gradient dynamics converge to the unique Stackelberg equilibrium of $x^* = (x_1^*, x_2^*) = (47.5, 23.75)$ that gives profit of $\pi(x_1^*, x_2^*) = (1128.13, 564.06)$. Thus, we observe that the natural learning dynamics for each interaction structure do in fact converge to the corresponding equilibrium notion. It is worth noting that the unique Stackelberg equilibrium is not a stationary point of the simultaneous gradient dynamics and similarly the unique

[9] Note this restriction was selected so that the production of each firm at each equilibrium is positive.



(a) (b)

(c)

Figure 2.7: (a) Production curves: sample learning paths for each firm showing the production evolution and convergence to the Nash equilibrium under the simultaneous gradient dynamics and convergence to the Stackelberg equilibrium under the Stackelberg dynamics. (b) Profit curves: evolution of each firm's profit under the simultaneous learning dynamics and the Stackelberg learning dynamics. Of note is the improved profit obtained by the leader in the Stackelberg equilibrium compared to the Nash equilibrium.

Nash equilibrium is not a stationary point of the Stackelberg gradient dynamics. Indeed, unlike in zero-sum games, the stationary points do not coincide for each set of dynamics and the set of Nash and Stackelberg equilibrium can be distinct. The general-sum structure elucidates how each of set of learning dynamics is best-suited for seeking the equilibrium notion corresponding to the given interaction structure and also how the equilibrium notions have enhanced predictive power of the outcomes of competitive interactions when they match up with the interaction structure that the learning algorithms are emulating.

2.6.2 Learning a Covariance Matrix

We now present an illustrative toy generative adversarial network training problem to learn an unknown covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$.[10] The generator is restricted to be a linear function of the latent input noise $z \sim \mathcal{N}(0, I)$ defined by $G_V(z) = Vz$ The discriminator is restricted to be a quadratic function of the real data generating process $x \sim \mathcal{N}(0, \Sigma)$ defined by $D_W(x) = x^\top Wx$. The matrices $V \in \mathbb{R}^{d\times d}$ and $W \in \mathbb{R}^{d\times d}$ are the parameters of the generator (player 1) and the discriminator (player 2), respectively. For the given generator and discriminator networks, the zero-sum Wasserstein generative adversarial network (Arjovsky et al., 2017) problem is defined by

[10]The following problem formulation was previously presented by Daskalakis et al. (2018).

the cost

$$f(V, W) = \mathbb{E}_{x\sim\mathcal{N}(0,\Sigma)}[x^\top W x] - \mathbb{E}_{z\sim\mathcal{N}(0,I)}[z^\top V^\top W V z].$$

As shown by Daskalakis et al. (2018), this cost function can be simplified to

$$f(V, W) = \sum_{i=1}^{d}\sum_{j=1}^{d} W_{ij}(\Sigma_{ij} - \sum_{k=1}^{d} V_{ik}V_{jk}).$$

The critical points of the game are given by (V, W) such that $VV^\top = \Sigma$ and $W + W^\top = 0$. Observe that this implies that at any critical point of the game, the generator has recovered the underlying data distribution.

Now consider a general-sum variant of this zero-sum game in which the cost of the follower is regularized. Specifically, let the generator's cost be defined by

$$f_1(V, W) = \sum_{i=1}^{d}\sum_{j=1}^{d} W_{ij}(\Sigma_{ij} - \sum_{k=1}^{d} V_{ik}V_{jk}).$$

Moreover, define the discriminator's cost by

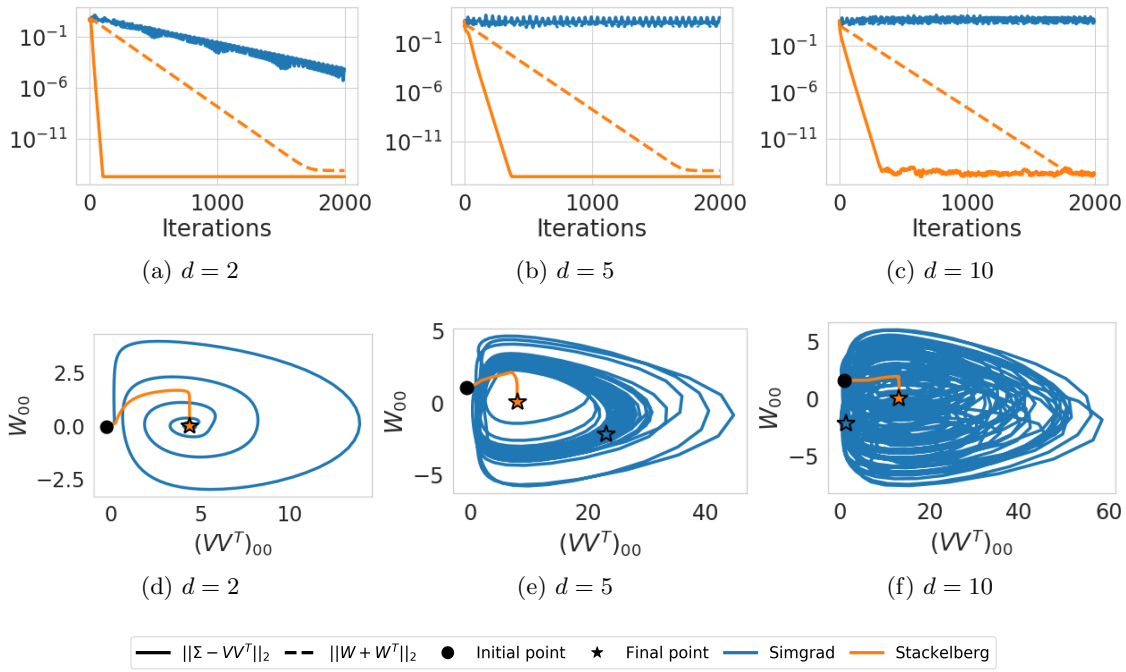
$$f_2(V, W) = -\sum_{i=1}^{d}\sum_{j=1}^{d} W_{ij}(\Sigma_{ij} - \sum_{k=1}^{d} V_{ik}V_{jk}) + \frac{\mu}{2}\operatorname{Tr}(W^\top W),$$

where $\mu > 0$ is a tunable regularization parameter. In this formulation, at any joint strategy (V^*, W^*) where $D_2 f_2(V^*, W^*) = 0$ then $W^* = 0$ and also $\det(D_2 f_2(V, W)) \neq 0$ for all (V, W) so that the Stackleberg gradient dynamics are always well-defined.

For this problem, we simulate both the Stackelberg gradient dynamics and the simultaneous gradient dynamics, and analyze the distance from the equilibrium as a function of time along with the trajectories of the dynamics. The learning rates are chosen as $\gamma_1 = \gamma_2/4 = 0.01$ and the regularization in the discriminator's cost is fixed to be $\mu = 0.5$. The covariance matrix is chosen to be $\Sigma = UU^T + I$ where $U \sim \mathcal{N}(0, 1)$. We plot $\|\Sigma - VV^\top\|_2$ for the generator's performance and $\|W + W^\top\|_2$ for the discriminator's performance in Figures 2.8a–2.8c for problems with $d \in \{2, 5, 10\}$. Moreover, we plot coordinates of $VV^\top$ against W in Figure 2.8d–2.8f for each problem. We observe that Stackelberg gradient dynamics converge to the solution in significantly faster. Furthermore, the simultaneous gradient dynamics rotational cyclic behavior around the solution. In contrast, the Stackelberg gradient dynamics do not cycle at all. For zero-sum games, our theory provides reasoning for this behavior since at any critical point the eigenvalues of the game Jacobian are purely real. Since at critical points, $W = 0$, this property carries over to this general-sum game. This is in contrast to simultaneous gradient descent, whose Jacobian can admit complex eigenvalues, which are known to cause rotational forces in the dynamics.

2.6.3 Parameterized Bimatrix Games

We now present a study of bimatrix games. To represent strategies with discrete actions, continuous probability distributions can be employed as mixed strategies over the discrete actions.



(a) $d = 2$ (b) $d = 5$ (c) $d = 10$

(d) $d = 2$ (e) $d = 5$ (f) $d = 10$

Figure 2.8: The Stackelberg gradient dynamics rapidly learn the covariance matrix Σ as compared to the simultaneous gradient dynamics. Errors given by $\|\Sigma - VV^T\|_2$ and $\|W + W^T\|_2$ are shown in (a)–(c). Trajectory plots of elements of W and VV^T demonstrating the cycling of the simultaneous gradient dynamics are presented in (d)–(f).

The gradient-based methods we study for continuous strategy spaces can thus be used to solve for equilibria in the parameterized strategy space. Consider a game (f_1, f_2) with costs given by

$$f_1(x_1, x_2) = \pi(x_1)^\top A \pi(x_2) + \frac{\mu_1}{2}\|x_1\|_2^2 \quad \text{and} \quad f_2(x_1, x_2) = \pi(x_1)^\top B \pi(x_2) + \frac{\mu_2}{2}\|x_2\|_2^2,$$

where

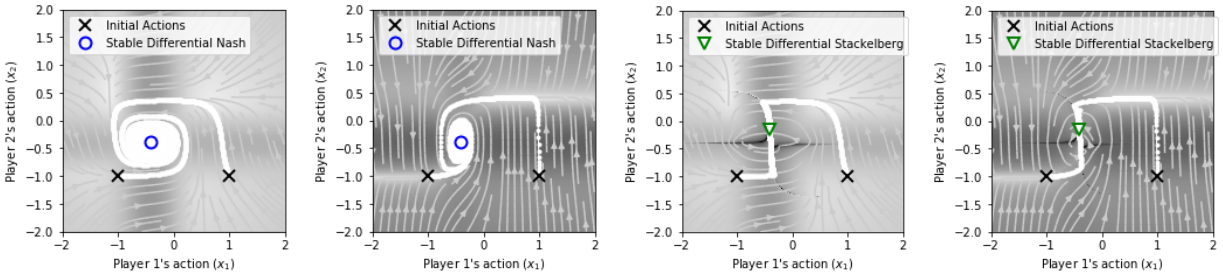
$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1/2 & 1 \\ 1 & 1/2 \end{bmatrix} \tag{2.18}$$

are the matrices representing the bimatrix game with player 1 as the row player and player 2 as the column player. We represent the mixed policy of two discrete actions with a sigmoid-based probability distribution on the simplex, $\pi : \mathbb{R} \to \Delta^1$, given by

$$\pi(x) = (e^{a_1 x + b_1},\ e^{a_2 x + b_2})/(e^{a_1 x + b_1} + e^{a_2 x + b_2})$$

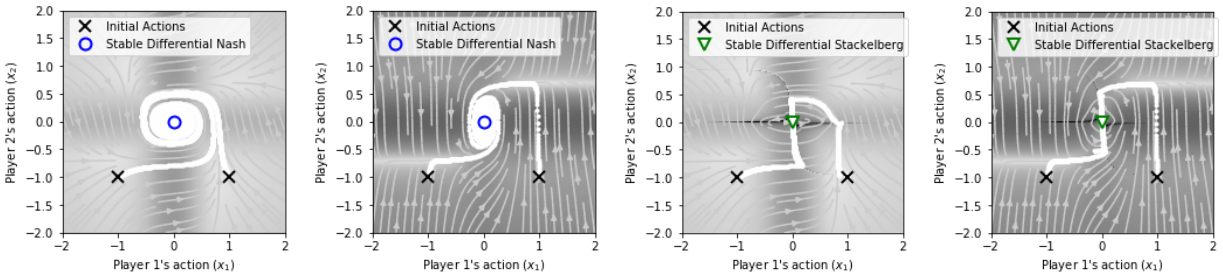
where the parameters a_1, b_1 and a_2, b_2 are constants that scale and shift the parameterization. This parameterization scheme can be extended to $d + 1$ actions using d variables. For two actions, we require that a_1 and a_2 have opposite signs. The 2-norm regularization of each player's individual action serves to regularize each player towards the interior of the simplex.

The bimatrix game admits a unique mixed Nash equilibrium of $(1/2, 1/2)$ for player 1 and



(a) simgrad ($\gamma_1 = \gamma_2$) (b) simgrad ($\gamma_1 = \gamma_2/10$) (c) Stackelberg ($\gamma_1 = \gamma_2$) (d) Stackelberg ($\gamma_1 = \gamma_2/10$)

Figure 2.9: **Parameterized bilinear game.** Parameters: $(a_1, a_2) = (2.5, -2.5)$ and $(b_1, b_2) = (1, -1)$. (a)–(b): For the simultaneous gradient dynamics, we observe parameter convergence to a neighborhood of $(x_1^*, x_2^*) = (-.4, -.4)$, which in policy space corresponds to the mixed Nash equilibrium of $(1/2, 1/2)$ for each player. (c)–(d): For the Stackelberg gradient dynamics, we observe convergence to a neighborhood of $(x_1^*, x_2^*) = (-.4, -.16)$, which in policy space corresponds to player 1 selecting the mixed Stackelberg equilibrium strategy of $(1/2, 1/2)$ and player 2 playing the distribution $(0.77, 0.23)$. The effects of time-scale separation is visualized as the light colored horizontal path, showing a low gradient norm along player 2's reaction curve.



(a) simgrad ($\gamma_1 = \gamma_2$) (b) simgrad ($\gamma_1 = \gamma_2/10$) (c) Stackelberg ($\gamma_1 = \gamma_2$) (d) Stackelberg ($\gamma_1 = \gamma_2/10$)

Figure 2.10: **Parameterized bilinear game.** Parameters: $(a_1, a_2) = (2.5, -2.5)$ and $(b_1, b_2) = (0, 0)$. (a)–(b): For the simultaneous gradient dynamics, we observe convergence to a neighborhood of $(x_1^*, x_2^*) = (0, 0)$, which in policy space corresponds to the mixed Nash equilibrium of $(1/2, 1/2)$ for each player. (c)–(d): For the Stackelberg gradient dynamics, we observe the equivalent final convergence characteristics. The effects of time-scale separation is visualized as the light colored horizontal path, showing a low gradient norm along player 2's reaction curve. Note that due to the choice of parameters (b_1, b_2), the regularization is penalizing for any deviation from the equilibrium at $(1/2, 1/2)$ in the policy space.

$(1/2, 1/2)$ for player 2. If the game is played in a hierarchical structure with the leader being player 1, the mixed Stackelberg equilibrium of the game is (π_1, π_2) with $\pi_1 = (1/2, 1/2)$ and any policy π_2 in the simplex for the follower. At this strategy, the cost the leader incurs is independent of the follower's strategy. We refer to Basar and Olsder (1998, §3.6) for discussion on the mixed Stackelberg

equilibrium of this bimatrix game. For the softmax parameterized policy class we consider, using $(a_1, a_2) = (2.5, -2.5)$, $(b_1, b_2) = (1, -1)$, $\pi(-0.4) = (1/2, 1/2)$. That is, the parameter $x = -0.4$ corresponds to the policy $(1/2, 1/2)$. On the other hand, if $(a_1, a_2) = (2.5, -2.5)$, $(b_1, b_2) = (0, 0)$, then $\pi(0) = (1/2, 1/2)$.

We plot the the vector fields g and $g_{\mathcal{S}}$ and their norms, along with simulations of both the simultaneous and Stackelberg gradient dynamics in Figures 2.9 and 2.10. We use parameters $(a_1, a_2) = (2.5, -2.5)$, and regularization $\mu = \mu_1 = \mu_2 = 0.1$. For the parameters (b_1, b_2), we explore two different pairs: $(b_1, b_2) = (1, -1)$ and $(b_1, b_2) = (0, 0)$. The latter is such that the regularization term is penalizing for any deviation from the equilibrium parameter values, while the former is such that the regularization is penalizing for any deviation from $(0, 0)$ while the equilibrium is at $(0.4, y)$ for any $y \in [0, 1]$. Observe that the simultaneous gradient dynamics cycle significantly in parameter space and in policy space they converge to the mixed Nash equilibrium in each example. In contrast, the Stackelberg gradient dynamics do not exhibit cyclic behavior in the parameter space, and in the cases of the parameters $(b_1, b_2) = (1, -1)$ and $(b_1, b_2) = (0, 0)$ converge to mixed Stackelberg and mixed Nash equilibrium, respectively.

The shading of the action space indicates the norm of the dynamics: darker has a larger norm. Different parameterization constants or regularization weights will affect the outcome of the gradient-based learning. The timescale separation improves the convergence properties of the Stackelberg gradient dynamics as it encourages the dynamics to converge to the follower's best-response curve. We observe the distinctly lighter path the shaded plots of Figure 2.9 (b) and (d), where the follower's response curve runs horizontal to the plot. Comparing plots Stackelberg gradient in Figures 2.9 (c) and (d), we observe that with timescale separation, the paths of the learning dynamics converges first to the manifold on the ridge, then towards the stationary point along the manifold. Doing so prevents the trajectory from being perturbed by the 'cliffs', visualized by the dark cusps with large gradient norm corresponding to area where the follower's individual second-order derivative is poorly conditioned.

2.6.4 Generative Adverarial Networks Parameterized by Neural Networks

We now move on to experiments training generative adversarial networks in which each player is parameterized by a neural network. The generator is always taken to be the leader and the discriminator the follower in this set of experiments. Prior to presenting the results, we provide details on the typical generative adversarial network formulation and some practical training modifications we make to the gradient-based learning dynamics that have been presented thus far.

Generative Adversarial Network Formulations. The standard generative adversarial network formulation (Goodfellow et al., 2014) (often known as the saturating objective) can be characterized by the zero-sum game

$$\min_{\theta} \max_{\omega} f(\theta, \omega) = \mathbb{E}_{x \sim p_{\mathcal{X}}}[\ell(\mathrm{D}_\omega(x))] + \mathbb{E}_{z \sim p_{\mathcal{Z}}}[\ell(-\mathrm{D}_\omega(\mathrm{G}_\theta(z)))]. \tag{2.19}$$

In this formulation $\mathrm{G}_\theta : \mathcal{Z} \to \mathcal{X}$ is the generator network parameterized by θ that maps from the latent space (noise distribution) $\mathcal{Z}$ to the input space (data distribution) $\mathcal{X}$, $\mathrm{D}_\omega : \mathcal{X} \to \mathbb{R}$ is the discriminator network parameterized by ω that maps from the input space $\mathcal{X}$ to real-valued logits, and $p_{\mathcal{X}}$ and $p_{\mathcal{Z}}$ are the distributions over the input space and the latent space respectively. The

loss function is defined by $\ell(x) = \log(\sigma(x))$ where $\sigma(x) = (1 + e^{-x})^{-1}$ is the logistic function that maps from a real-valued logit to a probability (real or fake data in this context). The goal of the problem formulation is for the generator to learn a map from the latent space to the input space that matches the underlying data distribution such that the discriminator cannot discern the real data from the fake data. A variant of this formulation is what is known as the non-saturating objective (Goodfellow et al., 2014). In the non-saturating formulation, the generator objective is to maximizing $\mathbb{E}_{z\sim p_{\mathcal{Z}}}[\ell(\mathrm{D}_\omega(\mathrm{G}_\theta(z)))]$, or equivalently minimize $\mathbb{E}_{z\sim p_{\mathcal{Z}}}[\ell(\mathrm{D}_\omega(\mathrm{G}_\theta(z)))]$. This results in the general-sum game defined by the costs:

$$\begin{aligned} f_1(\theta, \omega) &= \mathbb{E}_{p_{\mathcal{D}}(x)}[\ell(\mathrm{D}_\omega(x))] - \mathbb{E}_{p(z)}[\ell(\mathrm{D}_\omega(\mathrm{G}_\theta(z)))] \\ f_2(\theta, \omega) &= -\mathbb{E}_{p_{\mathcal{D}}(x)}[\ell(\mathrm{D}_\omega(x))] - \mathbb{E}_{p(z)}[\ell(-\mathrm{D}_\omega(\mathrm{G}_\theta(z)))]. \end{aligned} \tag{2.20}$$

The motivation for this reformulation is that it was observed (Goodfellow et al., 2014) that when training with the simultaneous gradient dynamics, $\mathbb{E}_{z\sim p_{\mathcal{Z}}}[\ell(-\mathrm{D}_\omega(\mathrm{G}_\theta(z)))]$ 'saturates' early in training so the gradient information is weak and hampers learning. The experiments that follow consider both objective formulations.

Practical Training Modifications. For this set of experiments, we implement some practical modifications to the basic learning dynamics studied in this chapter. To begin, for both the simultaneous gradient dynamics and the Stackelberg gradient dynamics we pass the gradient information of each player into the Adam optimizer (Kingma and Ba, 2015) using the default parameters of $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$ for the mixture of Gaussian experiments and $\beta_1 = 0.5$ for the MNIST experiment following the standard convention for the network architecture that is selected. Moreover, for the leader update in the Stackelberg gradient dynamics, since often $D_2 f_2(x_1, x_2)$ is ill-conditioned if not degenerate along the learning path, given regularization $\mu > 0$ we implement the (regularized) update

$$g_{\mathcal{S},1} = D_1 f_1(x) - D_{21} f_2(x)^\top \left(D_2^2 f_2(x) + \mu I\right)^{-1} D_2 f_1(x).$$

This update equation can be derived from viewing the leader as regularizing the conjectured behavior of the follower, so that the optimization problem for the leader is given by

$$\arg\min_{x_1} \left\{ f_1(x_1, x_2) |\ x_2 \in \arg\min_y f_2(x_1, y) + \frac{\mu}{2}\|y\|^2 \right\},$$

where the follower is actually aiming to solve the problem $\arg\min_{x_2} f_2(x_1, x_2)$. The leader then views the follower as descending the gradient $D_2 f_2(x_1, x_2) + \mu x_2$ so that the derivative of the implicit map is given by $Dr(x_1) = -(D_2^2 f_2(x_1, r(x_1)) + \mu I)^{-1} D_{21} f_2(x_1, r(x_1))$. It is possible to define sufficient conditions for a local Stackelberg equilibrium when considering the leader objective from above. Specifically, the conditions are equivalent as from the definition that has been provided for a differential Stackelberg equilibrium, but using the derivative of the implicit map as just provided. In the experiments that follow, we assess the eigenvalues of relevant game objects around critical points that the gradient-based dynamics converge towards and to determine if the strategy is a local Stackelberg equilibrium, we also use the regularized implicit map in the second-order total derivative when verifying the leader second-order sufficient condition. Specifics on the computations can be found in Section 2.D.4.

2.6.4.1 Mixture of Gaussians

In this subsection, we show results for training generative adversarial networks to learn a mixture of 2-dimensional Gaussian distributions. A pair of configurations are considered, a diamond configuration, and a circle configuration. The underlying data distribution that is to be learned for the diamond experiment consists of Gaussian distributions with means given by $\mu = [1.2\sin(\omega), 1.2\cos(\omega)]$ for $\omega \in \{k\pi/2\}_{k=0}^{3}$ and covariances $\sigma^2 I$ where $\sigma^2 = 0.15$. The underlying data distribution that is to be learned for the circle experiment consists of Gaussian distributions with means given by $\mu = [\sin(\omega), \cos(\omega)]$ for $\omega \in \{k\pi/4\}_{k=0}^{7}$ and covariances $\sigma^2 I$ where $\sigma^2 = 0.05$.

During training, the real data $x \in \mathbb{R}^2$ is selected uniformly at random from the set of Gaussian distributions and the latent data $z \in \mathbb{R}^{16}$ is drawn from a standard normal distribution with batch sizes of 256. The neural network for the generator contains two hidden layers, each of which contain 32 neurons. The generator and discriminator networks have two and one hidden layers, respectively; each hidden layer has 32 neurons. The activation function following the hidden layers in the generator network is the Tanh function and the ReLU function in the diamond and circle experiments, respectively. The objective for the game in the diamond experiment is the saturating generative adversarial network objective from (2.19) and in the circle experiment it is the non-saturating generative adversarial network objective from (2.20). The initial learning rates for each player and for each learning rule is 0.0001 and 0.0004 in the diamond and circle experiments, respectively. The learning rate for each player $i \in \mathcal{I}$ is decayed exponentially such that $\gamma_{i,k} = \gamma_i \nu_i^k$ where $\nu_1 = \nu_2 = 1{-}10^{-7}$ for the simultaneous gradient dynamics and $\nu_1 = 1{-}10^{-5}$ and $\nu_1 = 1{-}10^{-7}$ for the Stackelberg gradient dynamics. Finally, the implicit map of the follower is regularized as discussed in Section 2.6.4 using the parameter $\mu = 1$ and similarly in computing the eigenvalues of $D^2 f_1$ as detailed in Section 2.D.4. For both the diamond and circle configurations, 10 initial seeds were simulated for each set of learning dynamics and the behavior was generally consistent across the seeds for both algorithms. The experiments were run for 60k training batches and the eigenvalues of relevant game objectives evaluated at that stopping point.

Diamond Configuration. In Figures 2.11b–2.11c and Figures 2.11d–2.11e we show a sample of the generator and the discriminator for the simultaneous gradient dynamics and the Stackelberg gradient dynamics at the end of training. Specifically, for each set of learning dynamics we show the best run from 10 initial seeds when judged in terms of the KL-divergence between the real data and the generated data. Each learning rule converges so that the generator can create a distribution that is close to the ground truth and the discriminator is nearly at the optimal probability throughout the input space of 0.5, meaning that it cannot tell real from fake. In Figures 2.11f–2.11i and Figures 2.11j–2.11m for the simultaneous gradient dynamics and the Stackelberg gradient dynamics respectively, we show eigenvalues from the game objects evaluated at the final point that present a deeper view of the convergence behavior. Specifically for each relevant game object we show the 5 smallest and 15 largest real eigenvalue parts. We observe from the eigenvalues of J that both sets of dynamics converge to neighborhoods of points that are stable for the simultaneous dynamics and they appear to be in a neighborhood of a differential Stackelberg equilibrium since the eigenvalues of $D^2 f_1$ and $D_2^2 f_2$ are nearly all positive. Interestingly, however, since the eigenvalues of $D_1^2 f_1$ are nearly all zero and not all positive, it appears that the result may reflect the realizable assumption (refer to Section 2.3) as well as convergence to a differential Stackelberg equilibrium that is not a differential Nash equilibrium.

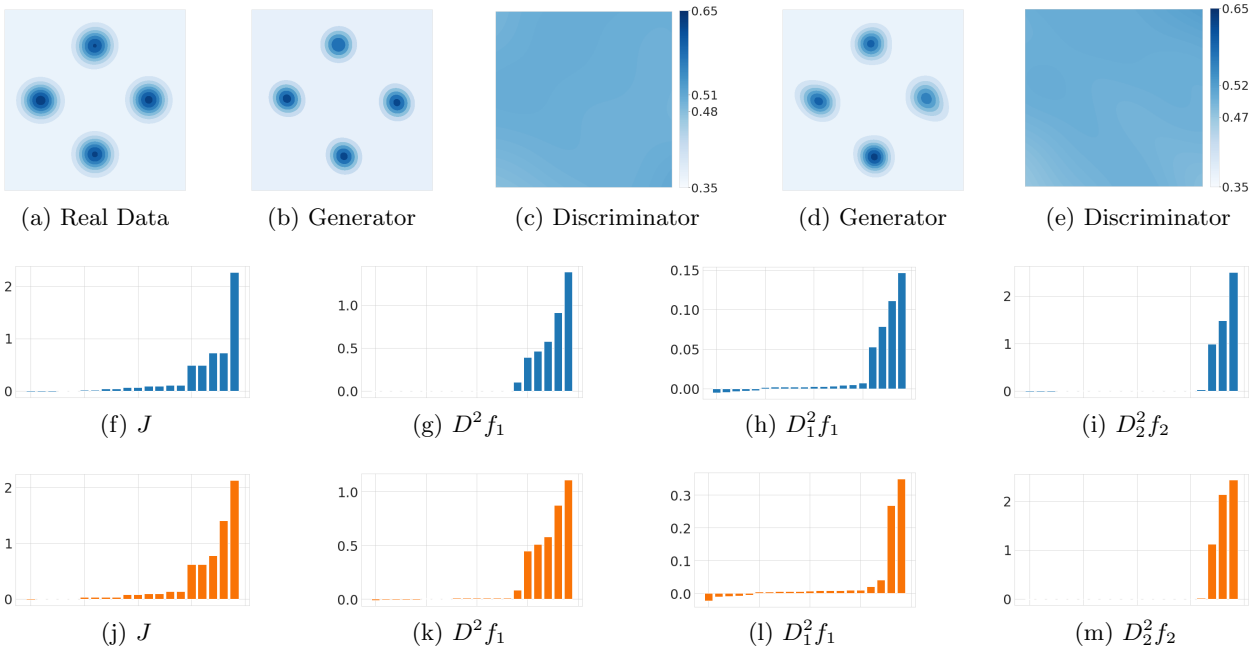
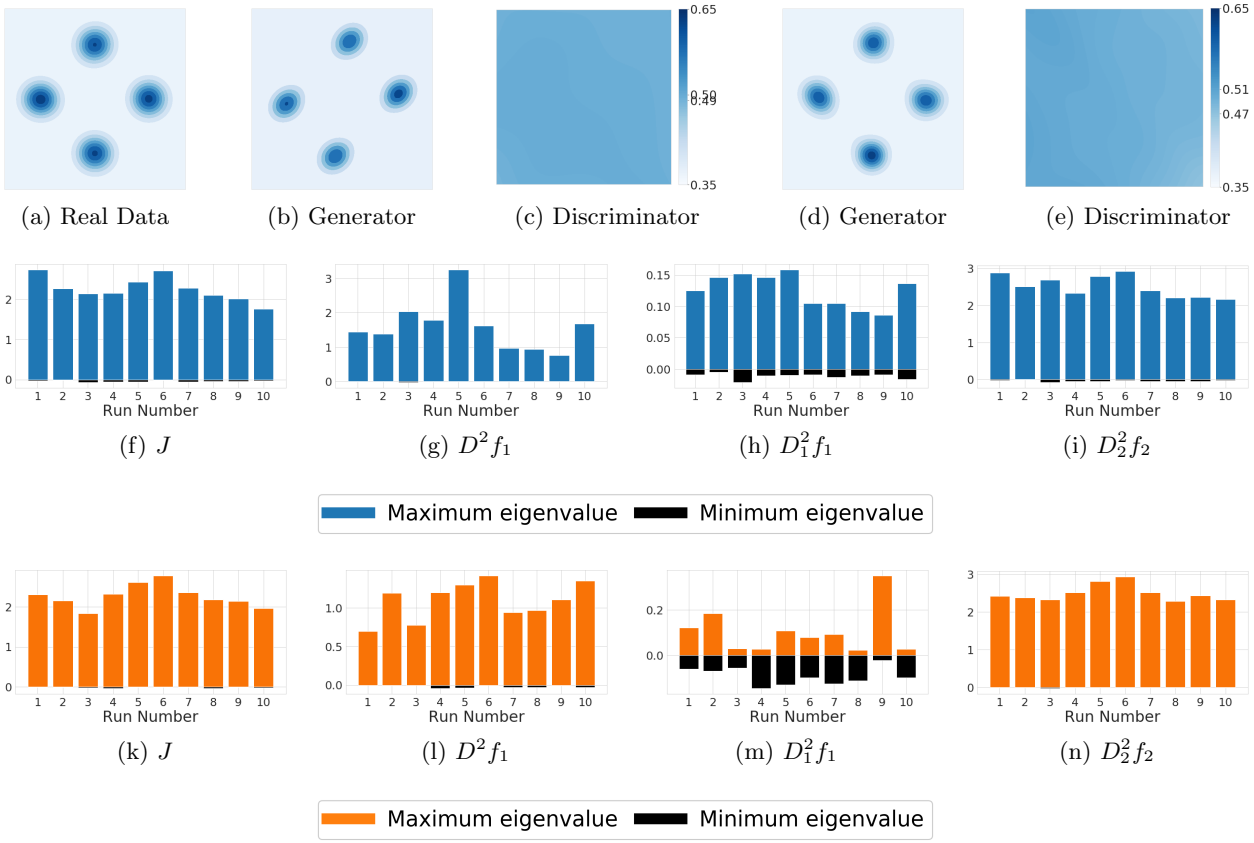


Figure 2.11: The generator and discriminator performances on the diamond configuration for the simultaneous gradient dynamics and the Stackelberg gradient dynamics from the best run of 10 seeds are shown in (b)–(c) and (d)–(e), respectively. We show the 5 smallest and 15 largest real eigenvalues parts of relevant game objects in (f)–(i) for the simultaneous gradient dynamics and (j)–(m) for the Stackelberg gradient dynamics.

In general, the conclusions that can be drawn from the best run are generally consistent across the random seeds. To demonstrate this, we provide additional simulation results for the diamond configuration in Figure 2.12. The generator and discriminator outputs in Figures 2.12b–2.12c and Figures 2.12d–2.12e correspond to the 5th best of the 10 runs when judged in terms of the KL-divergence between the real data and the generated data for the simultaneous gradient dynamics and the Stackelberg gradient dynamics, respectively. We again see reasonable performance for both the simultaneous gradient dynamics and the Stackelberg gradient dynamics in terms of the generator and the discriminator. For each of the 10 runs the minimum and maximum real eigenvalue parts for the relevant game objects are presented in Figures 2.12f–2.12i for the simultaneous gradient dynamics and in Figures 2.12k–2.12n for the Stackelberg gradient dynamics. The black bars show the minimum real parts of the eigenvalues and for several of the plots they are not visible since they are near zero. Observe that the eigenvalues of $D_1^2 f_1$ are consistently near zero and include negative values for both sets of dynamics, indicating they are not converging to a differential Nash equilibrium. Moreover, the results show consistent convergence to neighborhoods of stable critical points of the simultaneous gradient dynamics that are differential Stackelberg equilibrium since the real parts of the eigenvalues of J, $D^2 f_1$, and $D_2^2 f_2$ are positive.

Given the good generator and discriminator performance for this problem, it is worth further empirical investigation to determine if differential Stackelberg equilibrium that are not differential Nash equilibrium are desirable in generative adversarial networks and if successful training methods



(a) Real Data (b) Generator (c) Discriminator (d) Generator (e) Discriminator

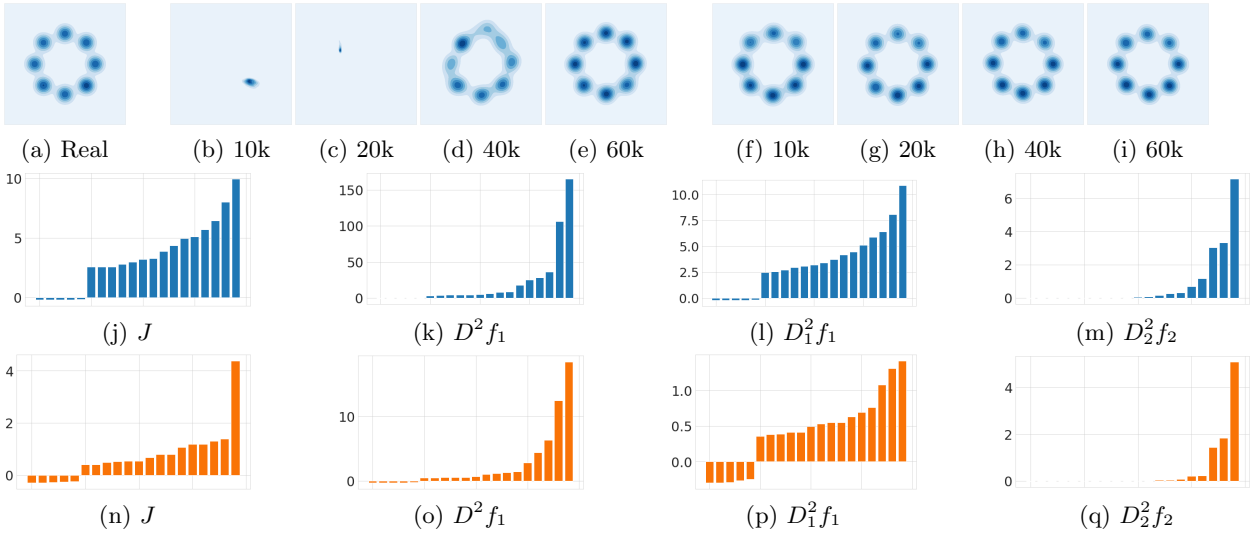
(f) J (g) $D^2 f_1$ (h) $D_1^2 f_1$ (i) $D_2^2 f_2$

(k) J (l) $D^2 f_1$ (m) $D_1^2 f_1$ (n) $D_2^2 f_2$

Figure 2.12: The generator and discriminator performances on the diamond configuration for the simultaneous gradient dynamics and the Stackelberg gradient dynamics from the 5th best run of 10 seeds are shown are shown in (b)–(c) and (d)–(e), respectively. The minimum and maximum real parts of the eigenvalues of the game objects (f)–(m) are shown for ten random initial seeds where from (f)–(i) is for the simultaneous gradient dynamics and (k)–(n) is for the Stackelberg gradient dynamics.

commonly are reaching such points.

Circle configuration. In Figures 2.13b–2.13e and Figures 2.13f–2.13i we show the generator along the learning path for the simultaneous gradient dynamics and the Stackelberg gradient dynamics, respectively. In particular, for each set of learning dynamics we show the best run from 10 initial seeds when evaluated in terms of the KL-divergence between the real data and the generated data. Observe that the simultaneous gradient dynamics cycle and perform poorly until the learning rates decay enough to stabilize the training process. In contrast, the Stackelberg gradient dynamics converge quickly to a solution that nearly matches the ground truth distribution. In a similar fashion as in the covariance example, the leader update is able to reduce rotational forces. In Figures 2.14j–2.14m and Figures 2.14o–2.14r for the simultaneous gradient dynamics and the Stackelberg gradient dynamics respectively, we show eigenvalues from the game objects evaluated at the final point. For each relevant game object we show the 5 smallest and 15 largest real eigenvalue



(a) Real (b) 10k (c) 20k (d) 40k (e) 60k (f) 10k (g) 20k (h) 40k (i) 60k

(j) J (k) $D^2 f_1$ (l) $D_1^2 f_1$ (m) $D_2^2 f_2$

(n) J (o) $D^2 f_1$ (p) $D_1^2 f_1$ (q) $D_2^2 f_2$

Figure 2.13: The generator of the simultaneous gradient dynamics along the learning path is shown in (b)–(e) and the generator of the Stackelberg gradient dynamics along the learning path is shown in (f)–(i) for the best runs on the circle configuration. We present the 5 smallest and 15 largest real eigenvalue parts of relevant game objects in (j)–(d) for the simultaneous gradient dynamics and (n)–(q) for the Stackelberg gradient dynamics.

parts. We observe from the eigenvalues of J that both sets of dynamics converge to neighborhoods of points that are stable for the simultaneous gradient dynamics and the simultaneous gradient dynamics converge to a neighborhood of a differential Nash equilibrium since the eigenvalues of both $D_1^2 f_2$ and $D_2^2 f_2$ are positive and the Stackelberg gradient dynamics converge to a neighborhood of a differential Stackelberg equilibrium that is not differential Nash equilibrium since the eigenvalues of $D_1^2 f_1$ are not all positive but the eigenvalues of both $D^2 f_1$ and $D_2^2 f_2$ are positive.

We again observed that the behavior of the learning dynamics was consistent across the random seeds. Additional simulation results for the circle configuration are provided in Figure 2.14 to demonstrate this fact. The generated distribution along the learning path for the 5th best of the 10 runs when evaluated in terms of the KL-divergence between the real data and the generated data is shown for the simultaneous gradient dynamics in Figures 2.14b–2.14e and the Stackelberg gradient dynamics in Figures 2.14f–2.14i. As with the best run, the simultaneous gradient dynamics cycle, while the Stackelberg gradient dynamics converge quickly to the real distribution. For each of the 10 runs the minimum and maximum real eigenvalue parts for the relevant game objects are presented in Figures 2.14j–2.14m for the simultaneous gradient dynamics and in Figures 2.14o–2.14r for the Stackelberg gradient dynamics. The black bars show the minimum real parts of the eigenvalues and for several of the plots they are not visible since they are near zero. From the eigenvalues, we can conclude that the simultaneous gradient dynamics consistently converge to a neighborhood of a differential Nash equilibrium since the eigenvalues of both $D_1^2 f_1$ and $D_2^2 f_2$ are positive, while the Stackelberg gradient dynamics consistently converge to a neighborhood of a differential Stackelberg equilibrium that is not a differential Nash equilibrium since $D_1^2 f_1$ has negative eigenvalues but both $D^2 f_1$ and $D_2^2 f_2$ have positive eigenvalues.

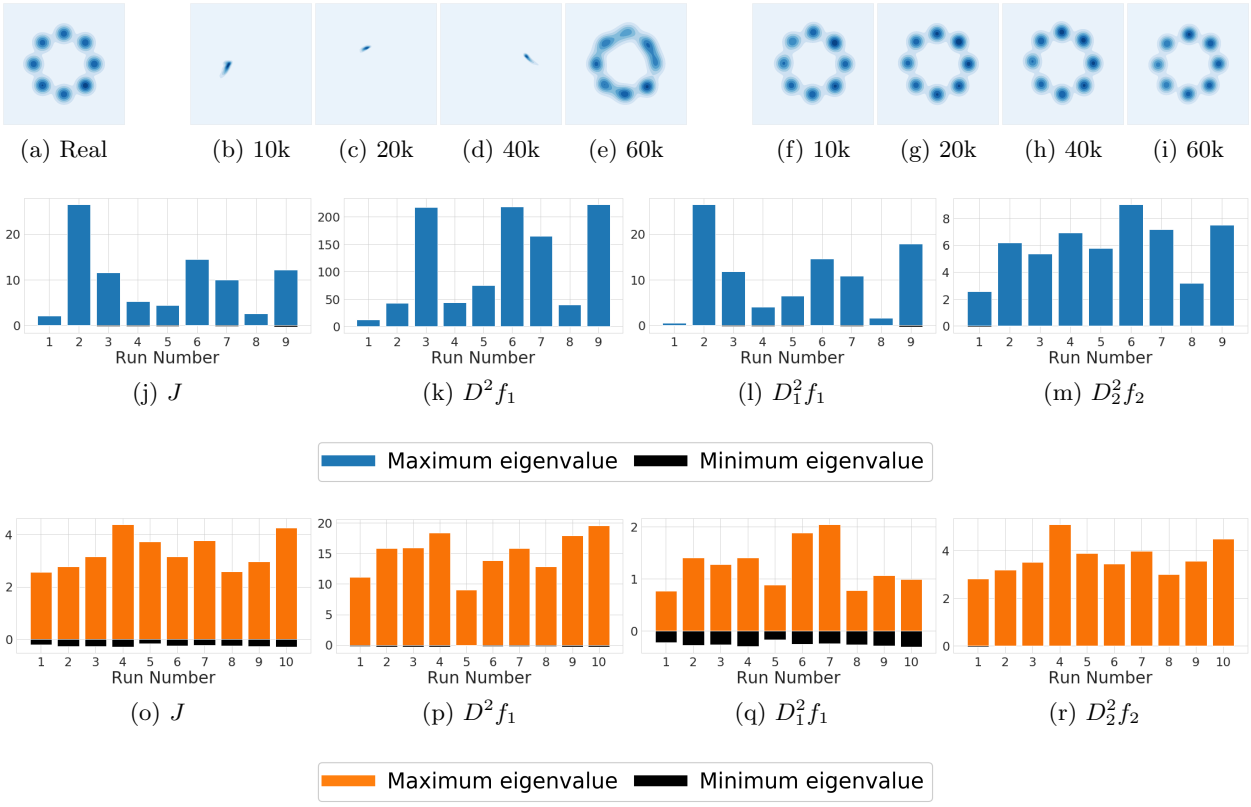


Figure 2.14: The generator performance along the learning path for the simultaneous gradient dynamics is shown in (b)–(e) and the generator performance along the learning path for the Stackelberg gradient dynamics is shown in (f)–(i) for the best runs on the circle configuration. The minimum and maximum real parts of the eigenvalues of the game objects (f)–(m) are shown for ten random initial seeds where from (j)–(m) is for the simultaneous gradient dynamics and (o)–(r) is for the Stackelberg gradient dynamics.

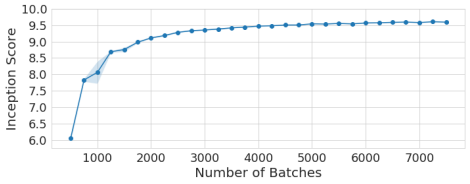
Together the circle configuration example shows that the Stackelberg gradient dynamics can mitigate cycling behavior and also the results provide further evidence that differential Stackelberg equilibrium may be easier to reach, and can provide suitable performance.

2.6.4.2 MNIST

The final generative adversarial network problem we consider is on the MNIST dataset using the DCGAN architecture (Radford et al., 2016) adapted to handle 28×28 images. The specific neural network architectures for both the generator and the discriminator are given in Tables 2.16 and 2.17. The implementation is in PyTorch and we describe the networks by the parameters passed into nn.Sequential class. The primary purpose of this experiment is to show that the Stackelberg gradient dynamics can scale to large-scale machine learning problems with millions of parameters despite needing to solve a linear equation at each iteration. In fact, the leader gradient can be computed with only a constant multiplicative cost compared to a normal gradient computation. Specifically,

(a) Real Images.

(b) Generated Images.



(c) Inception Scores.

Figure 2.15: Stackelberg gradient dynamics learning MNIST digits.

to solve the linear equation in the leader gradient, we call the conjugate gradient method (Martens et al., 2010). Each iteration of the conjugate gradient algorithms requires computing a Hessian-vector product and the cost of this computation is approximately twice as costly as a normal gradient computation. Warm-starting the solver at each iteration, we find that with $k = 5$ iterations, a solution to the linear equation within numerical precision could be found. Thus each, leader gradient computation is roughly 10 times as costly as a normal gradient. For further details on the implementation, see Section 2.D.4.

For the MNIST experiment, we only evaluate the Stackelberg gradient dynamics. During training, the real data (images) $x \in \mathbb{R}^{28\times 28}$ is selected sequentially from a shuffled version of the dataset and the latent data $z \in \mathbb{R}^{100}$ is drawn from a standard normal distribution with batch sizes of 256. The neural network parameter weights are initialized from a zero-mean normal distribution with standard deviation 0.02 following the standard DCGAN configuration. The initial learning rate for each player $i \in \mathcal{I}$ is set to be 0.0002 and then they are decayed exponentially such that $\gamma_{i,k} = \gamma_i \nu_i^k$ where $\nu_1 = 1 - 10^{-5}$ and $\nu_2 = 1 - 10^{-7}$. Finally, the implicit map of the follower is regularized as discussed in Section 2.6.4 using the parameter $\mu = 10000$. If we view the regularization as a linear function of the number of parameters in the discriminator, then this selection of regularization is nearly equal to that from the mixture of Gaussian experiments.

We repeat the training procedure with 10 random seeds. We evaluate the performance using the Inception score (Salimans et al., 2016). The basic idea of the Inception score is to train a classifier on the dataset and then take samples of the generator to compute a conditional label distribution $\mathbb{P}(y|x)$ over the generated data. When the generated data has a meaningful output, the conditional label distribution should have low entropy and also the model should generate diverse images so that the marginal distribution $\mathbb{P}(y)$ has high entropy. The metric is then given by $\exp(\mathbb{E}_x(\mathrm{KL}(\mathbb{P}(y|x)||\mathbb{P}(y))))$. The Inception score was calculated using a LeNet classifier following (Berard et al., 2020). Each time we calculated a score $N = 5000$ samples were generated. Figure 2.15c shows the mean Inception scores over the 10 repeats of the experiment along the training path with the shading showing the standard error of the mean. Note that the Inception score using real data was 9.9 and the maximum possible Inception score is always the number of classes in the dataset. Finally, we show a real sample of images in Figure 2.15a and a fake sample in Figure 2.15a after 7500 batches from the run with the 5th highest Inception score. From the results, it can be observed that the Stackelberg gradient dynamics successfully generate handwritten digits that are indistinguishable from the real handwritten digits. Moreover, the Inception scores confirm this quantitatively and also show that the dynamics converge to a desirable solution fast and in a stable manner.

Module	In Channels	Out Channels	Kernel Size	Stride	Padding	Bias
ConvTranspose2d, BatchNorm2d, ReLU	100	512	4	1	0	False
ConvTranspose2d, BatchNorm2d, ReLU	512	256	4	2	1	False
ConvTranspose2d, BatchNorm2d, ReLU	256	128	4	2	1	False
ConvTranspose2d, BatchNorm2d, ReLU	128	512	4	1	0	False
ConvTranspose2d, BatchNorm2d, Tanh	64	1	1	1	2	False

Figure 2.16: Generator Network PyTorch Parameters for the nn.Sequential class in the MNIST experiment.

Module	In Channels	Out Channels	Kernel Size	Stride	Padding	Bias
Conv2d, LeakyReLU(0.2)	1	64	4	2	1	False
Conv2d, BatchNorm2d, LeakyReLU(0.2)	64	128	4	2	1	False
Conv2d, BatchNorm2d, LeakyReLU(0.2)	128	256	4	2	1	False
Conv2d, Sigmoid	256	1	4	2	1	False

Figure 2.17: Discriminator Network PyTorch Parameters for the nn.Sequential class in the MNIST experiment.

2.7 Discussion

This chapter presents a study of nonconvex games on continuous strategy spaces from a Stackelberg game perspective. The motivation for this perspective stems from the fact that despite the surge of research on equilibrium and gradient-based learning in this class of games, it has almost unanimously been approached from a simultaneous play point of view and focused on the Nash equilibrium solution concept. The Stackelberg game viewpoint is well-motivated both from the emergence of machine learning problems formulated as nonconvex-nonconcave zero-sum games that admit an implicit order of play, as well as from common game-theoretic applications where a hierarchical interaction structure between the players naturally emerges. The basis of this chapter is grounded upon the introduction of a local Stackelberg equilibrium concept suitable for this class of games and a gradient-based characterization via sufficient conditions (differential Stackelberg equilibrium) relevant to analyzing the convergence of gradient-based learning algorithms. The results in this chapter give significant insights into the properties of local Stackelberg equilibrium (genericity and structural stability) and the connections with local Nash equilibrium, namely that the latter is a subset of the former in zero-sum games and the cost is lower in general-sum games in the former than the latter for the leader. Moreover, for the subclass of nonconvex-nonconcave zero-sum games, we have shown that critical points of the simultaneous gradient dynamics previously thought to lack-meaning can often be explained through the lens of the Stackelberg equilibrium concept. The Stackelberg gradient dynamics that are developed are reflective of the interaction structure of a Stackelberg game, and consequently the convergence properties give insights into what can be expected from competitive interactions in hierarchical decision structures. The strongest convergence results are in zero-sum games, where it is shown that the Stackelberg gradient dynamics only locally converge to critical points that are local Stackelberg equilibrium, and do so at a rate that depends on fundamental properties of the equilibrium. While weaker, the convergence results in general-sum games highlight that natural Stackelberg dynamics may not always necessarily reach Stackelberg equilibrium, but under some suitable assumptions meaningful results can be given.

Finally the experimental results highlight the benefits in terms of stability that is gained by giving more power to a player in the game, and also underscore that local Stackelberg equilibrium play an important role in the optimization landscape of generative adversarial networks.

The work presented in this chapter first appeared in workshop form (Fiez et al., 2019a) before being published as a conference paper (Fiez et al., 2020a), and prior to that there was a technical report version (Fiez et al., 2019b) that includes some stochastic convergence results that were not included in the conference publication or this thesis due to the length. That being said, the presentation given in this chapter has been significantly revised from the previous write-ups to improve the clarity and also set the stage for the chapters that follow in this part of the thesis, which build closely off of the methods and results in this chapter. There has been a significant amount of research in the past couple of years that is closely related or builds directly on the work in this chapter. To begin, we remark that concurrently, Jin et al. (2020) also adopt the perspective of viewing the underlying interaction structure as a Stackelberg game. However, the focus is only on nonconvex-nonconcave zero-sum games and general-sum games are not considered. Notably, the authors also define a local minmax equilibrium concept for nonconvex-nonconcave zero-sum games and study its properties with more focus given to the issue of existence. We remark that the local Stackelberg equilibrium definition presented in this chapter is a direct local refinement of the typical global definition from game theory (Basar and Olsder, 1998). The local minmax equilibrium concept developed by Jin et al. (2020) has some subtle difference in terms of the condition for the leader. Specifically, recall that as presented in Definition 2.2, the local Stackelberg definition for the leader is formulated in terms of the best-response set of the follower. That is, x_1^* is a local Stackelberg solution for the leader if for all $x_1 \in U_1 \subset X_1$ the condition $\sup_{x_2 \in R_{U_2}(x_1^*)} f(x_1^*, x_2) \leq \sup_{x_2 \in R_{U_2}(x_2)} f(x_1, x_2)$ holds where $R_{U_2}(\cdot)$ is the set of best-responses of the follower in the neighborhood $U_2 \subset \mathcal{X}_2$ to a given leader strategy. This definition says that no local deviation of the leader considering the response-map of the follower can lower the cost. In contrast, the condition for the leader in the definition of Jin et al. (2020) requires that $f(x_1^*, x_2^*) \leq f(x_1, x_2)$ for all $x_1 \in U_1 \subset \mathcal{X}_1$ and $x_2 \in U_2 \subset \mathcal{X}_2$. The key difference is that along the local deviations of the leader, the local deviation of the follower is contained in a ball around the candidate equilibrium instead of constrained along the best-response map. To give some perspective of how this changes the local equilibrium definition, it is useful to see consider Example 2.1 and Figure 2.1, specifically Figure 2.1b. While the leader condition from Definition 2.2 is characterized by x_1^* being a local minimum of $f(x_1, R_{U_2}(x_1))$ along the response curve shown by the black line in Figure 2.1b, the condition presented in Jin et al. (2020) is characterized by (x_1^*, x_2^*) being a local minimum of $f(x_1, x_2)$ in a neighborhood, which in this toy example would effectively correspond to a rectangle around the equilibrium. In this sense, the local minmax equilibrium definition of Jin et al. (2020) can be slightly more restrictive than the local Stackelberg equilibrium definition in zero-sum games that we have presented. This being said, equivalent sufficient conditions are presented that define a strict local minmax equilibrium or equivalently a differential Stackelberg equilibrium. Moreover, Jin et al. (2020) similarly show any local Nash equilibrium is a local minmax equilibrium under the definition they adopt in zero-sum games. Finally, we remark that Jin et al. (2020) also study the simultaneous gradient dynamics (gradient descent-ascent in zero-sum games), but in addition consider the role of timescale separation. Like in this chapter, connections are drawn between the stable critical points and the set of strict local minmax equilibrium. While we discuss the results in further detail in the chapter that follows, the key insight is that as the timescale

parameter $\tau \to \infty$, then the sets of stable critical points and differential Stackelberg equilibrium coincide. This is an analogous result for the simultaneous gradient dynamics as we have shown in Theorem 2.5 for the Stackelberg gradient dynamics. However, it in some sense does not result in an implementable version of the simultaneous gradient dynamics with the guarantee in discrete-time, since the learning rate would need to tend to zero to retain the stability property. Nonetheless, it is an interesting result that motivates the work presented in the following chapter.

The research in this chapter and the work of Jin et al. (2020) has arguably resulted in an enlightened perspective of equilibrium concepts in the machine learning literature on learning in games. Indeed, there has been several equilibrium concepts proposed and studied since. The proximal equilibrium definition proposed by Farnia and Ozdaglar (2020) is (informally) suggested to interpolate between the set of Nash and Stackelberg equilibrium in zero-sum games. Moreover, Zhang et al. (2020a) propose a local robust point as an equilibrium definition that contains both local minmax and local maxmin equilibrium in zero-sum games. The consistent conjectural variations equilibrium and sufficient conditions has been investigated in our own work along with collaborators Chasnov et al. (2019). Notably, this is an equilibrium concept defined in terms of the behavior that players are conjecturing about each other. Finally, a pair of recent works (Keswani et al., 2020; Mangoubi and Vishnoi, 2021) define relaxed equilibrium notions that essentially correspond to the maximizing player reaching an approximate local maximum, and the minimizing player reaching an approximate local minimum of a 'greedy approximation' of the function $\max_{x_2} f(x_1, x_2)$. The benefit of this type of equilibrium notion is that it is shown to both exist under mild assumptions for nonconvex-nonconcave zero-sum games, while also being amenable to global finite-time convergence guarantees (Keswani et al., 2020; Mangoubi and Vishnoi, 2021). In some sense, the aforementioned set of works can be viewed as putting computational restrictions on the maximizing player to yield stronger theoretical results (that is, global versus local convergence and a guarantee of equilibrium existence), but with respect to weaker solution notions. In our own recent work with collaborators (Fiez et al., 2021a), a related perspective is studied in which the maximizing player is assumed to deploy a smooth algorithm so that the underlying nonconvex-nonconcave zero-sum game reduces to a nonconvex-concave zero-sum game in which global guarantees to stationary points are achievable. These equilibrium concepts are all with respect to the underlying function space. However, there is another perspective that has gained traction of viewing the game in parameter or model space and definining equilibrium concepts in that space (Flokas et al., 2021; Gidel et al., 2021; Vlatakis-Gkaragkounis et al., 2019).

Beyond equilibrium concepts, there has been several works developing gradient-based learning algorithms motivated by the Stackelberg gradient dynamics and the convergence guarantees with respect to differential Stackelberg equilibrium in zero-sum games. Specifically, Wang et al. (2020a) study learning dynamics they call follow-the-ridge and show the equivalent guarantee that in zero-sum games the only locally stable critical points are differential Stackelberg equilibrium. It is pointed out by Zhang et al. (2020b), that the follow-the-ridge dynamics are effectively a 'transposed' version of the Stackelberg dynamics. In particular, at critical points, the follow-the-ridge dynamics have an upper-block-triangular structure whereas the Stackelberg dynamics have a lower-block-triangular structure. Zhang et al. (2020a) study truly Newton-style learning dynamics that have a quadratic versus linear rate of local convergence and are also only locally stable around differential Stackelberg equilibrium. Most recently, a multi-player extension in which there can be N layers of hierarchical interactions with multiple players at each level is studied by Li et al. (2021) also using

gradient-based learning dynamics that are derived based on the implicit function theorem.

Recently, there has also been several works drawing the connection between a hierarchical interaction structure (Stackelberg game) and reinforcement learning problems. Indeed, Rajeswaran et al. (2020) argue that model-based reinforcement learning can be viewed as a general-sum Stackelberg game. That being said, to perform optimization in the game formulation of the reinforcement learning problem, they rely on the simultaneous gradient dynamics (k steps of unrolling the inner problem) with timescale separation rather than the Stackelberg gradient dynamics. On the other hand, in our own work with collaborators (Zheng et al., 2021), actor-critic reinforcement learning algorithms are cast as a nonconvex general-sum Stackelberg game and the Stackelberg gradient dynamics are employed in the optimization procedure. Empirically, it is shown that this method can outperform the simultaneous gradient dynamics that are typically used with timescale separation in actor-critical algorithms. A similar idea is analyzed theoretically by Hong et al. (2020).

The research in this chapter leaves several open questions and the following chapters address some of them. In particular, given the local stability characterization of the Stackelberg gradient dynamics from Theorem 2.5, which ensures that the only locally stable critical points are differential Stackelberg equilibrium and also that all differential Stackelberg equilibrium are locally stable, it of interest to investigate if an equivalent guarantee can be given for gradient-based dynamics using only first-order information. As discussed above, in some sense this has already been shown for the simultaneous gradient dynamics with timescale separation, but since it only holds in the limit as the timescale separation approaches infinity, it does not necessarily apply to practical implementations for machine learning problems. The following chapter seeks to resolve this question. Another interesting question is deriving global convergence guarantees in semi-structured classes of games for the Stackelberg gradient dynamics. Since the results in this chapter have shown that in zero-sum games the all stable critical points of the Stackelberg gradient dynamics are differential Stackelberg equilibrium and also that the Stackelberg gradient dynamics avoid strict saddle points of the continuous-time system, it is then possible to obtain global convergence guarantees to only differential Stackelberg equilibrium if a potential function can be constructed. Perhaps the most promising general, yet achievable class where this would be possible would be zero-sum games in which the $\det(D_2^2 f(x)) \neq 0$ everywhere. This would hold in nonconvex-strongly-convex games for example. This perspective is actually studied in general-sum games by Hong et al. (2020) and Ghadimi and Wang (2018), but not with a focus of converging to equilibrium necessarily. We do not address this question in this thesis, and leave it to future work. That being said, in Chapter 4, we do obtain results of this natural for the simultaneous gradient dynamics with timescale separation in classes of nonconvex zero-sum games. Finally, we remark that it would be interesting to theoretically analyze the sucecssful unrolled generative adversarial network training method proposed by (Metz et al., 2017). In unrolled generative adversarial networks (Metz et al., 2017), learning dynamics are studied for generative adversarial networks in which prior to updating the generator parameters, the discriminator is simulated taking k individual gradient steps from the initial parameters, while the parameters of the generator are held fixed, and then the simulated discriminator parameters are substituted into the objective of the generator and backpropagated through. As $k \to \infty$, the effect of 'unrolling' the simulated discriminator is that a surrogate objective of $f(x_1, r(x_1))$ arises for the generator where $r(x_1)$ is defined by the simulated discriminator converging to a local optimum. The discriminator then optimizes its cost taking individual gradient descent steps. Thus, this procedure induces an update reminiscent of the Stackelberg gradient dynamics but using a

simulated best-response $r(x_1)$ instead of the current iterate x_2 for the leader update as $k \to \infty$. Intuitively, the stability properties of the Stackelberg gradient dynamics should carry over to the unrolling procedure, which would imply that the unrolled generative adversarial network procedure only locally converges to differential Stackelberg equilibria in zero-sum games. Given the success of this training method, it would give further backing to local Stackelberg equilibrium as a desirable solution for generative adversarial networks.

CHAPTER 2 APPENDIX

2.A Proof of Proposition 2.1

Consider $x = (x_1, x_2) \in \mathcal{X}$ such that $D_2 f_2(x) = 0$ and $D_2^2 f_2(x) \succ 0$. The implicit function theorem (Abraham et al., 1988, Thm. 2.5.7) implies that there exists an implicit map $r : x_1 \mapsto x_2$ defined on a neighborhood $U_1 \subset \mathcal{X}_1$ such that $x_2 = r(x_1)$ and $Dr(x_1) = -(D_2^2 f_2(x_1, r(x_1)))^{-1} D_{21} f_2(x_1, r(x_1))$. The Schur complement of $J_{\mathcal{S}}(x_1, r(x_1)))$ is

$$\begin{aligned}\mathtt{S}_1(J_{\mathcal{S}}(x_1, r(x_1))) &= D_1(Df_1(x_1, r(x_1))) - D_2(Df_1(x_1, r(x_1)))(D_2^2 f_2(x_1, r(x_1)))^{-1} D_{21} f_2(x_1, r(x_1)) \\ &= D_1(Df_1(x_1, r(x_1))) + D_2(Df_1(x_1, r(x_1)))Dr(x_1).\end{aligned}$$

Observe that

$$\begin{aligned}D_1(Df_1(x_1, r(x_1))) &= D_1\big(D_1 f_1(x_1, r(x_1)) + Dr(x_1)^\top D_2 f_1(x_1, r(x_1))\big) \\ &= D_1^2 f_1(x_1, r(x_1)) + D_{21} f_1(x_1, r(x_1))^\top Dr(x_1) + D^2 r(x_1) D_2 f_1(x_1, r(x_1))^{11}\end{aligned}$$

and

$$\begin{aligned}D_2(Df_1(x_1, r(x_1)))Dr(x_1) &= D_2\big(D_1 f_1(x_1, r(x_1)) + Dr(x_1)^\top D_2 f_1(x_1, r(x_1))\big)Dr(x_1) \\ &= D_{12} f_1(x_1, r(x_1))Dr(x_1) + Dr(x_1)^\top D_2^2 f_1(x_1, r(x_1))Dr(x_1).\end{aligned}$$

Thus,

$$\begin{aligned}\mathtt{S}_1(J_{\mathcal{S}}(x_1, r(x_1))) = D_1^2 f_1(x_1, r(x_1)) &+ D_{21} f_1(x_1, r(x_1))^\top Dr(x_1) + D^2 r(x_1) D_2 f_1(x_1, r(x_1)) \\ &+ D_{12} f_1(x_1, r(x_1))Dr(x_1) + Dr(x_1)^\top D_2^2 f_1(x_1, r(x_1))Dr(x_1).\end{aligned}$$

Now, we also have that the total derivative of $Df_1(x_1, r(x_1))$ is given by

$$\begin{aligned}D(Df_1(x_1, r(x_1))) &= D\big(D_1 f_1(x_1, r(x_1)) + Dr(x_1)^\top D_2 f_1(x_1, r(x_1))\big) \\ &= D\big(D_1 f_1(x_1, r(x_1))\big) + D\big(Dr(x_1)^\top D_2 f_1(x_1, r(x_1))\big),\end{aligned}$$

where

$$D\big(D_1 f_1(x_1, r(x_1))\big) = D_1^2 f_1(x_1, r(x_1)) + D_{12} f_1(x_1, r(x_1))Dr(x_1),$$

and

$$\begin{aligned}D\big(Dr(x_1)^\top D_2 f_1(x_1, r(x_1))\big) &= D^2 r(x_1) D_2 f_1(x_1, r(x_1)) + Dr(x_1)^\top D_{21} f_1(x_1, r(x_1)) \\ &\quad + Dr(x_1)^\top D_2^2 f_1(x_1, r(x_1))Dr(x_1).\end{aligned}$$

Thus,

$$D(Df_1(x_1, r(x_1))) = D_1^2 f_1(x_1, r(x_1)) + D_{21} f_1(x_1, r(x_1))^\top Dr(x_1) + D^2 r(x_1) D_2 f_1(x_1, r(x_1)) \\ + D_{12} f_1(x_1, r(x_1)) Dr(x_1) + Dr(x_1)^\top D_2^2 f_1(x_1, r(x_1)) Dr(x_1).$$

Hence, we have that

$$D^2 f_1(x_1, r(x_1))|_{r(x_1)=x_2} = D(Df_1(x_1, r(x_1)))|_{r(x_1)=x_2} = \mathtt{S}_1(J_{\mathcal{S}}(x, r(x_1)))|_{r(x_1)=x_2}.$$

Moreover, in zero-sum games, since $D_2 f_1(x_1, x_2) = D_2 f(x_1, x_2) = 0$, we have that

$$D^2 r(x_1) D_2 f_1(x_1, r(x_1)) = D^2 r(x_1) D_2 f(x_1, r(x_1)) = 0$$

and

$$\begin{aligned} &Dr(x_1)^\top D_{21} f_1(x_1, r(x_1)) + Dr(x_1)^\top D_2^2 f_1(x_1, r(x_1)) Dr(x_1) \\ &\quad = Dr(x_1)^\top \big(D_{21} f(x_1, r(x_1)) + D_2^2 f(x_1, r(x_1)) Dr(x_1)\big) \\ &\quad = Dr(x_1)^\top \big(D_{21} f(x_1, r(x_1)) - D_2^2 f(x_1, r(x_1)) (D_2^2 f(x_1, r(x_1)))^{-1} D_{21} f(x_1, r(x_1))\big) \\ &\quad = Dr(x_1)^\top \big(D_{21} f(x_1, r(x_1)) - D_{21} f(x_1, r(x_1))\big) \\ &\quad = 0. \end{aligned}$$

Thus,

$$D(Df(x_1, r(x_1)))|_{r(x_1)=x_2} = \mathtt{S}_1(J(x_1, r(x_1)))|_{r(x_1)=x_2} = \mathtt{S}_1(J(x_1, x_2)).$$

The statement regarding the first-order sufficient conditions is immediate from the fact that in zero-sum games $Df(x) = D_1 f(x) - D_{21} f(x)^\top (D_2^2 f(x))^{-1} D_2 f(x)$ so that $D_1 f(x) = 0$ and $D_2 f(x) = 0$ if and only if $Df(x) = 0$ and $D_2 f(x) = 0$ given that $\det(D_2^2 f(x)) \neq 0$.

2.B Structural Stability and Genericity in Zero-Sum Games

This section is dedicated to showing that for a generic zero-sum q-differentiable game, all local Stackelberg equilibrium of the game are differential Stackelberg equilibrium, and further they are structurally stable. It is worth noting that analogous results are known for local Nash equilibrium and differential Nash equilibrium (Mazumdar and Ratliff, 2019; Ratliff et al., 2014). We begin with mathematical preliminaries, then provide the proof of Theorem 2.1 (genericity), and finally the proof of Theorem 2.2 (structural stability).

2.B.1 Mathematical Preliminaries

We now provide some additional mathematical preliminaries; the interested reader should see standard references for a more detailed introduction (Abraham et al., 1988; Lee, 2012).

A *smooth manifold* is a topological manifold with a smooth atlas. Euclidean space, as considered in this chapter, is a smooth manifold. For a vector space E, we define the vector space of continuous

$(p+s)$–multilinear maps $T^p_s(E) = L^{p+s}(E^*, \ldots, E^*, E, \ldots, E; \mathbb{R})$ with s copies of E and p copies of E^* and where E^* denotes the dual. Elements of $T^p_s(E)$ are *tensors* on E, and $T^p_s(X)$ denotes the vector bundle of such tensors (Abraham et al., 1988, Definition 5.2.9).

Consider smooth manifolds X and Y of dimension d_x and d_y respectively. An *k–jet* from X to Y is an equivalence class $[x, f, U]_k$ of triples (x, f, U) where $U \subset X$ is an open set, $x \in U$, and $f : U \to Y$ is a C^k map. The equivalence relation satisfies $[x, f, U]_k = [y, g, V]_k$ if $x = y$ and in some pair of charts adapted to f at x, f and g have the same derivatives up to order k. We use the notation $[x, f, U]_k = j^k f(x)$ to denote the k–jet of f at x. The set of all k–jets from X to Y is denoted by $J^k(X, Y)$. The jet bundle $J^k(X, Y)$ is a smooth manifold (see Hirsch, 1976, Chapter 2 for the construction). For each C^k map $f : X \to Y$ we define a map $j^k f : X \to J^k(X, Y)$ by $x \mapsto j^k f(x)$ and refer to it as the *k–jet extension*.

Definition 2.6. *Let X, Y be smooth manifolds and $f : X \to Y$ be a smooth mapping. Let Z be a smooth submanifold of Y and u a point in X. Then f* intersects Z transversally at u *(denoted $f \pitchfork Z$ at u) if either $f(u) \notin Z$ or $f(u) \in Z$ and $T_{f(u)}Y = T_{f(u)}Z + (f_*)_u(T_u X)$.*

For $1 \leq k < s \leq \infty$ consider the jet map $j^k : C^s(X, Y) \to C^{s-k}(X, J^k(X, Y))$ and let $Z \subset J^k(X, Y)$ be a submanifold. Define

$$\bigcap^s(X, Y; j^k, Z) = \{h \in C^s(X, Y) |\ j^k h \pitchfork Z\}. \tag{2.21}$$

A subset of a topological space X is *residual* if it contains the intersection of countably many open–dense sets. We say a property is *generic* if the set of all points of X which possess this property is residual (Broer and Takens, 2010).

Theorem 2.12. *(Jet Transversality Theorem, Hirsch, 1976, Chapter 2). Let X, Y be C^∞ manifolds without boundary, and let $Z \subset J^k(X, Y)$ be a C^∞ submanifold. Suppose that $1 \leq k < s \leq \infty$. Then, $\bigcap^s(X, Y; j^k, Z)$ is residual and thus dense in $C^s(X, Y)$ endowed with the strong topology, and open if Z is closed.*

Proposition 2.12. *(Golubitsky and Guillemin, 1973, Chapter II.4, Proposition 4.2). Let X, Y be smooth manifolds and $Z \subset Y$ a submanifold. Suppose that $\dim(X) < \operatorname{codim}(Z)$. Let $f : X \to Y$ be smooth and suppose that $f \pitchfork Z$. Then, $f(X) \cap Z = \emptyset$.*

2.B.2 Genericity: Proof of Theorem 2.1

We show that local Stackelberg equilibria of zero-sum games are generically differential Stackelberg equilibria. Towards this end, we utilize the well-known fact that non-degeneracy of critical points is a generic property of sufficiently smooth functions.

Lemma 2.2 (Broer and Takens, 2010, Chapter 1)**.** *For $C^q(\mathbb{R}^d, \mathbb{R})$ functions with $q \geq 2$ it is a generic property that all the critical points are non-degenerate.*

The above lemma implies that for a generic function $f \in C^q(\mathbb{R}^d, \mathbb{R})$, the Hessian

$$H(x) = \begin{bmatrix} D^2_1 f(x) & \cdots & D_{1d} f(x) \\ \vdots & \ddots & \vdots \\ D_{d1} f(x) & \cdots & D^2_d f(x) \end{bmatrix}$$

is non-degenerate at critical points—that is, $\det(H(x)) \neq 0$. For the following result recall that the Jacobian $J(x)$ of $g(x) = (D_1 f_1(x), Df_2(x))$ is defined in (2.3) and that the Jacobian $J_{\mathcal{S}}(x)$ of $g_{\mathcal{S}}(x) = (Df_1(x), D_2 f_2(x))$ is defined in (2.5).

Lemma 2.3. *Consider $f \in C^q(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}, \mathbb{R})$, $q \geq 2$ and the corresponding zero-sum game $\mathcal{G} = (f, -f)$. For any critical point defined by $x \in \mathbb{R}^d$ such that $g(x) = 0$, $\det(H(x)) \neq 0 \iff \det(J(x)) \neq 0$ and, if $\det(-D_2^2 f(x)) \neq 0$, then $g_{\mathcal{S}}(x) = 0$ and $\det(H(x)) \neq 0 \iff \det(J_{\mathcal{S}}(x)) \neq 0$.*

Proof. Consider a fixed $x = (x_1, x_2) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Note that $H(x)$ is equal to $J(x)$ with the last d_2 rows scaled each by -1. Hence, $J(x) = PH(x)$ where $P = \text{blockdiag}(I_{d_1}, -I_{d_2})$ with each I_{d_i} the $d_i \times d_i$ identity matrix, so that $\det(H(x)) = (-1)^{d_2} \det(J(x))$, which in turn proves the first equivalence. For the second equivalence, suppose that $\det(-D_2^2 f(x)) \neq 0$ so that $J_{\mathcal{S}}(x)$ is well-defined. Then, by Lemma 2.1, $g_{\mathcal{S}}(x) = 0$ and using the Schur decomposition of $J(x)$ it is easily seen that $\det(J(x)) = \det(\mathtt{S}_1(J(x))) \det(-D_2^2 f(x)) = \det(J_{\mathcal{S}}(x))$ where the last equality holds since $J_{\mathcal{S}}(x)$ is a lower block triangular matrix with $\mathtt{S}_1(J(x))$ and $-D_2^2 f(x)$ on the diagonal at critical points (recall from the proof of Theorem 2.5). Hence, the result. □

This equivalence between the non-degeneracy of the Hessian and the game Jacobian J (and the relationship to the determinant of $J_{\mathcal{S}}$ via the Schur decomposition) allows us to lift the generic property of non-degeneracy of critical points of functions to critical points of the Stackelberg learning dynamics. The Jet Transversality Theorem and Proposition 2.12 can be used to show a subset of a jet bundle having a particular set of desired properties is generic. Indeed, consider the jet bundle $J^k(X, Y)$ and recall that it is a manifold that contains jets $j^k f : X \to J^k(X, Y)$ as its elements where $f \in C^k(X, Y)$. Let $Z \subset J^k(X, Y)$ be the submanifold of the jet bundle that *does not* possess the desired properties. If $\dim X < \text{codim } Z$, then for a generic function $f \in C^k(X, Y)$ the image of the k–jet extension is disjoint from Z implying that there is an open–dense set of functions having the desired properties. Without loss of generality, we let player 1 be the leader.

Lemma 2.4. *Consider $f \in C^q(\mathbb{R}^{d_1+d_2}, \mathbb{R})$, $q \geq 2$ such that $D_i^2 f \in \mathbb{R}^{d_i \times d_i}$. It is a generic property that $\det(D_i^2 f(x)) \neq 0$ for any $i \in \mathcal{I} = \{1, 2\}$.*

Proof. Let us start with $f \in C^q$ with $q \geq 3$. First, critical points of a function f are such that $D_i f(x) = 0$, $i = 1, 2$. Furthermore, the $J^2(\mathbb{R}^d, \mathbb{R})$ bundle associated to f is diffeomorphic to $\mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^{d(d+1)/2}$ and the 2-jet extension of f at any point $x \in \mathbb{R}^d$ is given by $(x, f(x), Df(x), D^2 f(x))$.

Now, let us denote by $S(k)$ the space of $k \times k$ symmetric matrices, and consider the subset of $J^2(\mathbb{R}^d, \mathbb{R})$ defined by

$$\mathcal{D}_i = \mathbb{R}^d \times \mathbb{R} \times \{0\} \times Z_i(d_i) \times \mathbb{R}^{d_1 \times d_2} \times S(d - d_i)$$

where $Z(d_i) = \{A \in S(d_i) |\ \det(A) = 0\}$. The set $Z(d_i)$ is algebraic; hence, we can use the Whitney stratification theorem (Gibson et al., 1976, Ch. 1, Thm. 2.7) to get that each $Z(d_i)$ is the union of submanifolds of co-dimension at least 1. Hence, it is the union of sub-manifolds of codimension at least one and, in turn, $\mathcal{D}_i$ is the union of sub-manifolds of codimension at least $d + 1$. Thus, it follows from the Jet Transversality Theorem 2.12 (by way of Proposition 2.12 since $d + 1 > d$) that for a generic function f, the image of the 2-jet extension $j^2 f$ is disjoint from $\mathcal{D}_i$. Hence, for such an f, for each x that is a critical point, the Hessian of f is such that $\det(D_i^2 f(x)) \neq 0$.

Furthermore, if we consider the subset $\mathcal{D} \subset J^2(\mathbb{R}^d, \mathbb{R})$ defined by

$$\mathcal{D} = \mathbb{R}^d \times \mathbb{R} \times \{0\} \times Z(d_i) \times \mathbb{R}^{d_1 \times d_2} \times Z(d - d_i),$$

then both $Z(d_i)$ and $Z(d - d_i)$ are algebraic, and so they each are of co-dimension at least one. In turn, $\mathcal{D}$ is the the union of sub-manifolds of codimension at least $d + 2$. Applying the Jet Transversality Theorem 2.12 again, we get that for such an f, for each x that is a critical point, the Hessian of f is such that $\det(D_i^2 f(x)) \neq 0$ for $i \in \{1, 2\}$.

The extension to the $q \geq 2$ setting follows directly from the fact that non-degeneracy is an open condition in the C^2 topology, and any function can be C^2 approximated by a C^3 function (see, e.g., Hirsch, 1976, Thm. 2.4), which can then be approximated by a function without critical points such that $\det(D_i^2 f(x)) = 0$ (by the above argument), which we will call coordinate degenerate. This, in turn, implies that functions without critical points that are coordinate-degenerate are dense in the C^2 space of functions. □

While the theorems we leverage from differential geometry and dynamical systems theory are similar, the architecture of the proof of the following theorem deviates quite a bit from (Mazumdar and Ratliff, 2019; Ratliff et al., 2014) due to the hierarchical structure of the game.

Proof of Theorem 2.1. Let $J^2(\mathbb{R}^d, \mathbb{R})$ denote the second-order jet bundle containing 2–jets $j^2 f$ such that $f : \mathbb{R}^d \to \mathbb{R}$. Then, $J^2(\mathbb{R}^d, \mathbb{R})$ is locally diffeomorphic to $\mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^{\frac{d(d+1)}{2}}$ and the 2–jet extension of f at any point $x \in \mathbb{R}^d$ is given by $(x, f(x), Df(x), D^2 f(x))$. By Lemma 2.4, we know that

$$\mathcal{D}_2 = \mathbb{R}^d \times \mathbb{R} \times \{0\} \times Z(d_2) \times \mathbb{R}^{d_1 + d_2} \times S(d_1)$$

has co-dimension at least $d+1$ in $J^2(\mathbb{R}^d, \mathbb{R})$ so that there exists an open dense set of functions $\mathcal{F}_2 \subset C^q(\mathbb{R}^d, \mathbb{R})$ such that $\det(D_2^2 f(x)) \neq 0$ at critical points (that is, where $(D_1 f(x), D_2 f(x)) = (0, 0)$).

Now, we also know that there is an open dense set of functions $\mathcal{F}_1$ in $C^q(\mathbb{R}^d, \mathbb{R})$ such that $\det(H(x)) \neq 0$ at critical points. The intersection of open dense sets is open dense. Let $\mathcal{F} = \mathcal{F}_1 \cap \mathcal{F}_2$. Now, for any $f \in \mathcal{F}$, we have that at critical points $\det(H(x)) \neq 0$ and $\det(D_2^2 f(x)) \neq 0$. Hence, by Lemma 2.3, $\det(J_{\mathcal{S}}(x)) \neq 0$ for all $f \in \mathcal{F}$, and in particular, $\det(\mathtt{S}_1(J(x))) \neq 0$.

For all functions $f \in \mathcal{F}$, the critical points of $g_{\mathcal{S}}(x) = (Df(x), -D_2 f(x))$ coincide with the critical points of the function f. Indeed,

$$(D_1 f(x), D_2 f(x)) = (0, 0) \iff (Df(x), -D_2 f(x)) = (0, 0)$$

since for all $f \in \mathcal{F}$, $\det(D_2^2 f(x)) \neq 0$ and $D_2 f(x) = 0$ so that the C^q implicit map at a critical point $D_2 f(x) = 0$ is well-defined.

Thus, we have constructed an open dense set $\mathcal{F} \subset C^q(\mathbb{R}^d, \mathbb{R})$ such that for all $f \in \mathcal{F}$, if $x \in \mathbb{R}^d$ is a local Stackelberg equilibrium for $(f, -f)$, then x is a differential Stackelberg equilibrium. Indeed, suppose $f \in \mathcal{F}$ and $x \in \mathbb{R}^d$ is a local Stackelberg equilibrium. Then, a necessary condition is that $-D_2 f(x) = 0$ and $-D_2^2 f(x) \succeq 0$. However, since $f \in \mathcal{F}$, we have that $\det(-D_2^2 f(x)) = (-1)^{d_2} \det(D_2^2 f(x)) \neq 0$ so that, in fact, $-D_2^2 f(x) \succ 0$. Hence, a local Stackelberg equilibrium such that $-D_2 f(x) = 0$ and $-D_2^2 f(x) \succ 0$ necessarily satisfies $g_{\mathcal{S}}(x) = 0$ and $\mathtt{S}_1(J(x)) \succeq 0$. However, again since $f \in \mathcal{F}$, and $\det(\mathtt{S}_1(J(x))) \neq 0$ so that, in fact, $\mathtt{S}_1(J(x))(x) \succ 0$. Furthermore, due to the lower block triangular structure of $J_{\mathcal{S}}(x)$, $\det(-D_2^2 f(x)) = (-1)^{d_2} \det(D_2^2 f(x)) \neq 0$ and

$\det(\mathtt{S}_1(J(x))) \neq 0$ also imply that $\det(J_{\mathcal{S}}(x)) \neq 0$, which completes the proof.

2.B.3 Structural Stability: Proof of Theorem 2.2

Let $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Define the smoothly perturbed cost function $\tilde{f} : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ by $\tilde{f}(x, y, t) = f(x, y) + t\zeta(x, y)$, and $\widetilde{g}_{\mathcal{S}} : \mathbb{R}^d \times \mathbb{R} \to T^*(\mathbb{R}^d)$ by $\widetilde{g}_{\mathcal{S}}(x, y, t) = (D\tilde{f}(x, y)), -D_2\tilde{f}(x, y))$, for all $t \in \mathbb{R}$ and $(x, y) \in \mathbb{R}^d$. Since (x_1, x_2) is a differential Stackelberg equilibrium, $D\tilde{g_{\mathcal{S}}}(x, y, 0)$ is necessarily non-degenerate. Invoking the implicit function theorem (Abraham et al., 1988, Thm. 2.5.7), there exists neighborhoods $V \subset \mathbb{R}$ of zero and $W \subset \mathbb{R}^d$ and a smooth function $\sigma \in C^r(V, W)$ such that for all $t \in V$ and $(x_1, x_2) \in W$, $\tilde{g_{\mathcal{S}}}(x_1, x_2, s) = 0 \iff (x_1, x_2) = \sigma(t)$. Since $\tilde{g_{\mathcal{S}}}$ is continuously differentiable, there exists a neighborhood $U \subset W$ of zero such that $D\tilde{g_{\mathcal{S}}}(\sigma(t), t)$ is invertible for all $t \in U$. Thus, for all $t \in U$, $\sigma(t)$ must be the unique local Stackelberg equilibrium of $(f + t\zeta|_W, -f - t\zeta|_W)$.

2.C Proofs of Propositions 2.6 and 2.7

Let us begin by recalling the notation for this set of results and a preliminary technical result.

Notation and Preliminaries. Let $x_1 \in \mathbb{R}^{d_1}$ and $x_2 \in \mathbb{R}^{d_2}$. For a stable critical point $x^* = (x_1^*, x_2^*)$ of $\dot{x} = -g(x)$ that is not a differential Nash equilibrium and is such that $-D_2^2 f(x^*) \succ 0$, let $\operatorname{spec}(D_1^2 f(x^*)) = \{\mu_j,\ j \in [d_1]\}$ where $\mu_1 \leq \cdots \leq \mu_m \leq 0 < \mu_{m+1} < \cdots \leq \mu_{d_1}$, and let $\operatorname{spec}(-D_2^2 f(x^*)^{-1}) = \{\alpha_j,\ j \in [d_2]\}$ where $\alpha_1 \geq \cdots \geq \alpha_{d_2} > 0$, and define $p = \dim(\ker(D_1^2 f(x^*)))$.[12] For a matrix W, let $W^\dagger$ denote the conjugate transpose, $|W| := (W^\top W)^{1/2}$, and $\lambda_j^{\downarrow}(\cdot)$ as an operator returning the j-th eigenvalue counted with multiplicities of a Hermitian matrix when arranged in a non-increasing order. To prove Propositions 2.6 and 2.7, we need the following well-known result from linear algebra.

Lemma 2.5 (Berger et al. 2019, Proposition 2.2)**.** *Let $W \in \mathbb{C}^{d_2 \times d_2}$ be Hermitian with k positive eigenvalues counted with multiplicities and let $U \in \mathbb{C}^{d_1 \times d_2}$. Then,*

$$\lambda_j^{\downarrow}(UWU^\dagger) \leq \|U\|^2 \lambda_j^{\downarrow}(W) \quad \textit{for all} \quad 1 \leq j \leq \min\{k, d_1, \operatorname{rank}(UWU^\dagger)\}.$$

2.C.1 Proof of Proposition 2.6

For the sake of presentation, define $A = D_1^2 f(x^*)$, $B = D_{12} f(x^*)$, and $C = D_2^2 f(x^*)$. Recall that $A - BC^{-1}B^\top \succ 0$ and $-C \succ 0$ by assumption. That is, x^* is a differential Stackelberg equilibrium. We now show that given there is a $\kappa > 0$ such that $\|B\| \leq \kappa$, it is necessary that $m \leq d_2$ and $\mu_j + \kappa^2 \alpha_j > 0$ for all $j \in [m - p]$.

Claim: $m \leq d_2$ is necessary. This proof follows the main arguments in Berger et al. (2019, Lemma 3.2) with some minor changes due to the nature of the problem. Assume for the sake of contradiction that $m > d_2$. That is, the number of negative eigenvalues of A is greater than the

[12] Note that the eigenvalues are being counted with multiplicities.

dimension of C. Observe that if $d_1 \leq d_2$, then this is not possible since $m \leq d_1$. Thus, in this scenario, the contradiction $m \leq d_2$ is automatically obtained. Now suppose that $d_1 \geq m > d_2$. Let $\mathcal{S}_1 = \ker(B(-C^{-1} + |C^{-1}|)B^\top)$ and observe by the rank-nullity theorem and using the fact that C is full-rank,

$$\dim(\mathcal{S}_1) = d_1 - \operatorname{rank}(B(-C^{-1} + |C^{-1}|)B^\top) \geq d_1 - \operatorname{rank}(-C^{-1} + |C^{-1}|) = d_1 - d_2.$$

Moreover, consider the subspace $\mathcal{S}_2$ of $\mathbb{C}^{d_1}$ spanned by all the eigenvectors of A corresponding to non-positive eigenvalues. By assumption, we have that $\dim(\mathcal{S}_2) = m > d_2$ so that

$$\dim(\mathcal{S}_1 + \mathcal{S}_2) + \dim(\mathcal{S}_1 \cap \mathcal{S}_2) = \dim(\mathcal{S}_1) + \dim(\mathcal{S}_2) \geq (d_1 - d_2) + m = d_1 + (m - d_2) > d_1.$$

Thus, $\mathcal{S}_1 \cap \mathcal{S}_2 \neq \{0\}$, since the previous equation implies that

$$\dim(\mathcal{S}_1 \cap \mathcal{S}_2) > d_1 - \dim(\mathcal{S}_1 + \mathcal{S}_2) \geq 0.$$

Consequently, for any non-trivial vector $v \in \mathcal{S}_1 \cap \mathcal{S}_2$,

$$B(-C^{-1} + |C^{-1}|)B^\top)v = 0 \implies BC^{-1}B^\top = B|C^{-1}|B^\top.$$

Hence, we have

$$\begin{aligned}\langle (A - BC^{-1}B^\top)v, v\rangle &= \langle Av, v\rangle - \langle BC^{-1}B^\top v, v\rangle \\ &= \langle Av, v\rangle - \langle B|C^{-1}|B^\top v, v\rangle \\ &\leq 0.\end{aligned}$$

Note that the inequality holds since the vector v is in the non-positive eigenspace of A and the second term is clearly non-positive since $|C^{-1}|$ has positive eigenvalues. Thus, $A - BC^{-1}B^\top$ cannot be positive definite, which gives a contradiction so that it must be $m \leq d_2$.

Claim: $\mu_j + \kappa^2\alpha_j > 0$ for $j \in [m]$ is necessary. This proof follows the main arguments in Berger et al. (2019, Theorem 3.3) with some minor changes. By the assumption that $A - BC^{-1}B^\top \succ 0$ and the Weyl theorem for Hermitian matrices,[13] we have that for all $j \in [d_1]$:

$$0 < \lambda^{\downarrow}_{d_1}(A - BC^{-1}B^\top) \leq \lambda^{\downarrow}_j(A) + \lambda^{\downarrow}_{d_1-j+1}(-BC^{-1}B^\top) = \mu_{d_1-j+1} + \lambda^{\downarrow}_{d_1-j+1}(-BC^{-1}B^\top). \quad (2.22)$$

Observe that this implies

$$\lambda^{\downarrow}_j(-BC^{-1}B^\top) > 0 \quad \forall\ j \in [d_1] : \mu_j \leq 0.$$

Hence, this directly gives a lower bound on the rank of the matrix $-BC^{-1}B^\top$. Specifically,

$$\operatorname{rank}(-BC^{-1}B^\top) \geq m.$$

---

[13] Note that this application of Weyl's theorem is stated with the eigenvalues arranged in a non-increasing order (Bhatia, 2013, Theorem III.2.1), whereas sometimes it is stated with the eigenvalues being ordered in a non-decreasing order (Horn and Johnson, 2011).

Moreover, applying Lemma 2.5 using the facts that $-C^{-1}$ has d_2 positive eigenvalues and that

$$\min\{d_2, d_1, \operatorname{rank}(-BC^{-1}B^\top)\} \geq \min\{d_2, d_1, m\} = m$$

since by definition $m \leq d_1$ and $m \leq d_2$ was shown to be necessary, along with the assumption $\|B\| \leq \kappa$, we have for all $1 \leq d_1 - j + 1 \leq m$:

$$\lambda^{\downarrow}_{d_1-j+1}(-BC^{-1}B^\top) \leq \|B\|^2 \lambda^{\downarrow}_{d_1-j+1}(-C^{-1}) \leq \kappa^2 \lambda^{\downarrow}_{d_1-j+1}(-C^{-1}) = \kappa^2 \alpha_{d_1-j+1}. \tag{2.23}$$

Observe that the the following conditions are equivalent:

$$1 \leq d_1 - j + 1 \leq m \iff d_1 + 1 - m \leq j \leq d_1.$$

Thus, combining (2.22) and (2.23), it follows that for all $j \in \{d_1 + 1 - m, \dots, d_1\}$:

$$0 < \mu_{d_1-j+1} + \kappa^2 \alpha_{d_1-j+1}.$$

Equivalently, for all $j \in \{1, \dots, m\}$:

$$0 < \mu_j + \kappa^2 \alpha_j.$$

This proves the claim. Since we have shown both the necessary conditions, this concludes the proof.

2.C.2 Proof of Proposition 2.7

Since $D_1^2 f(x^*)$ and $D_2^2 f(x^*)$ are both symmetric, let $D_1^2 f(x^*) = W_1 M W_1^\dagger$ with $W_1 W_1^\dagger = I_{d_1 \times d_1}$ and $M = \operatorname{diag}(\mu_1, \dots, \mu_{d_1})$, and $-D_2^2 f(x^*) = W_2 \Lambda W_2^\dagger$ with $W_2 W_2^\dagger = I_{d_2 \times d_2}$ and $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_{d_2})$. By assumption, there exists a diagonal matrix $\Sigma \in \mathbb{R}^{d_1 \times d_2}$ such that $D_{12} f(x^*) = W_1 \Sigma W_2^\dagger$ where W_1 are the orthonormal eigenvectors of $D_1^2 f(x^*)$ and W_2 are orthonormal eigenvectors of $-D_2^2 f(x^*)$. Thus,

$$\begin{aligned}
\mathtt{S}_1(J(x^*)) &= D_1^2 f(x^*) - D_{12} f(x^*)(D_2^2 f(x^*))^{-1} D_{12}^\top f(x^*) \\
&= W_1 M W_1^\dagger + W_1 \Sigma W_2^\dagger (W_2 \Lambda W_2^\dagger)^{-1} W_2 \Sigma^\dagger W_1^\dagger \\
&= W_1 (M + \Sigma \Lambda^{-1} \Sigma^\dagger) W_1^\dagger
\end{aligned}$$

Hence, to understand the eigenstructure, we simply need to compare the all negative eigenvalues of $D_1^2 f(x^*)$ in increasing order with the most positive eigenvalues of $-D_2^2 f(x^*)$ in decreasing order. Indeed, by assumption, $m \leq d_2$ and $\mu_j + \kappa^2 \lambda_j > 0$ for each $j \in [m - p]$. Thus,

$$D_1^2 f(x^*) - D_{12} f(x^*)(D_2^2 f(x^*))^{-1} D_{12}^\top f(x^*) \succ 0.$$

Combining this with the fact that $-D_2^2 f(x^*) \succ 0$, x^* is a differential Stackelberg equilibrium.

2.D Proofs of Deterministic Convergence Results

In the following subsections, we provide the proofs of the deterministic convergence guarantees in zero-sum games, then the convergence guarantees in general-sum games, and finally almost sure avoidance of strict saddles in zero-sum games.

2.D.1 Zero-Sum Convergence: Proofs of Theorem 2.7 and Corollary 2.1

Under appropriate choices on the step-size so that the local linearization of the update is a contraction, standard arguments from numerical analysis for dynamical systems give rise to a guarantee on local asymptotic convergence (including a rate of convergence), and a finite-time convergence guarantee to an ε–differential Stackelberg equilibrium.

2.D.1.1 Proof of Theorem 2.7

Fix $\gamma_1 = 1/(2\beta)$. The structure of the Jacobian $J_{\mathcal{S}_\tau}(x^*)$ is lower-block triangular, with symmetric components along the diagonal given by $\mathtt{S}_1(J(x))$ and $-\tau D_2^2 f(x)$. From this structure, we know that $\operatorname{spec}(J_{\mathcal{S}_\tau}(x)) = \operatorname{spec}(\mathtt{S}_1(J(x))) \cup \operatorname{spec}(-\tau D_2^2 f(x))$. Then, by the spectral mapping theorem and the fact that the eigenvalues of $J_{\mathcal{S}_\tau}(x^*)$ are real,

$$\max_{\lambda \in \operatorname{spec}(I - \gamma_1 J_{\mathcal{S}_\tau}(x^*))} |\lambda| = |1 - \gamma_1 \min\{\lambda_{\min}(\mathtt{S}_1(J(x^*))), \lambda_{\min}(-\tau D_2^2 f(x^*))\}| = |1 - \gamma_1 \alpha|.$$

so that $\rho(I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)) < 1$ since $\alpha \leq \beta$. We also note that $\lambda_{\min}(I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)) = 1 - \gamma_1 \max\{\lambda_{\max}(\mathtt{S}_1(J(x^*))), \lambda_{\max}(-\tau D_2^2 f(x^*))\} \geq 0$ for the choice of γ_1 so that, in fact, $\operatorname{spec}(I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)) \subset [0, 1 - \frac{\alpha}{2\beta}]$.

Since $\rho(I - \gamma_1 J_{\mathcal{S}_\tau}) \leq 1 - \frac{\alpha}{2\beta}$, given $\varepsilon > 0$, there exists a matrix norm $\|\cdot\|$ such that $\|I - \gamma_1 J_{\mathcal{S}_\tau}\| \leq 1 - \frac{\alpha}{2\beta} + \varepsilon$ (Horn and Johnson, 2011, Lemma 5.6.10).[14] Consider $\varepsilon = \frac{\alpha}{8\beta}$ so that there exists $\|\cdot\|$ such that $\|I - \gamma_1 J_{\mathcal{S}_\tau}\| \leq 1 - \frac{3\alpha}{8\beta}$. Taking the Taylor expansion of $I - \gamma_1 g_{\mathcal{S}_\tau}(x)$ in a neighborhood of x^*, we have

$$I - \gamma_1 g_{\mathcal{S}_\tau}(x) = (I - \gamma_1 g_{\mathcal{S}_\tau}(x^*)) + (I - \gamma_1 J_{\mathcal{S}_\tau}(x^*))(x - x^*) + R_1(x - x^*)$$

where $R_1(x - x^*)$ is the remainder term satisfying $R_1(x - x^*) = o(\|x - x^*\|)$.[15] This, in turn, implies that there is a $\delta > 0$ such that $\|R_1(x - x^*)\| \leq \frac{\alpha}{8\beta}\|x - x^*\|$ whenever $\|x - x^*\| < \delta$. Hence,

$$\|I - \gamma_1 g_{\mathcal{S}_\tau}(x) - (I - \gamma_1 g_{\mathcal{S}_\tau}(x^*))\| \leq \Big(\|I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)\| + \frac{\alpha}{8\beta}\Big)\|x - x^*\| \leq \Big(1 - \frac{\alpha}{4\beta}\Big)\|x - x^*\|$$

Thus, for any $x_0 \in \{x | \ \|x - x^*\| < \delta\}$,

$$\|x_k - x^*\| \leq \Big(1 - \frac{\alpha}{4\beta}\Big)^k \|x_0 - x^*\|. \tag{2.24}$$

---

[14] In fact, one can show that there is both a weighted induced 1-norm and weighted induced ∞-norm works for any ε so that the construction is not unreasonable.

[15] The notation $R_1(x - x^*) = o(\|x - x^*\|)$ as $x \to x^*$ means $\lim_{x \to x^*} \|R_1(x - x^*)\|/\|x - x^*\| = 0$.

Hence, the iteration complexity (or rate of convergence) is $O((1-\alpha/(4\beta))^k)$, since all finite dimensional norms are equivalent.

2.D.1.2 Proof of Corollary 2.1

The proof of Corollary 2.1 follows directly from the conclusion of Theorem 2.7. Following standard arguments, (2.24) in the proceeding proof implies a finite time convergence guarantee. Indeed, let $\varepsilon > 0$ be given. Since $0 < \frac{\alpha}{4\beta} < 1$ we have that $(1 - \frac{\alpha}{4\beta})^k < \exp(-\frac{k\alpha}{4\beta})$. Hence,

$$\|x_k - x^*\| \leq \exp(-k\alpha/(4\beta))\|x_0 - x^*\|$$

This, in turn implies that $x_k \in B_\varepsilon(x^*)$, meaning x_k is a ε-differential Stackelberg equilibrium for all $k \geq \lceil \frac{4\beta}{\alpha} \log(\|x_0 - x^*\|/\varepsilon) \rceil$ whenever $\|x_0 - x^*\| < \delta$.

Now, given that $f_i \in C^q(X, \mathbb{R})$ for $q \geq 2$, $I - \gamma_1 J_{\mathcal{S}_\tau}(x)$ is locally Lipschitz with constant L so that we can find an explicit expression for δ in terms of L. Indeed, recall that $R_1(x - x^*) = o(\|x - x^*\|)$ as $x \to x^*$ which means $\lim_{x \to x^*} \|R_1(x - x^*)\|/\|x - x^*\| = 0$ so that

$$\|R_1(x - x^*)\| \leq \int_0^1 \|I - \gamma_1 J_{\mathcal{S}_\tau}(x^* + \eta(x - x^*)) - (I - \gamma_1 J_{\mathcal{S}_\tau}(x^*))\|\|x - x^*\| \, d\eta \leq \frac{L}{2}\|x - x^*\|^2$$

Observing that

$$\|R_1(x - x^*)\| \leq \frac{L}{2}\|x - x^*\|^2 = \frac{L}{2}\|x - x^*\|\|x - x^*\|,$$

we have that the $\delta > 0$ such that $\|R_1(x - x^*)\| \leq \alpha/(8\beta)\|x - x^*\|$ is $\delta = \alpha/(4L\beta)$.

2.D.2 General-Sum Convergence: Proofs of Theorem 2.6 and Corollary 2.8

Consider a general sum setting defined by $f_i \in C^q(X, \mathbb{R})$ with $q \geq 2$ for $i \in \mathcal{I}$ and where, without loss of generality, player 1 is the leader and player 2 is the follower. Unlike the zero-sum case, the structure of the Jacobian $J_{\mathcal{S}_\tau}$ is not lower block triangular and hence, the convergence rate depends more abstractly on the spectral structure of $J_{\mathcal{S}_\tau}$ as opposed to the second-order sufficient conditions in the differential Stackelberg equilibrium definition.

Let $S(x^*) = \frac{1}{2}(J_{\mathcal{S}_\tau}(x^*)^\top + J_{\mathcal{S}_\tau}(x^*))$. Define constants $\alpha = \lambda_{\min}^2\big(S(x^*)\big)$ and $\beta = \lambda_{\max}(J_{\mathcal{S}_\tau}(x^*)^\top J_{\mathcal{S}_\tau}(x^*))$.

2.D.2.1 Proof of Theorem 2.8

Let $\gamma_1 = \sqrt{\alpha}/\beta$. Then to bound $\|I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)\|_2^2$ consider the following:

$$(I - \gamma_1 J_{\mathcal{S}_\tau}(x^*))^\top (I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)) \leq (1 - 2\gamma_1 \lambda_{\min}(S(x^*)) + \gamma_1^2 \lambda_{\max}(J_{\mathcal{S}_\tau}^\top(x^*) J_{\mathcal{S}_\tau}(x^*)))I \leq (1 - \alpha/\beta)I. \tag{2.25}$$

Moreover, $\rho(I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)) \leq \|I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)\|$ for any matrix norm (Horn and Johnson, 2011) so that $\rho(I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)) \leq (1 - \alpha/\beta)^{1/2}$. Taking the Taylor expansion of $I - \gamma_1 g_{\mathcal{S}_\tau}(x)$ around x^*, we have

$$I - \gamma_1 g_{\mathcal{S}_\tau}(x) = (I - \gamma_1 g_{\mathcal{S}_\tau}(x^*)) + (I - \gamma_1 J_{\mathcal{S}_\tau}(x^*))(x - x^*) + R_1(x - x^*)$$

where $R_1(x-x^*)$ is the remainder term satisfying $R_1(x-x^*) = o(\|x-x^*\|_2)$ as $x \to x^*$. This implies that there is a $\delta > 0$ such that $\|R_1(x-x^*)\|_2 \leq \frac{\alpha}{4\beta}\|x-x^*\|_2$ whenever $\|x-x^*\|_2 < \delta$. Hence,

$$\|I-\gamma_1 g_{\mathcal{S}_\tau}(x)-(I-\gamma_1 g_{\mathcal{S}_\tau}(x^*))\|_2 \leq \Big(\|I-\gamma_1 J_{\mathcal{S}_\tau}(x^*)\|_2+\frac{\alpha}{4\beta}\Big)\|x-x^*\|_2 \leq \Big(\Big(1-\frac{\alpha}{\beta}\Big)^{1/2}+\frac{\alpha}{4\beta}\Big)\|x-x^*\|_2$$

We claim that $c(z) = (1-z)^{1/2} + \frac{z}{4} - \big(1-\frac{z}{2}\big)^{1/2} \leq 0$ for any $z \in [0,1]$. Since $c(0) = 0$ and $c(1) = \frac{1}{4} - \frac{1}{\sqrt{2}} \leq 0$, we simply need to show that $c'(z) \leq 0$ on $(0,1)$ to get that $c(z)$ is a decreasing function on $[0,1]$, and hence negative on $[0,1]$. Indeed, $c'(z) = \frac{1}{4} + \frac{1}{2\sqrt{4-2z}} - \frac{1}{2\sqrt{1-z}} \leq 0$ since $(1-z)^{-1/2} - (4-2z)^{-1/2} \geq 1/2$ for all $z \in (0,1)$.

Note that $\alpha/\beta \in [0,1]$ since $\alpha \leq \beta$; indeed,

$$\alpha = \lambda^2_{\min}(S(x^*)) \leq \lambda^2_{\max}(S(x^*)) \leq \lambda_{\max}(J^\top_{\mathcal{S}_\tau}(x^*)J_{\mathcal{S}_\tau}(x^*)) = \beta.$$

Further, $\alpha > 0$ and $\beta > 0$ by assumption. Hence,

$$\|I-\gamma_1 g_{\mathcal{S}_\tau}(x)-(I-\gamma_1 g_{\mathcal{S}_\tau}(x^*))\|_2 \leq \Big(1-\frac{\alpha}{2\beta}\Big)^{1/2}\|x-x^*\|_2$$

Thus, for any $x_0 \in \{x|\ \|x-x^*\| < \delta\}$,

$$\|x_k-x^*\|_2 \leq \Big(1-\frac{\alpha}{2\beta}\Big)^{k/2}\|x_0-x^*\|_2 \tag{2.26}$$

so that the (local) rate of convergence is $O((1-\alpha/(2\beta))^{k/2})$, since finite dimensional norms are equivalent.

2.D.2.2 Proof of Corollary 2.2

Analogous to the zero-sum setting, the proof of the finite time convergence guarantee follows directly from the arguments in the proof of Theorem 2.8. Following standard arguments, (2.26) in the proceeding proof implies a finite time convergence guarantee. Indeed, let $\varepsilon > 0$ be given. Since $(1-\alpha/(2\beta))^{1/2} \leq \exp(-\alpha/(4\beta))$, we have that

$$\|x_k-x^*\|_2 \leq \exp(-k\alpha/(4\beta))\|x_0-x^*\|_2.$$

This, in turn, implies that $x_k \in B_\varepsilon(x^*)$ (i.e., x_k is an ε-differential Stackelberg equilibrium) for all $k \geq \lceil \frac{4\beta}{\alpha}\log(\|x_0-x^*\|_2/\varepsilon)\rceil$ whenever $\|x_0-x^*\| < \delta$.

Given that $f_i \in C^2(X,\mathbb{R})$ so that $I-\gamma_1 J_{\mathcal{S}_\tau}(x)$ is locally Lipschitz with constant L, we can find an explicit expression for δ in terms of L. Indeed, using similar arguments as in the proof of Corollary 2.1, $\delta = \alpha/(2L\beta)$.

2.D.3 Strict Saddle Avoidance: Proof of Theorem 2.6

We now provide the proof of strict saddle avoidance in the deterministic regime for the τ-Stackelberg gradient dynamics. To show this, we follow the arguments in Mazumdar et al. (2020) with slight

modifications. The proof requires the stable manifold theorem.

Theorem 2.13. *(Stable Manifold Theorem Shub, 1978, Theorem III.7). Let $x_0 \in X$ be a fixed point for the C^q local diffeomorphism $\phi : U \to \mathbb{R}^d$ where U is an open neighborhood of $x_0 \in \mathbb{R}^d$ and $q \geq 1$. Let $E^s \otimes E^c \otimes E^u$ be the invariant splitting of $\mathbb{R}^d$ into generalized eigenspaces of $D\phi(x_0)$ corresponding to the eigenvalues of absolute value less than one, equal to one, and greater than one. To the $D\phi(x_0)$ invariant subspace $E^s \otimes E^c$ there is an associated local ϕ-invariant C^q embedded disc W^{cs}_{loc} called the local stable center manifold of dimension $\dim(E^s \otimes E^c)$ and ball B (in an adapted norm) around x_0 such that $\phi(W^{cs}_{loc}) \cap B \subset W^{cs}_{loc}$, and if $\phi^t(x) \in B$ for all $t \geq 0$ then $x \in W^{sc}_{loc}$ where $\phi^t = \phi \circ \cdots \circ \phi$ is the t-times composition of the map ϕ.*

The proof technique is to show that $g_{\mathcal{S}_\tau}$ is a diffeomorphism, and then apply the center manifold theorem. We claim that $g_{\mathcal{S}_\tau} : \mathbb{R}^d \to \mathbb{R}^d$ is a diffeomorphism. If we can show that $g_{\mathcal{S}_\tau}$ is invertible and a local diffeomorphism, then the claim follows. Recall that the game Jacobian for the τ-Stackelberg update is denoted by $J_{\mathcal{S}_\tau}$. We first argue by contradiction that $g_{\mathcal{S}_\tau}$ is invertible. Consider $x \neq y$ and suppose $g_{\mathcal{S}_\tau}(y) = g_{\mathcal{S}_\tau}(x)$ so that $y - x = \gamma_1(g_{\mathcal{S}_\tau}(y) - g_{\mathcal{S}_\tau}(x))$. The assumption $\sup_{x \in \mathbb{R}^d} \|J_{\mathcal{S}_\tau}(x)\|_2 \leq L < \infty$ implies that $g_{\mathcal{S}_\tau}$ satisfies the Lipschitz condition on $\mathbb{R}^d$. Hence, $\|g_{\mathcal{S}_\tau}(y) - g_{\mathcal{S}_\tau}(x)\|_2 \leq L\|y - x\|_2$. Then, $\|y - x\|_2 \leq L\gamma_1\|y - x\|_2 < \|y - x\|_2$ since $\gamma_1 < 1/L$ which gives rise to a contradiction.

Now, observe that the Jacobian of the discrete-time dynamics is given by $Dg_{\mathcal{S}_\tau} = I - \gamma_1 J_{\mathcal{S}_\tau}(x)$. If $Dg_{\mathcal{S}_\tau}$ is invertible, then the implicit function theorem (Abraham et al., 1988, Thm. 2.5.7) implies that $g_{\mathcal{S}_\tau}$ is a local diffeomorphism. Hence, it suffices to show that $\gamma_1 J_{\mathcal{S}_\tau}(x)$ does not have an eigenvalue of 1. Indeed, letting $\rho(A)$ be the spectral radius of a matrix A, we know in general that $\rho(A) \leq \|A\|$ for any square matrix A and induced operator norm $\|\cdot\|$ so that

$$\rho(\gamma_1 J_{\mathcal{S}_\tau}(x)) \leq \|\gamma_1 J_{\mathcal{S}_\tau}(x)\|_2 \leq \gamma_1 \sup_{x \in \mathbb{R}^d} \|J_{\mathcal{S}_\tau}(x)\|_2 < \gamma_1 L < 1.$$

Of course, the spectral radius is the maximum absolute value of the eigenvalues, so that the above implies that all eigenvalues of $\gamma_1 J_{\mathcal{S}_\tau}(x)$ have absolute value less than 1. Since $g_{\mathcal{S}_\tau}$ is injective by the preceding argument, its inverse is well-defined and since $g_{\mathcal{S}_\tau}$ is a local diffeomorphism on $\mathbb{R}^d$, it follows that $g^{-1}_{\mathcal{S}_\tau}$ is smooth on $\mathbb{R}^d$. Thus, $g_{\mathcal{S}_\tau}$ is a diffeomorphism.

Consider all critical points to the game, given by $\mathcal{X}_c = \{x \in \mathcal{X} |\ g_{\mathcal{S}_\tau}(x) = 0\}$. For each $u \in \mathcal{X}_c$, let B_u, where u indexes the point, be the open ball derived from the center manifold theorem stated in Theorem 2.13 and let $\mathcal{B} = \cup_u B_u$. Since $\mathcal{X} \subseteq \mathbb{R}^d$, Lindelõf's lemma (Kelley, 1955)—every open cover has a countable subcover—gives a countable subcover of $\mathcal{B}$. That is, for a countable set of critical points $\{u_i\}_{i=1}^\infty$ with $u_i \in \mathcal{X}_c$, we have that $\mathcal{B} = \cup_{i=1}^\infty B_{u_i}$.

Starting from some point $x_0 \in \mathcal{X}$, if τ-Stackelberg converges to a strict saddle point, then there exists a t_0 and index i such that $g^t_{\mathcal{S}_\tau}(x_0) \in B_{u_i}$ for all $t \geq t_0$. Again, applying the center manifold theorem from Theorem 2.13 and using that $g_{\mathcal{S}_\tau}(\mathcal{X}) \subset \mathcal{X}$, which indeed holds if $\mathcal{X} = \mathbb{R}^d$, we get that $g^t_{\mathcal{S}_\tau}(x_0) \in W^{cs}_{\text{loc}} \cap \mathcal{X}$ where W^{cs}_{loc} is the local stable center manifold.

Using the fact that $g_{\mathcal{S}_\tau}$ is invertible, we can iteratively construct the sequence of sets defined by $W_1(u_i) = g^{-1}_{\mathcal{S}_\tau}(W^{cs}_{\text{loc}} \cap \mathcal{X})$ and $W_{k+1}(u_i) = g^{-1}_{\mathcal{S}_\tau}(W_k(u_i) \cap \mathcal{X})$. Then we have that $x_0 \in W_t(u_i)$ for all $t \geq t_0$. The set $\mathcal{X}_0 = \cup_{i=1}^\infty \cup_{t=0}^\infty W_t(u_i)$ contains all the initial points in $\mathcal{X}$ such that τ-Stackelberg converges to a strict saddle. Since u_i is a strict saddle, $I - \gamma_1 J_{\mathcal{S}_\tau}(u_i)$ has an eigenvalue greater

than 1. This implies that the co-dimension of the unstable manifold is strictly less than d—i.e., $\dim(W^{cs}_{\text{loc}}) < d$. Hence, $W^{cs}_{\text{loc}} \cap X$ has Lebesgue measure zero in $\mathbb{R}^d$.

Using again that $g_{\mathcal{S}_\tau}$ is a diffeomorphism, $g^{-1}_{\mathcal{S}_\tau} \in C^1$ so that it is locally Lipschitz and locally Lipschitz maps are null set preserving. Hence, $W_k(u_i)$ has measure zero for all k by induction so that $\mathcal{X}_0$ is a measure zero set since it is a countable union of measure zero sets.

2.D.4 Computing the Stackelberg Update and Schur Complement

The Stackelberg gradient dynamics involve computing an inverse-Hessian-vector product for the $D^2_2 f_2(x)$ inverse term and Jacobian-vector product for the $D_{21} f_2(x)$ term. These operations can be done efficiently in Python by utilizing Jacobian-vector products in auto-differentiation libraries combined with the `sparse.LinearOperator` class in `scipy`. These objects can also be used to compute their eigenvalues, inverses, or the Schur complement of the game dynamics using the `scipy.sparse.linalg` package.

The operators required for the leader update can be obtained by the following. Consider the Jacobian $J(x)$ of the simultaneous gradient dynamics at a critical point for the general sum game (f_1, f_2). Its block components consist of four operators $D_{ij} f_i(x) : X_j \to X_i,\ i, j \in \{1, 2\}$ that can be computed using forward-mode or reverse-mode Jacobian-vector products. Instantiating these operators as a linear operator in `scipy` allows us to compute the eigenvalues of each player's individual Hessian. Properties such as the real eigenvalues of a Hermitian matrix or complex eigenvalues of a square matrix can be computed using `eigsh` or `eigs` respectively. Selecting to compute the smallest or largest k eigenvalues—sorted by either magnitude, real or imaginary values—allows one to examine the positive-definiteness of the operators.

Operators can be combined to compute other operators relatively efficiently for large scale problems without requiring to compute their full matrix representation. For an example, take the Schur complement of the Jacobian above at fixed network parameters $x \in \mathcal{X}_1 \times \mathcal{X}_2$, $D^2_1(x) - D_{12} f_1(x)(D^2_2 f_2)^{-1}(x) D_{21} f_2(x)$. We create an operator $S_1(x) : \mathcal{X}_1 \to \mathcal{X}_1$ that maps a vector v to $p - q$ by performing the following four operations: $u = D_{21} f_2(x) v$, $w = (D^2_2 f_2)^{-1}(x) u$, $q = D_{12} f_1(x) w$, and $p = D^2_1(x) v$. Each of the operations can be computed using a single backward pass through the network except for computing w, since the inverse-Hessian requires an iterative method. It solves the linear equation $D^2_2 f_2(x) w = u$ and there are various available methods: we tested (bi)conjugate gradient methods, residual-based methods, or least-squares methods, and each of them provide varying amounts of error when compared with the exact solution. we found that computing the leader update using the conjugate gradient method with maximum of five iterations and warm-start works well. We compared using the real Hessian for smaller scale problems and found the estimate to be within numerical precision. A similar procedure is used to compute a variety higher-order derivatives. For instance, the regularized total derivative of the leader's update is the total derivative of $Df_1(x_1, r(x_1))$ evaluated at (x_1, x_2) where the implicit map variation is defined by $Dr_\mu(x_1) = -(D^2_2 f_2(x_1), r(x_1) + \mu I)^{-1} D_{21} f_2(x_1, r(x_1))$ for a regularization parameter $\mu > 0$. To compute the spectrum of such an operator, we create a function $v \mapsto D(Df_1(x_1, r(x_1)))v$ that takes

a vector $v \in \mathbb{R}^{d_1}$ and returns

$$\begin{aligned} D(Df_1(x_1, r(x_1)))v = & D_1^2 f_1(x_1, r(x_1))v + D_{12} f_1(x_1, r(x_1)) Dr_\mu(x_1)v \\ & + (D_{12} f_1(x_1, r(x_1)) + Dr_\mu(x_1)^\top D_2^2 f_1(x_1, r(x_1)) Dr_\mu(x_1)v \\ & + D^2 r_\mu(x_1) D_2 f_1(x_1, r(x_1))v \end{aligned}$$

where the last higher order term is then assumed to be zero. The above derivative can be written as a composition of Jacobian-vector product operators and least squares problems, thus can be computed efficiently with auto-differentiation tools.

Chapter 3

Nonconvex Zero-Sum Games: Gradient Descent-Ascent with Timescale Separation

In this previous chapter, a local Stackelberg equilibrium concept along with a gradient-based characterization in terms of sufficient conditions termed a differential Stackelberg equilibrium was developed. Moreover, the connections between the limiting points of a pair of gradient-based learning dynamics were studied. The results showed that a set of Stackelberg gradient dynamics have the property of only locally converging to differential Stackelberg equilibrium in nonconvex-nonconcave zero-sum games. On the other hand, it was shown that the canonical simultaneous gradient dynamics (gradient descent-ascent in zero-sum games) can also sometimes locally converge to differential Stackelberg equilibrium. For machine learning applications, a weakness of the Stackelberg gradient dynamics is the requirement of higher-order gradient information. Thus, it is of interest to obtain a equivalent equilibrium convergence guarantee using only first-order information. As remarked in the discussion at the end of the last chapter, gradient descent-ascent with timescale separation has also been studied recently (Jin et al., 2020). Denote the learning rate of player 1 by $\gamma_1 > 0$ and let $\tau > 0$ be the "timescale separation" parameter such that the learning rate of player 2 is given by $\gamma_2 = \tau\gamma_1$ and $\tau = \gamma_2/\gamma_1$ is the ratio of learning rates. The results of Jin et al. (2020) show that as $\tau \to \infty$, then the only locally stable critical points of gradient descent-ascent correspond to differential Stackelberg equilibrium. From a practical machine learning perspective, a weakness of this result is that it also requires $\gamma_1 \to 0$ to maintain stability of the discrete-time system. This motivates finer-grained characterizations of the timescale separation parameters that guarantee local stability of differential Stackelberg equilibrium and instability of critical points lacking game-theoretic meaning. This forms the basis of the work presented in this chapter. In particular, a non-asymptotic construction of the finite timescale separation parameter τ^* such that gradient descent-ascent for all $\tau \in (\tau^*, \infty)$ locally converges to x^* if and only if it is a differential Stackelberg equilibrium is given. Moreover, local convergence rates given the finite timescale separation ensuring local stability are presented. The convergence results are complemented by a non-convergence result: given a critical point x^* that is not a differential Stackelberg equilibrium, we provide a non-asymptotic construction of a finite timescale separation τ_0 such that gradient descent-ascent with timescale separation $\tau \in (\tau_0, \infty)$ does not converge to x^*. Finally, we extend the results to gradient penalty regularization methods for generative adversarial networks and empirically demonstrate the effects of timescale separation through an extensive set of experiments.

3.1 Introduction

This chapter considers gradient-based learning in zero-sum games of the form

$$\min_{x_1 \in \mathcal{X}_1} \max_{x_2 \in \mathcal{X}_2} f(x_1, x_2)$$

where the objective function of the game f is assumed to be sufficiently smooth and potentially nonconvex and nonconcave in the strategy spaces $\mathcal{X}_1$ and $\mathcal{X}_2$ respectively with each $\mathcal{X}_i = \mathbb{R}^{d_i}$ or more generally a precompact subset of $\mathbb{R}^{d_i}$ for each $i \in \mathcal{I} = \{1, 2\}$. The focus on zero-sum game is classical in game theory (Basar and Olsder, 1998) and in machine learning they have become perhaps more important that ever with the advent of generative adversarial networks (Goodfellow et al., 2014) and robust supervised learning tasks (Madry et al., 2018; Sinha et al., 2018).

The gradient descent-ascent learning dynamics are now both widely studied theoretically and empirically deployed as a gradient-based optimization method for zero-sum game formulations. However, in nonconvex-nonconcave zero-sum games, a number of past works highlight issues of convergence to critical points devoid of game theoretic meaning, where common notions of 'meaningful' equilibria include the local Nash and Stackelberg concepts along with the corresponding gradient-based (differential) characterizations (Fiez et al., 2020a; Jin et al., 2020; Ratliff et al., 2016). Indeed, it has been shown gradient descent-ascent with a shared learning rate is prone to reaching critical points that are neither a differential Nash equilibrium nor a differential Stackelberg equilibrium (Daskalakis and Panageas, 2018; Jin et al., 2020; Mazumdar et al., 2020). While an important negative result, it does not rule out the prospect that gradient descent-ascent may be able to guarantee equilibrium convergence as it fails to account for a key structural parameter of the dynamics, namely the ratio of learning rates between the players.

Motivated by the observation that the order of play between players is fundamental to the definition of the game as was highlighted in the previous chapter, the role of timescale separation in gradient descent-ascent has recently been explored theoretically (Chasnov et al., 2020; Heusel et al., 2017; Jin et al., 2020). On the empirical side, it has been widely demonstrated in machine learning that timescale separation in gradient descent-ascent is crucial to improving both stability and convergence. In particular, timescale separation or unrolling multiple steps of the maximization procedure are common heuristics both when training generative adversarial and performing adversarial training (Arjovsky et al., 2017; Goodfellow et al., 2014; Heusel et al., 2017; Madry et al., 2018). Denote the learning rate of player 1 by $\gamma_1 > 0$ and let $\tau > 0$ be the "timescale separation" parameter such that the learning rate of player 2 is given by $\gamma_2 = \tau\gamma_1$ and $\tau = \gamma_2/\gamma_1$ is the ratio of learning rates. Toward understanding the effect of timescale separation, Jin et al. (2020) show the stable critical points of gradient descent-ascent coincide with the set of differential Stackelberg equilibrium as $\tau \to \infty$. In other words, all 'bad critical points' (critical points lacking game-theoretic meaning) become unstable and all 'good critical points' (game-theoretically meaningful equilibria) remain or become locally exponentially stable as $\tau \to \infty$. Notably, this result is a characterization across the class of smooth nonconvex-nonconcave zero-sum games. While a promising theoretical development, it does not lead to a practical, implementable learning rule or necessarily provide an explanation for the satisfying performance in applications of gradient descent-ascent with a finite timescale separation. Importantly, it leaves open the problem of fully characterizing the stability or instability of gradient descent-ascent around fixed critical points (or across classes of games

admitting common structure at critical points) as a function of the timescale separation, and this chapter seeks to close this gap.

3.1.1 Contributions and Overview

To motivate our primary theoretical results, we present a self-contained description of what is known about the local stability of gradient descent-ascent around critical points in Section 3.3. The existing results primarily concern gradient descent-ascent without timescale separation and with a ratio of learning rates approaching infinity (see Figure 3.1 for a graphical depiction of known results in each regime). In contrast, this chapter is focused on characterizing the stability and convergence of gradient descent-ascent across a range of finite learning rate ratios. To hint at what is achievable in this realm, we present simple, illustrative examples in which finite timescale separation remedies undesirable local stability properties of the gradient descent-ascent dynamics in zero-sum games (see Examples 3.1 and 3.2, Section 3.4). Moreover, we connect to the literature on singularly perturbed systems (Kokotovic et al., 1986, Chapter 2 and the citations within) and show how the stability of gradient descent-ascent as $\tau \to \infty$ can be seen rather directly using these methods (see Proposition 3.2 and the proof sketch). This is instructive, as the analysis methods for our main results derive from the historical singularly perturbed systems literature.

Stability and Instability Characterizations. A relevant line of study on singularly perturbed systems is that of characterizing the range of perturbation parameters for which a system is stable (Kokotovic et al., 1986; Saydy, 1996; Saydy et al., 1990). Arguably introduced by Saydy et al. (1990), guardian or guard maps act as a certificate that the roots of a polynomial lie in a particular guarded domain for a range of parameter values. Historically, guard maps serve as a tool for studying the stability of parameterized families of dynamical systems. We bring this tool to learning in games and construct a map that guards a class of Hurwitz stable matrices parameterized by the timescale separation parameter τ in order to analyze the range of learning rate ratios for which a critical point is stable with respect to gradient descent-ascent. This technique leads to perhaps the main result of this chapter: Theorem 3.3 shows there exists a $\tau^* \in [0, \infty)$ such that a critical point x^* is stable for all $\tau \in (\tau^*, \infty)$ if and only if x^* is a differential Stackelberg equilibrium and provides an explicit construction of τ^*. In fact, the construction of τ^* in Theorem 3.3 is tight, and this is confirmed by our numerical experiments.

The latter implication of Theorem 3.3 says that there exists a finite learning rate ratio such that a non-equilibrium critical point of gradient descent-ascent is unstable. Building off of this, the stability result of Theorem 3.3 is complemented with an analogous instability result. Theorem 3.4 establishes that there exists a $\tau_0 \in [0, \infty)$ such that for all $\tau \in (\tau_0, \infty)$ a non-equilibrium critical point is unstable with respect to gradient descent-ascent and provides a construction of τ_0. Together, Theorem 3.3 and Theorem 3.4 answer affirmatively that gradient descent-ascent with finite timescale separation can guarantee equilibrium convergence with a fine-grained characterization.

The final stability result connects to the literature on generative adversarial networks. We show under common assumptions on generative adversarial networks (Mescheder et al., 2018; Nagarajan and Kolter, 2017) that the introduction of gradient penalty based regularization to the discriminator does not change the set of critical points for the dynamics and, further, that for any $\tau \in (0, \infty)$ and any non-negative, finite regularization parameter $\mu > 0$, the continuous time limiting regularized learning dynamics remain stable.

Convergence Analysis. The stability result of Theorem 3.3 nearly immediately implies there exists a $\tau^* \in [0, \infty)$ such that gradient descent-ascent converges locally asymptotically for all $\tau \in (\tau^*, \infty)$ if and only if x^* is a differential Stackelberg equilibrium given a suitably chosen learning rate and deterministic gradient feedback. We give an explicit local asymptotic rate of convergence in Theorem 3.6 and characterize the iteration complexity in Corollary 3.3. Moreover, we extend the convergence guarantees to stochastic gradient feedback in Theorem 3.7.

Experimental Results. The theoretical results we provide are complemented by extensive experiments. In simulation, we explore a number of interesting behaviors of gradient descent-ascent with timescale separation analyzed theoretically including differential Stackelberg equilibria shifting from being unstable to stable and non-equilibrium critical points moving from being stable to unstable. Furthermore, we examine how the vector field and the spectrum of the game Jacobian evolve as a function of the timescale separation and explore the relationships with the regions of attraction and rate of convergence. We experiment with gradient descent-ascent on the Dirac-GAN proposed by Mescheder et al. (2018) and illustrate the interplay between timescale separation, regularization, and rate of convergence. Building on this, we train generative adversarial networks on the CIFAR-10 and CelebA datasets with regularization and demonstrate that timescale separation can boost performance and stability. In the experiments we observe that regularization and timescale separation are intimately connected and there is an inherent tradeoff between them. This indicates that insights made on simple generative adversarial network formulations may carry over to the complex problems where players are parameterized by neural networks.

Collectively, the primary contribution of this chapter is the near-complete characterization of the behavior of gradient descent-ascent with finite timescale separation. Moreover, by introducing a novel set of analysis tools to this literature, our work opens a number of future research questions. As an aside, we believe these technical tools open up novel avenues for not only proving results about learning dynamics in games, but also for synthesizing algorithms.

3.1.2 Organization

The organization of this chapter is as follows. Preliminaries on game theoretic notions of equilibria, gradient-based learning algorithms, and dynamical systems theory are reviewed in Section 3.2. We remark that much of this material is presented in much greater depth in Chapter 2 and the reader should specifically refer to Section 2.2.3 for further details on local equilibrium concepts, Section 2.3.1 for additional context on notions of stability and instability along with methods for verifying local stability using the linearization, and the beginning of Section 2.5 for details on stochastic stability in terms of internally chain-transitive sets and the ordinary differential equation method. Background on know results regarding the local stability of gradient descent-ascent around critical points is given in Section 3.3. Section 3.4 presents the theoretical stability and instability results for the continuous-time limiting system of the gradient descent-ascent dynamics with timescale separation. Following this, in Sections 3.5 and 3.6, convergence guarantees are provided for the original discrete time dynamical system of interest with deterministic and stochastic gradient feedback, respectively. Section 3.7 contains an extensive numerical study. Related work is dicussed at greater depth in Section 3.8. We conclude in Section 3.9 with a discussion. An appendix that follows the chapter includes the majority of the detailed proofs in order to improve the overall clarity and readability.

3.2 Preliminaries

The setting considered in this chapter is nearly equivalent to that from Chapter 2, but restricted to the zero-sum games. A two–player zero-sum continuous strategy space game is defined by a collection of costs (f_1, f_2) where $f_1 \equiv f$ and $f_2 \equiv -f$ with $f \in C^q(\mathcal{X}, \mathbb{R})$ for some $q \geq 2$ and where $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$ with each $\mathcal{X}_i$ a precompact subset of $\mathbb{R}^{d_i}$ for $i \in \mathcal{I} = \{1, 2\}$ and $d = d_1 + d_2$. Each player $i \in \mathcal{I}$ seeks to minimize their cost $f_i(x_i, x_{-i})$ with respect to their choice variable x_i where x_{-i} is the vector of all other actions x_j with $j \neq i$. Note that we often use the shorthand $x = (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$. Let us begin by stating the mathematical notation used in this chapter, much of which is shared with the notation in Chapter 2, but restated here for clarity of presentation.

Mathematical Notation. We denote $D_i f_i(x_1, x_2) \in \mathbb{R}^{d_i \times 1}$ as the derivative of $f_i(x_1, x_2)$ with respect to x_i, $D_{ij} f_i(x_1, x_2) \in \mathbb{R}^{d_i \times d_j}$ as the partial derivative of $D_i f_i(x_1, x_2)$ with respect to x_j, and $D_i^2 f_i(x_1, x_2) \in \mathbb{R}^{d_i \times d_i}$ as the partial derivative of $D_i f_i(x_1, x_2)$ with respect to x_i. Given a symmetric matrix A, we indicate that it is positive definite using the notation $A \succ 0$. We denote by $\mathbb{C}_-^\circ$ and $\mathbb{C}_+^\circ$ the open left and right half complex planes (that is, the set of complex numbers for which the real parts are negative and positive respectively), $\rho(\cdot)$ as an operator returning the spectral radius (maximum modulus of the eigenvalues) of a matrix argument, $\text{spec}(\cdot)$ as an operator returning the set of eigenvalues of a matrix argument, $\det(\cdot)$ as an operator returning the determinant of a matrix argument, and $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ as operators returning the minimum and maximum eigenvalues of a matrix argument, respectively. Let $\lambda_{\max}^+(\cdot)$ be the largest positive real root of its argument if it exists and zero otherwise. Given a matrix $A \in \mathbb{R}^{d_1 \times d_2}$, let $\text{vec}(A) \in \mathbb{R}^{d_1 d_2}$ be the vectorization of A. That is, $\text{vec}(A)$ takes rows a_i of A, transposes them and stacks them vertically in order of their index. Let $\otimes$ and $\oplus$ denote the Kronecker product and sum respectively, where $A \oplus B = A \otimes I + I \otimes B$. Moreover, $\boxplus$ is an operator that generates an $\frac{1}{2}d(d+1) \times \frac{1}{2}d(d+1)$ matrix from a matrix $A \in \mathbb{R}^{d \times d}$ such that $A \boxplus A = H_d^+(A \oplus A)H_d$ where $H_d^+ = (H_d^\top H_d)^{-1} H_d^\top$ is the (left) pseudo-inverse of H_d, a full column rank duplication matrix. See Lancaster and Tismenetsky (1985) and Magnus (1988) for more detail on these operators.

Gradient Descent-Ascent Learning Dynamics. We study agents seeking equilibria of the game via a learning algorithm and consider arguably the most natural learning rule in zero-sum continuous games: gradient descent-ascent (`GDA`). Moreover, we investigate this learning rule with timescale separation between the players. Let $\tau = \gamma_2/\gamma_1$ be the learning rate ratio and define $\Lambda_\tau = \text{blockdiag}(I_{d_1}, \tau I_{d_2})$ where I_{d_i} is a $d_i \times d_i$ identity matrix. Consider the vector field

$$g(x) = (D_1 f(x), -D_2 f(x)).$$

The discrete-time τ-`GDA` dynamics are then given by

$$x_{k+1} = x_k - \gamma_1 \Lambda_\tau g(x_k). \tag{3.1}$$

To characterize the convergence of τ-`GDA`, we also study its continuous time limiting system

$$\dot{x} = -\Lambda_\tau g(x). \tag{3.2}$$

The Jacobian of the system from (3.2) is given by

$$J_\tau(x) = \Lambda_\tau J(x) = \begin{bmatrix} D_1^2 f(x) & D_{12} f(x) \\ -\tau D_{12}^\top f(x) & -\tau D_2^2 f(x) \end{bmatrix}, \tag{3.3}$$

where

$$J(x) = \begin{bmatrix} D_1^2 f(x) & D_{12} f(x) \\ -D_{12}^\top f(x) & -D_2^2 f(x) \end{bmatrix}. \tag{3.4}$$

Observe that critical points of the dynamics (x such that $g(x) = 0$) are shared between τ-GDA and its continuous-time limiting system in (3.2). Moreover, note that the gradient descent-ascent dynamics without timescale (that is, $\tau = 1$) are equivalent to the simultaneous gradient dynamics studied in Chapter 2, but restricted to only zero-sum games.

Equilibrium Concepts. As is discussed in depth in Section 2.2.3 of Chapter 2, there are natural equilibrium concepts depending on the interaction structure: the (local) Nash equilibrium concept in the case of simultaneous play and the (local) Stackelberg (equivalently minmax in zero-sum games) equilibrium concept in the case of hierarchical play (Basar and Olsder, 1998). Following the approach in Chapter 2, and what is now a common approach in the machine learning literature on game theory, we consider characterizations of these equilibrium notions defined in terms of gradient-based sufficient conditions. Notice that compared to the statements in Section 2.2.3 of Chapter 2, the statements are specialized to the context of zero-sum games.

The following definition is characterized by sufficient conditions for a local Nash equilibrium.

Definition 3.1 (Differential Nash Equilibrium, Ratliff et al. 2013)**.** *The joint strategy $x \in \mathcal{X}$ is a differential Nash equilibrium if $D_1 f(x) = 0$, $D_2 f(x) = 0$, $D_1^2 f(x) \succ 0$, and $D_2^2 f(x) \prec 0$.*

Let $\mathtt{S}_1(\cdot)$ denote the Schur complement of $(\cdot)$ with respect to the $d_2 \times d_2$ block in $(\cdot)$. The following definition is characterized by sufficient conditions for a local Stackelberg equilibrium (recall Proposition 2.1).

Definition 3.2 (Differential Stackelberg Equilibrium, Fiez et al. 2020a)**.** *The joint strategy $x \in \mathcal{X}$ is a differential Stackelberg equilibrium if $D_1 f(x) = 0$, $D_2 f(x) = 0$, $\mathtt{S}_1(J(x)) \succ 0$, $D_2^2 f(x) \prec 0$.*

It is worth remarking at this junction that with timescale separation in the gradient descent-ascent dynamics, the interaction structure begins to emulate a hierarchical decision-making order, which in turn motivates the study of the connections between the dynamics and the differential Stackelberg equilibrium concept.

Stability and Instability Characterizations. To conclude this preliminary section, we provide some background on stability and instability characterizations using the linearization of the gradient dynamics. The reader is referred to Section 2.3.1 of Chapter 2 for a detailed background section on notions of stability and instability, along with methods for verifying local stability using the linearization, both for the continuous-time system and the discrete-time system. Moreover, the beginning of Section 2.5 in Chapter 2 details stochastic stability.

A significant focus of this chapter is on determining the stability and instability of critical points. The primary technique for this will be to assess the eigenvalues of the local linearization. Specifically, recall from (3.2) that the continous-time τ-GDA system is given by $\dot{x} = -\Lambda_\tau g(x)$ and the

Jacobian of $\Lambda_\tau g(x)$ is $J_\tau(x)$ which is defined in (3.3). By analyzing the spectral properties of $J_\tau(x^*)$ at critical points (that is $x^* : g(x^*) = 0$), local stability and convergence properties of the nonlinear system can be determined. Consider the following stability and instability characterizations that are also presented in Chapter 2, but tailored here to the τ-GDA system.

Theorem 3.1 (Stability Characterizations, Khalil 2002, Theorem. 4.6, Corollary 4.3)**.** *Consider a critical point x^* of $\dot{x} = -\Lambda_\tau g(x)$. The following are equivalent: (a) x^* is a locally exponentially stable critical point of $\dot{x} = -\Lambda_\tau g(x)$; (b) $\text{spec}(-J_\tau(x^*)) \subset \mathbb{C}^\circ_-$; (c) Given any symmetric matrix $Q \in \mathbb{R}^{d\times d}$, there exists a unique symmetric positive-definite matrix $P \in \mathbb{R}^{d\times d}$ such that $PJ_\tau(x^*) + J_\tau(x^*)^\top P = Q$.*

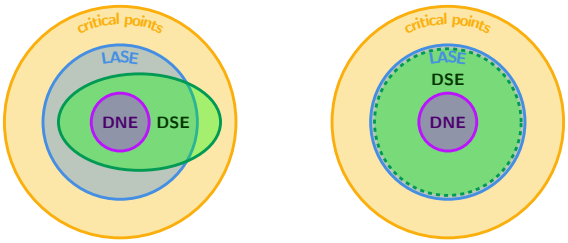
Theorem 3.2 (Instability Characterization, Sastry 1999, Theorem, 5.42)**.** *Consider a critical point x^* of $\dot{x} = -\Lambda_\tau g(x)$. If $-J_\tau(x^*)$ has at least one eigenvalue in $\mathbb{C}^\circ_+$, then x^* is an unstable critical point of $\dot{x} = -\Lambda_\tau g(x)$.*

Note that as detailed in Section 2.3.1 of Chapter 2, the analogous discrete-time stability and instability conditions can be determined by the spectral radius of the the linearization. Specifically, for the τ-GDA system, when $\rho(I - \gamma_1 J_\tau(x^*)) < 1$ then x^* is locally exponentially stable, and when $\rho(I - \gamma_1 J_\tau(x^*)) > 1$, then x^* is unstable (Ortega and Rheinboldt, 1970; Sastry, 1999).

3.3 Background Observations

In Figure 3.1 we present a graphical representation of known results on the stability of gradient descent-ascent with timescale separation in continuous-time, where we remark that such results nearly directly imply equivalent conclusions regarding the discrete time system τ-GDA with a suitable choice of learning rate γ_1. The primary focus of past work has been on the edge cases of $\tau = 1$ and $\tau \to \infty$. For $\tau = 1$, the set of differential Nash equilibrium are stable, but differential Stackelberg equilibrium may be stable or unstable, and non-equilibrium critical points can be stable. As $\tau \to \infty$, the set of differential Nash equilibrium remain stable, each differential Stackelberg equilibrium is guaranteed to become stable, and each non-equilibrium critical point must be unstable. That is, the sets of stable critical points and differential Stackelberg equilibrium coincide. It is worth noting that this is an equivalent property to what was shown for the Stackelberg gradient dynamics in Chapter 2 for zero-sum games with any $\tau \in (0, \infty)$.

We remark that the result regarding τ-GDA as $\tau \to \infty$ is in a sense tight across the spectrum of nonconvex-nonconcave zero-sum games in a way that we now explain. Jin et al. (2020) show (in Proposition 27) two interesting examples: (a) for an a priori fixed, finite τ, there exists a game with a differential Stackelberg equilibrium that is not stable and (b) for an a priori fixed, finite τ, there exists a game with a stable critical point that is not a differential Stackelberg equilibrium. However, (a) does not imply that for the constructed game, there does not exist another (finite) τ—independent of the game parameters—such the differential Stackelebrg equilibrium is stable for all larger τ. In simple language, the result summarized in (a) says the following: if a bad timescale separation is chosen, then convergence may not be guaranteed. Similarly, (b) does not imply that there is no τ such that for all larger τ for the constructed game instance, the critical point becomes unstable. Again, in simple language, the result summarized in (b) says the following: if a bad timescale separation is chosen, then non-game theoretically meaningful equilibria may



(a) $\tau = 1$ (b) $\tau \to \infty$

Figure 3.1: Graphical representation of the known stability results on τ-GDA in relationship to local equilibrium concepts with $\tau = 1$ and $\tau \to \infty$. The acronyms in the figure are differential Nash equilibria (DNE), differential Stackelberg equilibria (DSE), and locally asymptotically stable equilibria (LASE). Note that the terminology of locally asymptotically stable equilibrium refers to the set of stable critical points with respect to the system $\dot{x} = -\Lambda_\tau g(x)$ for the given τ. The subset relationship between differential Nash equilibrium and differential Stackelberg equilibrium is from Proposition 2.2 in Chapter 2. For the regime of $\tau = 1$, Daskalakis and Panageas (2018); Mazumdar et al. (2020) presented the relationship between the set of differential Nash equilibrium and the set of locally asymptotically stable equilibrium, and Jin et al. (2020) provided the relationship between the set of differential Stackelberg equilibrium and the set of locally asymptotically stable equilibrium. Finally, Jin et al. (2020) reported the characterization of the locally asymptotically stable equilibrium as $\tau \to \infty$. The missing pieces in the literature are results as a function of finite τ, which we seek to answer in this work.

persist. While at first glance this set of results may appear to indicate that the undesirable stability characteristics of gradient descent without timescale separation cannot be averted by any finite timescale separation, it is important to emphasize that these results do not answer the questions of whether there (a) exists a game with a differential Stackelberg equilibrium that is not stable for all finite timescale separation ratios or (b) exists a game with a critical point that is not a differential Stackelberg equilibrium which is stable with respect to gradient descent-ascent without timescale separation and remains stable for all finite timescale separation ratios. For more details on a comparison between the results of this chapter and Proposition 27 of Jin et al. (2020), see Appendix J of the paper associated with this chapter (Fiez and Ratliff, 2020). The preceding questions are left open from previous work and are exactly the focus of this chapter. Specifically, from a different perspective, we seek to show that (i) given a fixed game and differential Stackelberg equilibrium, there exists a range of finite learning rate ratios for which the equilibrium is stable and (ii) given a fixed game and a non-equilibrium critical point, there exists a range of finite learning rate ratios for which the critical point is not stable. With an eye toward this goal in the following section, we now provide further background on these existing results along with illustrative examples that motivate our theoretical results.

Let us begin by considering the set of differential Nash equilibrium. It is nearly immediate from the structure of the Jacobian that each differential Nash equilibrium is stable for $\tau = 1$ (Daskalakis and Panageas, 2018; Mazumdar et al., 2020). Moreover, Jin et al. (2020) showed that regardless of the value of $\tau \in (0, \infty)$, the set of differential Nash equilibrium remain stable. In other words,

the desirable stability characteristics of differential Nash equilibrium are retained for any choice of timescale separation. We state this result as a proposition for later reference.[1]

Proposition 3.1. *Consider a zero-sum game $(f_1, f_2) = (f, -f)$ defined by $f \in C^q(\mathcal{X}, \mathbb{R})$ for some $q \geq 2$. Suppose that x^* is a differential Nash equilibrium. Then,* $\text{spec}(-J_\tau(x^*)) \subset \mathbb{C}_-^\circ$ *for all* $\tau \in (0, \infty)$.

Recall from Proposition 2.2 in Chapter 2 that the set of differential Nash equilibrium is a subset of the set of differential Stackelberg equilibrium in zero-sum games. In other words, any differential Nash equilibrium is a differential Stackelberg equilibrium. Thus, this result shows that any differential Stackelberg equilibrium that is a differential Nash equilibrium is stable for any choice of $\tau \in (0, \infty)$, which is a positive result when seeking equilibrium.

Now let us move into the negative results for gradient descent-ascent without timescale separation and explore how finite timescale separation can remedy the issues. Jin et al. (2020) show that the result of Proposition 3.1 fails to extend from the set of differential Nash equilibrium to the broader class of differential Stackelberg equilibrium. Indeed, not all differential Stackelberg equilibrium are stable with respect to the τ-`GDA` continuous-time limiting dynamics with $\tau = 1$. However, as the following example demonstrates, differential Stackelberg equilibrium that are unstable without timescale separation can become stable for a range of finite timescale parameters.

Example 3.1. *Consider the quadratic zero-sum game defined by the cost*

$$f(x_1, x_2) = \tfrac{v}{2}(-x_{11}^2 + \tfrac{1}{2}x_{12}^2 - 2x_{11}x_{21} - \tfrac{1}{2}x_{21}^2 + x_{12}x_{22} - x_{22}^2)$$

where $x_1, x_2 \in \mathbb{R}^2$ and $v > 0$. The unique critical point of the game given by $x^* = (0, 0)$ is a differential Stackelberg equilibrium since $g(x^*) = 0$, $\mathtt{S}_1(J(x^*)) = \text{diag}(v, v/4) \succ 0$ and $-D_2^2 f(x^*) = \text{diag}(v/2, v) \succ 0$. The spectrum of the Jacobian of $-\Lambda_\tau g(x)$ is given by

$$\text{spec}(-J_\tau(x^*)) = \left\{ \frac{-v(2\tau + 1 \pm \sqrt{4\tau^2 - 8\tau + 1})}{4}, \frac{-v(\tau - 2 \pm \sqrt{\tau^2 - 12\tau + 4})}{4} \right\}.$$

Observe that for $\tau = 1$, $\text{spec}(-J_\tau(x^*)) \not\subset \mathbb{C}_-^\circ$ for any $v > 0$ so that the differential Stackelberg equilibrium x^* is never stable for the choice of τ. However, for any $v > 0$, $\text{spec}(-J_\tau(x^*)) \subset \mathbb{C}_-^\circ$ for all $\tau \in (2, \infty)$, meaning that the differential Stackelberg equilibrium x^* is stable with respect to the dynamics $\dot{x} = -\Lambda_\tau g(x)$ for a range of finite learning rate ratios.

We explore Example 3.1 further via simulations in Section 3.7.1. The key takeaway from Example 3.1 is that it is clearly not always necessary for the timescale separation τ to approach infinity in order to guarantee the stability of a differential Stackelberg equilibrium and instead there exists a sufficient finite learning rate ratio. Put simply, the undesirable property of differential Stackelberg equilibria not being stable with respect to gradient descent-ascent without timescale separation can potentially be remedied with only a finite timescale separation.

It is well-documented that some stable critical points of the continuous time gradient descent-ascent limiting dynamics without timescale separation can lack game-theoretic meaning, as they may be neither a differential Nash equilibria nor differential Stackelberg equilibria (Daskalakis and

[1] Also see Fiez and Ratliff (2020, Appendix F.3) for an alternative proof.

Panageas, 2018; Jin et al., 2020; Mazumdar et al., 2020). The following example demonstrates that such undesirable critical points that are stable without timescale separation can become unstable for a range of finite learning ratios.

Example 3.2. *Consider the quadratic zero-sum game defined by the cost*

$$f(x_1, x_2) = \tfrac{v}{4}(x_{11}^2 - \tfrac{1}{2}x_{12}^2 + 2x_{11}x_{21} + \tfrac{1}{2}x_{21}^2 + 2x_{12}x_{22} - x_{22}^2)$$

where $x_1, x_2 \in \mathbb{R}^2$ *and* $v > 0$*. The unique critical point* $x^* = (0, 0)$ *is not a differential Stackelberg or a Nash equilibrium since* $D_1^2 f(x^*) = \operatorname{diag}(v/2, -v/4) \not\succ 0$*,* $D_2^2 f(x^*) = \operatorname{diag}(v/4, -v/2) \not\prec 0$*. Moreover,*

$$\operatorname{spec}(-J_\tau(x^*)) = \{\tfrac{-v}{8}(2\tau - 1 \pm \sqrt{4\tau^2 - 12\tau + 1}), \tfrac{-v}{8}(2 - \tau \pm \sqrt{\tau^2 - 12\tau + 4})\}.$$

Observe that for any $v > 0$*,* x^* *is stable for* $\tau = 1$ *since* $\operatorname{spec}(-J_\tau(x^*)) \subset \mathbb{C}_-^\circ$*, but* x^* *is unstable for a range of learning rates since* $\operatorname{spec}(-J_\tau(x^*)) \not\subset \mathbb{C}_-^\circ$ *for all* $\tau \in (2, \infty)$*. This is not an artifact of the quadratic example: games can be constructed in which stable critical points lacking game-theoretic meaning become unstable for all* $\tau > \tau_0$ *even in the presence of multiple equilibria.*

We investigate a variant of the game in Example 3.2 that has multiple critical points with simulations in Section 3.7.2. In an analogous manner to Example 3.1, Example 3.2 demonstrates that it is not always necessary for the timescale separation τ to approach infinity in order to guarantee non-equilibrium critical points become unstable as there can exist a sufficient finite learning rate ratio. This is to say that the unwanted property of non-equilibrium critical points being stable without timescale separation can also potentially be remedied with only a finite timescale separation.

The examples of this section have provided evidence that there exists a range of finite learning rate ratios for which differential Stackelberg equilibrium are stable and a range of learning rate ratios for which non-equilibrium critical points are unstable. Yet, no result has appeared in the literature on gradient descent-ascent with timescale separation confirming this behavior in general. We focus on doing precisely that in the section that follows. Before doing so, we remark further on the closest existing result. As mentioned previously Jin et al. (2020) show that as $\tau \to \infty$, the set of stable critical points with respect to the dynamics $\dot{x} = -\Lambda_\tau g(x)$ coincide with the set of differential Stackelberg equilibrium. An equivalent result in the context of general singularly perturbed systems has been known in the literature (cf. Kokotovic et al. 1986, Chap. 2). We give a proof sketch based on this type of analysis in order to provide intuition and since it reveals a set of analysis tools that are not as common in the machine learning literature on games. A full proof can be found in Appendix I of the paper associated with this chapter (Fiez and Ratliff, 2020).

Proposition 3.2 (Jin et al. 2020; Kokotovic et al. 1986)**.** *Consider a zero-sum game* $(f_1, f_2) = (f, -f)$ *defined by* $f \in C^q(\mathcal{X}, \mathbb{R})$ *for some* $q \geq 2$*. Suppose that* x^* *is such that* $g(x^*) = 0$ *and* $\det(D_2^2 f_2(x^*)) \neq 0$*. Then, as* $\tau \to \infty$*,* $\operatorname{spec}(J_\tau(x^*)) \subset \mathbb{C}_+^\circ$ *if and only if* x^* *is a differential Stackelberg equilibrium.*

Proof Sketch. The basic idea in showing this result is that there is a (local) transformation of

coordinates from the linearized dynamics of $\dot{x} = -\Lambda_\tau g(x)$, which we write as

$$\dot{x} = \begin{bmatrix} A_{11} & A_{12} \\ -\tau A_{12}^\top & \tau A_{22} \end{bmatrix} x,$$

in a neighborhood of a critical point to an upper triangular system that depends parametrically on τ and hence, the asymptotic behavior is readily obtainable from the block diagonal components of the system in the new coordinates. Indeed, consider the change of variables $z = x_2 + L(\tau)x_1$ where $L(\tau) \in \mathbb{R}^{d_2 \times d_1}$ is a transformation such that

$$\begin{bmatrix} \dot{x}_1 \\ \dot{z} \end{bmatrix} = \begin{bmatrix} A_{11} - A_{12}L(\tau) & A_{12} \\ R(L,\tau) & A_{22} + \tau^{-1}L(\tau)A_{12} \end{bmatrix} \begin{bmatrix} x_1 \\ z \end{bmatrix} \tag{3.5}$$

where

$$R(L,\tau) = -A_{12}^\top - A_{22}L(\tau) + \tau^{-1}L(\tau)A_{11} - \tau^{-1}L(\tau)A_{12}L(\tau) = 0$$

A transformation of coordinates $L(\tau)$ such that $R(L,\tau) = 0$ always exists (see Lemma I.1, Appendix I, Fiez and Ratliff 2020). Hence, the characteristic equation of (3.5) can be expressed as

$$\chi(s,\tau) = \tau^d \chi_s(s,\tau)\chi_f(p,\tau) = 0$$

where $\chi_s(s,\tau) = \det(sI - (A_{11} - A_{12}L(\tau)))$ and $\chi_f(p,\tau) = \det(pI - (A_{22} + \tau^{-1}A_{12}L(\tau)))$ with $p = s\tau^{-1}$. As $\tau \to \infty$, $L(\tau) \to L(\infty) = -A_{22}^{-1}A_{12}^\top$. Consequently, d_1 of the eigenvalues of $\dot{x} = -\Lambda_\tau g(x)$, denoted by $\{\lambda_1, \ldots, \lambda_{d_1}\}$, are the roots of the slow characteristic equation $\chi_s(s,\tau) = 0$ and the rest of the eigenvalues $\{\lambda_{d_1+1}, \ldots, \lambda_{d_1+d_2}\}$ are denoted by $\lambda_i = \nu_j/\varepsilon$ for $i = d_1 + j$ and $j \in \{1, \ldots, d_2\}$ where $\{\nu_1, \ldots, \nu_{d_2}\}$ are the roots of the fast characteristic equation $\chi_f(p,\tau) = 0$. The roots of $\chi_s(s,\tau)$ are precisely those of the (first) Schur complement of $-J_\tau(x^*)$ while the roots of $\chi_f(p,\tau)$ are precisely those of $D_2^2 f(x^*)$. Thus, stability holds if and only if it is a differential Stackelberg equilibrium. □

This simple transformation of coordinates to an upper triangular dynamical system shown in (3.5) leads immediately to the asymptotic result in Proposition 3.2. It also shows that if the eigenavlues of $\mathtt{S}_1(J_\tau(x^*))$ are distinct[2] and similarly, so are those of $D_2^2 f(x^*)$ (although, $\mathtt{S}_1(J_\tau(x^*))$ and $D_2^2 f(x^*)$ are allowed to have eigenvalues in common), then the asymptotic results from Proposition 3.2 imply the following approximations for the elements of $\mathrm{spec}(J_\tau(x^*))$:

$$\lambda_i = \lambda_i(\mathtt{S}_1(J_\tau(x^*)) + O(\tau^{-1}),\ i = 1, \ldots, d_1,$$
$$\lambda_{j+d_1} = \tau(\lambda_j(-D_2^2 f(x^*)) + O(\tau^{-1})),\ j = 1, \ldots, d_2.$$

This follows simply by observing that when the eigenvalues are distinct, the derivatives $ds/d\tau$ and $dp/d\tau$ are well-defined by the implicit mapping theorem and the total derivative of $\chi_s(s,\tau)$ and $\chi_f(p,\tau)$, respectively. It is worth noting that this result shows that as $\tau \to \infty$, the Jacobian of the τ-`GDA` dynamics tends toward something that is locally equivalent to the Jacobian of the Stackelberg gradient dynamics studied extensively in Chapter 2. In the following section, we generalize the insights from the examples in this section and present the main theoretical results.

[2]Distinct eigenvalues is a generic property in the space of $d \times d$ real matrices (Hirsch et al., 2012, Chapter 5.6).

3.4 Stability of Continuous-Time GDA with Timescale Separation

The sections presents the main local stability and instability characterizations of this chapter for the continuous-time gradient descent-ascent system with timescale separation. We remark that these properties with respect to the continous-time dynamics can be translated into local convergence and non-convergence results as will be become clear in Sections 3.5 and 3.6.

3.4.1 Necessary and Sufficient Conditions for Stability

Before stating the main result of this subsection, let us pick up where we left off at the end of the last section. As was shown in Proposition 3.2 and discussed following the proof sketch, when $\tau \to \infty$, the spectrum of $-J_\tau(x^*)$ asymptotically splits as $\tau \to \infty$ such that d_1 eigenvalues tend to fixed positions defined by the eigenvalues of $-\mathtt{S}_1(J(x^*))$, while the remaining d_2 eigenvalues tend to infinity at a linear rate τ along asymptotes defined by the eigenvalues of $D_2^2 f(x^*)$. The result is due to Klimushchev and Krasovskii (1961) and further discussion can be found in Appendix F of Fiez and Ratliff (2020) and Chapter 2 of Kokotovic et al. (1986). This fact is specialized from the class of singularly perturbed linear systems to τ-GDA by Jin et al. (2020) which directly results in the connection between critical points of ∞–GDA and differential Stackelberg equilibrium. Specifically, the result of Jin et al. (2020) is showing that for the class of all sufficiently smooth zero-sum games, the stable critical points of ∞-GDA are exactly the differential Stackelberg equilibria. As a corollary of this fact, there exists a $\tau_1 < \infty$, such that τ-GDA is stable for all $\tau > \tau_1$ (Kokotovic et al., 1986, Chap. 2, Cor. 3.1); this can be inferred from the proof of Theorem 28 in Jin et al. (2020) as well. Indeed, Jin et al. (2020) gives an asymptotic expansion showing that n_1 eigenvalues of $-J_\tau(x^*)$ are in $\text{spec}(-\mathtt{S}_1(J(x^*))) + O(\tau^{-1})$ and the remaining n_2 eigenvalues are in $\tau(\text{spec}(D_2^2 f(x^*)) + O(\tau^{-1}))$. Using the limit definition for the asymptotic expansion, for any fixed game and a differential Stackelberg x^*, one can show that there exists a finite τ such that x^* is stable. Unfortunately, the finite τ_1 obtainable from the asymptotic expansion method can be arbitrarily large. From a practical perspective, this poses significant problems for the implementation and performance of τ-GDA. Indeed, the *eigenvalue gap* between $\text{spec}(-\mathtt{S}_1(J(x^*)))$ and $\text{spec}(\tau D_2^2 f(x^*))$ has a linear dependence on τ and, in turn, the problem may become highly ill-conditioned from a numerical perspective as τ becomes large (Kokotovic, 1975). In contrast, in a non-asymptotic way, we determine exactly the range of τ such that the spectrum of $-J_\tau(x)$ remains in $\mathbb{C}_-^\circ$, and hence, remedy this problem.

For the statement of the following theorem on the non-asymptotic construction of τ^*, we define the following matrices: for a critical point x^*, let $\mathtt{S}_1 = \mathtt{S}_1(-J_\tau(x^*)) = A_{11} - A_{12}A_{22}^{-1}A_{12}^\top$ and

$$-J_\tau(x^*) = \begin{bmatrix} -D_1^2 f(x^*) & -D_{12} f(x^*) \\ \tau D_{12}^\top f(x^*) & \tau D_2^2 f(x^*) \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ -\tau A_{12}^\top & \tau A_{22} \end{bmatrix}.$$

Theorem 3.3 (Non-Asymptotic Construction of Necessary and Sufficient Conditions for Stability)**.** *Consider a zero-sum game $(f_1, f_2) = (f, -f)$ defined by $f \in C^q(\mathcal{X}, \mathbb{R})$ for some $q \geq 2$. Suppose that x^* is such that $g(x^*) = 0$ and $\det(D_2^2 f_2(x^*)) \neq 0$. There exists a $\tau^* \in [0, \infty)$ such that $\text{spec}(-J_\tau(x^*)) \subset \mathbb{C}_-^\circ$ for all $\tau \in (\tau^*, \infty)$ if and only if x^* is a differential Stackelberg equilibrium.*

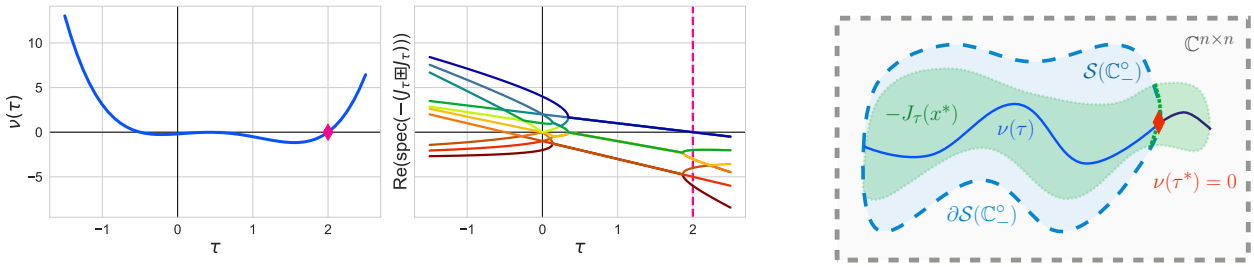


Figure 3.2: Guard map $\nu(\tau)$ and real parts of the eigenvalues of the vectorized Lyapunov operator $-(J_\tau(x^*) \boxplus J_\tau(x^*))$ using the reduction via the duplication matrix for the quadratic example given in Example 3.1. The largest real positive root of $\nu(\tau)$ is $\tau^* = 2$ and in the right plot, we see that all the real parts of the eigenvalues of the Lyapunov operator are negative indicating stability. The right most graphic is a cartoon visualization of the guard map method: the outer grey region represents $\mathbb{C}^{d\times d}$, the blue region represents the Hurwitz stable $d \times d$ matrices $\mathcal{S}(\mathbb{C}^\circ_-)$, the green region represents the parameterized class of matrices $\{J_\tau(x^*)\}_\tau$, and the curve cutting through the regions is the guard map $\nu(\tau)$. The goal is to fine the subset of $\{J_\tau(x^*)\}_\tau$ that lie within $\mathcal{S}(\mathbb{C}^\circ_-)$, which can be done by reducing the problem to finding the roots of $\nu(\tau)$.

Moreover, $\tau^* = \lambda^+_{\max}(Q)$ *where*

$$Q = 2\left[(A_{12}\otimes A_{22}^{-1})H_{d_2} \quad (I_{d_1}\otimes A_{22}^{-1}A_{12}^\top)H_{d_1}\right]\begin{bmatrix}\bar{A}_{22}^{-1}H^+_{d_2}(A_{12}^\top\otimes I_{d_2})\\ -\bar{\mathtt{S}}_1^{-1}H^+_{n_1}(\mathtt{S}_1\otimes A_{12}A_{22}^{-1})\end{bmatrix} - (A_{11}\otimes A_{22}^{-1})$$

with $\bar{A}_{22} = A_{22}\boxplus A_{22}$ *and* $\bar{\mathtt{S}}_1 = \mathtt{S}_1 \boxplus \mathtt{S}_1$.

While at first glance Q may appear difficult to understand, it is efficiently computable and can be used to understand the typical value for important classes of games. Indeed, many problems like generative adversarial networks have specific structure for the individual Hessians of each player and the interaction matrix $D_{12}f$ (as can be seen from Assumption 3.1, Section 3.4.3) and are in a sense subject to design via network architecture and loss function selection. This result opens up an interesting future direction of research on understanding and potentially designing the structure of Q. To take a step in this direction, we explore a number of games in Section 3.7 where we compute τ^* by the construction and validate it is tight empirically. Along the way, we discover that τ^* is typically a reasonable value that is amenable to practical implementations. We now give an outline of the proof technique. The full proof of Theorem 3.3 is deferred to Section 3.B in the appendix that follows this chapter.

Proof Sketch of Theorem 3.3. The key tools used in this proof are a combination of Lyapunov stability and the notion of a *guard map* (Saydy et al., 1990), a new tool to the learning community. Recall that a matrix is exponentially stable if and only if there exists a symmetric positive definite $P = P^\top \succ 0$ such that $PJ_\tau(x^*) + J_\tau^\top(x^*)P \succ 0$ (Khalil, 2002, Thm. 4.15). Hence, given a positive definite $Q = Q^\top \succ 0$, $-J_\tau(x^*)$ is stable if and only if there exists a unique solution $P = P^\top$ to

$$((J_\tau^\top(x^*)\otimes I) + (I\otimes J_\tau^\top(x^*)))\mathrm{vec}(P) = (J_\tau^\top(x^*)\oplus J_\tau^\top(x^*))\mathrm{vec}(P) = \mathrm{vec}(Q) \tag{3.6}$$

where $\otimes$ and $\oplus$ denote the Kronecker product and Kronecker sum, respectively.[3] The existence of a unique solution P occurs if and only if $J_\tau^\top$ and $-J_\tau^\top$ have no eigenvalues in common. Hence, using the fact that eigenvalues vary continuously, if we vary τ and examine the eigenvalues of the map $J_\tau^\top(x^*) \oplus J_\tau^\top(x^*)$, this tells us the range of τ for which $\text{spec}(-J_\tau(x^*))$ remains in $\mathbb{C}_-^\circ$.

This method of varying parameters and determining when the roots of a polynomial (or correspondingly, the eigenvalues of a map) cross the boundary of a domain uses a *guard map*; it provides a certificate that the roots of a polynomial lie in a particular guarded domain for a range of parameter values. Formally, let X be the set of all $d \times d$ real matrices or the set of all polynomials of degree d with real coefficients. Consider $\mathcal{S}$ an open subset of X with closure $\bar{\mathcal{S}}$ and boundary $\partial\mathcal{S}$. The map $\nu : X \to \mathbb{C}$ is said to be a guardian map for $\mathcal{S}$ if for all $x \in \bar{\mathcal{S}}$,

$$\nu(x) = 0 \iff x \in \partial\mathcal{S}.$$

Elements of $\mathcal{S}(\mathbb{C}_-^\circ) = \{A \in \mathbb{R}^{d\times d} : \text{spec}(A) \subset \mathbb{C}_-^\circ\}$ are (Hurwitz) stable. Given a pathwise connected set $U \subseteq \mathbb{R}$, the parameterized family $\{A(\tau) : \tau \in U\}$ is stable if and only if (i) it is nominally stable—meaning $A(\tau_1) \in \mathcal{S}(\mathbb{C}_-^\circ)$ for some $\tau_1 \in U$—and (ii) $\nu(A(\tau)) \neq 0$ for all $\tau \in U$ (Saydy et al., 1990, Prop. 1).

The map

$$\nu(\tau) = \det(2(-J_\tau(x^*) \odot I)) = \det(-(J_\tau(x^*) \oplus J_\tau(x^*)))$$

guards $\mathcal{S}(\mathbb{C}_-^\circ)$ where $\odot$ is the *bialternate product* and is defined by $A \odot B = \frac{1}{2}(A \oplus B)$ for matrices A and B (see Govaerts 2000, Sec. 4.4.4). For intuition, consider the case where each $x_1, x_2 \in \mathbb{R}$ so that

$$J_\tau(x^*) = \begin{bmatrix} a & b \\ -\tau b & \tau d \end{bmatrix} \in \mathbb{R}^{2\times 2}.$$

It is known that $\text{spec}(-J_\tau(x^*)) \subset \mathbb{C}_-^\circ$ if $\det(-J_\tau(x^*)) \succ 0$ and $\text{tr}(-J_\tau(x^*)) < 0$. Consequently $\nu(\tau) = \det(-J_\tau(x^*))\,\text{tr}(-J_\tau(x^*))$ is a guard map for the 2×2 stable matrices $\mathcal{S}(\mathbb{C}_-^\circ)$. Since the bialternate product generalizes the trace operator and $\det(-J_\tau(x^*)) = \tau^{d_2}\det(D_2^2 f(x^*))\det(-\mathtt{S}_1(J(x^*))) \neq 0$ for $\tau \neq 0$ by the facts ($\det(\mathtt{S}_1(J(x^*))) \neq 0$ and $\det(D_2^2 f(x^*)) \neq 0$) for a differential Stackelberg equilibrium x^*, a guard map in the general $d \times d$ case is $\nu(\tau) = \det(-(J_\tau(x^*) \oplus J_\tau(x^*)))$.

This guard map in τ is closely related to the vectorization in (3.6): for any symmetric positive definite $Q = Q^\top \succ 0$, there will be a symmetric positive definite solution $P = P^\top \succ 0$ of

$$-(J_\tau^\top(x^*) \oplus J_\tau^\top(x^*))\text{vec}(P) = \text{vec}(-Q)$$

if and only if $\det(-(J_\tau(x^*) \oplus J_\tau(x^*))) \neq 0$. Hence, to find the range of τ for which, given any $Q = Q^\top \succ 0$, the solution $P = P^\top$ is no longer positive definite, we need to find the value of τ such that $\nu(\tau) = \det(-(J_\tau(x^*) \oplus J_\tau(x^*))) = 0$—that is, where it hits the boundary $\partial\mathcal{S}(\mathbb{C}_-^\circ)$. Through algebraic manipulation, this problem reduces to an eigenvalue problem in τ, giving rise to an explicit construction of τ^*. Note that in the full proof, and to get the statement of the result,

---

[3] See Lancaster and Tismenetsky (1985); Magnus (1988) for more detail on the definition and properties of these mathematical operators.

we actually use the equivalent guard map

$$\nu(\tau) = \det(-(J_\tau(x^*) \boxplus J_\tau(x^*)))$$

since $A \boxplus A$ is a more computationally efficient expression of $A \oplus A$, and as such the eigenvalues of $A \boxplus A$ are those of $A \oplus A$ removing redundancies. We use $A \boxplus A$ specifically because of its computational advantages in computing τ^*. □

Selecting the maximum value of τ^* over the finite set of equilibria guarantees that the local linearization of $\dot{x} = -\Lambda_\tau g(x)$ around any differential Stackelberg equilibria is stable, and hence, the nonlinear system is locally stable around each of these critical points.

Corollary 3.1. *Consider a zero-sum game $\mathcal{G} = (f, -f)$ with $f \in C^q(\mathcal{X}, \mathbb{R})$ for some $q \geq 2$. Suppose that the assumptions of Theorem 3.3 hold and that the set of differential Stackelberg equilibria, denoted $\texttt{DSE}(\mathcal{G})$, is finite. Let $\tau^* = \max_{x^* \in \texttt{DSE}(\mathcal{G})} \tau(x^*)$ where $\tau(x^*)$ is the value of τ obtained via Theorem 3.3 for each individual critical point $x^* \in \texttt{DSE}(\mathcal{G})$. Then, for all $\tau \in (\tau^*, \infty)$ and $x^* \in \texttt{DSE}(\mathcal{G})$, $\text{spec}(-J_\tau(x^*)) \subset \mathbb{C}_-^\circ$.*

Before moving on, we remark on the utility of the algebraic tools we use in the proof for Theorem 3.3. Indeed, the guard map concept is extremely powerful for understanding stability of parameterized families of dynamical systems, and it is not limited to single parameter families. Hence, there is potential to extend the above results to games with more than two players or additional parameters. In fact, we do exactly this in Section 3.4.3 where we present results for generative adversarial network formulations with gradient-penalty type regularizers for the discriminator. Moreover, it is fairly easy to construct analogous guard maps for non-zero sum games. Many of the constructions readily extend, and these are interesting directions of future work.

3.4.2 Sufficient Conditions for Instability

Note that Theorem 3.3 implies that for any critical point which is not a differential Stackelberg equilibrium, there is no finite τ^* such that $\text{spec}(-J_\tau(x^*)) \subset \mathbb{C}_-^\circ$ for all $\tau \in (\tau^*, \infty)$. In particular, there exists at least one finite, positive value of τ such that $\text{spec}(-J_\tau(x^*)) \not\subset \mathbb{C}_-^\circ$. We can extend this result to answer the question of whether there exists a finite learning rate ratio τ_0 such that $-J_\tau(x^*)$ has at least one eigenvalue with strictly positive real part for all $\tau \in (\tau_0, \infty)$, thereby implying that x^* is unstable.

Theorem 3.4 (Non-Asymptotic Construction of Sufficient Condition for Instability.)**.** *Consider a zero-sum game $(f_1, f_2) = (f, -f)$ defined by $f \in C^q(\mathcal{X}, \mathbb{R})$ for some $q \geq 2$. Suppose that x^* is such that $g(x^*) = 0$, $D_2^2 f_2(x^*) \neq 0$, and x^* is not a differential Stackelberg equilibrium. Then $\text{spec}(-J_\tau(x^*)) \not\subset \mathbb{C}_-^\circ$ for all $\tau \in (\tau_0, \infty)$ with*

$$\begin{aligned}\tau_0 = \lambda_{\max}^+(&Q_2^{-1}((P_1 D_{12} f(x^*) + \mathtt{S}_1(-J(x^*))L_0^\top P_2)^\top Q_1^{-1}(P_1 D_{12} f(x^*) \\ &+ \mathtt{S}_1(-J(x^*))L_0^\top P_2) - P_2 L_0 D_{12} f(x^*) - (P_2 L_0 D_{12} f(x^*))^\top)).\end{aligned}$$

where P_1, P_2, Q_1, Q_2 are any non-singular Hermitian matrices such that (a) $Q_i \succ 0$ for each $i \in \mathcal{I}$, (b) $\mathtt{S}_1(-J(x^*))P_1 + P_1\mathtt{S}_1(-J(x^*)) = Q_1$ and $D_2^2 f(x^*)P_2 + P_2 D_2^2 f(x^*) = Q_2$, and (c) the following matrix pairs have the same inertia: $(P_1, \mathtt{S}_1(-J(x^*)))$ and $(P_2, D_2^2 f(x^*))$.

The proof of Theorem 3.4 is left to Section 3.C in the appendix that follows this chapter. We now provide a proof sketch highlighting the key techniques.

Proof Sketch. The key idea of the proof is to leverage the Lyapunov equation and Theorem 2 of Lancaster and Tismenetsky (1985, Chapter 13.1) to show that $-J_\tau(x^*)$ has at least one eigenvalue with strictly positive real part. Indeed, Theorem 2 of Lancaster and Tismenetsky (1985, Chapter 13.1) in the context of this problem states that if $\mathtt{S}_1(-J(x^*))$ has no zero eigenvalues, then there exists matrices $P_1 = P_1^\top$ and $Q_1 = Q_1^\top \succ 0$ such that $P_1\mathtt{S}_1(-J(x^*)) + \mathtt{S}_1(-J(x^*))P_1 = Q_1$ where P_1 and $\mathtt{S}_1(-J(x^*))$ have the same *inertia*—that is, the number of eigenvalues with positive, negative and zero real parts, respectively, are the same. An analogous statement applies to $-D_2^2 f(x^*)$ with some P_2 and Q_2. Since x^* is a non-equilibrium critical point, without loss of generality, let $\mathtt{S}_1(-J(x^*))$ have at least one strictly positive eigenvalue so that P_1 does as well. Next, we construct a matrix P that is *congruent* to blockdiag(P_1, P_2) and a matrix Q_τ such that $-PJ_\tau(x^*) - J_\tau^\top(x^*)P = Q_\tau$. Since P and blockdiag(P_1, P_2) are congruent, Sylvester's law of inertia implies that they have the same number of eigenvalues with positive, negative, and zero real parts, respectively. Hence, P has at least one eigenvalue with strictly positive real part. We then construct τ_0 via an eigenvalue problem such that for all $\tau > \tau_0$, $Q_\tau \succ 0$. Applying Theorem 2 of Lancaster and Tismenetsky (1985, Chapter 13.1) again, for any $\tau > \tau_0$, $-J_\tau(x^*)$ has at least one eigenvalue with strictly positive real part so that $\mathrm{spec}(-J_\tau(x^*)) \not\subset \mathbb{C}_-^\circ$. □

Unlike τ^* in Theorem 3.3, τ_0 in Theorem 3.4 is not tight in the sense that $-J_\tau(x^*)$ may become unstable for $\tau < \tau_0$ since the matrices P_1, P_2 and Q_1, Q_2 are not necessarily unique. Hence, the question of finding the exact value of τ beyond which a spurious critical point of `GDA` is unstable remains open. Nonetheless, no result has appeared previously showing that `GDA` with a finite timescale separation avoids such critical points.

3.4.3 Regularization with Applications to Adversarial Learning

In this subsection, we focus on generative adversarial networks with regularization and using the theory developed so far extend the results to provide a stability guarantee for a range of regularization parameters and learning rate ratios. Consider the training objective

$$f(\theta, \omega) = \mathbb{E}_{p(z)}[\ell(\mathrm{D}(\mathrm{G}(z;\theta);\omega))] + \mathbb{E}_{p_{\mathcal{D}}(x)}[\ell(-\mathrm{D}(x;\omega))] \tag{3.7}$$

where $\mathrm{D}_\omega(x)$ and $\mathrm{G}_\theta(z)$ are discriminator and generator networks, $p_{\mathcal{D}}(x)$ is the data distribution while $p(z)$ is the latent distribution, and $\ell \in C^2(\mathbb{R})$ is some real-value function.[4] Nagarajan and Kolter (2017) show, under suitable assumptions, that gradient-based methods for training generative adversarial networks are locally convergent assuming the data distributions are absolutely continuous. However, as observed by Mescheder et al. (2018), such assumptions not only may not be satisfied by many practical generative adversarial network training scenarios such as natural images, but often the data distribution is concentrated on a lower dimensional manifold. The latter characteristic leads to highly ill-conditioned problems and nearly purely imaginary eigenvalues.

Gradient penalties ensure that the discriminator cannot create a non-zero gradient which is orthogonal to the data manifold without suffering a loss. Introduced by Roth et al. (2017) and

[4]For example, $\ell(x) = -\log(1 + \exp(-x))$ gives the original formulation of Goodfellow et al. (2014).

refined in Mescheder et al. (2018), we consider training generative adversarial networks with one of two fairly natural gradient-penalties used to regularize the discriminator:

$$R_1(\theta,\omega)=\frac{\mu}{2}\mathbb{E}_{p_{\mathcal{D}}(x)}[\|\nabla_x \mathrm{D}(x;\omega)\|^2] \quad \text{and} \quad R_2(\theta,\omega)=\frac{\mu}{2}\mathbb{E}_{p_\theta(x)}[\|\nabla_x \mathrm{D}(x;\omega)\|^2],$$

where, by a slight abuse of notation, $\nabla_x(\cdot)$ denotes the partial gradient with respect to x of the argument $(\cdot)$ when the argument is the discriminator $\mathrm{D}(\cdot;\omega)$ in order prevent any conflation between the notation $D(\cdot)$ elsewhere for derivatives. We consider the relaxed assumptions—as compared to the work by Nagarajan and Kolter (2017)—which allow us to consider generative adversarial networks with data distributions that do not (locally) have the same support and hence, are concentrated on lower dimensional manifolds. Let $h_1(\theta)=\mathbb{E}_{p_\theta(x)}[\nabla_\omega \mathrm{D}(x;\omega)|_{\omega=\omega^*}]$ and $h_2(\omega)=\mathbb{E}_{p_{\mathcal{D}}(x)}[|\mathrm{D}(x;\omega)|^2+\|\nabla_x \mathrm{D}(x;\omega)\|^2]$. Define *reparameterization manifolds* $\mathcal{M}_{\mathrm{G}}=\{\theta:\ p_\theta=p_{\mathcal{D}}\}$ and $\mathcal{M}_{\mathrm{D}}=\{\omega:\ h_2(\omega)=0\}$ and let $T_{\theta^*}\mathcal{M}_{\mathrm{G}}$ and $T_{\omega^*}\mathcal{M}_{\mathrm{D}}$ denote their respective tangent spaces at θ^* and ω^*. As in Mescheder et al. (2018), we make the following assumption.

Assumption 3.1. *Consider a zero-sum game of the form given in* (3.7) *where* $f\in C^2(\mathbb{R}^{d_1}\times\mathbb{R}^{d_2},\mathbb{R})$ *and* $\mathrm{G}(\cdot;\theta)$ *and* $\mathrm{D}(\cdot;\omega)$ *are the generator and discriminator networks, respectively, and* $x=(\theta,\omega)\in\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}$. *Suppose that* $x^*=(\theta^*,\omega^*)$ *is an equilibrium. Then, (a) at* (θ^*,ω^*), $p_{\theta^*}=p_{\mathcal{D}}$ *and* $\mathrm{D}(x;\omega^*)=0$ *in some neighborhood of* $\mathrm{supp}(p_{\mathcal{D}})$, *(b) the function* $\ell\in C^2(\mathbb{R})$ *satisfies* $\ell'(0)\neq 0$ *and* $\ell''(0)<0$, *(c) there are* ϵ*–balls* $B_\epsilon(\theta^*)$ *and* $B_\epsilon(\omega^*)$ *centered around* θ^* *and* ω^*, *respectively, so that* $\mathcal{M}_{\mathrm{G}}\cap B_\epsilon(\theta^*)$ *and* $\mathcal{M}_{\mathrm{D}}\cap B_\epsilon(\omega^*)$ *define* C^1*-manifolds. Moreover, (i) if* $w\notin T_{\theta^*}\mathcal{M}_{\mathrm{G}}$, *then* $w^\top\nabla_w h_1(\theta^*)w\neq 0$, *and, (ii) if* $v\notin T_{\omega^*}\mathcal{M}_{\mathrm{D}}$, *then* $v^\top\nabla^2_\omega h_2(\omega^*)v\neq 0$.

We note that as explained by Mescheder et al. (2018), Assumption 3.1.c(i) implies that the discriminator is capable of detecting deviations from the generator distribution in equilibrium, and Assumption 3.1.c(ii) implies that the manifold $\mathcal{M}_{\mathrm{D}}$ is sufficiently regular and, in particular, its (local) geometry is captured by the second (directional) derivative of h_2. Under Assumption 3.1, we show that x^* is a differential Stackelberg equilibrium, and characterize the learning rate ratio and regularization parameter range for which x^* is (locally) stable with respect to τ-`GDA`. The proof of Theorem 3.5 is left to Section 3.D of the appendix that follows this chapter.

Theorem 3.5. *Consider training a generative adversarial network via a zero-sum game with generator network* G_θ, *discriminator network* D_ω, *and loss* $f(\theta,\omega)$ *with regularization* $R_j(\theta,\omega)$ *(for either* $j=1$ *or* $j=2$*) and any regularization parameter* $\mu\in(0,\infty)$ *such that Assumption 3.1 is satisfied for a critical point* $x^*=(\theta^*,\omega^*)$ *of the regularized dynamics. Then,* $x^*=(\theta^*,\omega^*)$ *is a differential Stackelberg equilibrium. Furthermore, for any* $\tau\in(0,\infty)$, $\mathrm{spec}(-J_{(\tau,\mu)}(x^*))\subset\mathbb{C}^\circ_-$.

The proof of this result follows arguments from Mescheder et al. (2018) to determine that x^* is a differential Stackelberg equilibrium and then the concept of quadratic numerical range (Tretter, 2008) to determine the stability conditions. We note that the stability conditions can also be obtained using Theorem 3.3 directly (see Appendix H of Fiez and Ratliff (2020)). Notably, this result shows that under typical theoretical generative adversarial network assumptions, any differential Stackelberg equilibrium is stable for any timescale parameter and regularization parameter. This result highlights the value of finer-grained characterizations of the timescale parameters that ensure stability of differential Stackelberg equilibrium since often common properties of critical points, the insights can extend to entire classes of games with much stronger characterizations than the asymptotic results.

The following proposition provides necessary conditions on the sizes of the network architectures for the discriminator and generator network for stability. Theorem A.7 of Mescheder et al. (2018) shows that matrices of the form

$$-J = \begin{bmatrix} 0 & -B \\ B^\top & -C \end{bmatrix} \tag{3.8}$$

are stable if B is full rank and $C \succ 0$. The result that follows provides conditions on the relationships between the dimensions of the players.

Proposition 3.3. *Consider training a generative adversarial network via a zero-sum game with generator network* G_θ*, discriminator network* D_ω*, and loss* $f(\theta, \omega)$ *with regularization* $R_j(\theta, \omega)$ *(for some* $j \in \{1, 2\}$*) such that Assumption 3.1 is satisfied for an equilibrium* $x^* = (\theta^*, \omega^*)$*. Independent of the learning rate ratio and the regularization parameter* μ*, for* x^* *to be stable it is necessary that the dimension of the discriminator network parameter vector is at least half as large as the corresponding generator network parameter vector:* $d_2 \geq d_1/2$ *where* $\theta \in \mathbb{R}^{d_1}$ *and* $\omega \in \mathbb{R}^{d_2}$*.*

The intuition for the why this proposition should hold follows immediately from observing the structure of the Jacobian: for any matrix of the form (3.8), at least one eigenvalue will be purely imaginary if $d_2 < d_1/2$ where $B \in \mathbb{R}^{d_1 \times d_2}$ and $C \in \mathbb{R}^{d_2 \times d_2}$. This proposition follows immediately from observing the structure of the Jacobian: for any matrix of the form

$$-J = \begin{bmatrix} 0 & -B \\ B^\top & -C \end{bmatrix}$$

at least one eigenvalue will be purely imaginary if $d_2 < d_1/2$ where $B \in \mathbb{R}^{d_1 \times d_2}$ and $C \in \mathbb{R}^{d_2 \times d_2}$. Indeed, by Lyapunov's stability theorem for linear systems (Hespanha, 2018, Theorem 8.2), a matrix A is Hurwitz stable if and only if for every symmetric positive definite $Q = Q^\top \succ 0$, there exists a unique symmetric positive definite $P = P^\top \succ 0$, such that $A^\top P + PA = -Q$. Hence, $-J$ is Hurwitz stable if and only if there exists a $P = P^\top \succ 0$ such that

$$\begin{aligned} 0 \prec Q &= \begin{bmatrix} 0 & -B \\ B^\top & C \end{bmatrix} \begin{bmatrix} P_1 & P_2 \\ P_2^\top & P_3 \end{bmatrix} + \begin{bmatrix} P_1 & P_2 \\ P_2^\top & P_3 \end{bmatrix} \begin{bmatrix} 0 & B \\ -B^\top & C \end{bmatrix} \\ &= \begin{bmatrix} -BP_2^\top - P_2 B^\top & -BP_3 + P_1 B + P_2 C \\ B^\top P_1 + CP_2^\top - P_3 B^\top & B^\top P_2 + CP_3 + P_2^\top B + P_3 C \end{bmatrix} \end{aligned}$$

Since this is a symmetric positive definite matrix, the block diagonal components must also be symmetric positive definite so that $-BP_2 - P_2 B^\top \succ 0$.[5] Recall that $B \in \mathbb{R}^{d_1 \times d_2}$ and $P_2 \in \mathbb{R}^{d_2 \times d_1}$. Hence, a necessary condition for this matrix to be positive definite is that $d_2 \geq d_1/2$ for $-BP_2 - P_2 B^\top$ to have full rank; of course this is not sufficient, but it is necessary. It is easy to see this argument is independent of whether a learning rate ratio $\tau \neq 0$ or regularization is incorporated.

[5] If a block matrix Q with block entries Q_{ij} for $i, j \in \{1, 2\}$ is positive definite symmetric, then $Q_{ii} \succ 0$ for $i = 1, 2$.

3.5 Convergence of GDA with Timescale Separation

For the full proofs of the results in this section, see Section 3.E in the appendix that follows this chapter. Given the stability and instability characterizations from the previous section with regard to the continuous-time system, analogous statements can be made about the discrete-time system with a proper choice of learning rate. Specifically, the stability of the discrete-time system around a critical point x^* requires selecting γ_1 such $\rho(I - \gamma_1 J_\tau(x^*)) < 1$. This is only possible when $J_\tau(x^*) \subset \mathbb{C}_-^\circ$, which leads to the following result that is a corollary of Theorem 3.3. That is, τ-GDA converges locally asymptotically for any sufficiently small $\gamma(\tau)$ and for all $\tau \in (\tau^*, \infty)$ if and only if x^* is a differential Stackelberg equilibrium.

Corollary 3.2. *Suppose the assumptions of Theorem 3.3 hold. Then, there exists a $\tau^* \in (0, \infty)$ such that τ-GDA with $\gamma_1 \in (0, \gamma(\tau))$ where $\gamma(\tau) = \arg\min_{\lambda \in \mathrm{spec}(J_\tau(x^*))} 2\mathrm{Re}(\lambda)/|\lambda|^2$ converges locally asymptotically for all $\tau \in (\tau^*, \infty)$ if and only if x^* is a differential Stackelberg equilibrium.*

Observe that this corollary can also be seen as an analogue for τ-GDA as the result provided in Proposition 2.9 of Chapter 2 for the Stackelberg gradient dynamics. Moreover, it will become shortly how the suitable learning rate in this corollary comes about. We remark that the avoidance of strict saddles is also known for τ-GDA and the implications of this will be discussed further at the end of this section.

For the remainder of this section, we characterize the asymptotic convergence rate for τ-GDA to differential Stackelberg equilibria, and provide a finite time guarantee for convergence to an ε–approximate equilibrium. The proof techniques generally mirror that which were described and developed for the Stackelberg gradient dynamics in Section 2.4.2 of Chapter 2 when combined with the stability characterizations of the previous section. However, a key technical challenge is optimizing the learning rate for the convergence rate since the Jacobian does not have a clean structure like it does in zero-sum games for the Stackelberg gradient dynamics from the previous chapter in zero-sum games. This is the analysis that we highlight below. Specifically, the asymptotic convergence rate result uses Theorem 3.3 to construct a finite $\tau^* \in (0, \infty)$ such that x^* is stable, meaning $\mathrm{spec}(-J_\tau(x^*)) \subset \mathbb{C}_-^\circ$, and then for any $\tau \in (\tau^*, \infty)$, a learning rate is chosen to ensure stability of the discrete time system and numerical analysis tools applied to get a local asymptotic convergence rate.

Theorem 3.6. *Consider a zero-sum game $(f_1, f_2) = (f, -f)$ defined by $f \in C^q(\mathcal{X}, \mathbb{R})$ for $q \geq 2$ and let x^* be a differential Stackelberg equilibrium of the game. There exists a $\tau^* \in (0, \infty)$ such that for any $\tau \in (\tau^*, \infty)$ and $\alpha \in (0, \gamma)$, τ-GDA with learning rate $\gamma_1 = \gamma - \alpha$ converges locally asymptotically at a rate of $\mathcal{O}((1 - \alpha/(4\beta))^{k/2})$ where $\gamma = \min_{\lambda \in \mathrm{spec}(J_\tau(x^*))} 2\mathrm{Re}(\lambda)/|\lambda|^2$, $\lambda_{\mathtt{m}} = \arg\min_{\lambda \in \mathrm{spec}(J_\tau(x^*))} 2\mathrm{Re}(\lambda)/|\lambda|^2$, and $\beta = (2\mathrm{Re}(\lambda_{\mathtt{m}}) - \alpha|\lambda_{\mathtt{m}}|^2)^{-1}$. Moreover, if x^* is a differential Nash equilibrium, $\tau^* = 0$ so that for any $\tau \in (0, \infty)$ and $\alpha \in (0, \gamma)$, τ-GDA with $\gamma_1 = \gamma - \alpha$ converges with a rate $\mathcal{O}((1 - \alpha/(4\beta))^{k/2})$.*

To build some intuition, consider a differential Stackelberg equilibrium x^* and its corresponding τ^* obtained via Theorem 3.3 so that for any fixed $\tau \in (\tau^*, \infty)$, $\mathrm{spec}(-J_\tau(x^*)) \subset \mathbb{C}_-^\circ$. For the discrete time system $x_{k+1} = x_k - \gamma_1 \Lambda_\tau g(x_k)$, if γ_1 is chosen such that the spectral radius of the local linearization of the discrete time map is a contraction, then x_k locally (exponentially) converges

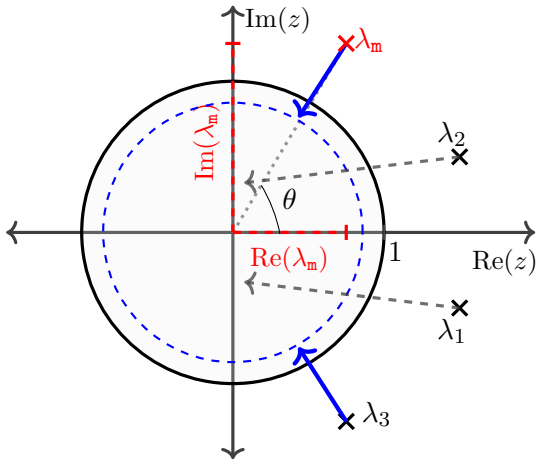


Figure 3.3: The inner maximization problem in (3.9) is over a finite set $\mathrm{spec}(J_\tau(x^*)) = \{\lambda_1, \ldots, \lambda_d\}$ where $J_\tau(x^*) \in \mathbb{R}^{d\times d}$. As $\gamma \to \infty$, $|1 - \gamma\lambda_i| \to 0$. The last λ_i such that $1 - \gamma\lambda_i$ hits the boundary of the unit circle in the complex plane—that is, $|1 - \gamma\lambda_i| = 1$—gives us the optimal value of $\gamma = 2\mathrm{Re}(\lambda_{\mathtt{m}})/|\lambda_{\mathtt{m}}|^2 = 2\cos(\theta)/|\lambda_{\mathtt{m}}|$ and the element of $\mathrm{spec}(J_\tau(x^*))$ that achieves it (see blue arrows).

to x^*. With this in mind, we formulate an optimization problem to find the upper bound γ on the learning rate γ_1 such that for all $\gamma_1 \in (0, \gamma)$, $\rho(I - \gamma_1 J_\tau(x^*)) < 1$; indeed, let

$$\gamma = \min_{\gamma>0} \left\{ \gamma : \max_{\lambda \in \mathrm{spec}(J_\tau(x^*))} |1 - \gamma\lambda| \leq 1 \right\}. \tag{3.9}$$

The intuition is as follows. The inner maximization problem is over a finite set $\mathrm{spec}(J_\tau(x^*)) = \{\lambda_1, \ldots, \lambda_d\}$ where $J_\tau(x^*) \in \mathbb{R}^{d\times d}$. As γ increases away from zero, each $|1 - \gamma\lambda_i|$ shrinks in magnitude. The last λ_i such that $1 - \gamma\lambda_i$ hits the boundary of the unit circle in the complex plane gives us the optimal γ and the $\lambda_{\mathtt{m}} \in \mathrm{spec}(J_\tau(x^*))$ that achieves it. Examining the constraint, we have that for each λ_i, $\gamma(\gamma|\lambda_i|^2 - 2\mathrm{Re}(\lambda_i)) \leq 0$ for any $\gamma > 0$. As noted this constraint will be tight for one of the λ, in which case $\gamma = 2\mathrm{Re}(\lambda)/|\lambda|^2$ since $\gamma > 0$. Hence, by selecting $\gamma = \min_{\lambda\in\mathrm{spec}(J_\tau(x^*))} 2\mathrm{Re}(\lambda)/|\lambda|^2$, we have that $|1-\gamma_1\lambda| < 1$ for all $\lambda \in \mathrm{spec}(J_\tau(x^*))$ and any $\gamma_1 \in (0, \gamma)$. From here, one can use standard arguments from numerical analysis to show that for the choice of α and β, the claimed asymptotic rate holds.

Theorem 3.6 directly implies a finite time convergence guarantee for obtaining an ε-differential Stackelberg equilibrium, that is, a point with an ε-ball around a differential Stackelberg equilibrium x^*.

Corollary 3.3. *Given $\varepsilon > 0$, under the assumptions of Theorem 3.6, τ-GDA obtains an ε–differential Stackleberg equilibrium in $\lceil (4\beta/\alpha)\log(\|x_0 - x^*\|/\varepsilon)\rceil$ iterations for any $x_0 \in B_\delta(x^*)$ with $\delta = \alpha/(4L\beta)$ where L is the local Lipschitz constant of $I - \gamma J_\tau(x^*)$.*

Moreover, the convergence rates and finite time guarantees extend to the gradient penalty regularized generative adversarial network described in the preceeding section.

Corollary 3.4. *Under the assumptions of Theorems 3.5 and 3.6, for any fixed $\mu \in (0, \infty)$ and*

$\tau \in (0, \infty)$, τ-GDA converges locally asymptotically at a rate of $O((1-\alpha/(4\beta))^{k/2})$, and achieves an ε-equilibrium in $\lceil (4\beta/\alpha)\log(\|x_0 - x^*\|/\varepsilon)\rceil$ iterations for any $x_0 \in B_\delta(x^*)$.

Comments on computing the neighborhood $B_\delta(x^*)$. We note that we have essentially given a proof that there exists a neighborhood on which τ-GDA converges. Of course, due to the non-convexity of the problem in general, this neighborhood could be arbitrarily small. We provide an estimate of the neighborhood size using the local Lipschitz constant of the local linearization $I - \gamma_1 J_\tau(x^*)$. One way to better understand the size of this neighborhood is to use Lyapunov analysis, a tool which is well explored in the singular perturbation theory (Kokotovic et al., 1986). In particular, Lyapunov methods can be applied directly to the nonlinear system if one can construct Lyapunov functions for the fast and slow subsystems individually—also known as the boundary layer model and reduced order model. With these Lyapunov functions in hand, one can "stitch" the two together (via convex combination) and show under some reasonable assumptions that this combined function is a Lyapunov function for the overall singularly perturbed system. The benefit of this analysis is that the Lyapunov function gives one an estimate of the region of attraction (via, e.g., the level sets); however, it is not easy to construct a Lyapunov function for a nonlinear system in general. We leave expanding such methods to learning in nonconvex-nonconcave zero-sum games to future work.

Comments on avoiding saddle points. Before turning to the stochastic setting, we comment on saddle point avoidance in the deterministic setting. It was shown by Mazumdar et al. (2020) that gradient-based learning in continuous games with heterogeneous learning rates avoids saddles on all but a set of measure zero initializations. Hence, τ-GDA avoids saddles for almost every initialization. We also know that all differential Nash equilibria are locally asymptotically stable for zero-sum settings. Hence, there are no differential Nash equilibria that are saddle points of the dynamics $\dot{x} = -\Lambda_\tau g(x)$. On the other hand, as Example 3.1 shows, there are differential Stackelberg equilibria which correspond to saddle points of the dynamics for some choices of τ—in particular, $\tau = 1$ in that example. Theorem 3.3 and Corollary 3.1, however, implies that for a given zero-sum game (or minmax problem), there exists a finite τ^* such that all locally asymptotically stable equilibria are differential Stackelberg equilibria. Hence, an 'almost sure' saddle point avoidance result together with the local convergence guarantee provided by Theorem 3.6 provides a strong characterization of long-run learning behavior. Avoidance of saddles nor the if and only if convergence guarantee of Theorem 3.3 are, however, enough to ensure avoidance of limit cycles. In fact, it is known that limit cycles can exist in zero sum games (Daskalakis et al., 2018; Mazumdar et al., 2020). Understanding when such complex phenomena exist in games and determining how to ascribe meaning the behavior is an active area of study (see, e.g., the work of Papadimitriou and Piliouras (2019)).

3.6 Convergence of Stochastic GDA with Timescale Separation

In this section, we analyze convergence when players do not have oracle access to their gradients but instead have an unbiased estimator in the presence of zero mean, finite variance noise. Specifically, we show that the agents will converge locally asymptotically almost surely to a differential Stackelberg equilibrium.

The key insight in this section is that due to Theorem 3.3, we know that a critical point x^* is stable for $\dot{x} = -\Lambda_\tau g(x)$ for a range of finite learning rates $\tau \in (\tau^*, \infty)$ if and only if x^* is a

differential Stackelberg equilibrium. Hence, treating $\dot{x} = -\Lambda_\tau g(x)$ as the continuous time limiting differential equation in the so-called ordinary differential equation (ODE) method in stochastic approximation (Borkar, 2008), we apply classical stochastic approximation analysis to conclude that the stochastic gradient descent-ascent update with timescale separation converges.

3.6.1 Asymptotic Convergence Guarantees via Stochastic Approximation

The stochastic form of the update is given by

$$x_{k+1} = x_k - \gamma_k(\Lambda_\tau g(x_k) + w_{k+1}) \tag{3.10}$$

where w_{k+1} is a zero mean, finite variance random variable and $\{\gamma_k\}$ is the learning rate sequence.

Assumption 3.2. *The stochastic process $\{w_k\}$ is a martingale difference sequence with respect to the increasing family of σ-fields defined by*

$$\mathcal{F}_k = \sigma(x_\ell, w_\ell, \ell \leq k),\ \forall k \geq 0,$$

so that $\mathbb{E}[w_{k+1}|\ \mathcal{F}_k] = 0$ almost surely (a.s.) for all $k \geq 0$. Moreover, w_k is square-integrable so that, for some constant $C > 0$,

$$\mathbb{E}[\|w_{k+1}\|^2|\ \mathcal{F}_k] \leq C(1 + \|x_k\|^2)\ \text{a.s.},\ \forall k \geq 0.$$

We note that this assumption has been relaxed in the literature (cf. Thoppe and Borkar (2019)), however simplicity, we state the theorem with the most accessible criteria. We remark below in the paragraph on extensions to concentration bounds on the nature of the relaxed assumptions.

Theorem 3.7. *Consider a zero-sum game $(f, -f)$ such that $f \in C^q(\mathcal{X}, \mathbb{R})$ for some $q \geq 2$. Suppose that Assumption 3.2 holds and that $\{\gamma_k\}$ is square summable but not summable—i.e., $\sum_k \gamma_k^2 < \infty$, yet $\sum_k \gamma_k = \infty$. For any $\tau \in (0, \infty)$, the sequence $\{x_k\}$ generated by (3.10) converges to a, possibly sample path dependent, internally chain transitive invariant set of $\dot{x} = -\Lambda_\tau g(x)$. Moreover, if x^* is a differential Stackelberg equilibrium, then there exists a finite $\tau^* \in (0, \infty)$ such that $\{x_k\}$ almost surely converges locally asymptotically to x^* for every $\tau \in (\tau^*, \infty)$.*

Proof. The convergence of $\{x_k\}$ to a, possibly sample path dependent, compact connected internally chain transitive invariant set of $\dot{x} = -\Lambda_\tau g(x)$ follows from classical results in stochastic approximation theory (Borkar, 2008, Chap. 2); (Benaim, 1996).

Suppose that x^* is a differential Stackelberg equilibrium. By Theorem 3.3, there exists a finite $\tau^* \in (0, \infty)$ such that for all $\tau \in (\tau^*, \infty)$, x^* is a locally exponentially stable equilibrium of the continuous time dynamics $\dot{x} = -\Lambda_\tau g(x)$—that is, $\text{spec}(-J_\tau(x^*)) \subset \mathbb{C}_-^\circ$ for all $\tau \in (\tau^*, \infty)$.

Fix arbitrary $\tau \in (\tau^*, \infty)$. Since $\text{spec}(-J_\tau(x^*)) \subset \mathbb{C}_-^\circ$, $\det(-J_\tau(x^*)) \neq 0$ so that x^* is an isolated critical point. Furthermore, exponentially stability of x^* implies that there exists a (local) Lyapunov function defined on a neighborhood of x^* by the converse Lyapunov theorem (cf. Sastry (1999, Thm. 5.17) or Krasovskii (1963, Thm. 4.3)). Let U be the neighborhood of x^* on which the

local Lyapunov function is defined, such that U contains no other critical points (which is possible since x^* is isolated). That is, let $\Phi : U \to [0, \infty)$ be the local Lyapunov function defined on U where $x^* \in U$, Φ is positive definite on U, and for all $x \in U$, $\frac{d}{dt}\Phi(x) \leq 0$ where equality holds for $z \in U$ if and only if $\Phi(z) = 0$. By Corollary 3 (Borkar, 2008, Chap. 2), $\{x_k\}$ converges to an internally chain transitive invariant set contained in U almost surely. The only internally chain transitive invariant set in U is x^*. □

The following corollary shows that if there is a finite τ^* such that x^* is stable for $\dot{x} = -\Lambda_\tau g(x)$, then by Theorem 3.3 x^* must be a differential Stackelberg equilibrium and in turn, $\{x_k\}$ almost surely converges locally asymptotically to x^* by the above theorem.

Corollary 3.5. *Consider a zero-sum game $(f, -f)$ such that $f \in C^2(\mathcal{X}, \mathbb{R})$. Suppose that Assumption 3.2 holds and that $\{\gamma_k\}$ is square summable but not summable: $\sum_k \gamma_k^2 < \infty$, yet $\sum_k \gamma_k = \infty$. If there exists a finite $\tau^* \in (0, \infty)$ such that $\mathrm{spec}(-J_\tau(x^*)) \subset \mathbb{C}_-^\circ$ for all $\tau \in (\tau^*, \infty)$, then x^* is a differential Stackelberg equilibrium and $\{x_k\}$ almost surely converges locally asymptotically to x^*.*

While (local) almost sure convergence in gradient descent-ascent (Chasnov et al., 2020; Heusel et al., 2017) to a critical point[6] in the stochastic setting, the result requires time varying learning rates with a sufficient separation in timescale. Specifically, the players need to be using learning rate sequences $\{\gamma_{i,k}\}$ for each $i \in \{1, 2\}$ such that (without loss of generality) not only is it assumed that $\gamma_{1,k} = o(\gamma_{2,k})$, but also $\sum_k \gamma_{1,k}^2 + \gamma_{2,k}^2 < \infty$ and $\sum_k \gamma_{i,k} = \infty$ for each $i \in \{1, 2\}$. The challenge with these assumptions on the learning rate sequences is that empirically the sequences that satisfy them result in poor behavior along the learning path such as getting stuck at saddle points or making no progress. This is, in essence, due to the fact that the faster player—i.e., player 2 if $\gamma_{1,k} = o(\gamma_{2,k})$—equilibriates too quickly causing progress to stall. This can result in undesirable behavior such as vanishing gradients (so that the discriminator does not provide enough information for the generator to make progress), mode collapse, or failure to converge in practical applications such as generative adversarial networks.

On the other hand, our convergence result gives a similar guarantee with less restrictive requirements on the stepsize sequence. In particular, only a single stepsize sequence is required (so that the algorithm can be viewed as a single timescale stochastic approximation update) as long as the fast player (who, without loss of generality, is taken to be player 2) scales their estimated gradient by $\tau \in (\tau^*, \infty)$ where τ^* is as in Theorem 3.3.

3.6.2 Extensions to concentration bounds and relaxed assumptions on stepsizes

It is possible to obtain concentration bounds and even finite time, high probability guarantees on convergence leveraging recent advances in stochastic approximation (Borkar, 2008; Kamal, 2010; Thoppe and Borkar, 2019). To our knowledge, the concentration bounds in (Thoppe and Borkar, 2019) require the weakest assumptions on learning rates—e.g., the stepsize sequence $\{\gamma_k\}$ needs only to satisfy $\sum_k \gamma_k = \infty$, $\lim_{k\to\infty} \gamma_k = 0$, and $\sum_k \gamma_k \leq 1$. Specifically, since it is assumed, for the zero sum game $(f, -f)$, that $f \in C^2(\mathcal{X}, \mathbb{R})$ and x^* is a differential Stackelberg equilibrium, Theorem 3.3 implies that x^* is a locally asymptotically stable attractor of $\dot{x} = -\Lambda_\tau g(x)$ for arbitrary

[6]To date it has not been shown that for a sufficient separation in timescale the only critical point attractors are local minmax.

fixed $\tau \in (\tau^*, \infty)$, and hence, the concentration bounds in Theorem 1.1 and 1.2 of (Thoppe and Borkar, 2019) directly apply.

Furthermore, we note that in applications such as generative adversarial networks, while it has been observed that timescale separation heuristics such as unrolling or annealing the stepsize of the discriminator work well, in the stochastic case, summmable/square-summable assumptions on stepsizes are generally too restrictive in practice since they lead to a rapid decay in the stepsize which, in turn, can stall progress. On the other hand, stepsize sequences such as $\gamma_k = 1/(k+1)^\beta$ for $\beta \in (0, 1]$—a sequence which satisfies the assumptions posed in (Thoppe and Borkar, 2019)—tend not to have this issue of decaying too rapidly for appropriately chosen β, while also maintaining the guarantees of the theoretical results. We state a convergence guarantee under these relaxed assumptions in Proposition 3.4 below.

Let $\tilde{x}(t)$ be the asymptotic pseudo-trajectories of the stochastic approximation process $\{x_k\}$. That is, $\tilde{x}(t)$ are linear interpolates between the sample points x_k generated by the stochastic τ-`GDA` process, and are defined by

$$\tilde{x}(t) = \tilde{x}(t_k) + \frac{(t - t_k)}{\gamma_k}(\tilde{x}(t_{k+1}) - \tilde{x}(t_k))$$

where $t_k = t_k + \gamma_k$ and $t_0 = 0$.

Assumption 3.3. *The stochastic process $\{w_k\}$ is a martingale difference sequence with respect to the increasing family of σ-fields defined by*

$$\mathcal{F}_k = \sigma(x_\ell, w_\ell, \ell \leq k),\ \forall k \geq 0,$$

so that $\mathbb{E}[w_{k+1}|\ \mathcal{F}_k] = 0$ almost surely for all $k \geq 0$. Furthermore, there exists $c_1, c_2 \in C(\mathbb{R}^d, \mathbb{R}_{>0})$ such that

$$\Pr\{\|w_{k+1}\| > v|\ \mathcal{F}_k\} \leq c_1(x_k)\exp(-c_2(x_k)v),\ n \geq 0$$

for all $v \geq \tilde{v}$ where $\tilde{v}$ is some sufficiently large, fixed number.

Proposition 3.4. *Suppose that Assumption 3.3 holds and that x^* is a differential Stackelberg equilibrium. Let $\gamma_k = 1/(k+1)^\beta$ where $\beta \in (0, 1]$. There exists a $\tau^* \in (0, \infty)$ and an $\epsilon_0 \in (0, \infty)$ such that for any fixed $\epsilon \in (0, \epsilon_0]$, there exists functions $h_1(\epsilon) = O(\log(1/\epsilon))$ and $h_2(\epsilon) = O(1/\epsilon)$ so that when $T \geq h_1(\epsilon)$ and $k_0 \geq K_\tau$ where K_τ is such that $1/\gamma_k \geq h_2(\epsilon)$ for all $k \geq K_\tau$, the stochastic iterates of τ-`GDA` with stepsize sequence γ_k and timescale separation $\tau \in (\tau^*, \infty)$ satisfy*

$$\Pr\{\|\tilde{x}(t) - x^*\| \leq \epsilon\ \forall t \geq t_{k_0} + T + 1|\ \tilde{x}(t_{k_0}) \in B_\epsilon(x^*)\} = 1 - O(k_0^{1-\beta/2}\exp(-C_\tau k_0^{\beta/2}))$$

for some constant $C_\tau > 0$.

The proof largely follows from the proofs of Theorem 1.1 and 1.2 in (Thoppe and Borkar, 2019), combined with the existence of a finite timescale separation parameter obtained via Theorem 3.3. Indeed, since x^* is a differential Stackelberg equilibrium, by Theorem 3.3 there exists a range of τ—namely, (τ^*, ∞)—such that for any $\tau \in (\tau^*, \infty)$, x^* is a locally asymptotically stable equillibrium for $\dot{x} = -\Lambda_\tau g(x)$. Hence, fixing any $\tau \in (\tau^*, \infty)$, a converse Lyapunov theorem can be applied to construct a local Lyapunov function. Let $V : \mathbb{R}^n \to \mathbb{R}$ be this Lyapunov function so that there

exists $r, r_0, \epsilon_0 > 0$ such that $r > r_0$, and

$$B_\epsilon(x^*) \subseteq V^{r_0} \subset \mathcal{N}_{\epsilon_0}(V^{r_0}) \subseteq V^r$$

for any $\epsilon \in (0, \epsilon_0]$ where, for a given $q > 0$, $V^q = \{x \in \text{dom}(V) : \ V(x) \leq q\}$ and $\mathcal{N}_{\epsilon_0}(V^{r_0})$ is an ϵ_0–neighborhood of V^{r_0}—i.e., $\mathcal{N}_{\epsilon_0}(V^{r_0}) = \{x \in \mathbb{R}^n | \ \exists y \in V^{r_0}, \ \|x - y\| \leq \epsilon_0\}$. From here, the result follows from an application of the results in the work by Thoppe and Borkar (2019).

The utility of this result is that it provides a guarantee in the stochastic setting for a more reasonable and practically useful stepsize sequence. However, constructing the constants such as K_τ, C_τ and ϵ_0 is highly non-trivial as can be seen in the work of Thoppe and Borkar (2019) and similar works in the area of stochastic approximation (Borkar, 2008). One direction of future work is examining the Lyapunov approach for directly analyzing the nonlinear singularly perturbed system; it is known, however, that the stochastic singularly perturbed systems have much weaker guarantees in terms of stability (Kokotovic et al., 1986, Chap. 4).

3.7 Experiments

In this section we present extensive experimental results. We numerically investigate Example 3.1 in Section 3.7.1 and a game similar to that from Example 3.2 in Section 3.7.2. After that, we investigate a polynomial game with multiple equilibria in Section 3.7.3. We study a torus game in Section 3.7.4 and examine the connection between timescale separation and the region of attraction. Then, in Sections 3.7.5 and 3.7.6, consider the Dirac-GAN game (Mescheder et al., 2018) and consider both the saturating and non-saturating objective functions. In Section 3.7.7, we explore a generative adversarial network formulation using the Wasserstein cost function with a linear generator and quadratic discriminator for the problem of learning a covariance matrix. We finish in Section 3.7.8 by presenting experiments training generative adversarial networks parameterized by neural networks on a mixture of Gaussians and image datasets.

3.7.1 Quadratic Game: Timescale Separation and Stackelberg Stability

We now revisit the game from Example 3.1 that demonstrated there exists differential Stackelberg equilibrium that are unstable for choices of the timescale separation τ. To be clear, we repeat the game construction and some characteristics of the game. Let us consider the quadratic zero-sum game defined by the cost

$$f(x_1, x_2) = \frac{1}{2} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^\top \begin{bmatrix} -v & 0 & -v & 0 \\ 0 & \frac{1}{2}v & 0 & \frac{1}{2}v \\ -v & 0 & -\frac{1}{2}v & 0 \\ 0 & \frac{1}{2}v & 0 & -v \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{3.11}$$

where $x_1, x_2 \in \mathbb{R}^2$ and $v > 0$. The unique critical point of the game given by $x^* = (x_1^*, x_2^*) = (0, 0)$ is a differential Stackelberg equilibrium. The spectrum of the Jacobian evaluated at the equilibrium

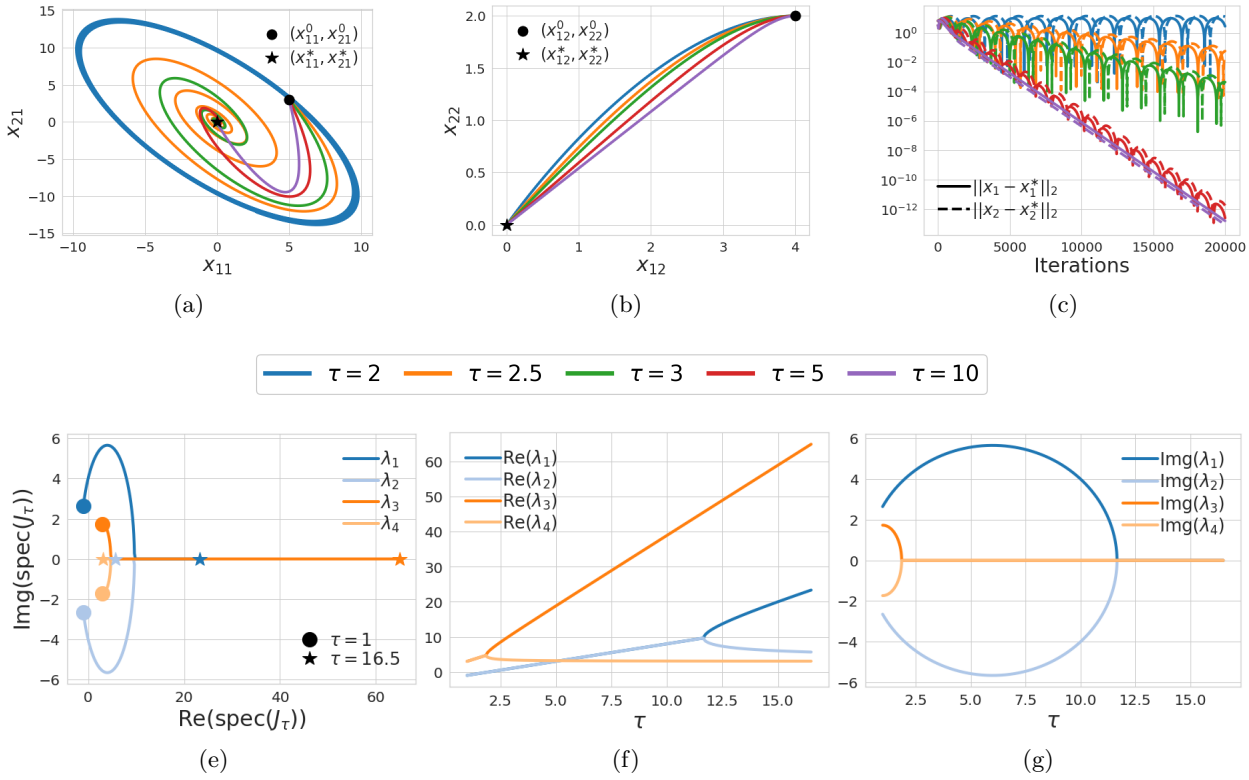


Figure 3.4: Experimental results for the quadratic game defined in (3.11) of Section 3.7.1 and presented in Example 3.1. Figures 3.4a and 3.4b show trajectories of the players coordinate pairs (x_{11}, x_{21}) and (x_{21}, x_{22}) for a range of learning rate ratios, respectively. Figures 3.4c shows the distance from the equilibrium along the learning paths. Figures 3.4e, 3.4f, and 3.4g show the trajectories of the eigenvalues, the real parts of the eigenvalues, and the imaginary parts of the eigenvalues for the $J_\tau(x^*)$ as a function of the τ, respectively.

is given by

$$\text{spec}(J_\tau(x^*)) = \left\{ \frac{v(2\tau + 1 \pm \sqrt{4\tau^2 - 8\tau + 1})}{4}, \frac{v(\tau - 2 \pm \sqrt{\tau^2 - 12\tau + 4})}{4} \right\}.$$

As mentioned in Example 3.1, it turns out that $\text{spec}(J_\tau(x^*) \subset \mathbb{C}_+^\circ$ only when $\tau \in (2, \infty)$. We remark that we computed τ^* using the theoretical construction from Theorem 3.3 and found that it recovered the precise value of $\tau^* = 2$ such that the equilibrium is stable for all $\tau \in (\tau^*, \infty)$ with respect to the dynamics $\dot{x} = -\Lambda_\tau g(x)$. In the experiments that follow, we consistently observe that the construction of τ^* from the theory is tight.

For this experiment, we select $v = 4$ and simulate τ-GDA from the initial condition $(x_1^0, x_2^0) = (5, 4, 3, 2)$ with $\gamma_1 = 0.0005$ and $\tau \in \{2, 2.5, 3, 5, 10\}$. In Figures 3.4a and 3.4b, we show the trajectories of the players coordinate pairs (x_{11}, x_{21}) and (x_{21}, x_{22}), respectively. We observe that τ-GDA cycles around the equilibrium with $\tau = 2$ since it is marginally stable with respect to the dynamics. For $\tau \in (2, \infty)$, the equilibrium is stable and τ-GDA ends up converging to it at a rate

that depends on the choice of τ. We demonstrate how the convergence rate depends on the choice of τ in Figure 3.4c by showing the distance from the equilibrium along the learning path for each of the trajectories. The primary observation is that the cyclic behavior of τ-GDA dissipates as τ grows and as a result the dynamics then rapidly converge to the equilibrium.

The behavior of the learning dynamics as a function of the timescale separation τ can be further explained by evaluating the eigenvalues of the game Jacobian at the equilibrium. We show the eigenvalues of the Jacobian at the equilibrium in several forms in Figures 3.4e, 3.4f, and 3.4g. Analyzing the spectrum, we are able to verify that for all $\tau \in (2, \infty)$ the equilibrium is indeed stable. Moreover, we see that the imaginary parts of the conjugate pairs of eigenvalues decay after $\tau = 1$ and $\tau = 6$, and then the eigenvalues of the conjugate pairs eventually become purely real at $\tau = 1.87$ and $\tau = 11.66$, respectively. After the eigenvalues of a conjugate pair become purely real, they split so that one of the eigenvalues asymptotically converges to an eigenvalue of $\mathtt{S}_1(J(x^*))$ by moving back along the real line, while the other eigenvalue tends toward an eigenvalue of $-\tau D_2^2 f(x^*)$. This occurrence is exactly what was described in Section 3.3 as an immediate implication of Proposition 3.2 when the eigenvalues of $\mathtt{S}_1(J(x^*))$ and $\tau D_2^2 f(x^*)$ are distinct. The convergence rate is in fact limited by the eigenvalues splitting since as τ grows, the spectrum of the Jacobian is limited by the eigenvalues of the Schur complement which remain constant. A related open question centers on finding the worst case convergence rate as a function of the spectral properties of $\mathtt{S}_1(J(x^*))$ and $D_2^2 f(x^*)$. Finally, the evolution of the eigenvalues as a function of the timescale separation τ demonstrates that the rotational dynamics in τ-GDA vanish as the ratio between the magnitude of the real and imaginary parts of the eigenvalues grows.

3.7.2 Polynomial Game: Timescale Separation and Non-Equilibrium Stability

We now return to a game similar to that from Example 3.2 with a non-equilibrium critical point which is stable without timescale separation and becomes unstable for a range of finite learning ratios with multiple equilibria in the vicinity. Consider a zero-sum game defined by the cost

$$\begin{aligned} f(x_1, x_2) = \tfrac{5}{4}\left(x_{11}^2 + 2x_{11}x_{21} + \tfrac{1}{2}x_{21}^2 - \tfrac{1}{2}x_{12}^2 + 2x_{12}x_{22} - x_{22}^2\right)(x_{11} - 1)^2 \\ + x_{11}^2\big(\textstyle\sum_{i=1}^{2}(x_{1i} - 1)^2 - (x_{2i} - 1)^2\big). \end{aligned} \tag{3.12}$$

This game has critical points at $(0, 0, 0, 0)$, $(1, 1, 1, 1)$, and $(-4.73, 0.28, -92.47, 0.53)$. Among the critical points, only $(1, 1, 1, 1)$ and $(-4.73, 0.28, -92.47, 0.53)$ are game-theoretically meaningful equilibrium. In fact, they are each differential Nash equilibrium and are locally stable for any choice of $\tau \in (0, \infty)$ as a result of Proposition 3.1. On the other hand, the critical point $x^* = (0, 0, 0, 0)$ is neither a differential Nash equilibrium nor a differential Stackelberg equilibrium. However, x^* is stable for $\tau \in (0, 2)$ and it is marginally stable for $\tau = 2$. In general, convergence to the non-equilibrium critical point x^* in the presence of multiple game-theoretically meaningful equilibrium would be viewed as undesirable. In fact, this is precisely the type of critical point that sophisticated schemes for converging to only differential Nash equilibria or only differential Stackelberg equilibria seek to avoid (Adolphs et al., 2019; Fiez et al., 2020a; Mazumdar et al., 2019; Wang et al., 2020a). We show in this example that the simple inclusion of timescale separation in gradient descent-ascent is sufficient to avoid x^* and instead converge to a differential Nash equilibrium.

Indeed, for all $\tau \in (2, \infty)$ the non-equilibrium critical point x^* is unstable with respect to $\dot{x} = -\Lambda_\tau g(x)$. We simulate τ-GDA from the initial condition $(x_1^0, x_2^0) = (-1.5, 2.5,\ 2.5, 3)$ with

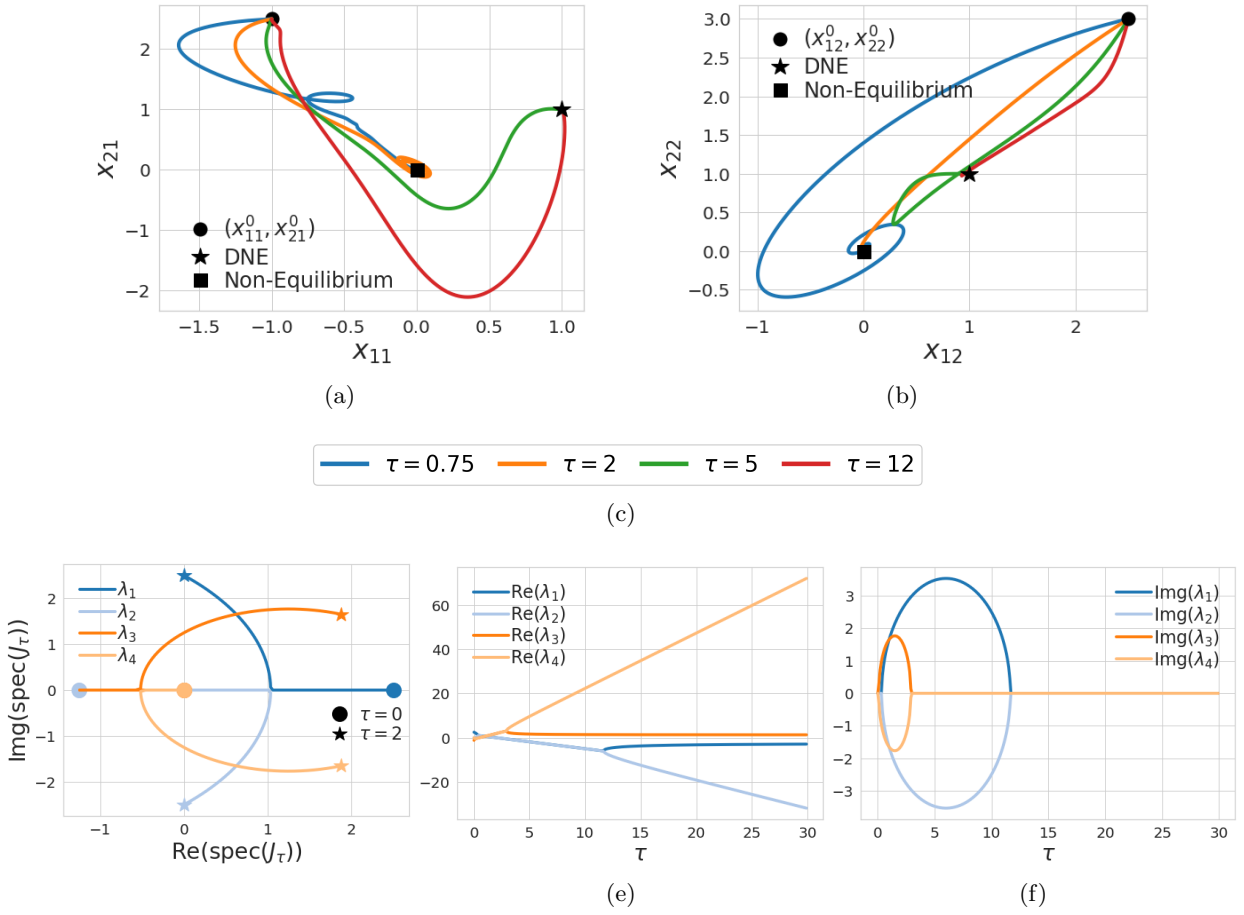
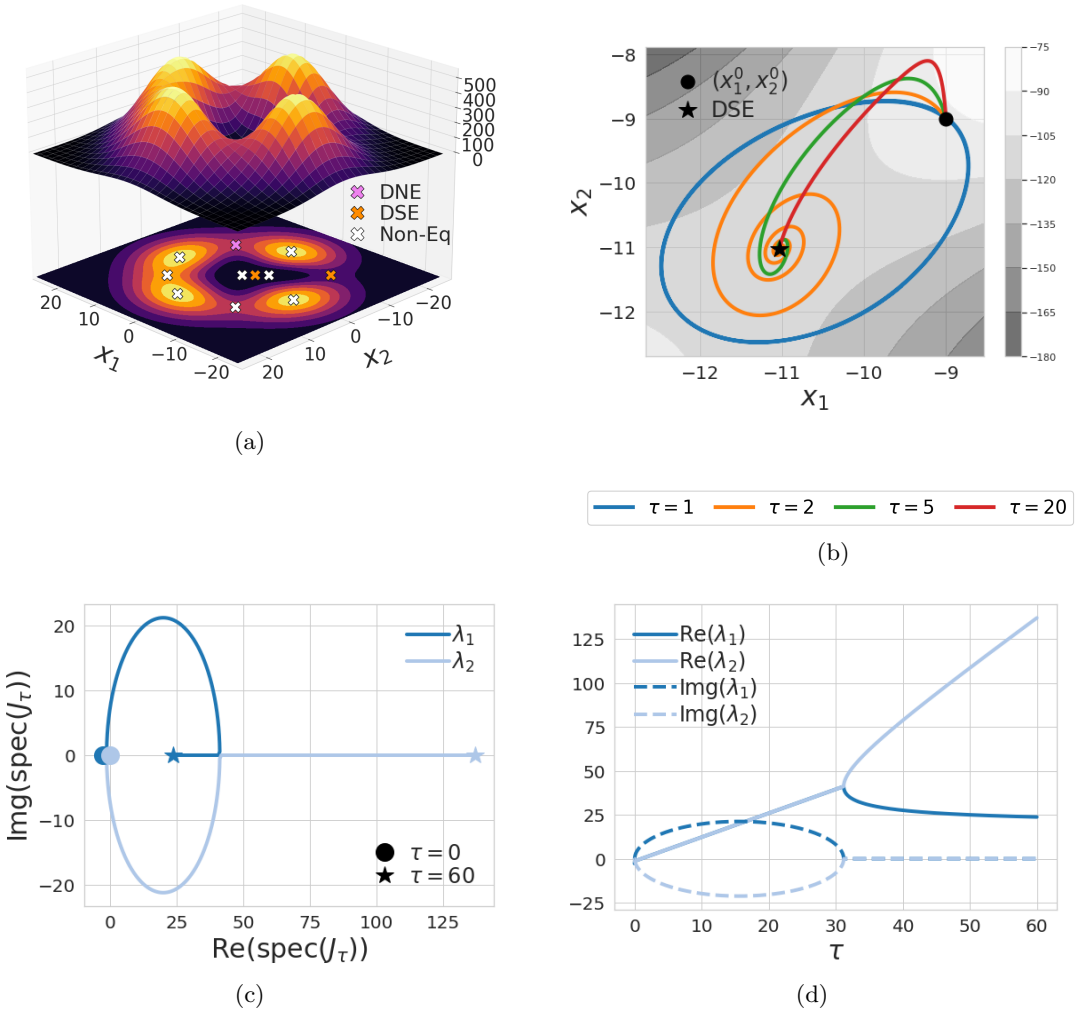


Figure 3.5: Experimental results for the polynomial game defined in (3.12) of Section 3.7.2 and presented in Example 3.2. Figures 3.5a and 3.5b show trajectories of the players coordinate pairs (x_{11}, x_{21}) and (x_{21}, x_{22}) for a range of learning rate ratios, respectively. Figures 3.5d, 3.5e, and 3.5f show the trajectories of the eigenvalues, the real parts of the eigenvalues, and the imaginary parts of the eigenvalues for $J_\tau(x^*)$ as a function of the τ, respectively where x^* is the non-equilibrium critical point.

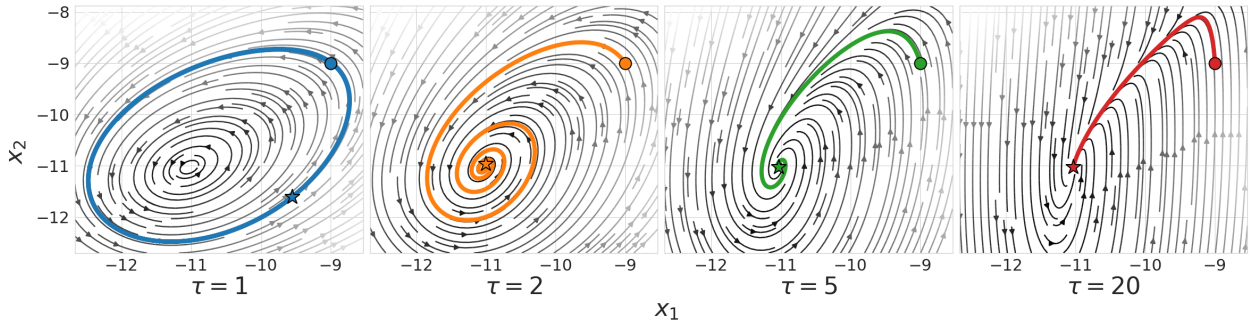
$\gamma_1 = 0.0005$ and $\tau \in \{0.75, 2, 5, 12\}$, where we use the superscript to denote the time index so as not to be confused with the multiple indexes for player choice variables. In Figures 3.5a and 3.5b, we show the trajectories of the players coordinate pairs (x_{11}, x_{21}) and (x_{21}, x_{22}), respectively. We observe that τ-GDA converges to the non-equilibrium critical point x^* with $\tau = 0.75$ as expected and the dynamics move near it and then cycle around it with $\tau = 2$ since the critical point becomes marginally stable. However, for $\tau = 5$ and $\tau = 12$, τ-GDA avoids the non-equilibrium critical point since it becomes unstable and instead the dynamics converge to the nearby differential Nash equilibrium. We show the eigenvalues of the Jacobian at the non-equilibrium critical point $x^* = (0, 0, 0, 0)$ in several forms in Figures 3.5d–3.5f. Again, we observe that the eigenvalues quickly become purely real as τ grows and then they split, and asymptotically converge toward the



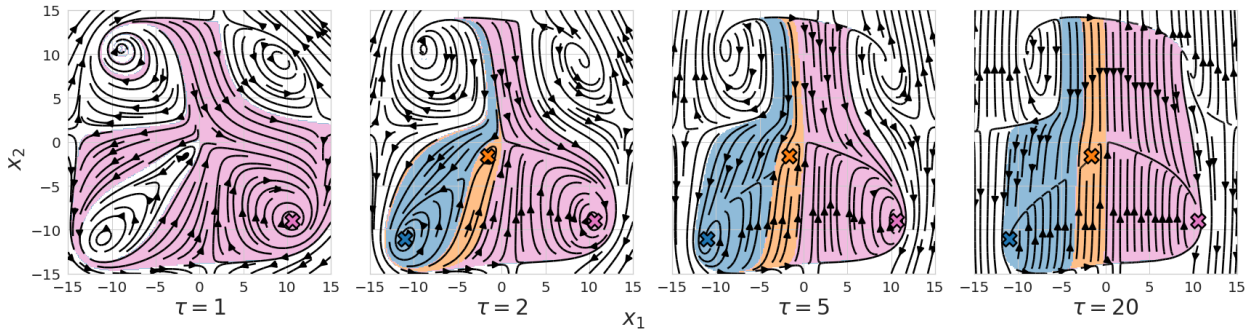
(a) (b) (c) (d)

Figure 3.6: Experimental results for the polynomial game defined in (3.13) of Section 3.7.3. Figures 3.6a provides a 3d view of the cost function $-f(x_1, x_2)$ along with the cost contours and critical point locations. Figure 3.6b shows trajectories of τ-GDA for a range of learning rate ratios given an initialization around the differential Stackelberg equilibrium $(x_1^*, x_2^*) = (-11.03, -11.03)$. Figures 3.6c and 3.6d show the evolution of the eigenvalues from $J_\tau(x^*)$ as a function of τ where x^* is the differential Stackelberg equilibrium $(x_1^*, x_2^*) = (-11.03, -11.03)$.

eigenvalues of $\mathtt{S}_1(J(x^*))$ and $-\tau D_2^2 f(x^*)$. Together, this example demonstrates that often there is a reasonable finite learning rate ratio such that non-meaningful critical points become unstable for τ-GDA.



(a)



(b)

Figure 3.7: Experimental results for the polynomial game defined in (3.13) of Section 3.7.3. In Figure 3.7a, we overlay the trajectories from Figure 3.6b produced by τ-GDA onto the vector field generated by the choice of timescale separation selection τ. The shading of the vector field is dictated by its magnitude so that lighter shading corresponds to a higher magnitude and darker shading corresponds to a lower magnitude. Figure 3.7b demonstrates the effect of timescale separation on the region of attractions around critical points by coloring points in the strategy space according to the equilibrium τ-GDA converges. We remark that areas without coloring indicate where τ-GDA did not converge in the time horizon.

3.7.3 Polynomial Game: Vector Field Warping and Region of Attraction

Consider a zero-sum game defined by the cost

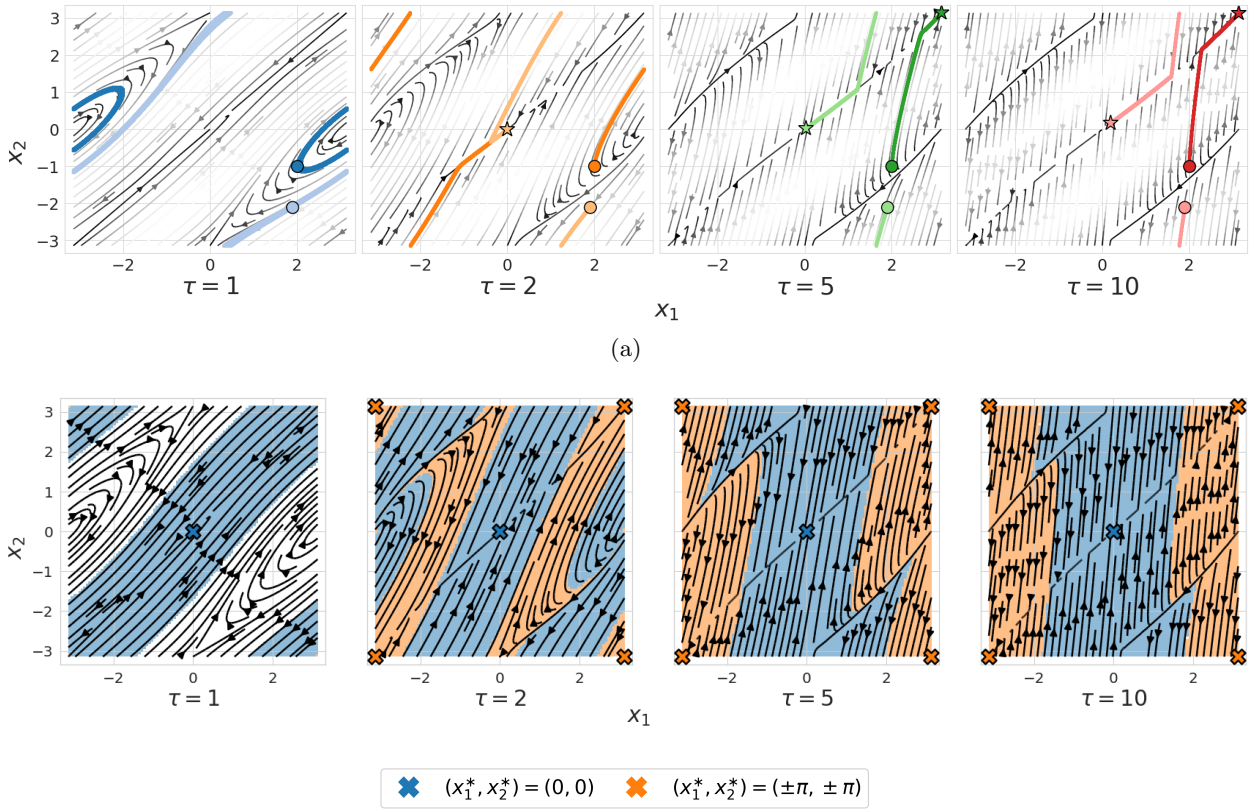
$$f(x_1, x_2) = -e^{-\left(0.01x_1^2 + 0.01x_2^2\right)} \left((0.3x_1 + x_2^2)^2 + (0.3x_2 + x_1^2)^2\right). \tag{3.13}$$

The cost structure of this game is visualized in Figure 3.6a, where we present a three dimensional view of $-f(x_1, x_2)$ along with the cost contours and the locations of critical points. This game has eleven critical points including one differential Nash equilibrium and two differential Stackelberg equilibria that are not a differential Nash equilibrium. The critical points that are neither a differential Nash equilibrium nor a differential Stackelberg equilibrium are unstable for any choice of timescale separation τ. The differential Nash equilibrium is at $(x_1, x_2) = (10.57, -8.95)$ and

it is stable for all $\tau \in (0, \infty)$ by Proposition 3.1. The differential Stackelberg equilibria are at $(x_1, x_2) = (-1.625, -1.625)$ and $(x_1^*, x_2^*) = (-11.03, -11.03)$; each is stable for all $\tau \in (1, \infty)$. We computed τ^* for the pair of differential Stackelberg equilibrium using the theoretical construction from Theorem 3.3 and observed that it properly recovered $\tau^* = 1$ for each equilibrium as the timescale separation such that the continuous time system is stable for all $\tau \in (\tau^*, \infty)$. Finally, we note that while the set of equilibrium follow a linear translation, this game is generic and the equilibria are in fact isolated.

In Figure 3.6b, we show the trajectories of τ-GDA with $\gamma_1 = 0.0001$ and $\tau \in \{1, 2, 5, 20\}$ given the initialization $(x_1^0, x_2^0) = (-9, -9)$ near the differential Stackelberg equilibrium at $(x_1^*, x_2^*) = (-11.03, -11.03)$. Moreover, in Figure 3.7a, we overlay the trajectories on the vector field generated by the respective timescale separation parameters. As expected, the choice of $\tau = 1$ results in a trajectory that cycles around the equilibrium in a closed curve since it is marginally stable and $J_\tau(x^*)$ has purely imaginary eigenvalues. Notably, as τ grows, the cyclic behavior dissipates as the timescale separation reshapes the vector field until the trajectory moves near directly to the zero derivative line of the maximizing player and then follows a path along that line toward the equilibrium and converges rapidly. The eigenvalues of $J_\tau(x^*)$ as a function of τ are presented in Figures 3.6c and 3.6d. As was the case for the previous experiments, we observe that after the eigenvalues become purely real as τ grows, they then split and asymptotically converge toward the eigenvalues of $\mathtt{S}_1(J(x^*))$ and $-\tau D_2^2 f(x^*)$. It is worth noting that much of the rotational behavior in the dynamics and vector field disappears as a result of timescale separation well before the eigenvalues become purely real; this seems to occur after the timescale separation is such that the magnitude of the real part of the eigenvalues is greater than that of the imaginary part.

Finally, in Figure 3.7b, we demonstrate how the choice of timescale separation τ not only warps the vector field but also shapes the regions of attraction around critical points. The vector field is again shown for each $\tau \in \{1, 2, 5, 20\}$, but now zoomed out to include each of the equilibria. The colors overlayed on the vector field indicate the equilibria that the dynamics converge to given an initialization at that position. Positions in the strategy space without color did not converge to an equilibrium in the fixed horizon of 75000 iterations with $\gamma_1 = 0.001$. This is explained by the fact that the dynamics are not guaranteed to be globally convergent and may get stuck in limit cycles or may simply move slowly for a long time in flat regions of the optimization landscape. We produced this experiment by running τ-GDA for a dense set of initial conditions chosen uniformly over the space of interest. It is clear from the experiment that the choice of timescale separation determines not only the stability of equilibria, but also has a fundamental impact on the equilibria the dynamics converge to from a given initial condition as a result of the warping of the vector field. As a concrete example, given an initialization of $(x_1, x_2) = (-10, -2)$, the dynamics with $\tau = 1$ converge to the differential Nash equilibria at $(x_1, x_2) = (10.57, -8.95)$. However, for any $\tau > 1$, the dynamics instead converge to the differential Stackelberg equilibrium at $(x_1, x_2) = (-11.03, -11.03)$ that is significantly closer to the initial condition. This example motivates future work on methods for obtaining accurate estimates of the regions of attraction around critical points and techniques to design τ in order to explicitly shape the region of attraction around an equilibrium of interest. We refer to the end of Section 3.5 for further discussion on potentially relevant analysis methods in this direction.



(a)

(b)

Figure 3.8: Experimental results for the torus game defined in (3.14) of Section 3.7.4. In Figure 3.8a, we overlay multiple trajectories produced by τ-GDA onto the vector field generated by the choice of timescale separation selection τ. The shading of the vector field is dictated by its magnitude so that lighter shading corresponds to a higher magnitude and darker shading corresponds to a lower magnitude. Figure 3.8b demonstrates the effect of timescale separation on the regions of attraction around critical points by coloring points in the strategy space according to the equilibrium τ-GDA converges. We remark that areas without coloring indicate where τ-GDA did not converge in the time horizon.

3.7.4 Location Game on the Torus

We use the example in this section to further study the role of timescale separation on the regions of attraction around critical points. Consider the zero-sum game defined by the cost

$$f(x_1, x_2) = -0.15\cos(x_1) + \cos(x_1 - x_2) + 0.15\cos(x_2). \tag{3.14}$$

This game can be interpreted as a location game on the torus. Specifically, the first player seeks to be far from the second player but near zero, while the second player seeks to be near the first player. This is a non-convex game on a non-convex strategy space. The critical points are given

by the set[7]:

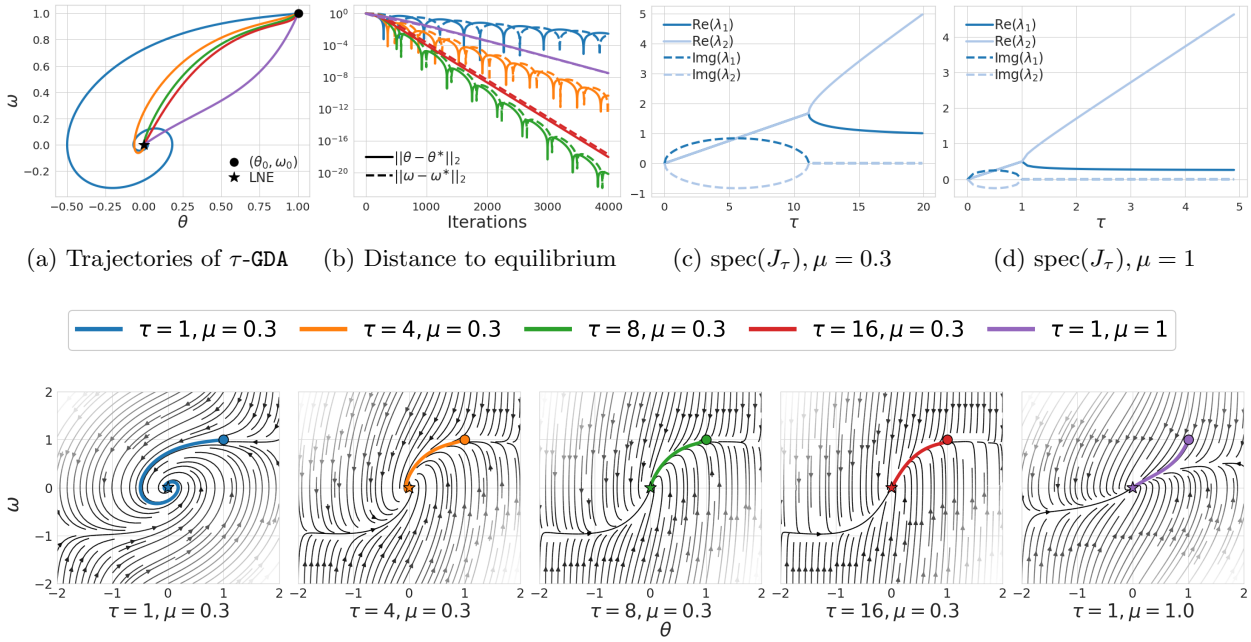
$$\{x:\ g(x)=0\}=\{(0,0),(\pi,\pi),(\pi,0),(0,\pi),(-1.646,-1.496),(1.646,1.496)\}.$$

The critical points $(0,0)$ and (π,π) are the only differential Stackelberg equilibrium and neither is a differential Nash equilibrium. The differential Stackelberg equilibrium at $(0,0)$ is stable for all $\tau\in(\tau^*,\infty)$ where $\tau^*=0.74$ and the differential Stackelberg equilibrium (π,π) is stable for all $\tau\in(\tau^*,\infty)$ where $\tau=1.35$. The rest of the critical points are unstable for any choice of τ. We remark that we computed τ^* for each differential Stackelberg equilibrium using the construction from Theorem 3.3 in Section 3.4 and it again gave the exact value of τ^* such that the system is stable for all $\tau>\tau^*$.

In Figure 3.8a, we show the trajectories of τ-GDA with $\gamma_1=0.001$ and $\tau\in\{1,2,5,10\}$ given the initializations $(x_1^0,x_2^0)=(2,-1)$ and $(x_1^0,x_2^0)=(1.9,-2.1)$ overlayed on the vector field generated by the respective timescale separation parameters. We observe that as the timescale separation τ grows, the rotational dynamics in the vector field dissipate and the directions of movement become sharp. As we mentioned in previous examples, τ-GDA moves directly to the zero line of $-D_2f(x_1,x_2)$ and then along that line to an equilibrium given sufficient timescale separation. The warping of the vector field that occurs as a result of timescale separation impacts the equilibrium that the dynamics converge to from a fixed initial condition and the neighborhood on which τ-GDA converges to an equilibrium. In other words, the *region of attraction* around critical points depends heavily on the timescale separation τ.

To illustrate this fact, in Figure 3.8b we show the regions of attraction for each choice of timescale separation. The vector fields are again shown for each $\tau\in\{1,2,5,10\}$, but now with colors overlayed indicating the equilibria that the dynamics converge to given an initialization at that position. This experiment was generated by running τ-GDA with a dense set of initial conditions chosen uniformly over the strategy space. Positions in the strategy space without color did not converge to an equilibrium in the fixed horizon of 20000 iterations with $\gamma_1=0.04$. This happens when τ-GDA is not initialized in the local neighborhood of attraction around a stable equilibrium. For the choice of $\tau=1$, $(0,0)$ is the only stable equilibrium. However, as demonstrated in Figure 3.8a, τ-GDA fails to converge to the equilibrium from the initial conditions $(x_1^0,x_2^0)=(2,-1)$ and $(x_1^0,x_2^0)=(1.9,-2.1)$. This behavior is further demonstrated over the strategy space in Figure 3.8b and highlights the local nature of the guarantees since convergence is only assured given an initialization in a suitable local neighborhood around a stable critical point. On the other hand, τ-GDA converges to an equilibrium from any initial condition for $\tau\in\{2,5,10\}$ as can be seen by Figure 3.8b. Notably, the equilibrium to which the learning dynamics converge depends on the timescale separation and initial condition. To give a concrete example, consider the initial conditions shown in Figure 3.8a of $(x_1^0,x_2^0)=(2,-1)$ and $(x_1^0,x_2^0)=(1.9,-2.1)$. For the initial condition $(x_1^0,x_2^0)=(2,-1)$, τ-GDA converges to the equilibrium at $(0,0)$ for each $\tau\in\{2,5,10\}$. Yet, for the initial condition $(x_1^0,x_2^0)=(1.9,-2.1)$, τ-GDA converges to the equilibrium at $\{(0,0),(\pi,\pi),(\pi,\pi)\}$ for the respective choices of $\tau\in\{2,5,10\}$. In other words, the region of attraction around the critical points changes so that from a fixed initial condition τ-GDA may converge to distinct equilibrium depending on the initial condition. From Figure 3.8b, we see that the region of attraction around $(x_1^0,x_2^0)=(1.9,-2.1)$ grows from $\tau=1$ to $\tau=2$ and $\tau=4$, but then shrinks at $\tau=10$. This example highlights that timescale

[7]Note that because the joint strategy space is a torus, $(\pm\pi,\pm\pi)=(\mp\pi,\pm\pi)$, $(\pi,0)=(-\pi,0)$, and $(0,-\pi)=(0,\pi)$.



(a) Trajectories of τ-GDA (b) Distance to equilibrium (c) spec$(J_\tau), \mu = 0.3$ (d) spec$(J_\tau), \mu = 1$

(f) Trajectories of τ-GDA overlayed on vector fields generated by choices of τ and μ.

Figure 3.9: Experimental results for the Dirac-GAN game of Section 3.7.5.

separation has a fundamental impact on the region of attraction around critical points and as τ grows it is possible for the region of attraction around an equilibrium to shrink. Collectively, this motivates explicit methods for trying to shape the region of attraction around desirable equilibria.

3.7.5 Dirac-GAN: Saturating Formulation

The Dirac-GAN introduced by Mescheder et al. (2018) consists of a univariate generator distribution $p_\theta = \delta_\theta$ and a linear discriminator $\mathrm{D}(x;\omega) = \omega x$, where the real data distribution $p_\mathcal{D}$ is given by a Dirac-distribution concentrated at zero. The resulting zero-sum game is defined by the cost $f(\theta,\omega) = \ell(\theta\omega) + \ell(0)$ and the unique critical point $(\theta^*,\omega^*) = (0,0)$ is a local Nash equilibrium. However, the eigenvalues of the Jacobian are purely imaginary regardless of the choice of timescale separation so that τ-GDA oscillates and fails to converge. This behavior is expected since the equilibrium is not hyperbolic and corresponds to neither a differential Nash equilibrium nor a differential Stackelberg equilibrium but it is undesirable nonetheless. The zero-sum game corresponding to the Dirac-GAN with regularization can be defined by the cost $f(\theta,\omega) = \ell(\theta\omega) + \ell(0) - \frac{\mu}{2}\omega^2$. The unique critical point remains unchanged, but for all $\tau \in (0,\infty)$ and $\mu \in (0,\infty)$ the equilibrium of the unregularized game is stable and corresponds to a differential Stackelberg equilibrium of the regularized game.

From Figures 3.9a and 3.9f, we observe that the impact of timescale separation with regularization $\mu = 0.3$ is that the trajectory is not as oscillatory since it moves faster to the zero line of $-D_2 f(\theta,\omega)$ and then follows along that line until reaching the equilibrium. We further see from Figure 3.9b that with regularization $\mu = 0.3$, τ-GDA with $\tau = 8$ converges faster to the equilibrium

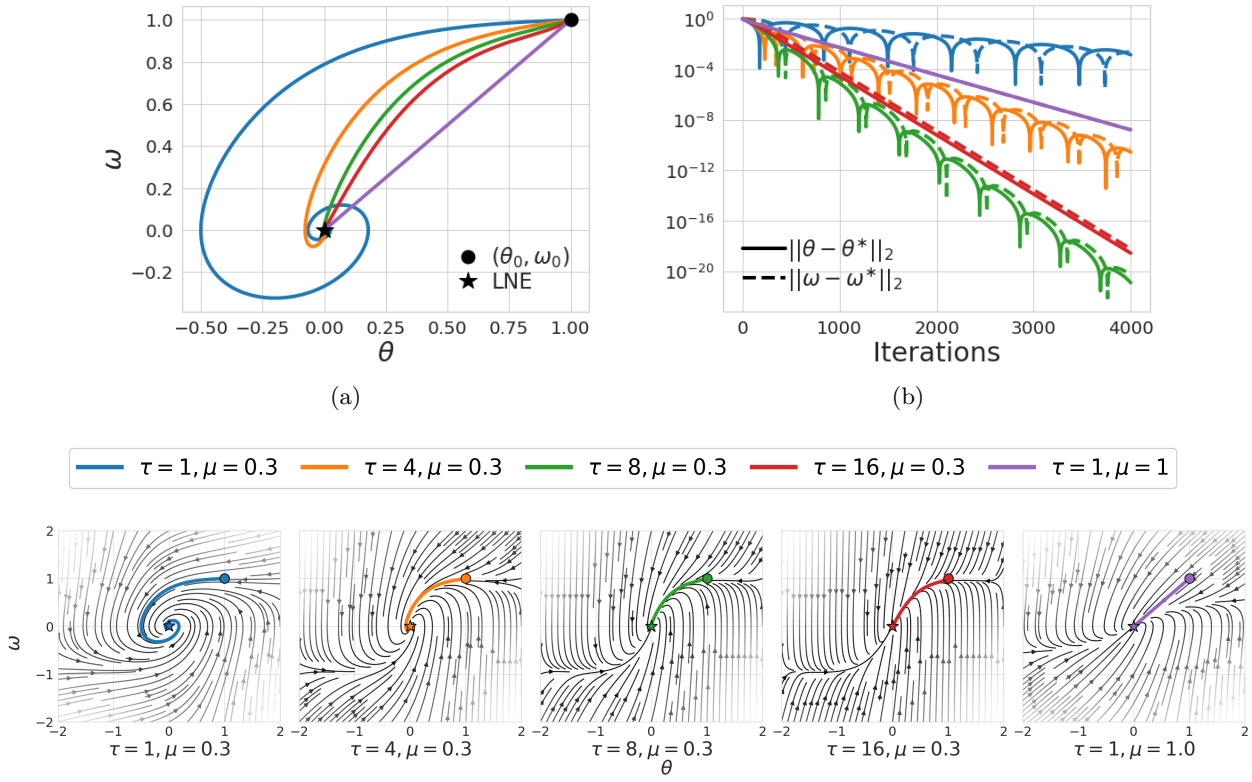


Figure 3.10: Experimental results for the Dirac-GAN game defined in (3.15) of Section 3.7.6. Figure 3.10a shows trajectories of τ-GDA for $\tau \in \{1, 4, 8, 16\}$ with regularization $\mu = 0.3$ and $\tau = 1$ with regularization $\mu = 1$. Figure 3.10b shows the distance from the equilibrium along the learning paths. Figure 3.10d shows the trajectories of τ-GDA overlayed on the vector field generated by the respective timescale separation and regularization parameters. The shading of the vector field is dictated by its magnitude so that lighter shading corresponds to a higher magnitude and darker shading corresponds to a lower magnitude.

than τ-GDA with $\tau = 16$, despite the fact that the former exhibits some cyclic behavior in the dynamics while the latter does not. The eigenvalues of the Jacobian with regularization $\mu = 0.3$ presented in Figure 3.9c explains this behavior since the imaginary parts are non-zero with $\tau = 8$ and zero with $\tau = 16$, but the eigenvalue with the minimum real part is greater at $\tau = 8$ than at $\tau = 16$. This highlights that some oscillatory behavior in the dynamics is not always harmful for convergence and it can even speed up the rate of convergence. For $\mu = 1$ and $\tau = 1$, Figures 3.9a and 3.9b show that even though τ-GDA follows a direct path toward the equilibrium and does not cycle since the eigenvalues of the Jacobian are purely real, the trajectory converges slowly to the equilibrium. Indeed, for each regularization parameter, the eigenvalues of $J_\tau(\theta^*, \omega^*)$ split after becoming purely real and then converge toward the eigenvalues of $\mathtt{S}_1(J(\theta^*, \omega^*))$ and $-\tau D_2^2 f(\theta^*, \omega^*)$. Since $\mathtt{S}_1(J(\theta^*, \omega^*)) \propto 1/\mu$ and $-\tau D_2^2 f(\theta^*, \omega^*) \propto \tau\mu$, there is a trade-off between the choice of regularization μ and the timescale separation τ on the conditioning of the Jacobian matrix that

dictates the convergence rate.

3.7.6 Dirac-GAN and Regularization: Non-Saturating Formulation

In Section 3.7.5, we presented experiments for the Dirac-GAN game studied by Mescheder et al. (2017) using the original generative adversarial network formulation of Goodfellow et al. (2014). In this section, we revisit the Dirac-GAN game using the non-saturating generative adversarial network formulation also proposed by Goodfellow et al. (2014). Recall that the zero-sum game which arises from the original objective with regularization $\mu > 0$ is defined by the cost

$$f(\theta, \omega) = \ell(\theta\omega) + \ell(0) - \frac{\mu}{2}\omega^2.$$

As discussed in Section 3.7.5, the unique critical point of the game is $(\theta^*, \omega^*) = (0, 0)$ and it corresponds to the local Nash equilibrium of the unregularized game and a differential Stackelberg equilibrium of the regularized game. Moreover, the equilibrium is stable with respect to the continous time dynamics for all $\tau > 0$ and $\mu > 0$ so that the discrete time update τ-GDA converges with a suitable learning rate γ_1.

The non-saturating generative adversarial network formulation proposed by Goodfellow et al. (2014) in the context of the Dirac-GAN game corresponds to player 1 maximizing $\ell(-\theta\omega)$ instead of minimizing $\ell(\theta\omega)$. This results in the general-sum game defined by the costs

$$(f_1(\theta, \omega), f_2(\theta, \omega)) = (-\ell(-\theta\omega) + \ell(0) - \frac{\mu}{2}\omega^2, -\ell(\theta\omega) - \ell(0) + \frac{\mu}{2}\omega^2). \tag{3.15}$$

As shown by Mescheder et al. (2018), the unique critical point of the game remains at $(\theta^*, \omega^*) = (0, 0)$. Moreover, it can be observed that $J_\tau(\theta^*, \omega^*)$ in this formulation is identical to the game Jacobian for the Dirac-GAN, which is given by

$$J_\tau(\theta^*, \omega^*) = \begin{bmatrix} 0 & \ell'(0) \\ -\tau\ell'(0) & \tau\mu \end{bmatrix}, \tag{3.16}$$

so this game is locally equivalent to the zero-sum game that arises from the original objective proposed by Goodfellow et al. (2014). This is despite the fact that the non-saturating objective was motivated by global concerns (vanishing gradients early in the training process) rather than local considerations. In Figure 3.10 we present experiments with τ-GDA for the regularized Dirac-GAN game with the non-saturating objective and $\ell(t) = -\ell(1 + \exp(-t))$. We observe similar behavior as the experiments with the standard objective and refer back to Section 3.7.5 for the insights we draw from the simulation. This experiment is primarily included for completeness and to motivate our use of the non-saturating objective in the generative adversarial networks experiments parameterized by neural networks.

3.7.7 Generative Adversarial Network: Learning a Covariance Matrix

We now consider a generative adversarial network formulation presented by Daskalakis et al. (2018) for learning a covariance matrix. This is a simple example with degeneracies much like the Dirac-GAN game, but it can be generalized to arbitrary dimensional strategy spaces and has served

as a benchmark for comparing convergence rates in a number of recent papers on learning in games. Often, the example is used to show that gradient descent-ascent cycles and converges slowly. However, by and large, timescale separation is not considered. We show that gradient descent-ascent converges fast in this game with suitable timescale separation and further explore the interplay between timescale separation, regularization, and rate of convergence. We primarily follow the notation of Daskalakis et al. (2018) when describing the problem.

The objective of this problem is to learn a covariance matrix using the Wasserstein GAN formulation. The real data x is drawn from a mean-zero multivariate normal distribution with an unknown covariance matrix Σ. The generator is restricted to be a linear function of the random input noise $z \sim \mathcal{N}(0, I)$ and is of the form $G_V(z) = Vz$. The discriminator is restricted to the set of all quadratic functions, which we represent by $D_W(x) = x^\top W x$. The parameters of the generator and the discriminator are given by $W \in \mathbb{R}^{d\times d}$ and $V \in \mathbb{R}^{d\times d}$, respectively. For the given generator and discriminator classes the Wasserstein GAN game is defined by the cost

$$f(V, W) = \mathbb{E}_{x\sim\mathcal{N}(0,\Sigma)}[x^\top W x] - \mathbb{E}_{z\sim\mathcal{N}(0,I)}[z^\top V^\top W V z].$$

As shown by Daskalakis et al. (2018), the cost function can be simplified to be expressed as

$$f(V, W) = \sum_{i=1}^{d}\sum_{j=1}^{d} W_{ij}\Big(\Sigma_{ij} - \sum_{k=1}^{d} V_{ik}V_{jk}\Big).$$

With this cost, the individual gradients for gradient descent-ascent are given by

$$g(V, W) = (-(W + W^\top)V, -(\Sigma - VV^\top)).$$

From the individual gradients, it is clear that the critical points of the game are given by (V, W) such that $VV^\top = \Sigma$ and $W + W^\top = 0$. Moreover, given the form of $g(V, W)$, the game Jacobian at any critical point (V^*, W^*) is of the form

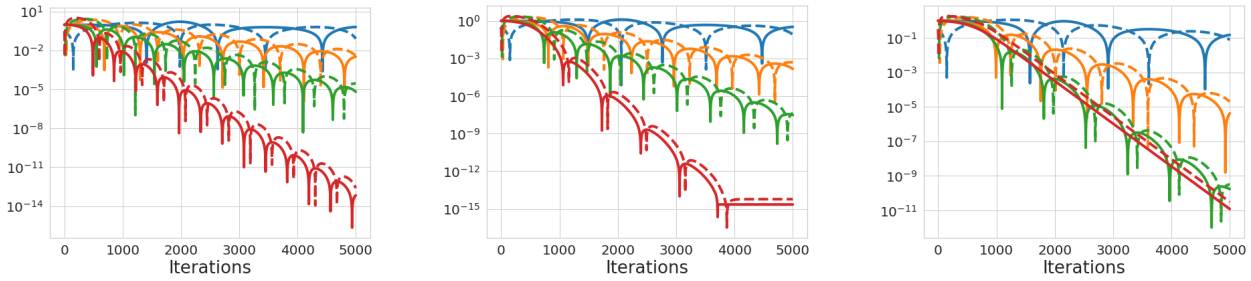
$$J_\tau(V^*, W^*) = \begin{bmatrix} 0 & D_{12}f(V^*, W^*) \\ -\tau D_{12}^\top f(V^*, W^*) & 0 \end{bmatrix}.$$

Consequently, the eigenvalues of the game Jacobian are purely imaginary and the critical points are not stable. To fix this problem, Daskalakis et al. (2018) regularized both the generator and discriminator. We only regularize the discriminator in this example. The cost function of the zero-sum game with regularization $\mu > 0$ is given by

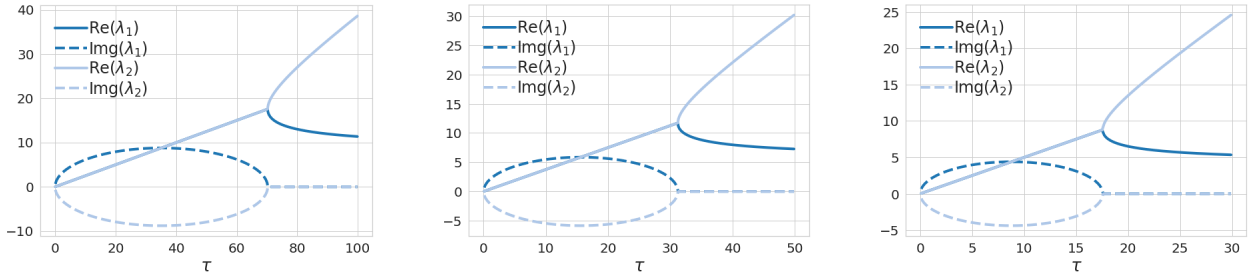
$$f(V, W) = \sum_{i=1}^{d}\sum_{j=1}^{d} W_{ij}\Big(\Sigma_{ij} - \sum_{k=1}^{d} V_{ik}V_{jk}\Big) - \frac{\mu}{2}\mathrm{Tr}(W^\top W). \tag{3.17}$$

The individual gradients for gradient descent-ascent in this regularized game are then

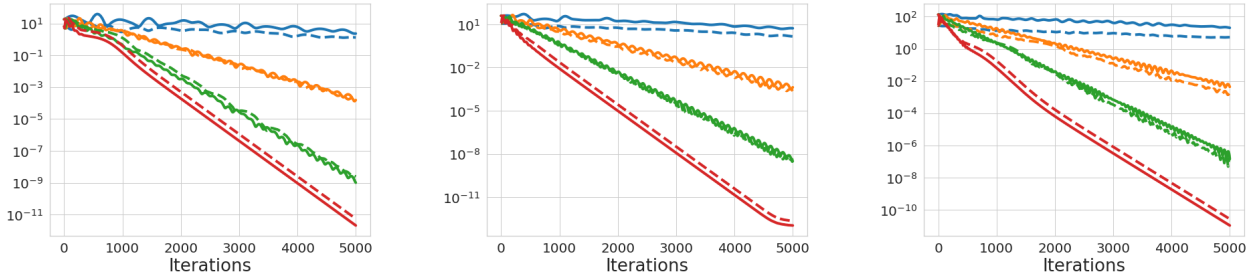
$$g(V, W) = (-(W + W^\top)V, -(\Sigma - VV^\top) + \frac{\mu}{2}W).$$



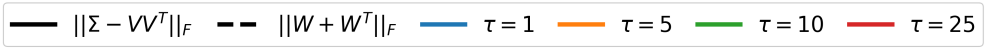
(a) $d = 1, \mu = 0.5$ (b) $d = 1, \mu = 0.75$ (c) $d = 1, \mu = 1$



(d) (e) (f)



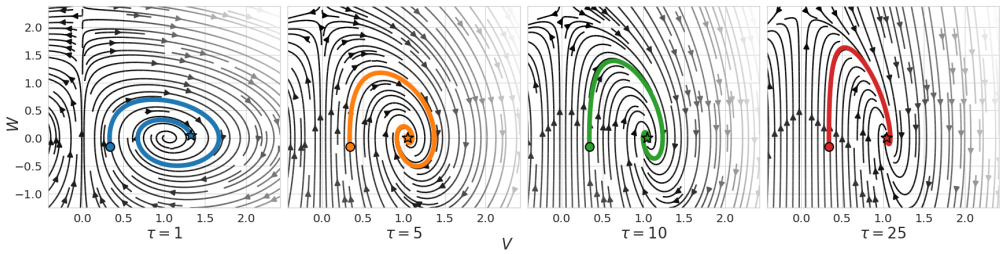
(g) $d = 5, \mu = 1$ (h) $d = 10, \mu = 1$ (i) $d = 20, \mu = 1$



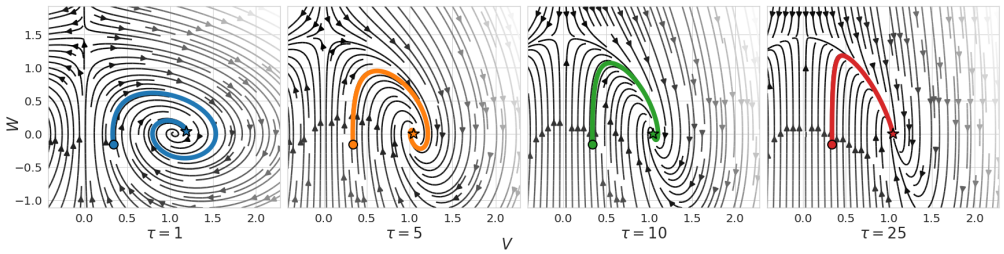
(j) Legend for Figures 3.11a, 3.11b, 3.11c, 3.11g, 3.11h, and 3.11i.

Figure 3.11: Experimental results for the generative adversarial network formulation for learning a covariance matrix defined by the cost from (3.17) of Section 3.7.7. Figures 3.11a, 3.11b, and 3.11c show the distance from the equilibrium along the learning paths of τ-GDA with $d = 1$. Figures 3.11d, 3.11e, and 3.11f show the trajectories of the eigenvalues of $J_\tau(x^*)$ as a function of the τ, respectively. Figures 3.11g, 3.11h, and 3.11i show the distance from the equilibrium along the learning paths of τ-GDA with $d = 5, 10, 20$.

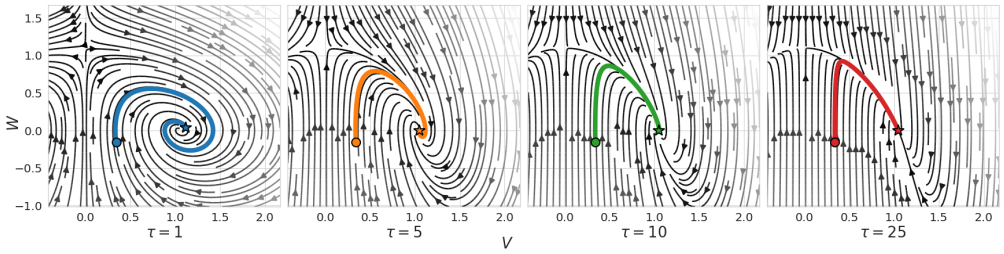
We begin by considering the simplest form of this problem, which is that $d = 1$. The critical points with this restriction are $(V^*, W^*) = (\sigma, 0)$ and $(V^*, W^*) = (-\sigma, 0)$ and the game Jacobian



(a) $d = 1, \mu = 0.5$



(b) $d = 1, \mu = 0.75$



(c) $d = 1, \mu = 1$

Figure 3.12: Experimental results for learning a covariance matrix defined by the cost from (3.17) of Section 3.7.7. We overlay the trajectories produced by τ-GDA onto the vector field generated by the choices of τ and μ. The shading of the vector field is dictated by its magnitude so that lighter shading corresponds to a higher magnitude and darker shading corresponds to a lower magnitude.

evaluated at them is

$$J_\tau(V^*, W^*) = \begin{bmatrix} 0 & -2\sigma \\ 2\tau\sigma & \tau\mu \end{bmatrix}.$$

Each critical point is a local Nash equilibrium of the unregularized game and a differential Stackelberg equilibrium of the regularized game since $-D_2^2 f(V^*, W^*) = \mu > 0$ and $\mathtt{S}_1(J(V^*, W^*)) = 4\sigma^2/\mu > 0$. Furthermore, $\text{spec}(J_\tau(V^*, W^*)) = \{(\tau\mu \pm \sqrt{\tau^2\mu^2 - 16\tau\sigma^2})/2\}$ so that each critical point is stable for all $\tau \in (0, \infty)$ and $\mu \in (0, \infty)$ since $\text{spec}(J_\tau(\theta^*, \omega^*)) \subset \mathbb{C}_+^\circ$. Thus, given a suitably chosen learning rate γ_1, the discrete time update τ-GDA locally converges to an equilibrium. For this reason, we focus on studying the rate of convergence for the problem as a function of timescale separation and regularization. Figures 3.11a, 3.11b, and 3.11c show the distance from an equilibrium along the learning path of τ-GDA with $\tau \in \{1, 5, 10, 25\}$ given a fixed initial condition with learning rate $\gamma_1 = 0.001$ and regularization $\mu \in \{0.5, 0.75, 1\}$, respectively. Moreover, Fig-

ures 3.11d, 3.11e, and 3.11f show the trajectories of the eigenvalues for $J_\tau(V^*, W^*)$ as a function of τ for the regularization parameters $\mu \in \{0.5, 0.75, 1\}$. Finally, Figures 3.12a, 3.12b, and 3.12c show the trajectories of τ-GDA overlayed on the vector field generated by the respective timescale separation and regularization parameters.

From the eigenvalue trajectories, we see that as μ grows, the eigenvalues become purely real at a smaller value of τ. Moreover, as μ increases, the magnitude of the real and imaginary parts of the eigenvalues decreases. We observe the effect of this on the convergence, where the dynamics do not cycles as much for larger μ. Again, we see the trade-off between timescale separation, regularization, and convergence. For example, despite the eigenvalues being purely real with $\mu = 1$ and $\tau = 25$ so that there is no rotational dynamics, the convergence is slower than for $\mu = 0.75$ where there is some non-zero imaginary piece of the eigenvalues.

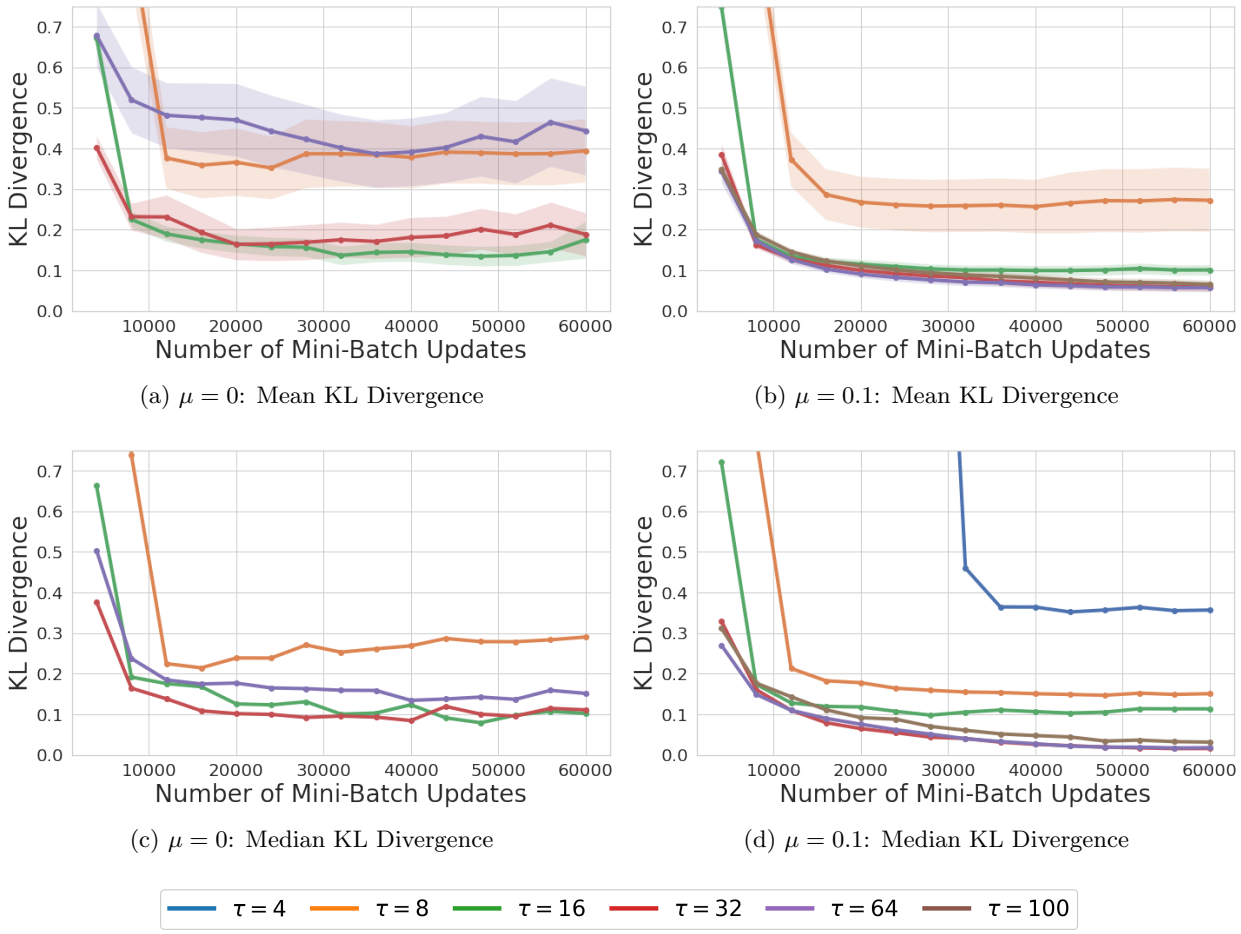
Figures 3.11g, 3.11h, and 3.11i show the distance from a critical point along the learning path of τ-GDA with $\tau \in \{1, 5, 10, 25\}$ given a fixed initial condition with learning rate $\gamma_1 = 0.001$, regularization $\mu = 1$, and the dimension of the problem d among the set $\{5, 10, 20\}$, respectively. The primary purpose of showing this set of results is simply to be clear that the behavior for $d = 1$, which is easier to explain and visualize, transfers over to higher dimensional formulations of this problem. This is to be expected since the problem dimension is not necessarily fundamental to the convergence rate, but rather it depends on the conditioning of Σ and each Σ was chosen so that the behavior was comparable for each choice of dimension.

3.7.8 Generative Adversarial Networks Parameterized by Neural Networks

In this section, we provide results for training generative adversarial networks parameterized by neural networks. This includes experiments with a mixture of Gaussians and image datasets.

Background. The empirical benefits of training with a timescale separation have been documented previously. For example, Heusel et al. (2017) showed on a number of image datasets that a timescale separation between the generator and discriminator improves generation performance as measured by the Frechet Inception Distance (FID). Since then a significant number of papers have presented results training generative adversarial networks with timescale separation. Moreover, it is common in the literature for the discriminator to be updated multiple times between each update of the generator (Arjovsky et al., 2017). Indeed, it has been widely demonstrated that this heuristic improves the stability and convergence of the training process and locally it has a similar effect as including a timescale separation between the generator and discriminator. The disadvantage of this approach is that the number of gradient calls per generator update increases and consequently the convergence is then slower in terms of wall clock time when a similar effect could potentially be achieved by a learning rate separation between the generator and discriminator. We remark that it appears to be reasonably common for practitioners to fix a shared learning rate for the generator and discriminator along with a pre-selected number of discriminator updates per generator update and not thoroughly investigate the impact timescale separation has on the training process.

The goal of our generative adversarial network experiments is to reinforce the importance of the timescale separation between the generator and the discriminator as a hyperparameter in the training process, demonstrate how it changes the behavior along the learning path, and show that it is compatible with a number of common training heuristics. This is to say that our goal is not necessarily to show state-of-the art performance, but rather to perform experiments that allow us



(a) $\mu = 0$: Mean KL Divergence

(b) $\mu = 0.1$: Mean KL Divergence

(c) $\mu = 0$: Median KL Divergence

(d) $\mu = 0.1$: Median KL Divergence

Figure 3.13: KL-divergence between generated and real data for a mixture of Gaussians.

to make insights relevant to the theory in this chapter. We remark that our empirical work on training generative adversarial networks is distinct from and complimentary to that of Heusel et al. (2017) in several ways. The theory given by Heusel et al. (2017) only applies to stochastic stepsizes, however in the experiments they implemented constant step sizes. We train with mini-batches and decaying stepsizes in our image dataset experiments, which does satisfy the theory we provide as detailed in Section 3.6. Moreover, by and large, the experiments by Heusel et al. (2017) compare a fixed learning rate ratio between the generator and discriminator to multiple fixed shared learning rates for the generator and discriminator. In contrast, we fix a learning rate for the generator and explore the behavior of the training process as the timescale parameter τ is swept over a given range.

3.7.8.1 Generative Adversarial Networks: Mixture of Gaussians

We now provide the results from training generative adversarial networks to learn a mixture of Gaussians. The underlying data distribution consists of Gaussian distributions configured in a circle

arrangement with means given by $\mu = [\sin(\omega), \cos(\omega)]$ for $\omega \in \{k\pi/4\}_{k=0}^{7}$, each with covariance $\sigma^2 I$ where $\sigma^2 = 0.05$. Each sample of real data given to the discriminator is selected uniformly at random from the set of Gaussian distributions. We train the generator using latent vectors $z \in \mathbb{R}^{16}$ sampled from a standard normal distribution in each training batch. The batch size for each player in the game is 512. The network for the generator and discriminator contain two and one hidden layers respectively, each which contain 32 neurons and ReLU activation functions. The training objective is the non-saturating objective and we run experiments without and with the R_1 gradient penalty proposed by Mescheder et al. (2018) using parameter $\mu = 0.1$. The generator learning rate is fixed to be $\gamma_1 = 0.005$ and the discriminator learning rate is fixed as $\gamma_2 = \tau\gamma_1$ where we experiment with $\tau \in \{4, 8, 16, 32, 64, 100\}$. For each parameter choice (timescale separation τ and regularization μ), the experiment is repeated with 50 random seeds. The training does not rely on any adaptive gradient methods (Adam, RMSprop, etc.) and is the 'vanilla' stochastic τ-`GDA` dynamics. We evaluate the performance along the learning path by computing the KL-divergence between the generated data and the real data, where we sample 4096 data points from each.

The results of this experiment are presented in Figure 3.13. We show the mean of the KL-divergence and the standard error of the means across the runs along the learning path without ($\mu = 0$) and with regularization ($\mu = 0.1$) in Figures 3.13a and 3.13b, respectively. Moreover, Figures 3.13c and 3.13d show the medians of the KL-divergence across the runs without ($\mu = 0$) and with regularization ($\mu = 0.1$), respectively. From Figure 3.13a, we observe that the choices of $\tau = 4$ and $\tau = 100$ do not show on the plot since they perform poorly, which may be a result of equilibrium not being stable for $\tau = 4$ and numerically conditioning for $\tau = 100$. Furthermore, we see that timescale separation improves the results up to a reasonable timescale parameter and after which the performance degrades. Furthermore, Figure 3.13b reveals that the results are improved with regularization and we see that $\tau = 100$ ends up performing well, potentially since the regularization can alleviate some of the problems of numerical stability. In general, we draw similar conclusions from the median scores as reported in Figures 3.13c and 3.13d.

The primary purpose of this experiment is to train generative adversarial networks parameterized by neural networks using τ-`GDA` without heuristics such as adaptive gradient methods or parameter averaging as is employed on the image dataset experiment that follow. Notably, we see that consistent themes emerge that timescale separation improves convergence until hitting a limiting value and regularization can improve the rate of convergence but there is an interplay with the timescale separation.

3.7.8.2 Generative Adversarial Networks: Image Datasets

We now provide the experiments training generative adversarial networks with image datasets.

Methods. The experiments in this section are based on the methods and implementations of Mescheder et al. (2018) and used the publicly available code from the paper available at `https://github.com/LMescheder/GAN_stability`. We effectively only changed the learning rates, retained multiple exponential averages at once, and modified the updates to be simultaneous in the code. In Figure 3.18 we provide the network architectures from our experiments and in Figure 3.19 we include the hyperparameters that were selected. The architectures are analogous to that reported in Mescheder et al. (2018), but scaled down since we run experiments with $32 \times 32 \times 3$ images. For evaluation, we computed the Frechet Inception Distance using

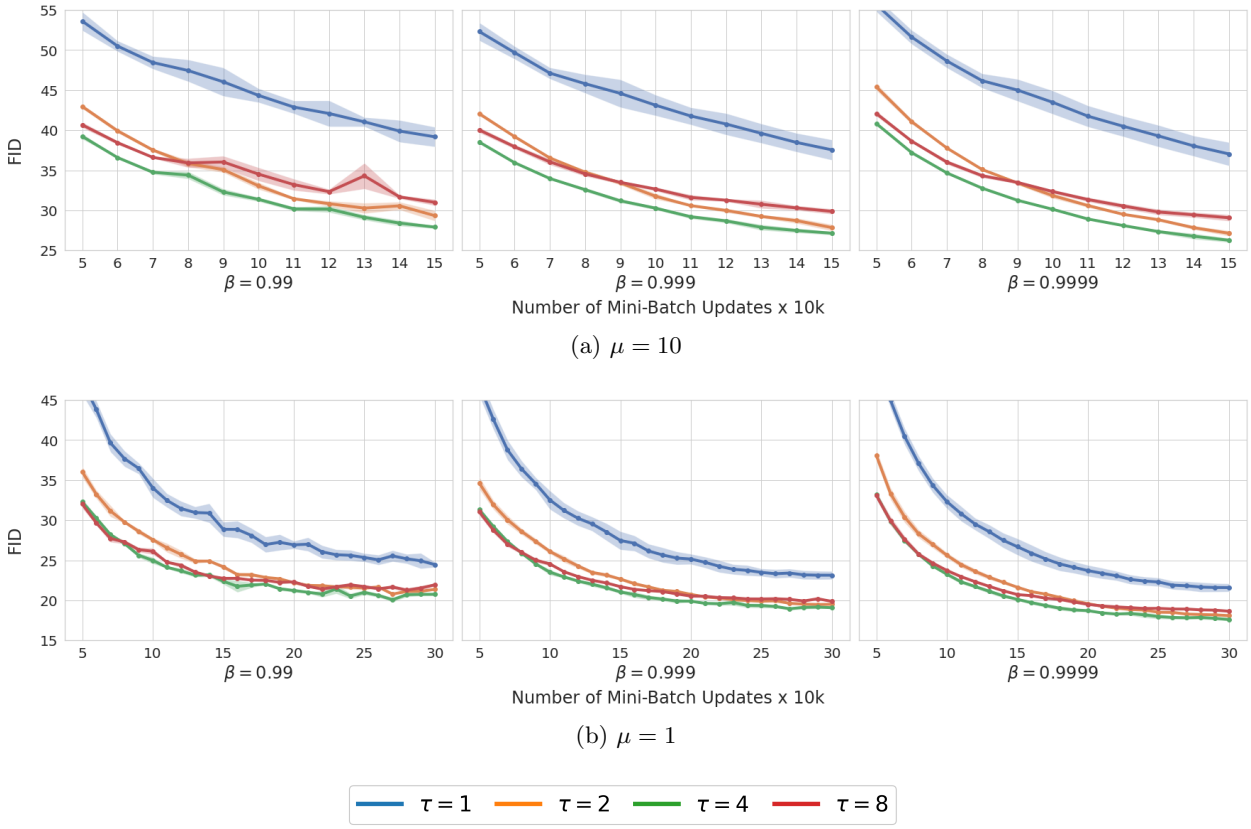
10k samples from the real and generated data. For both experiments and across the set of hyperparameters we performed the evaluation using a fixed random noise vector to make for an equal comparison and a fixed set of real images which were randomly selected. The evaluation was done using the training data. We used the FID score implementation in pytorch available at `https://github.com/mseitzer/pytorch-fid`.

We train the generative adversarial networks with the non-saturating objective function and the R_1 gradient penalty proposed by Mescheder et al. (2018) with regularization parameters $\mu \in \{1, 10\}$. We note that the non-saturating objective results in a game that is not zero-sum, however it is commonly used in practice and under the realizable assumptions it can be locally equivalent to the zero-sum objective as discussed in Section 3.7.6. The theory we provide does not apply to using RMSprop, but it is ubiquitous in practice for training generative adversarial networks and we are interested in exploring the interplay of timescale separation with common heuristics to understand if similar conclusions hold when using them as from the previous experiments regarding timescale separation with the 'vanilla' τ-`GDA` dynamics. Moreover, we note that similarly Heusel et al. (2017) and Mescheder et al. (2018) also rely upon Adam or RMSprop in generative adversarial experiments. A final heuristic and hyperparameter that we explore in conjunction with the timescale separation τ is that of using an exponential moving average to produce the model that is evaluated. This means that at each update k, given that the parameters of the generator are given by $x_{1,k}$, the moving average $\bar{x}_k = x_{1,k}\beta + \bar{x}_{1,k-1}(1-\beta)$ is kept where $\beta \in (0, 1)$. Experimental studies have shown that this heuristic can yield a significant improvement in terms of both the inception score and the FID (Gidel et al., 2019; Yazici et al., 2019). The success of this method is thought to be a result of dampening both rotational dynamics and the noise from the randomness in the mini-batches of data.

Experimental Results. We run the training algorithm with the learning rate ratio τ belonging to the set $\{1, 2, 4, 8\}$ for CIFAR-10 and $\{1, 2, 4, 8, 16\}$ for CelebA along with the regularization parameter μ belonging to the set $\{1, 10\}$. For each choice of τ and μ, we retain exponential moving averages of the generator parameters for $\beta \in \{0.99, 0.999, 0.9999\}$. The training process is repeated 3 times for each hyperparameter configuration. The performance is evaluated along the learning path at every 10,000 updates in terms of the FID score. We report the mean scores and the standard error of the mean over the repeated experiments for each dataset. The FID score is such that a lower score beats a higher score. The experiments are computationally intensive which limits the number of repeats of experiments that can be simulated, however, we observed that the scores were quite consistent between random seeds particularly with exponential averaging of the parameters. We run the experiments with $\mu = 1$ for 150k mini-batch updates and the experiments with $\mu = 10$ for 300k mini-batch updates.

The results for each dataset across the hyperparameter configurations are presented in numeric form in Figure 3.16. Figure 3.17 shows some generated samples selected at random for each dataset with the hyperparameter configuration that performed best in terms of the FID score at the end of the training process. We now describe the key observations from the experiments for each dataset.

CIFAR-10. The FID scores along the learning path for CIFAR-10 with $\mu = 10$ and $\mu = 1$ are presented in Figures 3.14a and 3.14b, respectively. The corresponding scores in numeric form are given in Figures 3.16a, 3.16c, and 3.16e for $\mu = 10$ at 150k iterations and $\mu = 1$ at 150k and 300k iterations, respectively. To begin, we observe that the exponential moving average significantly



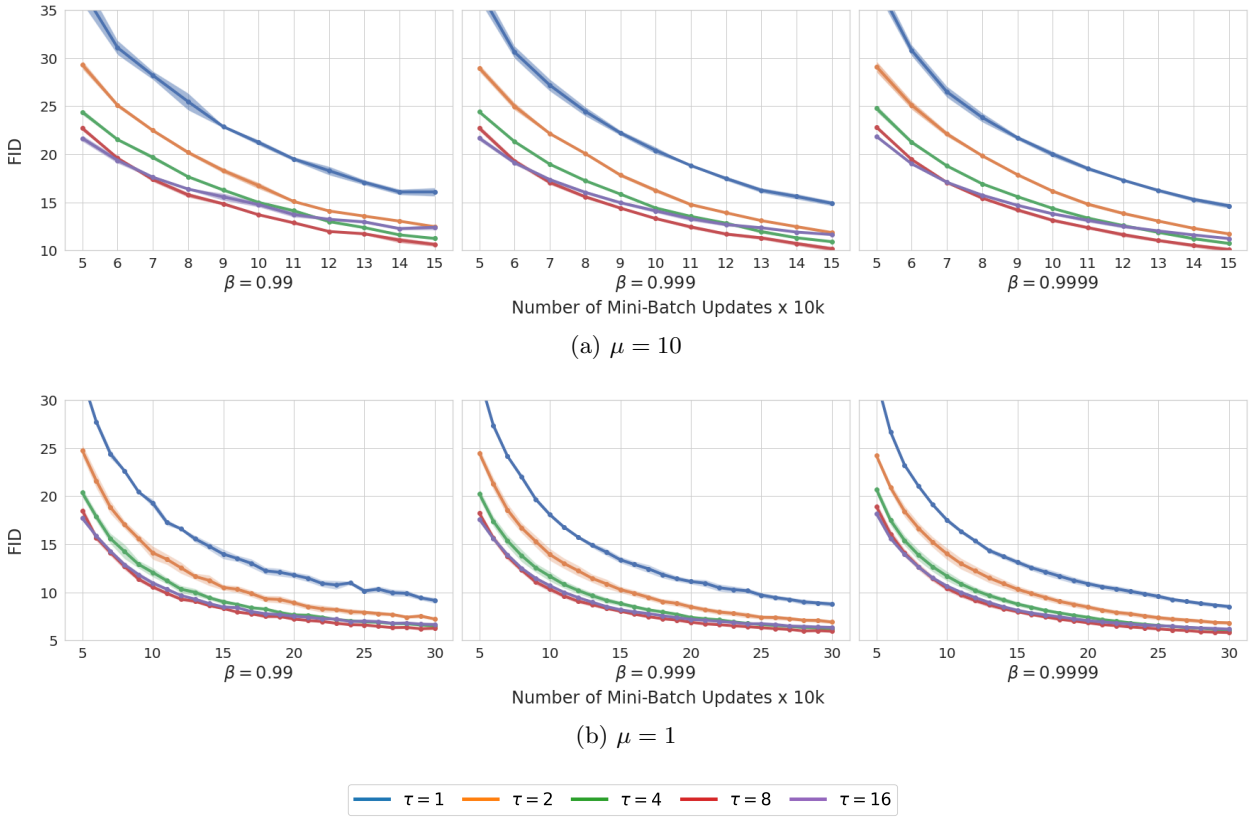
(a) $\mu = 10$

(b) $\mu = 1$

Figure 3.14: CIFAR-10 FID scores with regularization $\mu = 10$ in Figure 3.14a and $\mu = 1$ in Figure 3.14b.

improves performance, and of the parameters considered, $\beta = 0.9999$ performed best. This may be a result of removing noise as mentioned previously or potentially it could be from dampening oscillatory behavior in the dynamics. Moreover, we that timescale separation also has a significant impact on the FID score of the training process. Indeed, even selecting $\tau = 2$ versus $\tau = 1$ can yield an impressive performance gain. In this experiment for each regularization parameter, $\tau = 4$ converges fastest and performs the best. We see that $\tau = 2$ outperforms $\tau = 8$ when $\mu = 10$ and the relationship is flipped when viewing the evaluation at 150k updates with $\mu = 1$ and then returns back when looking at the evaluation at 300k updates. The choice of $\tau = 1$ performs the worst for each regularization parameter by a wide margin. Finally, observe that the performance with regularization $\mu = 1$ is much better than with regularization $\mu = 10$ for each timescale separation parameter and exponential averaging parameter.

CelebA. The FID scores along the learning path for CIFAR-10 with $\mu = 10$ and $\mu = 1$ are presented in Figures 3.15a and 3.15b, respectively. The corresponding scores in numeric form are given in Figures 3.16b, 3.16d, and 3.16f for $\mu = 10$ at 150k iterations and $\mu = 1$ at 150k and 300k iterations, respectively. In this experiment we observe that while the exponential moving average helps performance, the gain is not as drastic as it was for CIFAR-10. It is not entirely clear if this is a consequence of the scores being lower or something fundamental to the optimization landscape and



(a) $\mu = 10$

(b) $\mu = 1$

Figure 3.15: CelebA FID scores with regularization $\mu = 10$ in Figure 3.15a and $\mu = 1$ in Figure 3.15b.

dynamics for the dataset. The timescale separation in combination with the regularization again has a major effect on the the FID score of the training process in this experiment. For regularization $\mu = 10$, the timescale parameters of $\tau = 4$ and $\tau = 8$ outperform $\tau = 1$, $\tau = 2$, and $\tau = 16$ by a wide margin, again highlighting that timescale separation can speed up convergence until a certain point where it can potentially slow it down owing to the effect on the conditioning of the problem locally. A similar trend can be observed with regularization $\mu = 1$, but with $\tau = 16$ performing closer to $\tau = 4$ and $\tau = 8$. For each regularization parameter and timescale parameter, we see that $\tau = 8$ performs the best. We again observe in this experiment that for all timescale separation parameters, the performance is significantly improved with regularization $\mu = 1$ as compared with $\mu = 10$. This once again highlights the importance of considering how this the hyperparameters of regularization and timescale interact and dictate the local convergence rates.

Summary. In summary, we took a well-performing method and implementation for training generative adversarial networks and demonstrated that timescale separation is an extremely important, and easy to implement, hyperparameter that is worth careful consideration since it can have a major impact on the convergence speed and final performance of the training process. Interestingly, the conclusions we draw are in line with the insights drawn from the simple Dirac-GAN experiment in Section 3.7.5 and from the mixture of Gaussian experiments. In particular, timescale separation

$\tau \backslash \beta$	0.99	0.999	0.9999
1	39.18 ± 1.23	37.55 ± 1.27	37.04 ± 1.46
2	29.33 ± 0.65	27.84 ± 0.4	27.14 ± 0.34
4	**27.91 ± 0.17**	**27.14 ± 0.22**	**26.26 ± 0.25**
8	30.99 ± 0.39	29.86 ± 0.34	29.07 ± 0.4

(a) CIFAR-10 FID at 150k updates with $\mu = 10$

$\tau \backslash \beta$	0.99	0.999	0.9999
1	16.08 ± 0.44	14.9 ± 0.26	14.63 ± 0.26
2	12.46 ± 0.05	11.85 ± 0.11	11.72 ± 0.11
4	11.24 ± 0.13	10.9 ± 0.12	10.72 ± 0.13
8	**10.62 ± 0.22**	**10.16 ± 0.25**	**10.08 ± 0.25**
16	12.4 ± 0.28	11.64 ± 0.05	11.22 ± 0.07

(b) CelebA FID at 150k updates with $\mu = 10$

$\tau \backslash \beta$	0.99	0.999	0.9999
1	28.87 ± 0.92	27.47 ± 1.14	26.69 ± 1.02
2	24.18 ± 0.28	22.65 ± 0.15	21.63 ± 0.12
4	**22.38 ± 0.36**	**21.05 ± 0.21**	**20.12 ± 0.13**
8	22.74 ± 0.15	21.71 ± 0.11	20.72 ± 0.08

(c) CIFAR-10 FID at 150k updates with $\mu = 1$

$\tau \backslash \beta$	0.99	0.999	0.9999
1	13.98 ± 0.51	13.38 ± 0.33	13.13 ± 0.29
2	10.51 ± 0.28	10.29 ± 0.32	10.33 ± 0.36
4	9.01 ± 0.27	8.83 ± 0.25	8.78 ± 0.26
8	**8.32 ± 0.1**	**8.04 ± 0.14**	**7.98 ± 0.15**
16	8.47 ± 0.04	8.19 ± 0.07	8.13 ± 0.07

(d) CelebA FID at 150k updates with $\mu = 1$

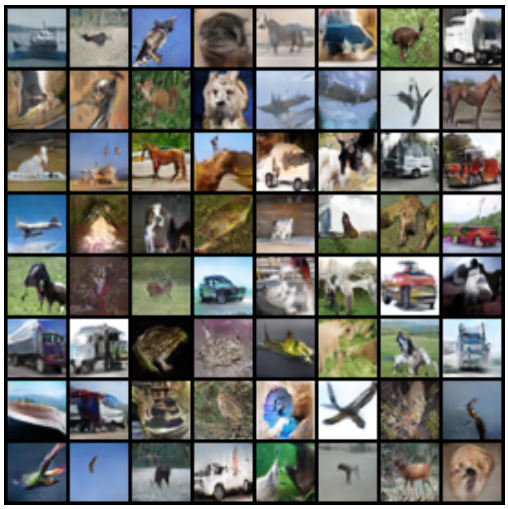
$\tau \backslash \beta$	0.99	0.999	0.9999
1	24.46 ± 0.32	23.14 ± 0.5	21.59 ± 0.5
2	21.37 ± 0.11	19.51 ± 0.07	18.08 ± 0.19
4	**20.75 ± 0.19**	**19.08 ± 0.08**	**17.61 ± 0.18**
8	21.94 ± 0.16	19.87 ± 0.1	18.64 ± 0.08

(e) CIFAR-10 FID at 300k updates with $\mu = 1$

$\tau \backslash \beta$	0.99	0.999	0.9999
1	9.16 ± 0.29	8.77 ± 0.25	8.52 ± 0.22
2	7.22 ± 0.11	6.91 ± 0.19	6.82 ± 0.18
4	6.47 ± 0.13	6.2 ± 0.11	6.07 ± 0.10
8	**6.25 ± 0.02**	**5.95 ± 0.05**	**5.81 ± 0.05**
16	6.65 ± 0.12	6.35 ± 0.11	6.17 ± 0.06

(f) CelebA FID at 300k updates with $\mu = 1$

Figure 3.16: FID Scores on CIFAR-10 and CelebA.



(a) CIFAR-10 generated sample images

(b) CelebA generated sample images

Figure 3.17: Generated sample images with $\tau = 4$ and $\beta = 0.9999$

only speeds up to convergence until hitting a limiting value and there is a key interplay between timescale separation, regularization, and convergence rate.

Layer	Output Size	Filter
Fully Connected	$512 \cdot 4 \cdot 4$	$256 \to 512 \cdot 4 \cdot 4$
Reshape	$512 \times 4 \times 4$	
Resnet-Block	$256 \times 4 \times 4$	$512 \to 256 \to 256$
NN-Upsampling	$256 \times 8 \times 8$	
Resnet-Block	$128 \times 8 \times 8$	$256 \to 128 \to 128$
NN-Upsampling	$128 \times 16 \times 16$	
Resnet-Block	$64 \times 16 \times 16$	$128 \to 64 \to 64$
NN-Upsampling	$64 \times 32 \times 32$	
Resnet-Block	$64 \times 32 \times 32$	$64 \to 64 \to 64$
Conv2D	$3 \times 64 \times 64$	$64 \to 3$

(a) Generator Network Architecture

Layer	Output Size	Filter
Conv2D	$64 \times 32 \times 32$	$3 \to 64$
Resnet-Block	$64 \times 32 \times 32$	$64 \to 64 \to 64$
Avg-Pool2D	$64 \times 16 \times 16$	
Resnet-Block	$128 \times 16 \times 16$	$64 \to 64 \to 128$
Avg-Pool2D	$128 \times 8 \times 8$	
Resnet-Block	$256 \times 8 \times 8$	$128 \to 128 \to 256$
Avg-Pool2D	$256 \times 4 \times 4$	
Resnet-Block	$512 \times 4 \times 4$	$256 \to 256 \to 512$
Fully Connected	$512 \cdot 4 \cdot 4$	$512 \cdot 4 \cdot 4 \to 1$

(b) Discriminator Network Architecture

Figure 3.18: Network Architectures for GAN experiments on CIFAR-10 and CelebA

Hyperparameter	Value(s)
Objective	NSGAN
Batch Size	64
Latent Distribution	$z \in \mathbb{R}^{256}$
Generator Learning Rate	CIFAR-10: 0.0001; CelebA: 0.00005
Timescale Separation τ	CIFAR-10: {1, 2, 4, 8}; CelebA: {1, 2, 4, 8, 16}
Learning Rate Decay	$(1+x)^{-0.005}$
Optimizer	RMSprop
RMSprop Smoothing Constant α	0.99
RMSprop ϵ	10^{-8}
Regularization μ	{1, 10}
EMA Parameter β	{0.99, 0.999, 0.9999}

Figure 3.19: Hyperparameters for GAN experiments on CIFAR-10 and CelebA

3.8 Historical Perspective: Dynamical Systems and Control

In this section, we provide a review of related work from a historical perspective on dynamical systems theory and control. The study of gradient descent-ascent dynamics with timescale separation between the minimizing and maximizing players is closely related to that of singularly perturbed dynamical systems (Kokotovic et al., 1986). Such systems arise in classical control and dynamical systems in the context of physical systems that either have multiple states which evolve on different timescales due to some underlying immutable physical process or property, or a single dynamical system which evolves on a sub-manifold of the larger state-space. For example, robot manipulators or end effectors often have have slower mechanical dynamics than electrical dynamics. On the other hand, in electrical circuits or mechanical systems, certain resistor-capacitor circuits or spring-mass systems have a state which evolves subject to a constraint equation (Lagerstrom and Casten, 1972; Sastry and Desoer, 1981). Due to their prevalence, singularly perturbed systems have been studied extensively with one of the outcomes being a number of works on determining the range of perturbation parameters for which the overall system is stable (Kokotovic et al., 1986; Saydy, 1996; Saydy et al., 1990). We exploit these results and analysis techniques to develop novel results for learning in games. One of contributions of this work is the introduction of the algebraic analysis techniques to the machine learning and game theory communities. These tools open up new avenues for algorithm synthesis; we comment on potential directions in the concluding discussion section.

This being said, there are a couple key difference between the present setting and that of the

classical literature including the following:

1. **The perturbation parameter is no longer an immutable characteristic of the physical system, but rather a hyperparameter subject to design.** Indeed, in singular perturbation theory, the typical dynamical system studied takes the form

$$\dot{x} = g_1(x, y) \quad \epsilon\dot{y} = g_2(x, y) \tag{3.18}$$

where ϵ is a small parameter that abstracts some physical characteristics of the state variables. On the other hand, in learning in games, the continuous time limiting dynamical system of gradient descent-ascent for a zero-sum game defined by $f \in C^2(\mathcal{X} \times Y, \mathbb{R})$ takes the form

$$\dot{x} = -D_1 f(x, y) \quad \dot{y} = \tau D_2 f(x, y) \tag{3.19}$$

where the x–player seeks to minimize f with respect to x and the y–player seeks to maximize f with respect to y, and τ is the ratio of learning rates (without loss of generality) of the maximizing to the minimizing player. These learning rates—and hence the value of τ—are hyperparameters subject to design in most machine learning and optimization applications. Another feature of (3.19) as compared to (3.18), is that the dynamics $D_i f$ are partial derivatives of a function f, which leads to the second key difference.

2. **There is structure in the dynamical system that arises from gradient-play which reflects the underlying game theoretic interactions between players.** This structure can be exploited in obtaining convergence guarantees in machine learning and optimization applications of game theory. For instance, minmax optimization is analogous to a zero sum game for which the local linearization of gradient descent-ascent dynamics has the structure

$$J = \begin{bmatrix} A & B \\ -\tau B^\top & -\tau C \end{bmatrix}$$

where $A = A^\top$ and $C = C^\top$ and τ is the learning rate ratio or timescale separation parameter. Such block matrices have very interesting properties. In particular, second order optimality conditions for a minmax equilibrium correspond to positive definiteness of the first Schur complement $\mathtt{S}_1(J) = A - BC^{-1}B^\top \succ 0$, and of $-C \succ 0$ (Fiez et al., 2020a). This turns out to be keenly important for understanding convergence of gradient descent-ascent. Furthermore, due to the structure of J, tools from the theory of block operators (see, e.g., works by Lancaster and Tismenetsky (1985); Magnus (1988); Tretter (2008)) such as the quadratic numerical range can be exploited (and combined with singular perturbation theory) to understand the effects of hyperparameters such as τ (the learning rate ratio) and regularization (which is common in applications such as generative adversarial networks) on convergence.

3.9 Discussion

In this chapter, we prove gradient descent-ascent locally converges to a critical point for a range of finite learning rate ratios if and only if the critical point is a differential Stackelberg equilibrium. Moreover, we exactly characterize this range of learning ratios. On the other hand, we show

that gradient descent-ascent is unstable around critical points that are not differential Stackelberg equilibria for a range of finite learning rate ratios. Together, with known saddle avoidance results regarding gradient descent-ascent with timescale separation (Mazumdar et al., 2020), the results provide a near complete characterization of the local stability of the algorithm. To obtain our theoretical results, we rely on the notion of a guard map, which is a novel tool to the machine learning literature and has potential to find relevance elsewhere. In addition, we provide results on iteration complexity and convergence rates. Finally, the extensive experimental results shed insights into a number of practical considerations.

A significant contribution of this chapter is the fact that we introduce tools that are arguably new to the machine learning and optimization communities and expose interesting new directions of research. In particular, the notion of a guard map, which is arguably even an obscure tool in modern control and dynamical systems theory, is 'rediscovered' here. There is potential to leverage this concept in not only providing certificates for performance (e.g., beyond stability to robustness) but also in synthesizing algorithms with performance guarantees. Moreover, it is worth investigating using the guard map tools with multiple parameters.

As commented on earlier in this chapter, an alternative but related technique to using the linearization when determining stability is to analyze the nonlinear system directly. The downside of this technique is that one needs to have in hand (or be able to construct) Lyapunov functions for both the *boundary layer model* (that is, the system that arises from treating the choice variable of the slow player as being 'static') and the *reduced order model* (that is, the system that arises from plugging in the implicit mapping from the fast player's action to the slow player's action into the slow player's dynamics). A convex combination of these functions provides a Lyapunov function for the original system $\dot{x} = -\Lambda_\tau g(x)$. The level sets of this combined Lyapunov function then give a better sense of the region of attraction and, in fact, one can optimize over the weighting in the convex combination in order to obtain better estimates of the region of attraction. For some background on this, see the book of Kokotovic et al. (1986). This is an interesting avenue to explore in the context of learning in games with lots of intrinsic structure that can potentially be exploited to improve both the rate of convergence and the region on which convergence is guaranteed.

Another set of related open questions center on practical considerations for the efficient use of first-order methods. For instance, with respect to generative adversarial networks, the exponential moving average is known to empirically reduce the negative effects of cycling. Additionally, increasing the learning rate ratio does lead to predominantly real eigenvalues which in turn reduces cycling. Understanding the trade offs between not only these two hyperparameters but also regularization is very important for practical implementations. Empirically, we study the tradeoffs between the learning rate ratio, regularization parameter, and the parameter controlling the degree of "smoothness" in the exponential moving average, another common heuristic that performs well in practice. Developing a theoretical understating of the interplay between these three hyper-parameters would be valuable.

Perhaps most importantly, the characterization of the timescale parameter τ^* that ensures the local stability of a differential Stackelberg equilibrium for all $\tau > \tau^*$ is with respect to a specific equilibrium. However, as was shown in the study of generative adversarial networks under some assumptions, there are classes of problems where a characterization of τ^* can be developed to apply to all differential Stackelberg equilibrium in a game. A similar comment can be made about the instability parameter τ_0 with respect to a critical point that is not a differential Stackelberg

equilibrium. An important question for future work is discovering more classes of games where uniform characterizations can be drawn. In fact, this will be a focus of the chapter that follows.

Finally, we remark that the stability and instability characterizations are with respect to the τ-GDA system. This system has many variants including alternating updates along the introduction of optimistic and extra-gradient methods. It would be interesting to explore stability of these algorithms further along with the connections to the τ-GDA system. Intuitively, one would expect that similar results could at least be given for the alternating gradient descent-ascent system since the set of critical points are unchanged and the alternating updates (or multiple unrolling steps of gradient ascent) approximate a form of timescale separation.

CHAPTER 3 APPENDIX

3.A Mathematical Preliminaries

We begin by reviewing some mathematical preliminaries needed for the technical details of the proofs. We also include some short technical lemmas from algebra that are used in the proofs.

3.A.1 Numerical and Quadratic Numerical Range.

The numerical range and quadratic numerical range of a block operator matrix are particularly useful for proving results about the spectrum of a block operator matrix as they are supersets of the spectrum (Tretter, 2008). Given a matrix $A \in \mathbb{R}^{d\times d}$, the numerical range is defined by

$$\mathcal{W}(A) = \{z \in \mathbb{C}^d : \ \langle Az, z\rangle, \ \|z\| = 1\},$$

and is a convex subset of $\mathbb{C}$. Define spaces $W_i = \{z \in \mathbb{C}^{d_i} : \ \|z\| = 1\}$ for each $i \in \{1, 2\}$. Consider a block operator

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

where $A_{ii} \in \mathbb{R}^{d_i\times d_i}$ and $A_{ij} \in \mathbb{R}^{d_i\times d_j}$ for each $i, j \in \{1, 2\}$. Given $v \in W_1$ and $w \in W_2$, let $A^{v,w} \in \mathbb{C}^{2\times 2}$ be defined by

$$A^{v,w} = \begin{bmatrix} \langle A_{11}v, v\rangle & \langle A_{12}w, v\rangle \\ \langle A_{21}v, w\rangle & \langle A_{22}w, w\rangle \end{bmatrix}.$$

The quadratic numerical range of A is defined by

$$\mathcal{W}^2(A) = \bigcup_{v\in W_1, w\in W_2} \text{spec}(A^{v,w})$$

where spec$(\cdot)$ denotes the spectrum of its argument.

The quadratic numerical range can be described as the set of solutions of the characteristic polynomial

$$\lambda^2 - \lambda(\langle A_{11}v, v\rangle + \langle A_{22}w, w\rangle) + \langle A_{11}v, v\rangle\langle A_{22}w, w\rangle - \langle A_{12}v, w\rangle\langle A_{21}w, v\rangle = 0 \tag{3.20}$$

for $v \in W_1$ and $w \in W_2$. We use the notation $\langle Av, w\rangle = \bar{v}^\top Aw$ to denote the inner product. Note that $\mathcal{W}^2(A)$ is a (potentially non-convex) subset of $\mathcal{W}(A)$ and contains spec(A).

3.A.2 Technical Lemmas

The following technical lemma is used in proving an upper bound on the spectral radius of the linearization of the discrete time update τ-`GDA` a requirement for obtaining the convergence rate results.

Lemma 3.1. *The function* $c(z) = (1-z)^{1/2} + \frac{z}{4} - (1-\frac{z}{2})^{1/2}$ *satisfies* $c(x) \leq 0$ *for all* $z \in [0,1]$.

Proof. Since $c(0) = 0$ and $c(1) = \frac{1}{4} - \frac{1}{\sqrt{2}} \leq 0$, we simply need to show that $c'(z) \leq 0$ on $(0,1)$ to get that $c(z)$ is a decreasing function on $[0,1]$, and hence negative on $[0,1]$. Indeed, $c'(z) = \frac{1}{4} + \frac{1}{2\sqrt{4-2z}} - \frac{1}{2\sqrt{1-z}} \leq 0$ since $(1-z)^{-1/2} - (4-2z)^{-1/2} \geq 1/2$ for all $z \in (0,1)$. □

The following technical lemma, due to Mustafa and Davidson (1994), is used in constructing the finite learning rate ratio.

Lemma 3.2 (Mustafa and Davidson 1994, Lemma 15)**.** *Let* $V, Z \in \mathbb{R}^{d_1 \times d_1}$, $W \in \mathbb{R}^{d_1 \times d_2}$ *and* $Y \in \mathbb{R}^{d_2 \times d_2}$. *If* V *and* $Y - XV^{-1}W$ *are non-singular, then*

$$\det \begin{pmatrix} V+Z & W \\ X & Y \end{pmatrix} = \det(V) \det(Y - XV^{-1}W) \det(I + V^{-1}(I + W(Y - XV^{-1}W)^{-1}XV^{-1})Z)$$

For completeness (and because there is a typo in the original manuscript), we provide the proof.

Proof. Suppose that V and $Y - XV^{-1}W$ are non-singular so that the partial Schur decomposition

$$\begin{bmatrix} V & W \\ X & Y \end{bmatrix} = \begin{bmatrix} V & 0 \\ X & Y - XV^{-1}W \end{bmatrix} \begin{bmatrix} I & V^{-1}W \\ 0 & I \end{bmatrix}$$

holds, and

$$\det \left(\begin{bmatrix} V & W \\ X & Y \end{bmatrix} \right) = \det(V) \det(Y - XV^{-1}W). \tag{3.21}$$

Further,

$$\begin{bmatrix} V & W \\ X & Y \end{bmatrix}^{-1} = \begin{bmatrix} I & -V^{-1}W \\ 0 & I \end{bmatrix} \begin{bmatrix} V^{-1} & 0 \\ -(Y - XV^{-1}W)^{-1}XV^{-1} & (Y - XV^{-1}W)^{-1} \end{bmatrix}.$$

Applying the determinant operator, we have that

$$\det \left(\begin{bmatrix} V+Z & W \\ X & Y \end{bmatrix} \right) = \det \left(\begin{bmatrix} V & W \\ X & Y \end{bmatrix} \right) \det \left(\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} + \begin{bmatrix} V & W \\ X & Y \end{bmatrix}^{-1} \begin{bmatrix} Z & 0 \\ 0 & 0 \end{bmatrix} \right) \tag{3.22}$$

so that

$$\det \left(\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} + \begin{bmatrix} V & W \\ X & Y \end{bmatrix}^{-1} \begin{bmatrix} Z & 0 \\ 0 & 0 \end{bmatrix} \right) = \det(V^{-1}(I + W(Y - XV^{-1}W)^{-1}XV^{-1})Z + I). \tag{3.23}$$

Combining (3.21) with (3.23) in (3.22) gives exactly the claimed result. □

The following lemma is Theorem 2 Lancaster and Tismenetsky (1985, Chap. 13.1). We use this lemma several times in the proofs of Theorem 3.3 and 3.4 so we include it here for ease of reference. For a given matrix A, $\upsilon_+(A)$, $\upsilon_-(A)$, and $\zeta(A)$ are the number of eigenvalues of the argument that have positive, negative and zero real parts, respectively.

Lemma 3.3. *Consider a matrix $A \in \mathbb{R}^{d\times d}$.*

(a) If P is a symmetric matrix such that $AP + PA^\top = Q$ where $Q = Q^\top \succ 0$, then P is nonsingular and P and A have the same inertia, meaning that

$$\upsilon_+(A) = \upsilon_+(P),\ \upsilon_-(A) = \upsilon_-(P),\ \zeta(A) = \zeta(P). \tag{3.24}$$

(b) On the other hand, if $\zeta(A) = 0$, then there exists a matrix $P = P^\top$ and a matrix $Q = Q^\top \succ 0$ such that $AP + PA^\top = Q$ and P and A have the same inertia ((3.24) *holds).*

3.B Proof of Theorem 3.3: Stability of τ–GDA

To prove Theorem 3.3 and Corollary 3.2, we introduce some techniques that are arguably new to the machine learning and artificial intelligence communities. The first is the notion of a guard map. A guard map can be used to provide a certificate of a particular behavior for a dynamical system as a parameter(s) varies. A critical point of a dynamical systems is known to be stable if the spectrum of the Jacobian at the critical point lies in the open left-half complex plane, denoted $\mathbb{C}^\circ_-$. Hence, we construct a guard map as a function of τ and show that it guards $\mathbb{C}^\circ_-$. Specifically we show that the existence of a $\tau^* \in (0,\infty)$ such that $\nu(\tau^*) = 0$ and $\nu(\tau) \neq 0$ for all $\tau \in (\tau^*, \infty)$ is equivalent to $\mathtt{S}_1(J(x^*)) \succ 0$ and $-D_2^2 f(x^*) \succ 0$ where

$$\mathtt{S}_1(J(x^*)) = \mathtt{S}_1(J_\tau(x^*)) = D_1^2 f(x^*) - D_{12} f(x^*)(D_2^2 f(x^*))^{-1} D_{21} f(x^*).$$

Towards this end, we need to introduced some notation as well as formal definitions for important concepts such as the guard map.

3.B.1 Notation and Preliminaries

Given a matrix $A \in \mathbb{R}^{d_1\times d_2}$, let $\mathrm{vec}(A) \in \mathbb{R}^{d_1 d_2}$ be the vectorization of A. We use the convention that rows are transposed and stacked in order. That is,

$$\mathrm{vec}: \begin{bmatrix} — & a_1 & — \\ & \vdots & \\ — & a_{d_1} & — \end{bmatrix} \mapsto \begin{bmatrix} a_1^\top \\ \vdots \\ a_{d_1}^\top \end{bmatrix}$$

Let $\otimes$ and $\oplus$ denote the Kronecker product and Kronecker sum respectively. Recall that $A \oplus B = A \otimes I + I \otimes B$. A less common operator, we define $\boxplus$ as an operator that generates an $\frac{1}{2}d(d+1) \times \frac{1}{2}d(d+1)$ matrix from a matrix $A \in \mathbb{R}^{d\times d}$ such that

$$A \boxplus A = H_d^+ (A \oplus A) H_d$$

where $H_d^+ = (H_d^\top H_d)^{-1} H_d^\top$ is the (left) pseudo-inverse of H_d, a full column rank duplication matrix. A duplication matrix $H_d \in \mathbb{R}^{d^2 \times d(d+1)/2}$ is a clever linear algebra tool for mapping a $\frac{d}{2}(d+1)$ vector to a d^2 vector generated by applying $\mathrm{vec}(\cdot)$ to a symmetric matrix and it is designed to respect the

vectorization map $\text{vec}(\cdot)$. In particular, if $\text{vech}(X)$ is the half-vectorization map of any symmetric matrix $X \in \mathbb{R}^{d\times d}$, then $\text{vec}(X) = H_d\text{vech}(X)$ and $\text{vech}(X) = H_d^+\text{vec}(X)$.

Given a square matrix A, let $\lambda^+_{\max}(A)$ be the largest positive real eigenvalue of A and if A does not have a positive real eigenvalue then it is zero.

Guardian maps. The use of guardian maps for studying stability of parameterized families of dynamical systems was arguably introduced by Saydy et al. (1990). Guardian or guard maps act as a certificate for a performance criteria such as stability.

Formally, let $\mathcal{X}$ be the set of all $d \times d$ real matrices or the set of all polynomials of degree d with real coefficients. Consider $\mathcal{S}$ an open subset of $\mathcal{X}$ with closure $\bar{\mathcal{S}}$ and boundary $\partial\mathcal{S}$.

Definition 3.3. *The map $\nu : \mathcal{X} \to \mathbb{C}$ is said to be a guardian map for $\mathcal{S}$ if for all $x \in \bar{\mathcal{S}}$, $\nu(x) = 0 \iff x \in \partial\mathcal{S}$.*

Consider an open subset Ω of the complex plane that is symmetric with respect to the real axis. Then, elements of $\mathcal{S}(\Omega) = \{A \in \mathbb{R}^{d\times d} : \text{spec}(A) \subset \Omega\}$ are said to be stable relative to Ω.

The following result gives a necessary and sufficient condition for stability of parameterized families of matrices relative to some open subset of the complex plane.

Proposition 3.5 (Proposition 1 (Saydy et al., 1990); Theorem 2 (Abed et al., 1990))**.** *Consider U to be a pathwise connected subset of $\mathbb{R}$ and $A(\tau) \in \mathbb{R}^{d\times d}$ a matrix which depends continuously on τ. Let $\mathcal{S}(\Omega)$ be guarded by the map ν. The family $\{A(\tau) : \tau \in U\}$ is stable relative to Ω if and only if (i) it is nominally meaning $A(\tau_1) \in \mathcal{S}(\Omega)$ for some $\tau_1 \in U$—and (ii) $\nu(A(\tau)) \neq 0$ for all $\tau \in U$.*

In proving Theorem 3.3, we define a guard map for the space of $d \times d$ Hurwitz stable matrices which is denoted by $\mathcal{S}(\mathbb{C}^\circ_-)$.

Lemma 3.4. *The map $\nu : A \mapsto \det(A \boxplus A)$ guards the set of non-singular $d \times d$ Hurwitz stable matrices $\mathcal{S}(\mathbb{C}^\circ_-)$.*

Proof. This follows from the following observation: for $A \in \mathbb{R}^{d\times d}$,

$$\text{vech}(AX + XA^\top) = H_d^+\text{vec}(AX + XA^\top) = H_d^+(A \oplus A)\text{vec}(X) = H_d^+(A \oplus A)H_d\text{vech}(X)$$

from which it can be shown that the eigenvalues of $A \boxplus A$ are $\lambda_i + \lambda_j$ for $1 \le j \le i \le d$ where λ_i for $i = 1, \ldots, d$ are the eigenvalues of A.

Indeed, let S be a non-singular matrix such that $S^{-1}AS = M$ where M is upper triangular with $\lambda_1, \ldots, \lambda_d$ on its diagonal. Observe that for any $d \times d$ matrix P, $H_dH_d^+(P \otimes P)H_d = (P \otimes P)H_d$ and $H_d^+(P \otimes P)H_dH_d^+ = H_d^+(P \otimes P)$. Hence, using properties of the Kronecker product (namely, that $(A_1 \otimes A_2)(B_1 \otimes B_2) = (A_1B_1 \otimes A_2B_2)$), we have that

$$H_d^+(S^{-1} \otimes S^{-1})H_dH_d^+(I \otimes A + A \otimes I)H_dH_d^+(S \otimes S)H_d = H_d^+(I \otimes M + M \otimes I)H_d$$

so that the spectrum of $H_d^+(I \otimes A + A \otimes I)H_d$ and $H_d^+(I \otimes M + M \otimes I)H_d$ coincide. Now, since M is upper triangular, $H_d^+(I \otimes M + M \otimes I)H_d$ is upper triangular with diagonal elements $\lambda_i + \lambda_j$ $(1 \le j \le i \le d)$ which can be verified by direct computation and using the definition of H_d. This implies that $\lambda_i + \lambda_j$ $(1 \le j \le i \le d)$ are exactly the eigenvalues of $H_d^+(I \otimes A + A \otimes I)H_d$. □

We note that there are several other guard maps for the space of Hurwtiz stable matrices including $\nu : A \mapsto \det(A \oplus A)$. To give some intuition for this map, it is fairly straightforward to see that the Kronecker sum $A \oplus A = A \otimes I + I \otimes A$ has spectrum $\{\lambda_j + \lambda_i\}$ where $\lambda_i, \lambda_j \in \text{spec}(A)$. The operator $A \boxplus A$ is simply a more computationally efficient expression of $A \oplus A$, and as such the eigenvalues of $A \boxplus A$ are those of $A \oplus A$ removing redundancies. We use $A \boxplus A$ specifically because of its computational advantages in computing τ^*.

3.B.2 Proof of Theorem 3.3

We first prove that if x^* is a differential Stackelberg equilibrium (that is, $\mathtt{S}_1(J_\tau(x^*)) \succ 0$ and $-D_2^2 f(x^*) \succ 0$), then there exists a finite $\tau^* \in (0, \infty)$ such that for all $\tau \in (\tau^*, \infty)$, x^* is locally exponentially stable for $\dot{x} = -\Lambda_\tau g(x)$ (that is, $\text{spec}(-J_\tau(x^*)) \subset \mathbb{C}_-^\circ$). Towards this end, we construct a guard map for the space of $d \times d$ Hurwtiz stable matrices and explicitly construct the τ^* using it.

Then we prove the other direction. That is, if there exists a finite $\tau^* \in (0, \infty)$ such that for all $\tau \in (\tau^*, \infty)$, x^* is exponentially stable for $\dot{x} = -\Lambda_\tau g(x)$, then x^* is a differential Stackelberg equilibrium. We prove this by contradiction.

3.B.2.1 Proof that if x^* is a differential Stackelberg then finite τ^* exists

For a critical point x^*, let

$$-J_\tau(x^*) = \begin{bmatrix} -D_1^2 f(x^*) & -D_{12} f(x^*) \\ \tau D_{12}^\top f(x^*) & \tau D_2^2 f(x^*) \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ -\tau A_{12}^\top & \tau A_{22} \end{bmatrix}$$

and define

$$\mathtt{S}_1 = \mathtt{S}_1(-J_\tau(x^*)) = A_{11} - A_{12} A_{22}^{-1} A_{12}^\top.$$

Note that this is equivalent to the first Schur complement of $-J(x^*)$ (i.e., when $\tau = 1$) since the τ and τ^{-1} cancel, and by assumption the first Schur complement of $-J(x^*)$ is positive definite. Suppose that x^* is a differential Stackelberg equilibrium so that $-\mathtt{S}_1 \succ 0$ and $-A_{22} \succ 0$.

Polynomial guard map with family of matrices parameterized by τ. By Lemma 3.4, $\nu : A \mapsto \det(A \boxplus A)$ is a guard map for $\mathcal{S}(\mathbb{C}_-^\circ)$. Indeed, using the fact that the determinant is the product of the eigenvalues of a matrix and the fact that $\text{spec}(A \boxplus A) = \{\lambda_i + \lambda_j, 1 \le i \le j \le d, \lambda_i, \lambda_j \in \text{spec}(A)\}$, we have that

$$\det(A \boxplus A) = \prod_{1 \le j \le i \le d} (\lambda_i + \lambda_j) = \prod_{1 \le i \le d} 2\text{Re}(\lambda_i)(4\text{Re}^2(\lambda_i) + 4\text{Im}^2(\lambda_i)) \prod_{\substack{1 < i < j < d: \\ \lambda_i \neq \bar{\lambda}_j}} (\lambda_i + \lambda_j).$$

Hence, consider $\bar{\mathcal{S}}(\mathbb{C}_-^\circ)$, $\det(A \boxplus A) = 0$ if and only if $A \boxplus A$ is singular if and only if A has a purely imaginary eigenvalue—that is, if and only if $A \in \partial\mathcal{S}(\mathbb{C}_-^\circ)$.[8] Now, consider the parameterized family of matrices $-J_\tau(x^*)$, parameterized by τ. By an abuse of notation, let $\nu(\tau) = \det(-J_\tau(x^*) \boxplus$

[8] Indeed, this holds since the only scenarios in which $\det(A \boxplus A) = 0$ are such that the eigenvalues of A do not lie in $\bar{\mathcal{S}}(\mathbb{C}_-^\circ)$.

$-J_\tau(x^*)$). If we consider the subset of this family of matrices that lies in $\mathcal{S}(\mathbb{C}_-^\circ)$ (this subset could a priori be empty thought we show it is not), then for any τ such that $-J_\tau(x^*)$ is in this subset, we have that $\nu(\tau) = 0$ if and only if $-J_\tau(x^*) \boxplus (-J_\tau(x^*))$ is singular if and only if $-J_\tau(x^*) \in \partial\mathcal{S}(\mathbb{C}_-^\circ)$. Hence, $\nu(\tau) = \det(-J_\tau(x^*) \boxplus -J_\tau(x^*))$ guards $\mathcal{S}(\mathbb{C}_-^\circ)$.

In particular, if we envision $-J_\tau(x^*)$ as the input to $\nu : A \mapsto \det(A \boxplus A)$ and simply vary τ (holding all the entries of $-J_\tau(x^*)$ otherwise fixed), then $\nu : \tau \mapsto \det(-J_\tau(x^*) \boxplus (-J_\tau(x^*)))$ can be thought of simply as a function of τ which guards the set of Hurwitz stable matrices via the reasoning describe above. Indeed, by slightly overloading the notation for ν,

$$\nu(\tau) := \nu_0 + \nu_1\tau + \cdots + \nu_{p-1}\tau^{p-1} + \nu_p\tau^p = \nu(-J_\tau(x^*))$$

Hence, for intuition, observe that as τ decreases (towards zero) stability is first lost when at least one eigenvalue of $-J_\tau(x^*)$ reaches the imaginary axis, at which point $\nu(\tau) = 0$.

There are two cases to consider:

Case 1: *$\nu(\tau)$ is an identically zero polynomial.* In this case, $-J_\tau(x^*)$ is in the interior of the complement of the set of Hurwitz stable matrices for all values of $\tau > 0$—that is, $-J_\tau(x^*) \in \text{int}(\mathcal{S}^{\mathsf{c}}(\mathbb{C}_-^\circ))$ for all $\tau \in \mathbb{R}_+ = (0, \infty)$.

Case 2: *$\nu(\tau)$ is not an identically zero polynomial.* In this case, $\nu(\tau)$ has finitely many zeros. If $\nu(\tau)$ has no positive real roots, then as τ varies in $\mathbb{R}_+$, $-J_\tau(x^*)$ does not cross $\partial\mathcal{S}(\mathbb{C}_-^\circ$—i.e., the boundary of the space of $d \times d$ Hurwitz stable matrices. Hence, $\{-J_\tau(x^*) : \ \tau \in \mathbb{R}_+\} \subset \mathcal{S}^{\mathsf{c}}(\mathbb{C}_-^\circ)$ or $\{-J_\tau(x^*) : \ \tau \in \mathbb{R}_+\} \subset \text{int}(\mathcal{S}^{\mathsf{c}}(\mathbb{C}_-^\circ))$. It suffices to check $-J_\tau(x^*) \in \mathcal{S}^{\mathsf{c}}(\mathbb{C}_-^\circ)$ or $-J_\tau(x^*) \in \text{int}(\mathcal{S}^{\mathsf{c}}(\mathbb{C}_-^\circ))$ for an arbitrary $\tau \in \mathbb{R}_+$.

On the other hand, if $\nu(\tau)$ has $\ell \geq 1$ real positive zeros, say $0 < \tau_1 < \cdots < \tau_\ell = \tau^*$, then by Proposition 3.5, $-J_\tau(x^*) \in \mathcal{S}(\mathbb{C}_-^\circ)$ for all $\tau > \tau^*$ if and only if $-J_\tau(x^*) \in \mathcal{S}(\mathbb{C}_-^\circ)$ for arbitrarily chosen $\tau > \tau^*$. We choose the largest positive root τ_ℓ because we are guaranteed that $\nu(\tau)$ stops changing sign for $\tau > \tau^*$. Further, the largest neighborhood in $\mathbb{R}_+$ for which $-J_\tau(x^*) \in \mathcal{S}(\mathbb{C}_-^\circ)$ is (τ_ℓ, ∞).

Recall that we have assumed that x^* is a differential Stackelberg equilibrium (that is, $\mathtt{S}_1 \succ 0$ and $-A_{22} \succ 0$). We will show next (by way of explicit construction of τ^*) that we are always in case 2.

Construction of τ^*. We note that there are more elegant, simpler constructions, but to our knowledge this construction gives the tightest bound on the range of τ for which $-J_\tau(x^*)$ is guaranteed to be Hurwitz stable. Recall that

$$-J_\tau(x^*) = \begin{bmatrix} -D_1^2 f(x^*) & -D_{12} f(x^*) \\ \tau D_{12}^\top f(x^*) & \tau D_2^2 f(x^*) \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ -\tau A_{12}^\top & \tau A_{22} \end{bmatrix}$$

and

$$\mathtt{S}_1 = A_{11} - A_{12}A_{22}^{-1}A_{12}^\top.$$

Let I_m denote the $m \times m$ identity matrix for some m.

Claim 3.1. *The finite learning rate ratio is $\tau^* = \lambda_{\max}^+(Q)$ where*

$$Q = 2\begin{bmatrix}(A_{12} \otimes A_{22}^{-1})H_{d_2} & (I_{d_1} \otimes A_{22}^{-1}A_{12}^\top)H_{d_1}\end{bmatrix}\begin{bmatrix}\bar{A}_{22}^{-1}H_{d_2}^+(A_{12}^\top \otimes I_{d_2}) \\ -\bar{\mathtt{S}}_1^{-1}H_{d_1}^+(\mathtt{S}_1 \otimes A_{12}A_{22}^{-1})\end{bmatrix} - (A_{11} \otimes A_{22}^{-1}) \tag{3.25}$$

with $\bar{A}_{22} = A_{22} \boxplus A_{22}$ *and* $\bar{\mathtt{S}}_1 = \mathtt{S}_1 \boxplus \mathtt{S}_1$.

Proof. Recall that $\nu(\tau) = \det(-J_\tau(x^*) \boxplus (-J_\tau(x^*)))$ is a guard map for $\mathcal{S}(\mathbb{C}^\circ_-)$.

We apply basic properties of the Kronecker product and sum as well as Schur's determinant formula to obtain a reduced form of the guard map. To this end, we have that

$$-J_\tau(x^*) \boxplus (-J_\tau(x^*)) = \begin{bmatrix} A_{11} \boxplus A_{11} & 2H_{d_1}^+(I_{d_1} \otimes A_{12}) & 0 \\ \tau(I_{d_1} \otimes (-A_{12}^\top))H_{d_1} & A_{11} \oplus \tau A_{22} & (A_{12} \otimes I_{d_2})H_{d_2} \\ 0 & 2\tau H_{d_2}^+(-A_{12}^\top \otimes I_{d_2}) & \tau(A_{22} \boxplus A_{22}) \end{bmatrix}$$

Now, we apply Schur's determinant formula to get that

$$\nu(\tau) = \tau^{d_2(d_2+1)/2} \det(A_{22} \boxplus A_{22}) \det\left(\begin{bmatrix} A_{11} \boxplus A_{11} & 2H_{d_1}^+(I_{d_1} \otimes A_{12}) \\ \tau(I_{d_1} \otimes (-A_{12}^\top))H_{d_1} & A_{11} \oplus \tau A_{22} + M_1 \end{bmatrix} \right) \tag{3.26}$$

where

$$M_1 = -2H_{d_2}^+(-A_{12}^\top \otimes I_{d_2})(A_{22} \boxplus A_{22})^{-1}(A_{12} \otimes I_{d_2})H_{d_2}$$

From here, we apply Lemma 3.2 to further reduce the guard map. First, note that

$$A_{11} \oplus \tau A_{22} = A_{11} \otimes I_{d_2} + I_{d_1} \otimes \tau A_{22}.$$

Let $V = I_{d_1} \otimes \tau A_{22}$, $Z = A_{11} \otimes I_{d_2} + M_1$, $Y = A_{11} \boxplus A_{11}$, $W = -\tau(I_{d_1} \otimes A_{12}^\top)H_{d_1}$, and $X = 2H_{d_1}^+(I_{d_1} \otimes A_{12})$. Using the two properties of the Kronecker product $(B_1 \otimes B_2)(B_3 \otimes B_4) = (B_1B_3 \otimes B_2B_4)$ and $(B_1 \otimes B_2)^{-1} = (B_1^{-1} \otimes B_2^{-1})$, we have that

$$\begin{aligned} Y - XV^{-1}W &= A_{11} \boxplus A_{11} + 2H_{d_1}^+(I_{d_1} \otimes A_{12})(I_{d_1} \otimes A_{22})^{-1}(I_{d_1} \otimes A_{12}^\top)H_{d_1} && (3.27) \\ &= A_{11} \boxplus A_{11} + 2H_{d_1}^+(I_{d_1} \otimes A_{12}A_{22}^{-1}A_{12}^\top)H_{d_1} && (3.28) \\ &= A_{11} \boxplus A_{11} + H_{d_1}^+((I_{d_1} \otimes A_{12}A_{22}^{-1}A_{12}^\top) + (A_{12}A_{22}^{-1}A_{12}^\top \otimes I_{d_1}))H_{d_1} && (3.29) \\ &= S_1 \boxplus S_1 && (3.30) \end{aligned}$$

where (3.29) holds since $H_{d_1}^+(I_{d_1} \otimes A_{12}A_{22}^{-1}A_{12}^\top)H_{d_1} = H_{d_1}^+(A_{12}A_{22}^{-1}A_{12}^\top \otimes I_{d_1})H_{d_1}$. Now, define $V^{-1} + V^{-1}W(Y - XV^{-1}W)^{-1}XV^{-1} = \tau^{-1}M_2$ where

$$M_2 = I_{d_1} \otimes A_{22}^{-1} - 2(I_{d_1} \otimes A_{22}^{-1}A_{12}^\top)H_{d_1}(\mathtt{S}_1 \boxplus \mathtt{S}_1)^{-1}H_{d_1}^+(I_{d_1} \otimes A_{12}A_{22}^{-1})$$

so that applying Lemma 3.2 we have

$$\nu(\tau) = \tau^{d_2(d_2+1)/2} \det(A_{22} \boxplus A_{22}) \det(\mathtt{S}_1 \boxplus \mathtt{S}_1) \det(I_{d_1} \otimes A_{22}) \det(\tau I_{d_1d_2} + M_2(A_{11} \otimes I_{d_2} + M_1)) \tag{3.31}$$

The assumptions that $\mathtt{S}_1 \succ 0$ and $-A_{22} \succ 0$ together imply that $\det(\mathtt{S}_1 \boxplus \mathtt{S}_1) \neq 0$ and $\det(I_{d_1} \otimes A_{22}) \neq 0$. Hence, $\nu(\tau) = 0$ if and only if $\det(\tau I_{d_1d_2} + M_2(A_{11} \otimes I_{d_2} + M_1)) = 0$ since $0 < \tau < \infty$. The determinant expression is exactly an eigenvalue problem.

Since by assumption the Schur complement of $J(x^*)$ and the individual Hessian $-D_2^2 f(x^*)$ are positive definite (that is, x^* is a differential Stackelberg equilibrium), Thus, the largest positive real

root of $\nu(\tau) = 0$ is

$$\tau^* = \lambda^+_{\max}(-M_2(A_{11} \otimes I_{d_2} + M_1))$$

where $\lambda^+_{\max}(\cdot)$ is the largest positive real eigenvalue of its argument if one exists and otherwise its zero. Using properties of the Kronecker product and duplication matrices, it can easily be seen that Q, as defined in (3.25), is equivalent to $-M_2(A_{11} \otimes I_{d_2} + M_1)$. □

The result of this claim concludes the proof that if x^* is a differential Stackelberg, then there exists a finite $\tau^* \in [0, \infty)$ such that for all $\tau \in (\tau^*, \infty)$, $\text{spec}(-J_\tau(x^*)) \subset \mathbb{C}^\circ_-$.

3.B.2.2 Proof that existence of finite τ^* implies that x^* is a differential Stackelberg

To begin, consider a critical point x^* such that $g(x^*) = 0$ and $\det(D_2^2 f(x^*)) \neq 0$. Then, $\det(-J_\tau(x^*)) = \tau^{d_2} \det(D_2^2 f(x^*)) \det(-\mathtt{S}_1(J(x^*)))$ so that $\det(-J_\tau(x^*)) = 0$ if and only if $\det(-\mathtt{S}_1(J(x^*))) = 0$ which implies $-J_\tau(x^*)$ is unstable for all $\tau \in (0, \infty)$ when $\det(-\mathtt{S}_1(J(x^*))) = 0$. As a result, we are left to consider when $\det(\mathtt{S}_1(J(x^*))) \neq 0$ for the remainder of the proof.

We proceed by arguing a contradiction. Let $-C \equiv -D_2^2 f(x^*)$ and $S_1 \equiv \mathtt{S}_1(J(x^*)) = D_1^2 f(x^*) - D_{12} f(x^*)(D_2^2 f(x^*))^{-1} D_{12}^\top f(x^*)$ have no zero eigenvalues—that is, $\det(S_1) \neq 0$ and $\det(C) \neq 0$.

The proof of this direction is argued by contradiction. Consider a critical point x^* (that is, where $g(x^*) = 0$ such that $-C \equiv -D_2^2 f(x^*)$ and $S_1 \equiv \mathtt{S}_1(J(x^*)) = D_1^2 f(x^*) - D_{12} f(x^*)(D_2^2 f(x^*))^{-1} D_{12}^\top f(x^*)$ have no zero eigenvalues—that is, $\det(S_1) \neq 0$ and $\det(C) \neq 0$.

Suppose that there exists a $\tau^* \in (0, \infty)$ such that for all $\tau \in (\tau^*, \infty)$, $\text{spec}(-J_\tau(x^*)) \subset \mathbb{C}^\circ_-$, yet x^* is not a differential Stackelberg equilibrium. That is, either $-S_1$ or C have at least one positive eigenvalue. Without loss of generality, let $-S_1$ have at least one positive eigenvalue.

Since $\det(S_1) \neq 0$ and $\det(C) \neq 0$, by Lemma 3.3.b, there exists non-singular Hermitian matrices P_1, P_2 and positive definite Hermitian matrices Q_1, Q_2 such that $-S_1 P_1 - P_1 S_1 = Q_1$ and $CP_2 + P_2 C = Q_2$. Further, $-S_1$ and P_1 have the same inertia, meaning

$$\upsilon_+(-S_1) = \upsilon_+(P_1),\ \upsilon_-(-S_1) = \upsilon_-(P_1),\ \zeta(-S_1) = \zeta(P_1)$$

where for a given matrix A, $\upsilon_+(A)$, $\upsilon_-(A)$, and $\zeta(A)$ are the number of eigenvalues of the argument that have positive, negative and zero real parts, respectively. Similarly, C and P_2 have the same inertia:

$$\upsilon_+(C) = \upsilon_+(P_2),\ \upsilon_-(C) = \upsilon_-(P_2),\ \zeta(C) = \zeta(P_2).$$

Since $-S_1$ has at least one strictly positive eigenvalue, $\upsilon_+(P_1) = \upsilon_+(-S_1) \geq 1$.

Define

$$P = \begin{bmatrix} I & L_0^\top \\ 0 & I \end{bmatrix} \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix} \begin{bmatrix} I & 0 \\ L_0 & I \end{bmatrix} \tag{3.32}$$

where $L_0 = (D_2^2 f(x^*))^{-1} D_{12}^\top f(x^*) = C D_{12}^\top f(x^*)$. Since P is congruent to $\text{blockdiag}(P_1, P_2)$, by Sylvester's law of inertia (Horn and Johnson, 2012, Thm. 4.5.8), P and $\text{blockdiag}(P_1, P_2)$ have the same inertia, meaning that $\upsilon_+(P) = \upsilon_+(\text{blockdiag}(P_1, P_2))$, $\upsilon_-(P) = \upsilon_-(\text{blockdiag}(P_1, P_2))$, and $\zeta(P) = \zeta(\text{blockdiag}(P_1, P_2))$. Consider the matrix equation $-PJ_\tau(x^*) - J_\tau^\top(x^*)P = Q_\tau$ for $-J_\tau(x^*)$ where

$$Q_\tau = \begin{bmatrix} I & L_0^\top \\ 0 & I \end{bmatrix} B_\tau \begin{bmatrix} I & 0 \\ L_0 & I \end{bmatrix}$$

with

$$B_\tau = \begin{bmatrix} Q_1 & P_1 D_{12} f(x^*) - S_1 L_0^\top P_2 \\ (P_1 D_{12} f(x^*) - S_1 L_0^\top P_2)^\top & P_2 L_0 D_{12} f(x^*) + (P_2 L_0 D_{12} f(x^*))^\top + \tau Q_2 \end{bmatrix}$$

which can be verified by straightforward calculations.

Observe that $Q_\tau \succ 0$ is equivalent to $B_\tau \succ 0$ and both matrices are symmetric so that $B_\tau \succ 0$ if and only if $Q_1 \succ 0$ and $\mathtt{S}_2(B_\tau) \succ 0$ where

$$\begin{aligned} \mathtt{S}_2(B_\tau) = {} & P_2 L_0 D_{12} f(x^*) + (P_2 L_0 D_{12} f(x^*))^\top + \tau Q_2 \\ & - (P_1 D_{12} f(x^*) - S_1 L_0^\top P_2)^\top Q_1^{-1} (P_1 D_{12} f(x^*) - S_1 L_0^\top P_2). \end{aligned}$$

Now, $\mathtt{S}_2(B_\tau)$ is also a real symmetric matrix, and hence, it is positive definite if and only if all its eigenvalues are positive. To determine the range of τ such that $\mathtt{S}_2(B_\tau)$ is positive definite, we can formulate an eigenvalue problem to determine the value of τ such that the matrix $\mathtt{S}_2(B_\tau)$ becomes singular. This is analogous to the guard map approach used in the proof in the previous subsection for the other direction of the proof, and in this case, we are varying τ from zero to infinity and finding the point such that for all larger τ, $\mathtt{S}_2(B_\tau)$ is positive definite. Intuitively, such an argument works since τ scales the positive definite matrix Q_2. Towards this end, consider the eigenvalue problem in τ given by

$$\begin{aligned} 0 = \det \Big(& \tau I - Q_2^{-1} \big((P_1 D_{12} f(x^*) - S_1 L_0^\top P_2)^\top Q_1^{-1} (P_1 D_{12} f(x^*) - S_1 L_0^\top P_2) \\ & - P_2 L_0 D_{12} f(x^*) - (P_2 L_1 D_{12} f(x^*))^\top \big) \Big). \end{aligned}$$

Let τ_0 be the maximum positive eigenvalue, and zero otherwise. Then, since eigenvalues vary continuously, for all $\tau \in (\tau_0, \infty)$, $Q_\tau \succ 0$ so that by Lemma 3.3.a we conclude that P and $-J_\tau(x^*)$ have the same inertia, but this contradicts the stability of $-J_\tau(x^*)$ for all $\tau \in (\tau^*, \infty)$ since $\upsilon_+(P) \geq 1$.

Remark on the tightness of τ^*. The construction of τ^* is *tight* in the following sense. While it is possible to construct multiple guard maps for a domain, all guard maps have the same positive real roots by definition (Saydy, 1996, Remark 2). Hence, independent of the guard map choice, we will get the same value of τ^*. Moreover, τ^* tells us exactly when the eigenvalues move into the open left-half complex plane $\mathbb{C}_-^\circ$ and remain there. Hence, this gives us the precise, tight lower bound on the value of τ^*.

3.C Proof of Theorem 3.4: Instability of τ–GDA

To begin, consider a critical point x^* defined by $g(x^*) = 0$ which is not a differential Stackelberg equilibria such that $\det(D_2^2 f(x^*)) \neq 0$. Then, $\det(-J_\tau(x^*)) = \tau^{d_2} \det(D_2^2 f(x^*)) \det(-\mathtt{S}_1(J(x^*)))$ so that $\det(-J_\tau(x^*)) = 0$ if and only if $\det(-\mathtt{S}_1(J(x^*))) = 0$ which implies $-J_\tau(x^*)$ is unstable for all $\tau \in (\tau_0, \infty)$ where $\tau_0 = 0$ when $\det(-\mathtt{S}_1(J(x^*))) = 0$. We now consider when $\det(\mathtt{S}_1(J(x^*))) \neq 0$ for the remainder of the proof.

Let x^* be a stable critical point of 1-GDA (without loss of generality) which is not a differential Stackelberg equilibrium. Without loss of generality, suppose that $\mathtt{S}_1(-J(x^*))$ has at least one strictly positive eigenvalue. Note that both $\mathtt{S}_1(-J(x^*))$ and $-D_2^2 f(x^*)$ are symmetric matrices and

hence, have purely real eigenvalues.

Since both $\mathtt{S}_1(-J(x^*))$ and $D_2^2 f(x^*)$ have no zero valued eigenvalues, by Lemma 3.3.b, there exists non-singular Hermitian matrices P_1, P_2 and positive definite Hermitian matrices Q_1, Q_2 such that $\mathtt{S}_1(-J(x^*))P_1 + P_1\mathtt{S}_1(-J(x^*)) = Q_1$ and $D_2^2 f(x^*)P_2 + P_2 D_2^2 f(x^*) = Q_2$. Further, $\mathtt{S}_1(-J(x^*))$ and P_1 have the same inertia, meaning

$$\upsilon_+(\mathtt{S}_1(-J(x^*))) = \upsilon_+(P_1),\ \upsilon_-(\mathtt{S}_1(-J(x^*))) = \upsilon_-(P_1),\ \zeta(\mathtt{S}_1(-J(x^*))) = \zeta(P_1)$$

where for a given matrix A, $\upsilon_+(A)$, $\upsilon_-(A)$, and $\zeta(A)$ are the number of eigenvalues of the argument that have positive, negative and zero real parts, respectively. Similarly, $D_2^2 f(x^*)$ and P_2 have the same inertia:

$$\upsilon_+(D_2^2 f(x^*)) = \upsilon_+(P_2),\ \upsilon_-(D_2^2 f(x^*)) = \upsilon_-(P_2),\ \zeta(D_2^2 f(x^*)) = \zeta(P_2).$$

Recall that we assumed $\mathtt{S}_1(-J(x^*))$ has at least one eigenvalue with strictly positive real part. Hence, $\upsilon_+(P_1) = \upsilon_+(\mathtt{S}_1(-J(x^*))) \geq 1$.

Define

$$P = \begin{bmatrix} I & L_0^\top \\ 0 & I \end{bmatrix} \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix} \begin{bmatrix} I & 0 \\ L_0 & I \end{bmatrix}$$

where $L_0 = (D_2^2 f(x^*))^{-1} D_{12}^\top f(x^*)$. Since P is congruent to blockdiag(P_1, P_2), by Sylvester's law of inertia (Horn and Johnson, 2012, Thm. 4.5.8), P and blockdiag(P_1, P_2) have the same inertia, meaning that $\upsilon_+(P) = \upsilon_+(\text{blockdiag}(P_1, P_2))$, $\upsilon_-(P) = \upsilon_-(\text{blockdiag}(P_1, P_2))$, and $\zeta(P) = \zeta(\text{blockdiag}(P_1, P_2))$. Consider now the Lyapunov equation $-PJ_\tau(x^*) - J_\tau^\top(x^*)P = Q_\tau$ for $-J_\tau(x^*)$ where

$$Q_\tau = \begin{bmatrix} I & L_0^\top \\ 0 & I \end{bmatrix} B_\tau \begin{bmatrix} I & 0 \\ L_0 & I \end{bmatrix}$$

with

$$B_\tau = \begin{bmatrix} Q_1 & P_1 D_{12} f(x^*) + \mathtt{S}_1(-J(x^*))L_0^\top P_2 \\ (P_1 D_{12} f(x^*) + \mathtt{S}_1(-J(x^*))L_0^\top P_2)^\top & P_2 L_0 D_{12} f(x^*) + (P_2 L_0 D_{12} f(x^*))^\top + \tau Q_2 \end{bmatrix}$$

which can be verified by straightforward calculations.

Since $\upsilon_+(P_1) \geq 1$, we have that $\upsilon_+(P) \geq 1$. Now, we find the value of τ_0 such that for all $\tau > \tau_0$, $Q_\tau \succ 0$ so that, in turn, we can apply Lemma 3.3.a, to conclude that spec$(-J_\tau(x^*)) \not\subset \mathbb{C}_-^\circ$. Indeed, observe that $Q_\tau \succ 0$ is equivalent to $B_\tau \succ 0$ and both matrices are symmetric so that $B_\tau \succ 0$ if and only if $Q_1 \succ 0$ and $\mathtt{S}_2(B_\tau) \succ 0$ where

$$\begin{aligned} \mathtt{S}_2(B_\tau) = {} & P_2 L_1 D_{12} f(x^*) + (P_2 L_1 D_{12} f(x^*))^\top + \tau Q_2 \\ & - (P_1 D_{12} f(x^*) + \mathtt{S}_1(-J(x^*))L_0^\top P_2)^\top Q_1^{-1} (P_1 D_{12} f(x^*) + \mathtt{S}_1(-J(x^*))L_0^\top P_2). \end{aligned}$$

Now, $\mathtt{S}_2(B_\tau)$ is also a real symmetric matrix, and hence, it is positive definite if and only if all its eigenvalues are positive. To determine the range of τ for which $Q_\tau > 0$, we simply need to solve

the eigenvalue problem

$$0 = \det(\tau I - Q_2^{-1}((P_1 D_{12} f(x^*) + \mathtt{S}_1(-J(x^*))L_0^\top P_2)^\top Q_1^{-1}(P_1 D_{12} f(x^*) + \mathtt{S}_1(-J(x^*))L_0^\top P_2) - P_2 L_0 D_{12} f(x^*) - (P_2 L_0 D_{12} f(x^*))^\top)).$$

and extract the maximum eigenvalue, namely,

$$\tau_0 = \lambda_{\max}(Q_2^{-1}((P_1 D_{12} f(x^*) + \mathtt{S}_1(-J(x^*))L_0^\top P_2)^\top Q_1^{-1}(P_1 D_{12} f(x^*) + \mathtt{S}_1(-J(x^*))L_0^\top P_2) - P_2 L_0 D_{12} f(x^*) - (P_2 L_0 D_{12} f(x^*))^\top)).$$

Hence, as noted previously, by Lemma 3.3.a, we conclude that for all $\tau \in (\tau_0, \infty)$, $\mathrm{spec}(-J_\tau(x^*)) \not\subset \mathbb{C}_-^\circ$. This concludes the proof.

Additional context/intuition for the proof approach. To provide some context for the proof approach, we remark that it follows the same idea as the proof of Theorem 3.3 in Appendix 3.B.2.2. Indeed, to determine the range of τ such that $\mathtt{S}_2(B_\tau)$ is positive definite, we can formulate an eigenvalue problem to determine the value of τ such that the matrix $\mathtt{S}_2(B_\tau)$ becomes singular. We vary τ from zero to infinity in order to find the point such that for all larger τ, $\mathtt{S}_2(B_\tau)$ is positive definite. Intuitively, such an argument works since τ scales the positive definite matrix Q_2.

3.D Proof of Theorem 3.5: Stability of τ–GDA in Regularized GANs

As in Mescheder et al. (2018), we only apply the regularization to the discriminator. In the following proof, we use $\nabla_x(\cdot)$ to denote the partial gradient with respect to x of the argument $(\cdot)$ when the argument is the discriminator $\mathrm{D}(\cdot;\omega)$ in order prevent any confusion between the notation $D(\cdot)$ which we use elsewhere for derivatives.

To prove the first part of this result, we following similar arguments to Theorem 4.1 of (Mescheder et al., 2018). To prove the second part, we leverage the concept of the quadratic numerical range. For both components of the proof, we will use the following form of the Jacobian of the regularized game. Indeed, first observe that the structural form of $J_{(\tau,\mu)}(x^*)$ is

$$J_{(\tau,\mu)}(x^*) = \begin{bmatrix} 0 & B \\ -\tau B^\top & \tau(C + \mu R) \end{bmatrix} \tag{3.33}$$

where $B = D_{12} f(x^*)$, $C = -D_2^2 f(x^*)$ and $R = D_2^2 R_j(x^*)$ where R_j is either gradient penalty indexed by $j = 1, 2$.[9] This follows from Assumption 3.1-a., which implies that $\mathrm{D}(x;\omega^*) = 0$ in some neighborhood of $\mathrm{supp}(p_{\mathcal{D}})$ and hence, $\nabla_x \mathrm{D}(x;\omega^*) = 0$ and $\nabla_x^2 \mathrm{D}(x;\omega^*) = 0$ for $x \in \mathrm{supp}(p_{\mathcal{D}})$. In turn, we have that $D_1^2 f(x^*) = 0$.

Proof that $x^* = (\theta^*, \omega^*)$ is a differential Stackelberg equilibrium. For any fixed $\mu \in (0, \infty)$, then we first observe that x^* is also a critical point of the unregularized dynamics. Indeed, by Assumption 3.1-a., $\mathrm{D}(x;\omega^*) = 0$ in some neighborhood of $\mathrm{supp}(p_{\mathcal{D}})$ and hence, $\nabla_x \mathrm{D}(x;\omega^*) = 0$ and $\nabla_x^2 \mathrm{D}(x;\omega^*) = 0$ for $x \in \mathrm{supp}(p_{\mathcal{D}})$. Further, $D_2 R_j(\theta,\omega) = \mu \mathbb{E}_{p_i(x)}[D_2(\nabla_x \mathrm{D}(x;\omega))\nabla_x \mathrm{D}(x,\omega)]$

[9] Mescheder et al. (2018) imply that their results hold for a convex combination of the two gradient penalties, which would in turn imply our results will hold in this case. However, we have not included the details here.

for $j = 1, 2$ where $p_1(x) = p_{\mathcal{D}}(x)$ and $p_2(x) = p_\theta(x)$. Thus, using the above observation that $\nabla_x \mathrm{D}(x; \omega^*) = 0$, we have that $D_2 R_i(\theta^*, \omega^*) = 0$ for $i = 1, 2$ meaning that the derivative of the regularizer with respect to ω is zero at $x^* = (\theta^*, \omega^*)$ which in turn implies that $D_1 f(x^*) = 0$ and $-D_2 f(x^*) = 0$. Hence, x^* is a critical point of the unregularized dynamics as claimed. Further, $C + \mu R \succ 0$ which follows from Lemma D.5 in (Mescheder et al., 2018). From Lemma D.6 in (Mescheder et al., 2018), due to Assumption 3.1-c., if $v \neq 0$ and $v \notin T_{\theta^*}\mathcal{M}_{\mathrm{G}}$, then $Bv \neq 0$ which implies that B can only be rank deficient on $T_{\theta^*}\mathcal{M}_{\mathrm{G}}$. Using this fact along with the structure of the Jacobian as in (3.33), we have that the Schur complement of $J_{(\tau,\mu)}(x^*)$ is equal to $B^\top (C+\mu R)^{-1} B \succ 0$ since $C + \mu R \succ 0$. Hence, $x^* = (\theta^*, \omega^*)$ is a differential Stackelberg equilibrium.

Proof of stability. Examining (3.33), it is straightforward to see that the quadratic numerical range $\mathcal{W}^2(J_{(\tau,\mu)})$ has eigenvalues of the form

$$\lambda_{\tau,\mu} = \tfrac{1}{2}(\tau(c + \mu r)) \pm \tfrac{1}{2}\sqrt{(-\tau(c + \mu r))^2 - 4\tau|b|^2}$$

where $b = \langle D_{12} f(x^*)v, w\rangle$, $c = \langle -D_2^2 f(x^*)w, w\rangle$ and $r = \langle D_2^2 R_i(x^*)w, w\rangle$ for vectors $v \in W_1 \cap (T_{\theta^*}\mathcal{M}_{\mathrm{G}})^\perp$ and $w \in W_2 \cap (T_{\omega^*}\mathcal{M}_{\mathrm{D}})^\perp$ where $U^\perp$ denotes the orthogonal complement of U. We claim that for any value of $\mu \in (0, \infty)$ and any $\tau \in (0, \infty)$, $\mathrm{Re}(\lambda_{\tau,\mu}) > 0$. Indeed, we argue this by considering the two possible cases: (1) $(\tau(c + \mu r))^2 \leq 4|b|^2\tau$ or (2) $(\tau(c + \mu r))^2 > 4\tau|b|^2$.

- **Case 1**: Suppose that $(\tau(c + \mu r))^2 \leq 4|b|^2\tau$. Then, $\mathrm{Re}(\lambda_{\tau,\mu}) = \frac{1}{2}(\tau(c + \mu r)) > 0$ trivially since $c + \mu r > 0$.

- **Case 2**: Suppose that $(\tau(c + \mu r))^2 > 4\tau|b|^2$. In this case, we want to ensure that

  $$\mathrm{Re}(\lambda_\tau) > \tfrac{1}{2}(\tau(c + \mu r)) - \tfrac{1}{2}\sqrt{(-\tau(c + \mu r))^2 - 4\tau|b|^2} > 0.$$

  which holds since

  $$(\tau(c + \mu r))^2 > (-\tau(c + \mu r))^2 - 4\tau|b|^2 \iff 0 > -4\tau|b|^2$$

This concludes the proof.

3.E Proofs of τ–GDA Convergence Results: Theorem 3.6, Corollary 3.2, and Corollary 3.3

This section contains the proofs of Theorem 3.6 (convergence rates), Corollary 3.2 (stability of the discrete-time system around differential Stackelberg equilibrium) and Corollary 3.3 (finite-time convergence to differential Stackelberg equilibrium).

3.E.1 Proof of Theorem 3.6

We begin by stating and proving a pair of lemmas that will be invoked to prove Theorem 3.6.

Lemma 3.5. *Consider a zero-sum game $(f_1, f_2) = (f, -f)$ defined by $f \in C^q(\mathcal{X}, \mathbb{R})$ for some $q \geq 2$. Suppose that x^* is a differential Stackelberg equilibrium and that given $\tau > 0$, $\text{spec}(-J_\tau(x^*)) \subset \mathbb{C}_-^\circ$. Let $\gamma = \min_{\lambda \in \text{spec}(J_\tau(x^*))} 2\text{Re}(\lambda)/|\lambda|^2$. For any $\gamma_1 \in (0, \gamma)$, τ-GDA converges locally asymptotically.*

Proof. Suppose that x^* is a differential Stackelberg or Nash equilibrium and that $0 < \tau < \infty$ is such that $\text{spec}(-J_\tau(x^*)) \subset \mathbb{C}_-^\circ$. For the discrete time dynamical system $x_{k+1} = x_k - \gamma_1 \Lambda_\tau g(x_k)$, it is well known that if γ_1 is chosen such that $\rho(I - \gamma_1 J_\tau(x^*)) < 1$, then x_k locally (exponentially) converges to x^* (Ortega and Rheinboldt, 1970). With this in mind, we formulate an optimization problem to find the upper bound γ on the learning rate γ_1 such that for all $\gamma_1 \in (0, \gamma)$, the spectral radius of the local linearization of the discrete time map is a contraction which is precisely $\rho(I - \gamma_1 J_\tau(x^*)) < 1$. The optimization problem is given by

$$\gamma = \min_{\gamma > 0} \left\{ \gamma : \max_{\lambda \in \text{spec}(J_\tau(x^*))} |1 - \gamma\lambda| \leq 1 \right\}. \tag{3.34}$$

The intuition is as follows. The inner maximization problem is over a finite set $\text{spec}(J_\tau(x^*)) = \{\lambda_1, \ldots, \lambda_d\}$ where $J_\tau(x^*) \in \mathbb{R}^{d \times d}$. As γ increases away from zero, each $|1 - \gamma\lambda_i|$ shrinks in magnitude. The last λ_i such that $1 - \gamma\lambda_i$ hits the boundary of the unit circle in the complex plane (that is, $|1 - \gamma\lambda_i| = 1$) gives us the optimal value of γ and the element of $\text{spec}(J_\tau(x^*))$ that achieves it. Examining the constraint, we have that for each λ_i, $\gamma(\gamma|\lambda_i|^2 - 2\text{Re}(\lambda_i)) \leq 0$ for any $\gamma > 0$. As noted this constraint will be tight for one of the λ, in which case $\gamma = 2\text{Re}(\lambda)/|\lambda|^2$ since $\gamma > 0$. Hence, by selecting $\gamma = \min_{\lambda \in \text{spec}(J_\tau(x^*))} 2\text{Re}(\lambda)/|\lambda|^2$, we have that $|1 - \gamma_1\lambda| < 1$ for all $\lambda \in \text{spec}(J_\tau(x^*))$ and any $\gamma_1 \in (0, \gamma)$.

To see this is the case, let

$$\gamma = \min_{\lambda \in \text{spec}(J_\tau(x^*))} 2\text{Re}(\lambda)/|\lambda|^2$$

and

$$\lambda_{\mathtt{m}} = \arg \min_{\lambda \in \text{spec}(J_\tau)} 2\text{Re}(\lambda)/|\lambda|^2.$$

Using the expression for γ, we have that

$$1 - 2\gamma\text{Re}(\lambda) + \gamma^2(\text{Re}(\lambda)^2 + \text{Im}(\lambda)^2) = 1 - 2\frac{2\text{Re}(\lambda_{\mathtt{m}})}{|\lambda_{\mathtt{m}}|^2}\text{Re}(\lambda) + \left(\frac{2\text{Re}(\lambda_{\mathtt{m}})}{|\lambda_{\mathtt{m}}|^2}\right)^2 |\lambda|^2.$$

Now, using the fact that $\text{Re}(\lambda)/|\lambda|^2 > \text{Re}(\lambda_{\mathtt{m}})/|\lambda_{\mathtt{m}}|^2$, we have

$$\begin{aligned} 1 - 4\frac{\text{Re}(\lambda_{\mathtt{m}})}{|\lambda_{\mathtt{m}}|^2}\text{Re}(\lambda) + \left(\frac{2\text{Re}(\lambda_{\mathtt{m}})}{|\lambda_{\mathtt{m}}|^2}\right)^2 |\lambda|^2 &\leq 1 - 2\frac{2\text{Re}(\lambda_{\mathtt{m}})}{|\lambda_{\mathtt{m}}|^2}\text{Re}(\lambda) + \left(\frac{2\text{Re}(\lambda_{\mathtt{m}})}{|\lambda_{\mathtt{m}}|^2}\right)^2 \frac{|\lambda_{\mathtt{m}}|^2\text{Re}(\lambda)}{\text{Re}(\lambda_{\mathtt{m}})} \\ &= 1 - 4\frac{\text{Re}(\lambda_{\mathtt{m}})}{|\lambda_{\mathtt{m}}|^2}\text{Re}(\lambda) + 4\frac{\text{Re}(\lambda_{\mathtt{m}})}{|\lambda_{\mathtt{m}}|^2}\text{Re}(\lambda) \\ &= 1 \end{aligned}$$

as claimed. From this argument, it is clear that for any $\gamma_1 \in (0, \gamma)$, $|1 - \gamma_1\lambda| < 1$ for all $\lambda \in \text{spec}(J_\tau(x^*))$.

Now, consider any $\alpha \in (0, \gamma)$ and let $\beta = (2\text{Re}(\lambda_{\mathtt{m}}) - \alpha|\lambda_{\mathtt{m}}|^2)^{-1}$. Observe that $\gamma_1 = \gamma - \alpha$ so

that $\gamma_1 \in (0, \gamma)$. Hence,

$$\begin{aligned}
|1-(\gamma-\alpha)\lambda_{\mathtt{m}}|^2 &= \left(1-\left(\frac{2\mathrm{Re}(\lambda_{\mathtt{m}})}{|\lambda_{\mathtt{m}}|^2}-\alpha\right)\mathrm{Re}(\lambda_{\mathtt{m}})\right)^2 + \left(\frac{2\mathrm{Re}(\lambda_{\mathtt{m}})}{|\lambda_{\mathtt{m}}|^2}-\alpha\right)^2 \mathrm{Im}(\lambda_{\mathtt{m}})^2 \\
&= 1 - 4\frac{\mathrm{Re}(\lambda_{\mathtt{m}})^2}{|\lambda_{\mathtt{m}}|^2} + 2\alpha\mathrm{Re}(\lambda_{\mathtt{m}}) + 4\frac{\mathrm{Re}(\lambda_{\mathtt{m}})^2}{|\lambda_{\mathtt{m}}|^2} - 4\alpha\mathrm{Re}(\lambda_{\mathtt{m}}) + \alpha^2|\lambda_{\mathtt{m}}|^2 \\
&= 1 - 2\alpha\mathrm{Re}(\lambda_{\mathtt{m}}) + \alpha^2|\lambda_{\mathtt{m}}|^2 \\
&= 1 - \frac{\alpha}{\beta}
\end{aligned}$$

so that

$$\rho(I-\gamma_1 J_\tau(x^*)) < \left(1-\frac{\alpha}{\beta}\right)^{1/2}.$$

Hence, the $\rho(I-\gamma_1 J_\tau(x^*)) < 1$ so that x^* is asymptotically stable (Argyros, 1999; Ortega and Rheinboldt, 1970). □

Lemma 3.6. *Consider a zero-sum game $(f_1, f_2) = (f, -f)$ defined by $f \in C^q(\mathcal{X}, \mathbb{R})$ for some $q \geq 2$. Suppose that x^* is a differential Stackelberg equilibrium and that given τ, $\mathrm{spec}(-J_\tau(x^*)) \subset \mathbb{C}_-^\circ$. Let $\gamma = \min_{\lambda \in \mathrm{spec}(J_\tau(x^*))} 2\mathrm{Re}(\lambda)/|\lambda|^2$, and $\lambda_{\mathtt{m}} = \arg\min_{\lambda \in \mathrm{spec}(J_\tau(x^*))} 2\mathrm{Re}(\lambda)/|\lambda|^2$. For any $\alpha \in (0, \gamma)$, τ-GDA with learning rate $\gamma_1 = \gamma - \alpha$ converges locally asymptotically at a rate of $O((1-\frac{\alpha}{4\beta})^{k/2})$ where $\beta = (2\mathrm{Re}(\lambda_{\mathtt{m}}) - \alpha|\lambda_{\mathtt{m}}|^2)^{-1}$.*

Proof. To prove this lemma, we build directly on the conclusion of the proof of Lemma 3.5. Indeed, since

$$\rho(I-\gamma_1 J_\tau(x^*)) < \left(1-\frac{\alpha}{\beta}\right)^{1/2},$$

given $\varepsilon = \frac{\alpha}{4\beta} > 0$ there exists a norm $\|\cdot\|$ (cf. Lemma 5.6.10 in Horn and Johnson (2012))[10] such that

$$\|I-\gamma_1 J_\tau(x^*)\| \leq \left(1-\frac{\alpha}{\beta}\right)^{1/2} + \frac{\alpha}{4\beta} \leq \left(1-\frac{\alpha}{2\beta}\right)^{1/2}$$

where the last inequality holds by Lemma 3.1. Taking the Taylor expansion of $I-\gamma_1 g_\tau(x)$ around x^*, we have

$$I-\gamma_1 g_\tau(x) = (I-\gamma_1 g_\tau(x^*)) + (I-\gamma_1 J_\tau(x^*))(x-x^*) + R_2(x-x^*)$$

where $R_2(x-x^*)$ is the remainder term satisfying $R_2(x-x^*) = o(\|x-x^*\|)$ as $x \to x^*$.[11] This

[10]The norm that exists can easily be constructed as essentially a weighted induced 1-norm. Note that the norm construction is not unique. The proof in Horn and Johnson (2012) is by construction and the construction of this norm can be found there.

[11]The notation $R_2(x-x^*) = o(\|x-x^*\|)$ as $x \to x^*$ means $\lim_{x\to x^*} \|R_2(x-x^*)\|/\|x-x^*\| = 0$.

implies that there is a $\delta > 0$ such that $\|R_2(x - x^*)\| \leq \frac{\alpha}{8\beta}\|x - x^*\|$ whenever $\|x - x^*\| < \delta$. Hence,

$$\begin{aligned}\|I - \gamma_1 g_\tau(x) - (I - \gamma_1 g_\tau(x^*))\| &\leq \left(\|I - \gamma_1 J_\tau(x^*)\| + \frac{\alpha}{4\beta}\right)\|x - x^*\| \\ &\leq \left(\left(1 - \frac{\alpha}{2\beta}\right)^{1/2} + \frac{\alpha}{8\beta}\right)\|x - x^*\| \\ &\leq \left(1 - \frac{\alpha}{4\beta}\right)^{1/2}\|x - x^*\|\end{aligned}$$

where the last inequality holds again by Lemma 3.1. Hence,

$$\|x_k - x^*\| \leq \left(1 - \frac{\alpha}{4\beta}\right)^{k/2}\|x_0 - x_*\| \tag{3.35}$$

whenever $\|x_0 - x^*\| < \delta$ which verifies the claimed convergence rate. □

Proof of Theorem 3.6. To prove Theorem 3.6, we apply Theorem 3.3 to construct τ^* via the guard map $\nu(\tau) = \det(-J_\tau(x^*) \boxplus -J_\tau(x^*))$ such that for all $\tau \in (\tau^*, \infty)$, $\text{spec}(J_\tau(x^*)) \subset \mathbb{C}_+^\circ$. This guarantees that $\text{spec}(-J_\tau(x^*)) \subset \mathbb{C}_-^\circ$ for any $\tau \in (\tau^*, \infty)$ and hence the nonlinear dynamical system

$$\dot{x} = -\Lambda_\tau g(x)$$

is locally asymptotically (in fact, exponentially) stable by the Hartman-Grobman (Sastry, 1999). Therefore, for any $\tau \in (\tau^*, \infty)$, by Lemma 3.6, τ-GDA converges with a rate of $O((1 - \frac{\alpha}{4\beta})^{k/2})$. Finally, $\tau^* = 0$ by Proposition 3.1 if x^* is a differential Nash equilibrium. This concludes the proof.

3.E.2 Proof of Corollary 3.2

Suppose that x^* is a differential Stackelberg equilibrium so that by Theorem 3.3, there exists a $\tau^* \in (0, \infty)$ such that $\text{spec}(-J_\tau(x^*)) \subset \mathbb{C}_-^\circ$ for all $\tau \in (\tau^*, \infty)$. Now that we have a guarantee that $-J_\tau(x^*)$ is Hurwitz stable for any $\tau \in (\tau^*, \infty)$, we apply Hartman-Grobman to get that the nonlinear system $\dot{x} = -\Lambda_\tau g(x)$ is stable in a neighborhood of x^*. Fix any $\tau \in (\tau^*, \infty)$ and let $\gamma = \arg\min_{\lambda \in \text{spec}(J_\tau(x^*))} 2\text{Re}(\lambda)/|\lambda|^2$. Then, applying Lemma 3.5, for any $\gamma_1 \in (0, \gamma)$, τ-GDA converges locally asymptotically to x^*.

On the other hand, suppose that there exists a $\tau^* \in (0, \infty)$ such that $\text{spec}(-J_\tau(x^*)) \subset \mathbb{C}_-^\circ$ for all $\tau \in (\tau^*, \infty)$. Then by Theorem 3.3, if x^* is a differential Stackelberg equilibrium. Furthermore, since $\text{spec}(-J_\tau(x^*)) \subset \mathbb{C}_-^\circ$ for all $\tau \in (\tau^*, \infty)$, if we let $\gamma = \arg\min_{\lambda \in \text{spec}(J_\tau(x^*))} 2\text{Re}(\lambda)/|\lambda|^2$, then by Lemma 3.5 τ-GDA converges locally asymptotically to x^* for any choice of $\gamma_1 \in (0, \gamma)$.

3.E.3 Proof of Corollary 3.3

Let $\|\cdot\|$ be the norm that exists (via construction a la Horn and Johnson (2012, Lem. 5.6.10)) in the proof of Lemma 3.6 which is given in Appendix 3.E. Following standard arguments, (3.35) in the proof of Lemma 3.6 implies a finite time convergence guarantee. Indeed, let $\varepsilon > 0$ be given.

Since $0 < \frac{\alpha}{4\beta} < 1$ we have that $(1 - \alpha/(4\beta))^k < \exp(-k\alpha/(4\beta))$. Hence,

$$\|x_k - x^*\| \leq \exp(-k\alpha/(4\beta))\|x_0 - x^*\|.$$

In turn, this implies that $x_k \in B_\varepsilon(x^*)$, meaning that x_k is a ε-differential Stackelberg equilibrium for all $k \geq \lceil \frac{4\beta}{\alpha} \log(\|x_0 - x^*\|/\varepsilon) \rceil$ whenever $\|x_0 - x^*\| < \delta$.

Now, given that $f_i \in C^q(\mathcal{X}, \mathbb{R})$ for $r \geq 2$, $I - \gamma_1 J_\tau(x)$ is locally Lipschitz with constant L so that we can find an explicit expression for δ in terms of L. Indeed, recall that $R_2(x - x^*) = o(\|x - x^*\|)$ as $x \to x^*$ which means $\lim_{x \to x^*} \|R_2(x - x^*)\|/\|x - x^*\| = 0$ so that

$$\|R_2(x - x^*)\| \leq \int_0^1 \|I - \gamma_1 J_\tau(x^* + \eta(x - x^*)) - (I - \gamma_1 J_\tau(x^*))\| \|x - x^*\| \, d\eta \leq \frac{L}{2}\|x - x^*\|^2$$

Observing that

$$\|R_2(x - x^*)\| \leq \frac{L}{2}\|x - x^*\|^2 = \frac{L}{2}\|x - x^*\|\|x - x^*\|,$$

we have that the $\delta > 0$ such that $\|R_2(x - x^*)\| \leq \alpha/(8\beta)\|x - x^*\|$ is $\delta = \alpha/(4L\beta)$.

Chapter 4

Nonconvex Zero-Sum Games: Global Convergence Guarantees

For the final chapter of this section, we again study the gradient descent-ascent learning dynamics with timescale separation in unconstrained continuous action zero-sum games, but restricted to problems where the minimizing player faces a nonconvex optimization problem and the maximizing player optimizes a Polyak-Łojasiewicz (PŁ) or strongly-concave (SC) objective. These are classes of games that when been shown to be tractable in the sense that global convergence guarantees to near-stationary points can be achieved in finite-time, unlike the classes of nonconvex games that have been studied thus far. In contrast to past work in these classes of games, we assess convergence in relation to game-theoretic equilibrium notions instead of only notions of stationarity. In pursuit of this goal, we prove that the only locally stable points of the continuous-time limiting system correspond to differential Stackelberg equilibrium in each class of games. This characterization is obtained by strengthening the results of the previous chapter using the structure of the class of games. For the class of nonconvex-PŁ games, we exploit timescale separation to construct a potential function that when combined with the stability characterization and an asymptotic saddle avoidance result gives a global asymptotic almost-sure convergence guarantee to the set of differential Stackelberg equilibrium. For the class of nonconvex-SC games, we show the surprising property that the function of the game can be made a potential with timescale separation. Combining this insight with the stability characterization allows us to generalize methods for efficiently escaping saddle points in nonconvex optimization to obtain a global finite-time convergence guarantee to approximate differential Stackelberg equilibrium.

4.1 Introduction

In this chapter, we study continuous action zero-sum games of the form

$$\min_{x\in\mathcal{X}_1}\max_{y\in\mathcal{X}_2} f(x,y)$$

where $f \in C^2(\mathcal{X},\mathbb{R})$ with $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$ and $\mathcal{X}_1 = \mathbb{R}^{d_1}$ and $\mathcal{X}_2 = \mathbb{R}^{d_2}$ denote the individual action spaces and $d = d_1 + d_2$. In particular, we focus on zero-sum games in which $f(\cdot, y)$ is potentially nonconvex in $x \in \mathcal{X}_1$ and $f(x,\cdot)$ satisfies the Polyak-Łojasiewicz (PŁ) condition or is strongly-concave (SC) in $y \in \mathcal{X}_2$. We refer to these classes of games as nonconvex-PŁ and nonconvex-SC zero-sum games. Note that in this chapter, we switch notation and use y in place of x_2 for the choice variable of player 2 since it will be more clear notationally throughout this chapter. Moreover, in this chapter, we commonly denote joint strategies pairs by the shorthand $z = (x, y) \in \mathcal{X}_1 \times \mathcal{X}_2$.

This general formulation has a broad spectrum of applications such as fair classifica-

tion (Nouiehed et al., 2019), distributionally robust optimization (Namkoong and Duchi, 2016; Rafique et al., 2021), and adversarial learning (Madry et al., 2018). Consequently, there has been a surge of interest in recent years toward developing methods for solving these problems efficiently. So far, existing work on gradient-based learning in nonconvex-PŁ/SC zero-sum games has exclusively focused on providing global convergence rates to approximate stationary points with no attention given to the characterization in terms of game-theoretic equilibrium concepts (Lin et al., 2020a,b; Lu et al., 2020; Nouiehed et al., 2019; Rafique et al., 2021; Yang et al., 2020).

In contrast, a common theme in the study of general nonconvex-nonconcave zero-sum games is to assess the *types* of stationary points an algorithm locally converges toward in terms of their higher order structure (Daskalakis and Panageas, 2018; Fiez and Ratliff, 2020; Fiez et al., 2020a; Jin et al., 2020; Mazumdar et al., 2020, 2019; Wang et al., 2020a). The purpose of this analysis is generally to determine whether commonly deployed algorithms can guarantee local convergence to *only* game-theoretic equilibria or to design algorithms that achieve this objective. Indeed, this was exactly the focus of the previous chapters in this part of the thesis.

The goal of this chapter is to close the gap between the two problem classes and determine whether gradient-based learning algorithms in nonconvex-PŁ/SC zero-sum games can be shown to globally converge to *only* game-theoretically meaningful equilibria (local Nash or Stackelberg equilibrium). We focus our attention on the canonical gradient descent-ascent learning dynamics with timescale separations between players and a stochastically perturbed variant. In these algorithms, timescale separation is manifested in the different (yet constant) learning rates of the maximizing and minimizing players. The descriptions of these systems, which we refer to as τ-`GDA` and τ-`PGDA` where $\tau > 0$ parameterizes the timescale separation between players, are provided in Algorithm 4.1 and Algorithm 4.2, respectively. Simply put, τ-`GDA` (Algorithm 4.1) corresponds to each player following their individual gradient in a noiseless setting, while τ-`PGDA` (Algorithm 4.2) describes each player following their individual (potentially stochastic) gradient with artificial noise injections. Notably, both τ-`GDA` and τ-`PGDA` only require first-order gradients and as a result are computationally efficient methods for machine learning problems formulated as zero-sum games in this class.

4.1.1 Contributions

We show that the τ-`GDA` and τ-`PGDA` dynamics have *global* convergence guarantees in nonconvex-PŁ/SC zero-sum games to the natural game-theoretic solution concept for this problem class of differential Stackelberg equilibrium. The contributions of this work are summarized as follows.

1) In Theorem 4.1, we prove the only critical points that are locally stable with respect to the τ-`GDA` continuous-time limiting system are differential Stackelberg equilibrium in nonconvex-PŁ/SC zero-sum games.

2) In Theorem 4.2, we combine Theorem 4.1 with a potential function construction and an asymptotic saddle avoidance result to prove that τ-`GDA` (Algorithm 4.1) globally asymptotically converges to differential Stackelberg equilibrium almost surely in nonconvex-PŁ/SC zero-sum games. In Corollary 4.1, we show a corresponding local convergence rate of $\widetilde{\mathcal{O}}(\varepsilon^{-2})$ to an ε–differential Stackelberg equilibrium.

3) In Theorem 4.3, for the class of nonconvex-SC zero-sum games, we provide high-probability finite-time rates showing that τ-PGDA (Algorithm 4.2) globally converges to ε–local minmax equilibria with a complexities of $\widetilde{\mathcal{O}}(\varepsilon^{-4})$ and $\widetilde{\mathcal{O}}(\varepsilon^{-2})$ in stochastic and deterministic problems, respectively.

To our knowledge, our results provide the broadest existing global convergence guarantee for gradient-based algorithms to *game-theoretically meaningful* equilibria in zero-sum continuous games.

4.1.2 Practical Motivation

The study of nonconvex-PŁ/SC zero-sum games has often been motivated by machine learning problems formulated as games. We remark that given the problem formulations, it is natural to seek notions of minmax equilibria as solutions. Indeed, notions of stationarity are not guaranteed to reflect a meaningful solution since they could even correspond to maxmin solutions. Consequently, it is critical to give convergence guarantees to minmax equilibrium as we pursue in this work. We now provide examples of machine learning applications that belong to the class of games we consider.

Example 4.1. *In the fair classification problem (Nouiehed et al., 2019), the objective is to minimize the maximum loss over multiple categories. An example formulation is*

$$\min_{W} \max_{i=1,\dots,m} \{\mathcal{L}_i(W)\} \tag{4.1}$$

where $\mathcal{L}_i$ represents the loss on category i and W denotes parameter weights of a neural network.

Example 4.2. *To train a robust neural network against adversarial attacks, a common approach is to formulate training as a robust min-max optimization problem of the form*

$$\min_{w} \sum_{i=1}^{N} \max_{j=1,\dots,m} \{\ell(f(\hat{x}_{ij}(w)), y_i)\} \tag{4.2}$$

where w is the parameter vector of the neural network and each $\hat{x}_{ij}(w)$ is the result of a targeted attack on the sample x_i seeking to change the output of the network for label j (Madry et al., 2018; Nouiehed et al., 2019).

Example 4.3. *Distributionally robust optimization and robust learning from multiple distributions is commonly formulated as a minmax optimization problem of the form*

$$\min_{x\in\mathbb{R}^d} \max_{y\in\mathcal{Y}} \sum_{i=1}^{n} y_i f(x) - r(y) \tag{4.3}$$

where $f_i(x)$ is the loss of a model x on the i-th data point, $\mathcal{Y}$ is the simplex in $\mathbb{R}^n$, and $r(y)$ is carefully selected regularizer (Madry et al., 2018; Namkoong and Duchi, 2016; Rafique et al., 2021).

A common approach in each problem is to transform the inner maximization problem through relaxations and regularization methods to give an unconstrained problem in the parameter space.

Table 4.1: The gradient complexity of gradient descent (GD), gradient descent-ascent with timescale separation (τ-GDA), and alternating (including multi-step) gradient descent-ascent (AGDA) or perturbed variants (PGD, τ-PGDA) in deterministic and stochastic nonconvex optimization, nonconvex-PŁ zero-sum games, and nonconvex-SC zero-sum games. We state the complexity in terms of the ε tolerance of the guarantee and the dimension d with the notation $\widetilde{\mathcal{O}}(\cdot)$ hiding logarithmic factors in ε and d.

Problem	Algorithm & Reference	Complexity		Guarantee
		Deterministic	Stochastic	
Nonconvex Optimization	GD (Jin et al., 2021)	$\widetilde{\mathcal{O}}(\varepsilon^{-2})$	$\widetilde{\mathcal{O}}(\varepsilon^{-4})$	ε–Stationarity
	PGDA (Jin et al., 2021)	$\widetilde{\mathcal{O}}(\varepsilon^{-2})$	$\widetilde{\mathcal{O}}(\varepsilon^{-4})$	ε–Local Min
Nonconvex SC Zero-Sum	τ-GDA, τ-AGDA (Lin et al., 2020a)	$\widetilde{\mathcal{O}}(\varepsilon^{-2})$	$\widetilde{\mathcal{O}}(\varepsilon^{-4})$	ε–Stationarity
	Theorem 4.3 (τ-PGDA)	$\widetilde{\mathcal{O}}(\varepsilon^{-2})$	$\widetilde{\mathcal{O}}(\varepsilon^{-4})$	ε–**DSE**
Nonconvex PŁ Zero-Sum	AGDA (Nouiehed et al., 2019), (Yang et al., 2020, Appendix D)	$\widetilde{\mathcal{O}}(\varepsilon^{-2})$	–	ε–Stationarity
	Theorem 4.2 (τ-GDA)	Asymptotic	–	**DSE**

4.2 Related Work

We now cover the most relevant related work.

Nonconvex-Nonconcave Zero-Sum Games. A common theme in analyzing gradient descent-ascent with or without timescale separation in nonconvex-nonconcave zero-sum games has been to assess the local stability around critical points of the continuous-time limiting system and draw connections to the differential Nash and Stackelberg equilibrium notions (see Definition 4.4) (Daskalakis and Panageas, 2018; Fiez and Ratliff, 2020; Fiez et al., 2020a; Jin et al., 2020; Mazumdar et al., 2020; Mescheder et al., 2018; Nagarajan and Kolter, 2017; Zhang et al., 2020a). Importantly, unless the timescale separation is chosen very carefully, the stable critical points of gradient descent-ascent may not be game-theoretically meaningful (Fiez and Ratliff, 2020; Jin et al., 2020). We obtain stronger stability characterizations in our analysis of the continuous-time system using the structure of nonconvex-PŁ/SC zero-sum games. Further discussion of this topic is given in Section 4.4.

Nonconvex-PŁ and Nonconvex-SC Zero-Sum Games. Table 4.1 provides a comprehensive comparison between our results and existing results for gradient descent variants in nonconvex optimization and nonconvex-PŁ/SC zero-sum games. The key distinction between this chapter and past work is that instead of assessing convergence in terms of only reaching an approximate stationary point of the dynamics or a surrogate function, we obtain convergence guarantees in regards to differential Stackelberg equilibrium. Despite this being a much stricter and meaningful notion of solving the problem, we show that global asymptotic convergence guarantees remain obtainable in nonconvex-PŁ zero-sum games (Theorem 4.2) and that the local convergence rate (Corollary 4.1) is comparable to existing results. Moreover, for the subclass of nonconvex-SC zero-sum games, we provide novel global finite-time convergence guarantees to differential Stackelberg equilibria with rates comparable to existing results for finding a local minimum in nonconvex optimization or any stationary point in this class of games (Theorem 4.3).

Escaping Saddle Points in Nonconvex Optimization. A key contribution of this work is in regards to escaping saddles of the dynamics in classes of nonconvex zero-sum games. We build on methods from nonconvex optimization to obtain analogous results. In general, saddle avoidance results for variants of gradient descent in nonconvex optimization are asymptotic or finite-time. The former states that almost surely the algorithm does not converge to saddles points (Lee et al., 2016, 2019), while the latter gives rates of escape to conclude convergence to approximate local minimum (Ge et al., 2015; Jin et al., 2017, 2021). A key assumption in the aforementioned works is what is known as the strict saddle property, which ensures directions of escape exists from a saddle point. We make an analogous assumption refined for games. Further discussion of how we extend the methods from nonconvex optimization to the zero-sum game problem along with the challenges is included in Section 4.6.

4.3 Preliminaries

In the zero-sum games we study, we refer to the minimizing player controlling x as player 1 and the maximizing player controlling y as player 2. We consider objective functions $f \in C^2(\mathcal{X}, \mathbb{R})$ where the joint strategy space is denoted by $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$ where $\mathcal{X}_1 = \mathbb{R}^{d_1}$ and $\mathcal{X}_2 = \mathbb{R}^{d_2}$ denote the individual action spaces and $d = d_1 + d_2$. We often denote a joint strategy by $z = (x, y) \in \mathcal{X}$.

Notation. We denote by $D_i f(x, y) \in \mathbb{R}^{d_i \times 1}$ the derivative of $f(x, y)$ with respect to the choice variable of player i, $D_{ij} f(x, y) \in \mathbb{R}^{d_i \times d_j}$ as the partial derivative of $D_i f(x, y)$ with respect to the choice variable of player j, and $D_i^2 f(x, y) \in \mathbb{R}^{d_i \times d_i}$ as the partial derivative of $D_i f(x, y)$ with respect to the choice variable of player i. We let $\|\cdot\|$ denote the 2-norm of vectors unless otherwise specified, $\mathrm{spec}(\cdot)$ denote the set of eigenvalues of a matrix, $\mathrm{Re}(\cdot)$ denote the real part of a complex number, and $\mathbb{C}_-^\circ$ and $\mathbb{C}_+^\circ$ denote the open left-half and right-half complex plane, respectively. Let $\lambda_{\min}(A)$ denote the eigenvalue of A with the minimum real part, and $\lambda_{\max}(A)$ the eigenvalue of A with the maximum real part.

Classes of Games. We study and analyze both nonconvex-PŁ and nonconvex-SC zero-sum games. Throughout, we make the following assumptions on the class of functions that define the games.

Assumption 4.1. *Given a zero-sum game $(f, -f)$ defined by $f \in C^2(\mathcal{X}, \mathbb{R})$, $D_1 f(x, y)$ and $D_2 f(x, y)$ are L_1 and L_2 Lipschitz, respectively, that is,*

$$\|D_1 f(x, y) - D_1 f(x', y')\| \le L_1(\|x - x'\| + \|y - y'\|),$$
$$\|D_2 f(x, y) - D_2 f(x', y')\| \le L_2(\|x - x'\| + \|y - y'\|).$$

This assumption immediately implies that the vector of individual gradients given by

$$g(x, y) = (D_1 f(x, y), -D_2 f(x, y))$$

is also Lipschitz with parameter $L \le L_1 + L_2$.

Assumption 4.2. *Given a zero-sum game $(f, -f)$ defined by $f \in C^2(\mathcal{X}, \mathbb{R})$, $D^2 f(\cdot, \cdot)$ is β-Lipschitz with respect to the induced 2-norm. That is, for all $z, z' \in \mathcal{X}$, $\|D^2 f(z) - D^2 f(z')\| \le \beta \|z - z'\|$.*

We now define a nonconvex-PŁ zero-sum game. This class of games allows for the objective function to be nonconvex in $x \in \mathcal{X}_1$, but it needs to satisfy the Polyak-Łojasiewicz (PŁ) condition in $y \in \mathcal{X}_2$.

Definition 4.1 (Nonconvex-PŁ Game)**.** *Consider a zero-sum game $(f, -f)$ defined by $f \in C^2(\mathcal{X}, \mathbb{R})$. The game is called nonconvex-PŁ if $f(x, \cdot)$ is μ-PŁ with respect to the argument $y \in \mathcal{X}_2$. That is, for $\mu > 0$ and for all $(x, y) \in \mathcal{X}$,*

$$\|D_2 f(x, y)\|^2 \geq 2\mu(\max_{y' \in \mathcal{X}_2} f(x, y') - f(x, y)).$$

The following provides a definition for what we call a nonconvex-SC zero-sum game. In this class of games, the function that defines the game may be nonconvex in $x \in \mathcal{X}_1$, but it must be SC in $y \in \mathcal{X}_2$.

Definition 4.2 (Nonconvex-SC Game)**.** *Consider a zero-sum game $(f, -f)$ defined by $f \in C^2(\mathcal{X}, \mathbb{R})$. The game is nonconvex-SC if $f(x, \cdot)$ is μ-strongly-concave with respect to the argument $y \in \mathcal{X}_2$. That is, given any $x \in \mathcal{X}_1$ and for all $y, y' \in \mathcal{X}_2$,*

$$f(x, y') \leq f(x, y) + \langle D_1 f(x, y), y' - y \rangle - \tfrac{\mu}{2}\|y - y'\|_2^2.$$

It is worth noting that nonconvex-SC zero-sum games are nonconvex-PŁ zero-sum games, but the converse does not necessarily hold as a result of the relationship between PŁ and SC functions (Karimi et al., 2016).

Learning Dynamics. We study gradient descent-ascent with timescale separation and a perturbed variant. Let γ be the learning rate of player one and $\tau > 0$ be the timescale parameterization. The deterministic τ-`GDA` dynamics we study are presented in Algorithm 4.1 and the potentially stochastic variant with injected noise perturbations called τ-`PGDA` is given in Algorithm 4.2 (see Section 4.6 for the relevant notation). The key distinction between this study of gradient descent-ascent with timescale separation and past work is how we assess convergence as we now begin to formalize.

Stationarity Notions. A critical point corresponds to a joint strategy profile at which the individual gradient of each player is equal to zero.

Definition 4.3 (Critical Point)**.** *A point $(x, y) \in \mathcal{X}$ is a critical point if $D_1 f(x, y) = 0$ and $D_2 f(x, y) = 0$.*

Critical points correspond to stationary points of the τ-`GDA` dynamics. The nonconvex-PŁ/SC zero-sum game literature has generally assessed convergence in terms of the complexity of finding an approximate stationary point (see, e.g., (Nouiehed et al., 2019; Yang et al., 2020)). That is, joint strategies $(x, y) \in \mathcal{X}$ such that $\|D_1 f(x, y)\| \leq \varepsilon$ and $\|D_2 f(x, y)\| \leq \varepsilon$. A related and common notion of convergence in this body of work (see e.g., Lin et al. 2020a) is that of finding a stationary point of the function $\max_{y \in \mathcal{X}_2} f(\cdot, y)$. This criterion amounts to seeking to achieve the condition $\|D_1 \max_{y \in \mathcal{X}_2} f(x, y)\| \leq \varepsilon$.

In contrast, we assess convergence with connections to the equilibrium notions that are commonly studied in the nonconvex-nonconcave zero-sum game literature. Since either stationarity notion may lack any game-theoretic meaning, we consider a strictly harder notion of solving a game.

Algorithm 4.1 τ-GDA

Input: $x_0 \in \mathbb{R}^{d_1}$, $y_0 \in \mathbb{R}^{d_2}$
for $k = 0, 1, \ldots$ **do**
$x_{k+1} \leftarrow x_k - \gamma D_1 f(x_k, y_k)$
$y_{k+1} \leftarrow y_k + \gamma\tau D_2 f(x_k, y_k)$
end for

Algorithm 4.2 τ-PGDA

Input: $x_0 \in \mathbb{R}^{d_1}$, $y_0 \in \mathbb{R}^{d_2}$
for $k = 0, 1, \ldots$ **do**
Sample $\theta_{i,k} \sim \mathcal{D}_i$, $i = 1, 2$; $(\xi_{1,k}, \xi_{2,k}) \sim \mathcal{N}(0, (r^2/d)I)$
$x_{k+1} \leftarrow x_k - \gamma(g_1(x_k, y_k; \theta_{1,k}) + \xi_{1,k})$,
$y_{k+1} \leftarrow y_k + \gamma\tau(g_2(x_k, y_k; \theta_{2,k}) + \tau^{-1}\xi_{2,k})$
end for

Equilibrium Notions. The typical solution concept in game theory when an implicit or explicit order of play is present in the structure of the game is the (local) Stackelberg (equivalently minmax in zero-sum games) equilibrium concept (Basar and Olsder, 1998).[1] Informally, in nonconvex-PŁ/SC zero-sum games, a local Stackelberg equilibrium corresponds to a strategy pair $(x^*, y^*) \in \mathcal{X}$ such that x^* is a local minimum of the function $f(x, y_*(x))$ where $y_*(x) \in \arg\max_{y \in \mathcal{X}_2} f(x, y)$ and y^* is a local maximum of the function $f(x^*, y)$. When the function f is bounded or when $f(\cdot, y)$ is bounded and $f(x, \cdot)$ is strongly concave, a Stackelberg equilibrium is guaranteed to exist.

We characterize the local Stackelberg equilibrium notion in terms of sufficient conditions on player costs as is in previous chapters and is now typical in learning in games (Fiez and Ratliff, 2020; Fiez et al., 2020a; Jin et al., 2020; Mazumdar et al., 2020; Wang et al., 2020a). Toward presenting this definition, we denote by $J(x, y)$ the Jacobian of the vector of individual gradients $g(x, y)$ that is given by

$$J(x, y) = \begin{bmatrix} D_1^2 f(x, y) & D_{12} f(x, y) \\ -D_{12}^\top f(x, y) & -D_2^2 f(x, y) \end{bmatrix}. \tag{4.4}$$

Let $\mathtt{S}_1(\cdot)$ denote the Schur complement of $(\cdot)$ with respect to the $d_2 \times d_2$ block in $(\cdot)$. The following definition is characterized by sufficient conditions for a local Stackelberg equilibrium in zero-sum games and is equivalent to that presented in the previous chapters.

Definition 4.4 (Differential Stackelberg/Strict Local Minmax Equilibrium (Fiez et al., 2020a))**.** *The joint strategy $(x^*, y^*) \in \mathcal{X}$ is a differential Stackelberg equilibrium if the first-order conditions $Df(x^*, y^*) = D_1 f(x^*, y^*) - D_{12} f(x^*, y^*)[D_2^2 f(x^*, y^*)]^{-1} D_2 f(x^*, y^*) = 0$, $D_2 f(x^*, y^*) = 0$ hold and the second-order conditions $\mathtt{S}_1(J(x^*, y^*)) \succ 0$ and $D_2^2 f(x^*, y^*) \prec 0$ also hold.*

A differential Stackelberg equilibrium corresponds to a joint strategy at which the minimizing player is at a local optimum with respect to its choice variable along the best response curve of the maximizing player and the maximizing player is at a local optimum with respect to its choice variable. In the next section, we also provide a definition for an ε-differential Stackelberg equilibrium.

[1] By implicit order of play, we mean the min and max order are not interchangeable.

4.4 Local Stability Analysis

To characterize the convergence of τ-GDA, we begin by studying its continuous-time limiting system

$$\dot{z} = -\Lambda_\tau g(z) \tag{4.5}$$

where $\dot{z} = (\dot{x}, \dot{y})$, $\Lambda_\tau = \text{blockdiag}(I_{d_1}, \tau I_{d_2})$, and $g(z)$ is the vector of individual gradients. The Jacobian of this system is given by

$$J_\tau(z) = \Lambda_\tau J(z). \tag{4.6}$$

We analyze the stability of the continuous-time system around critical points z^* as a function of the timescale separation τ using the Jacobian $J_\tau(z^*)$ in this section toward drawing conclusions about the stability and convergence of the discrete time system τ-GDA. A critical point is said to be locally (exponentially) stable when the spectrum of $-J_\tau(z^*)$ is in the open left-half complex plane $\mathbb{C}_-^\circ$ (Khalil, 2002; Sastry, 1999). Simply put, a critical point z^* is locally exponentially stable if and only if the real parts of the eigenvalues of $-J_\tau(z^*)$ are strictly negative. Throughout, we use the broader term "stable" to mean the following in a consistent manner with the previous chapters.

Definition 4.5 (Stability)**.** *A critical point $z^* = (x^*, y^*) \in \mathcal{X}$ is locally exponentially stable for $\dot{z} = -\Lambda_\tau g(z)$ if and only if* $\text{spec}(-J_\tau(z^*)) \subset \mathbb{C}_-^\circ$ *($\equiv$ $\text{spec}(J_\tau(z^*)) \subset \mathbb{C}_+^\circ$).*

Stability with respect to the continuous-time τ-GDA dynamics guarantees that the system asymptotically converges at an exponential rate to the critical point in a local neighborhood. Moreover, given a suitable choice of learning rates, equivalent insights hold for the discrete-time dynamics (Chasnov et al., 2020).

Stability in Nonconvex-Nonconcave Zero-Sum Games. Given the implications regarding convergence, a number of papers in the past several years study the stability of τ-GDA around critical points and the connections to differential Nash and Stackelberg equilibrium in zero-sum games (Daskalakis and Panageas, 2018; Fiez and Ratliff, 2020; Fiez et al., 2020a; Jin et al., 2020; Mazumdar et al., 2020). However, this body of research focuses on general nonconvex-nonconcave zero-sum games. In general across the spectrum of nonconvex-nonconcave zero-sum games, the stable critical points of τ-GDA coincide with the set of differential Stackelberg equilibria only when the timescale separation $\tau \to \infty$ (Jin et al., 2020). Given that such a choice of timescale separation requires the learning rate $\gamma \to 0$ in order to retain stability of the discrete-time system, it is not clear how to derive a practical algorithm from this insight.

Toward remedying this problem, the previous chapter provides stability results in terms of the timescale separation concerning a given critical point, rather than across the space of nonconvex-nonconcave zero-sum games. Indeed, we proved a stability and instability result as a function of the timescale separation in the τ-GDA dynamics in Chapter 3. The stability results say that given a differential Stackelberg equilibrium z^*, there exists a finite $\tau^* \in (0, \infty)$ that can be constructed such that z^* is stable for all $\tau \in (\tau^*, \infty)$. On the other hand, the instability results says that given a critical point which is not a differential Stackelberg equilibrium, there exists a finite $\tau_0 \in (0, \infty)$ that can be constructed such that z^* is not stable for all $\tau \in (\tau_0, \infty)$.

Stability in Nonconvex-PŁ/SC Zero-Sum Games. To our knowledge, the connection between the stability (and instability) of critical points with respect to τ-GDA dynamics and game-theoretic equilibrium notions in the semi-structured problems of nonconvex-PŁ and nonconvex-SC

has not been fully characterized. We show in the following result that when a nonconvex-nonconcave game is specialized to a nonconvex-PŁ game as from Definition 4.1, significantly more general stability characterizations can be obtained. Notably, any critical point z^* that is not a differential Stackelberg equilibrium is unstable for all $\tau \geq 0$. Meaning that τ-GDA does not admit spurious stable points in nonconvex-PŁ game. This is in stark contrast to the known fact that 1-GDA admits spurious stable points in the more general class of nonconvex-nonconcave games as has been shown in previous literature (Daskalakis and Panageas, 2018; Jin et al., 2020; Mazumdar et al., 2020). Moreover, if z^* is a differential Stackelberg equilibrium, then z^* is stable for all τ larger than the minimum τ_* for which z^* is stable and such a finite τ_* is guaranteed to exist. This result implies that in practice, one can select a finite value of τ to run τ-GDA with, and all stable critical points (if they exist) will be differential Stackelberg equilibria and if τ is scaled up and the set of stable points grows, then only differential Stackelberg equilibria can be introduced.

Theorem 4.1. *Consider a nonconvex-PŁ zero-sum game $(f, -f)$ where $f \in C^2(\mathcal{X}, \mathbb{R})$. Then, the following hold: 1) Any critical point z^* that is not a differential Stackelberg equilibrium is unstable for all $\tau \in (0, \infty)$; 2) If z^* is a differential Stackelberg equilibrium, then $\text{spec}(-J_\tau(z^*)) \subset \mathbb{C}_-^\circ$ for all $\tau \in [\tau_*, \infty)$ where τ_* is the minimum $\tau \in (0, \infty)$ such that $\text{spec}(-J_\tau(z^*)) \subset \mathbb{C}_-^\circ$ and a finite τ_* is guaranteed to exist.*

In comparison to the stability results for nonconvex-nonconcave zero-sum games from the previous chapter, we obtain the stronger results in nonconvex-PŁ zero-sum games that (i) $\tau_0 = 0$ for any critical point that is not a differential Stackelberg equilibria and (ii) a differential Stackelberg equilibria is never unstable after it becomes stable as a function of the timescale separation τ.

The stability characterization also allows us to define a natural approximate differential Stackelberg equilibrium notion. The first-order conditions in Definition 4.4 can equivalently be reformulated as $D_1 f(x^*, y^*) = 0$ and $D_2^2 f(x^*, y^*) = 0$. We presented the differential Stackelberg equilibrium using the total derivative for the minimizing player in the first-order condition to mirror the presentation of proper approximate notions. In particular, let $\bar{g}(x, y) = (Df(x, y), -D_2 f(x, y))$ denote the vector containing the total derivative for the minimizing player and the individual derivative for the maximizing player. We have the following ε–differential Stackelberg equilibrium.

Definition 4.6. *A point $z^* = (x^*, y^*) \in \mathcal{X}_1 \times \mathcal{X}_2$ is an ε-differential Stackelberg equilibrium for a nonconvex–PŁ zero-sum game $f \in C^2(\mathcal{X}, \mathbb{R})$ satisfying Assumptions 4.1 and 4.2 if $\|\bar{g}(z^*)\| \leq \varepsilon$ and $\text{Re}(\lambda_{\min}(J_\tau(z^*))) \geq -\sqrt{\varepsilon\beta}$.*

4.5 Global Asymptotic Convergence Analysis

Theorem 4.1 in the previous section completely characterizes the behavior of the continuous-time τ-GDA dynamics. However, ultimately we are concerned with the convergence properties of the discrete-time τ-GDA dynamics. We begin our study of this system by providing a global asymptotic analysis in this section of the deterministic τ-GDA dynamics presented in Algorithm 4.1.

We prove in Theorem 4.2 of this section that the deterministic τ-GDA (Algorithm 4.1) almost surely converges to a differential Stackelberg equilibrium in nonconvex-PŁ zero-sum games, a class of games that subsumes nonconvex-SC zero-sum games. Despite this result being asymptotic in nature, to our knowledge it is the most general class of zero-sum games in which a global convergence guarantee to established game-theoretic equilibria has been given.

To begin, observe that the τ-GDA dynamics do not necessarily correspond to a gradient flow a function, which can be observed from the fact that the Jacobian J_τ is not guaranteed to be symmetric. As a consequence, there may exist complex limiting behavior beyond convergence to a critical point such as non-trivial limit cycles or periodic orbits. To rule out this phenomenon, we prove that there exists potential function that decreases along the iterates of τ-GDA. The best response map is well-defined by the implicit mapping theorem since the PŁ condition on the maximizing player's problem implies quadratic growth (Karimi et al., 2016) which in turn implies that at critical points not only is $D_2^2 f$ is non-degenerate, but also it is negative definite. In order to construct a potential function for nonconvex, μ-PŁ zero-sum games, we need the best-response map of the maximizing player, $y_*(x) \in \arg\max_y f(x,y)$, defined implicitly by $D_2 f(x,y) = 0$ to be L_3-Lipschitz; we show this in Lemma 4.7 in Section 4.A.1.[2]

Lemma 4.1. *Consider a non-convex, μ-PŁ zero sum game defined by $f \in C^2(\mathcal{X}, \mathbb{R})$ which satisfies Assumptions 4.1 and 4.2, and has condition number $\kappa = L_2/\mu$. Suppose that $\tau > 7\kappa^2$, and $\gamma < \min\{\frac{1}{3L_2\tau}, \frac{1}{2(L_1+L_3L_2)}\}$ then, for any $\Gamma \in (0, 1/7]$, $\Phi(x,y) = f(x, y_*(x)) - \Gamma f(x,y)$ is a potential function for τ-GDA.*

The function $f(x, y_*(x))$ can be seen as the function the x player would minimize if the y player was playing a best-response $y_*(x)$. The potential function $\Phi(x,y)$ essentially captures that along trajectories of τ-GDA, the function $f(x,y)$ should either decrease the value of $f(x, y_*(x))$, or decrease the value of $f(x, y_*(x)) - f(x,y)$ since the y-player converges at a fast rate to $y_*(x)$ given the time-scale separation. Indeed, this potential function implicitly guarantees that the maximizing player *tracks* the best response set, and that the minimizing player essentially ends up minimizing $f(x, y_*(x))$ as desired. The choice of τ and the learning rate γ allow us to guarantee that this occurs.

Given the potential function in Lemma 4.1, we can conclude that Algorithm 4.1 converges to critical points. The critical points may correspond to stable points of the dynamics (Definition 4.5) or saddle points of the dynamics, which are defined as follows.[3]

Definition 4.7 (Saddle Point)**.** *The critical point $z^* = (x^*, y^*) \in \mathcal{X}$ is a saddle point of the dynamics $\dot{z} = -\Lambda_\tau g(z)$ if $\mathrm{Re}(\lambda_{\max}(-J_\tau(z^*))) = 0$ and a strict saddle if $\mathrm{Re}(\lambda_{\max}(-J_\tau(z^*))) > 0$.*

Recall that Theorem 4.1 indicates that the only stable points of the continuous-time τ-GDA dynamics are differential Stackelberg equilibria. Thus, if we can show that the discrete-time τ-GDA dynamics avoid saddle points of the continuous-time τ-GDA dynamics, then we can conclude that Algorithm 4.1 converges to only differential Stackelberg equilibria. The following result of Mazumdar et al. (2020) states that the discrete-time τ-GDA dynamics do indeed avoid saddle points of the continuous-time system.[4]

---

[2]This mapping is $L(\kappa+1)$-Lipschitz in the case where $f(x,\cdot)$ is strongly concave (Lin et al., 2020a, Lemma 4.3).

[3]The terminology of saddle point in this chapter is with respect to the dynamics and should not be conflated with the terminology of a saddle point of a function.

[4]The saddle avoidance result of Mazumdar et al. (2020) holds for both general-sum n-player nonconvex games. For simplicity, we only state it in the context of this work.

Lemma 4.2 (Theorem 4.1 Mazumdar et al. 2020)**.** *Consider a zero sum game defined by the non-convex, non-concave function $f \in C^2(\mathcal{X}, \mathbb{R})$. The set of initial conditions $z \in \mathcal{X}$ from which τ-GDA converges to strict saddle points is of measure zero.*

Lemma 4.2 can be combined with the existence of a potential function (Lemma 4.1) and the guarantee that the only stable points of τ-GDA are differential Stackelberg, we obtain the following global convergence guarantee for non-convex, PŁ-games. To obtain this result, we need the additional commonly assumed strict saddle point property (Ge et al., 2015; Jin et al., 2017; Mazumdar et al., 2020).

Assumption 4.3 (Strict Saddle Property)**.** *Given a zero-sum game $(f, -f)$ defined by $f \in C^2(\mathcal{X}, \mathbb{R})$, we assume the strict saddle property that every saddle point of the dynamics is a strict saddle point.*

We are now ready to the state main result of this section.

Theorem 4.2. *Consider a nonconvex-PŁ zero-sum game $(f, -f)$ defined by $f \in C^2(\mathcal{X}, \mathbb{R})$ that satisfies Assumptions 4.1–4.3. Then, τ-GDA with $\tau > 7\kappa^2$, and stepsize $\gamma < \min\{\frac{1}{3L_2\tau}, \frac{1}{2(L_1+L_3L_2)}\}$ asymptotically converges to the set of differential Stackelberg equilibrium that are stable for $\dot{z} = -\Lambda_\tau g(z)$ almost surely. That is, for almost all initial conditions, τ-GDA will converge to a differential Stackelberg equilibrium.*

If we fix a τ satisfying the assumptions of the above theorem, then if we consider the set of stable critical points for $\dot{z} = -\Lambda_\tau(z)$, we know that by Theorem 4.1 this set only contains differential Stackelberg equilibrium. It is precisely this set of differential Stackelberg equilibrium to which τ-GDA converges almost surely. Moreover, by Theorem 4.1, as we increase τ, no new spurious non-equilibrium points are introduced to the set of stable critical points; only additional differential Stackelberg equilibrium points can be added to this set.

We complement Theorem 4.2 with the following characterization of the convergence rate.

Corollary 4.1. *Consider a nonconvex-PŁ zero-sum game $(f, -f)$ defined by $f \in C^2(\mathcal{X}, \mathbb{R})$ that satisfies Assumptions 4.1–4.3. Given an initialization in the region of attraction of a differential Stackelberg equilibrium, then after $\widetilde{\mathcal{O}}(\varepsilon^{-2})$ iterations at least one iterate is an ε-differential Stackelberg equilibrium.*

Observe that this corollary is a local convergence rate guarantee. The reason for this is that it is possible that the dynamics could get stuck around a strict saddle point for a long period of time. We deal with this problem in the next section.

4.6 Finite Time Convergence Results

In this section, we show that in nonconvex-SC games not only does gradient descent-ascent provably converge in an almost sure asymptotic sense, but also that there is a finite time escape time from saddle points. The game dynamics can get stuck in a region around a non-game theoretically meaningful saddle, and while this problem has been observed in the nonconvex single player setting, it appears more of nuanced issue in the game setting since saddles of the dynamics are not always saddles of the function defining the game and vice versa. We show the result for the stochastic case

in which players have unbiased stochastic gradients. The deterministic case is immediate by taking the stochasticity to zero. Due to the length of proving these results, we omit the proofs from this section from the thesis and refer the interested reader to the associated paper (Fiez et al., 2021b).

Players do not have access to their exact gradients $D_1 f(x,y)$ and $D_2 f(x,y)$. Instead, for any (x,y) a gradient query to an oracle return the individual gradient estimates $g_1(x,y;\theta_1)$ and $g_2(x,y;\theta_2)$ to the two players respectively, where each θ_i is a random variables drawn from distribution $\mathcal{D}_i$ with $i \in \mathcal{I}$ and the stochastic gradients satisfy the following assumptions.[5]

Assumption 4.4. *For any $(x,y) \in \mathcal{Z}$, the stochastic gradient vector*

$$g(x,y;\theta) = (g_1(x,y;\theta_1), g_2(x,y;\theta_2))$$

satisfies $D_i f(x,y) = \mathbb{E}_{\theta_i \sim \mathcal{D}_i}[g_i(x,y;\theta_i)]$ and $\forall\ t \in \mathbb{R}$ and each $i \in \mathcal{I}$:

$$\mathbb{P}(\|g_i(x,y;\theta) - D_i f(x,y)\| \geq t) \leq 2\exp(-\tfrac{t^2}{2\sigma_i^2}).$$

Assumption 4.5. *For each $i \in \mathcal{I}$ and any $\theta_i \in \operatorname{supp}(\mathcal{D}_i)$, $g_i(\cdot,\cdot;\theta_i)$ is $\tilde{\ell}_i$-Lipschitz: that is, for all $(x_1,y_1),(x_2,y_2) \in \mathcal{X}$, $\|g_i(x_1,y_1;\theta_i) - g_i(x_2,y_2;\theta_i)\| \leq \tilde{\ell}_i \|(x_1,y_1) - (x_2,y_2)\|$.*

We also make the following assumption on the properties of strict saddle points.

Assumption 4.6. *The eigenvalue of $J_\tau(z)$ with minimum real part for any strict saddle z is simple. That is, the algebraic multiplicity $m_A(-\eta) = 1$ and geometric multiplicity of $m_G(-\eta) = 1$ where $-\eta = \operatorname{Re}(\lambda_{\min}(J_\tau(z)))$.*

Let $\ell = L_1 + L_2$ if $\kappa \leq 1/2$, and $\ell = L_1 + 2\kappa L_2$ otherwise. In either case, $\lambda_{\max}(D^2 f(x,y)) \leq \ell$. In the remainder, we work in the case for which $\ell \geq \sqrt{\beta\varepsilon}$; otherwise the problem of finding ε-local minmax is straightforward since all ε-stationary points will be ε-local minmax points.

The steps in showing that τ-`PGDA` (Algorithm 4.2) escapes saddles with high probability largely follow those in Daneshmand et al. (2018); Ge et al. (2015); Jin et al. (2017, 2021), with several modifications due to the fact that we consider a zero-sum game as opposed to a single player optimization problem. The challenge in the minmax (or zero-sum game) setting as compared to the single player optimization setting is that the gradient descent-ascent dynamics in continuous-time as given in (4.5) are not a gradient flow, and hence, the eigenvalues of the local linearization from (4.6) may be complex-valued. In the key step of many of the proofs on saddle avoidance, escaping from the *stuck regions* around a saddle (with high probability) is achieved by analyzing the local linearization around the saddle point, which crucially, is symmetric. Hence, the problem reduces to analyzing the escape time along the eigendirection of the minimum eigenvalue of the Hessian which relies heavily on the orthogonality of eigenvectors of symmetric matrices. In zero-sum games, the linearization is no longer symmetric and these key results have to be substantially modified.

One of the other key distinctions is in constructing a potential function. In the existing literature on convergence in the nonconvex-SC literature (Liu et al., 2020; Wang et al., 2020b; Yang et al.,

[5]Note that by slight abuse of notation, we are overloading notation by using $g_1(x,y;\theta_1)$ and $g_2(x,y;\theta_2)$ to denote stochastic individual gradients for each player, whereas previously $g(x,y)$ has denoted the vector of each players individual gradients.

2020), the proposed potential function has a structure closely related to the potential function in Lemma 4.1; in particular, it takes the form $f(x,y) + r(x,y)$ where $r(x,y)$ is a tracking term of the form $\|y - y_*(x)\|^2$ or $f(x, y_*(x))$. Yet, perhaps surprisingly, despite the fact that f depends on to sequences generated by two separate gradient updates—one from minimizing f and one from maximizing it—the descent lemma below (Lemma 4.3) shows that the function f can be decomposed into a component that is decreasing along trajectories of τ-PGDA (Algorithm 4.2) and a component that exhibits a possible increase due to randomness in the stochastic gradients and injected noise. The primary reason for the decrease is large gradients in addition to the strongly concave structure of the maximizing player's problem and the timescale separation which we exploit in the proof to ensure sufficient decrease of f along trajectories of τ-PGDA.

Lemma 4.3 (Descent Lemma)**.** *Consider a non-convex, μ-SC zero-sum game defined by $f \in C^2(\mathcal{X}, \mathbb{R})$. Under Assumptions 4.1–4.5, if $\frac{3}{2\mu\tau} < \gamma < 1/\ell$, then with probability at least $1 - \delta$ for some $\delta > 0$, $f(x_k, y_k)$ is a potential function for τ-PGDA in both the stochastic and deterministic settings.*

The descent lemma enables us to argue that with high probability, either the function value decreases a sufficient amount or the iterates remain in a small region. Using this property of τ-PGDA together with the descent lemma, we then show that if two sequences that are *coupled* (i.e., two sequences starting from the same initial condition and having equivalent noise except along the direction of escape) and are in the region around a saddle point, then with high probability, at least one of them escapes.

Putting these results together, we have the following finite time guarantee on convergence to ε-differential Stackelberg equilibria (ε–local minmax points). Let $f^* = \min_{x,y} f(x,y)$ be the minimum value of the function f, which we assume is finite. These results provide novel convergence guarantees to not just ε-stationary points but those that are game theoretically meaningful.

Theorem 4.3. *Consider a non-convex, μ-strongly concave zero-sum game defined by $f \in C^2(\mathcal{X}, \mathbb{R})$ and suppose that Assumptions 4.1–4.6 hold. For any $\varepsilon, \delta > 0$, there exists γ and τ such that, with probability $1 - \delta$ for some $\delta > 0$, starting from any $z_0 = (x_0, y_0)$, at least half the iterates of τ-PGDA will be ε-differential Stackelberg equilibria after $\widetilde{\mathcal{O}}(\varepsilon^{-4})$ iterations and $\widetilde{\mathcal{O}}(\varepsilon^{-2})$ in the stochastic and deterministic settings, respectively.*

4.7 Discussion

To the best of our knowledge, our results are the first to guarantee global convergence of gradient-based algorithms to game theoretically meaningful equilibria in such a general class of games. We believe that a more detailed analysis could make use of the potential function defined for nonconvex-PŁ in Section 4.5 to prove similar finite-time results for τ-PGDA, though the proof appears to be tedious and hence, we leave this for future work. We also believe future work can be done to obtain global guarantees in this class of games to game-theoretically meaningful equilibria without gradient information and also in the setting with constraints extending recent saddle avoidance results for the single player nonsmooth, constrained setting.

CHAPTER 4 APPENDIX

4.A Preliminaries

We now review preliminaries on PŁ functions and nonconvex-PŁ zero-sum games, along with a linear algebra needed for the proofs.

4.A.1 Polyak-Łojasiewicz Functions and Nonconvex-Polyak-Łojasiewicz Zero-Sum Games

In this section, we state properties of PŁ functions in the context of nonconvex-PŁ zero-sum games. Specifically, we characterize the curvature around critical points of PŁ functions and also state Lipschitz and smoothness properties that follow from our assumptions. These properties will be used in both the proofs for the local stability and the global convergence in nonconvex-PŁ zero-sum games.

We begin by stating a known property (Karimi et al., 2016) that μ-PŁ functions satisfy a quadratic growth condition also with parameter μ. For clarity of presentation, we present this condition in the context of the nonconvex-PŁ zero-sum games we study.

Lemma 4.4 (Karimi et al. 2016)**.** *Consider a non-convex, μ-PŁ zero-sum game defined by $f \in C^2(\mathcal{X}, \mathbb{R})$. For all $x \in \mathcal{X}_1$, the function $f(x, \cdot)$ satisfies the following quadratic growth condition:*

$$\max_{y' \in \mathcal{X}_2} f(x, y') - f(x, y) \geq \frac{\mu}{2} \|y_p - y\|^2, \ \forall \ y \in \mathcal{X}_2 \tag{4.7}$$

where y_p is the projection onto the set $\arg\max_{y \in \mathcal{X}_2} f(x, y)$.

We now show that the quadratic growth property of PŁ functions implies that $D_2^2 f(x^*, y^*) \preceq -\frac{\mu}{2} I$ at any critical point (x^*, y^*) of a μ-PŁ zero-sum game. This means that at any critical point of the game, player 2 must be at a maximum. The following property will be used in the proof of Theorem 4.1 in Section 4.B.

Lemma 4.5. *Consider a non-convex, μ-PŁ zero-sum game defined by $f \in C^2(\mathcal{X}, \mathbb{R})$. At any critical point (x^*, y^*) of the game, that is where $D_1 f(x^*, y^*) = 0$ and $D_2 f(x^*, y^*) = 0$, the individual Hessian of the maximizing player given by $D_2^2 f(x^*, y^*)$ is negative definite and eigenvalues bounded above by $-\mu/2$.*

Proof. Let us consider any critical point (x^*, y^*) of the game so that $D_1 f(x^*, y^*) = 0$ and $D_2 f(x^*, y^*) = 0$. Taking a Taylor expansion of $f(x^*, \cdot)$ about the point y^*, we and get that

$$\begin{aligned} f(x^*, y) &\geq f(x^*, y^*) + D_2 f(x^*, y^*)^\top (y - y^*) + \frac{1}{2}(y - y^*)^\top D_2^2 f(x^*, y^*)(y - y^*) \\ &= f(x^*, y^*) + \frac{1}{2}(y - y^*)^\top D_2^2 f(x^*, y^*)(y - y^*) \end{aligned}$$

Hence, from the quadratic growth condition of Lemma 4.4, we have that

$$\frac{\mu}{2}\|y^* - y\|^2 \leq f(x^*, y^*) - f(x^*, y) \leq -(y - y^*)^\top D^2 f(x^*, y^*)(y - y^*)$$

and consequently

$$\frac{\mu}{2}\|y^* - y\|^2 I \succeq -D_2^2 f(x^*, y^*)\|y^* - y\|^2 \implies D_2^2 f(x^*, y^*) \succeq -\frac{\mu}{2}.$$

Since this holds for any critical point, the conclusion follows. □

We now state several more properties that will be used in most of the proofs in Section 4.C regarding the results for nonconvex-PŁ zero-sum games from Section 4.5.

Danskin's theorem in optimization provides conditions under which $Df(x, y_*(x))$ where $y^*(x) = \arg\max_{y \in \mathcal{X}_2} f(x, y)$ is equivalent to $D_1 f(x, y_*)$. That is, it gives conditions when the gradient of the function $f(x, y_*(x))$ is equal to the gradient of $f(x, y_*)$ evaluated directly at the optimum. Typically this requires the maximizer to be unique. However, it has been show that for nonconvex-PŁ zero-sum games, this property carries over even without a unique solution.

Lemma 4.6 (Danskin-Type Property for PŁ functions Nouiehed et al. 2019, Lemma A.5)**.** *Consider a non-convex, μ-PŁ zero-sum game defined by $f \in C^2(\mathcal{X}, \mathbb{R})$ satisfying Assumption 4.1. Then,*

$$Df(x, y_*(x)) = D_1 f(x, y_*(x)) \quad \textit{where} \quad y_*(x) \in \arg\max_{y \in \mathcal{X}_2} f(x, y).$$

We now show that the mapping $y_*(x)$, defined implicitly by $D_2 f(x, y) = 0$ is Lipschitz in nonconvex-PŁ zero-sum games.

Lemma 4.7. *Consider a non-convex, μ-PŁ zero-sum game defined by $f \in C^2(\mathcal{X}, \mathbb{R})$ satisfying Assumptions 4.1 and 4.2. The best-response map $y_*(x) \in \arg\max_y f(x, y)$, defined implicitly by $D_2 f(x, y) = 0$ is $L_3 = \frac{\beta}{\mu}$-Lipschitz. That is, for all $x, x' \in \mathcal{X}_1$,*

$$\|y_*(x) - y_*(x')\| \leq L_3\|x - x'\|.$$

Proof. For all $x, x' \in \mathcal{X}_1$, we have that

$$\begin{aligned}
\|y_*(x) - y_*(x')\| &\leq \max_z \|Dy_*(z)\|\|x - x'\| \\
&\leq \max_\nu \|(D_y^2 f(\nu, y_*(\nu)))^{-1}\|\|D_{yx} f(\nu, y_*(\nu))\|\|x - x'\| \\
&\leq \frac{\beta}{\mu}\|x - x'\| = L_3\|x - x'\|.
\end{aligned}$$

□

Finally, we assume that total derivative $Df(x, y)$ is Lipschitz in nonconvex-PŁ zero-sum games.

Assumption 4.7. *The total derivative* $Df(x,y) = D_1 f(x,y) - D_{12} f(x,y)[D_2^2 f(x,y)]^{-1} D_2 f(x,y)$ *is* L_5*-Lipschitz. That is, for any* $x \in \mathcal{X}_1$ *and all* $y, y' \in \mathcal{X}_2$*,*

$$\|Df(x,y) - Df(x,y')\| \leq L_5 \|y - y'\|.$$

Note that if f is Lipschitz in y, then the above assumption follows directly from the proceeding assumptions on the smoothness of f.

4.A.2 Linear Algebra

In this section, we state a property regarding matrix inertia that is important for the proof of Theorem 4.1 in Section 4.B.

4.A.2.1 Matrix Inertia

The following result from Lancaster and Tismenetsky (1985, Theorem 2, Chapter 13.1) is needed for the proof of Theorem 4.1 given in Section 4.B. We include it here for ease of reference. For a given matrix $A \in \mathbb{R}^{n \times n}$, $\upsilon_+(A)$, $\upsilon_-(A)$, and $\zeta(A)$ are the number of eigenvalues of the argument that have positive, negative and zero real parts, respectively.

Lemma 4.8 (Lancaster and Tismenetsky 1985, Theorem 2, Chapter 13.1)**.** *Consider a matrix* $A \in \mathbb{R}^{n \times n}$*.*

(a) If P *is a symmetric matrix such that* $AP + PA^\top = Q$ *where* $Q = Q^\top \succ 0$*, then* P *is nonsingular and* P *and* A *have the same inertia, meaning that*

$$\upsilon_+(A) = \upsilon_+(P),\ \upsilon_-(A) = \upsilon_-(P),\ \zeta(A) = \zeta(P). \tag{4.8}$$

(b) On the other hand, if $\zeta(A) = 0$*, then there exists a matrix* $P = P^\top$ *and a matrix* $Q = Q^\top \succ 0$ *such that* $AP + PA^\top = Q$ *and* P *and* A *have the same inertia so that* (4.8) *holds.*

4.B Stability Analysis: Proof of Theorem 4.1

We begin by proving the first claim of the theorem statement.

Proof of 1. Let us first consider the case that a given a critical point z^* could be such that $\mathtt{S}_1(J_1(z^*))$ or $-D_2^2 f(z^*)$ are singular. By Lemma 4.5 we know that for any critical point z^*, $-D_2^2 f(z^*) \succ 0$ so that $-D_2^2 f(z^*)$ is non-singular. Observe that at any critical point z^*, since $-\tau D_2^2 f(z^*)$ is positive definite for all $\tau \in (0, \infty)$, the following identity holds for any $\tau \in (0, \infty)$:

$$\det(J_\tau(z)) = \det(\mathtt{S}_1(J(z^*)))\det(-\tau D_2^2 f(z^*)).$$

From the fact that $\det(-\tau D_2^2 f(z^*))$, it then easily follows that $\det(J_\tau(z^*)) = 0$ if and only if $\det(\mathtt{S}_1(J(z^*))) = 0$. Note that $\det(J_\tau(z^*)) = 0$ if and only if $0 \in \operatorname{spec}(J_\tau(z^*))$ since eigenvalues of

a real square matrix are either purely real or come in complex conjugate pairs. Hence, given any critical point such that $\det(\mathtt{S}_1(J(z^*))) = 0$, then $0 \in \operatorname{spec}(J_\tau(z^*))$ and $\operatorname{spec}(-J_\tau(z^*)) \not\subset \mathbb{C}_-^\circ$ for all $\tau \in (0,\infty)$. Hence, any critical point such that $\mathtt{S}_1(J_1(z^*))$ is singular is unstable for all $\tau \in (0,\infty)$ and such a point is not a differential Stackelberg equilibrium.

Now, suppose that z^* is a critical point such that $\mathtt{S}_1(J_1(z^*))$ and $-D_2^2 f(z^*)$ are non-singular. Let $\operatorname{spec}(-J_{\tau_0}(z^*)) \subset \mathbb{C}_-^\circ$ for some $\tau_0 \in (0,\infty)$. We know that $-D_2^2 f(z^*) \succ 0$ by Lemma 4.5. We argue by contradiction that $\mathtt{S}_1(J_1(z^*)) \succ 0$. Towards this end, suppose not.

Since $\det(\mathtt{S}_1(J_1(z^*))) \neq 0$ and $\det(-D_2^2 f(z^*)) \neq 0$, by Lemma 4.8.b, there exists non-singular Hermitian matrices P_1, P_2 and positive definite Hermitian matrices Q_1, Q_2 such that $-\mathtt{S}_1(J_1(z^*))P_1 - P_1\,\mathtt{S}_1(J_1(z^*)) = Q_1$ and $D_2^2 f(z^*)P_2 + P_2 D_2^2 f(z^*) = Q_2$.

Furthermore, $-\mathtt{S}_1(J_1(z^*))$ and P_1 have the same inertia, meaning

$$\upsilon_+(-\mathtt{S}_1(J_1(z^*))) = \upsilon_+(P_1),\ \upsilon_-(-\mathtt{S}_1(J_1(z^*))) = \upsilon_-(P_1),\ \zeta(-\mathtt{S}_1(J_1(z^*))) = \zeta(P_1)$$

where for a given matrix A, $\upsilon_+(A)$, $\upsilon_-(A)$, and $\zeta(A)$ are the number of eigenvalues of the argument that have positive, negative and zero real parts, respectively. Similarly, $D_2^2 f(z^*)$ and P_2 have the same inertia:

$$\upsilon_+(D_2^2 f(z^*)) = \upsilon_+(P_2),\ \upsilon_-(D_2^2 f(z^*)) = \upsilon_-(P_2),\ \zeta(D_2^2 f(z^*)) = \zeta(P_2).$$

Since $-\mathtt{S}(J_1(z^*))$ has at least one strictly positive eigenvalue, $\upsilon_+(P_1) = \upsilon_+(-\mathtt{S}(J_1(z^*))) \geq 1$.

Define

$$P = \begin{bmatrix} I & L_0^\top \\ 0 & I \end{bmatrix} \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix} \begin{bmatrix} I & 0 \\ L_0 & I \end{bmatrix} \tag{4.9}$$

where $L_0 = (D_2^2 f(z^*))^{-1} D_{12}^\top f(z^*)$. Since P is congruent to $\operatorname{blockdiag}(P_1, P_2)$, by Sylvester's law of inertia (Horn and Johnson, 2012, Thm. 4.5.8), P and $\operatorname{blockdiag}(P_1, P_2)$ have the same inertia, meaning that $\upsilon_+(P) = \upsilon_+(\operatorname{blockdiag}(P_1, P_2))$, $\upsilon_-(P) = \upsilon_-(\operatorname{blockdiag}(P_1, P_2))$, and $\zeta(P) = \zeta(\operatorname{blockdiag}(P_1, P_2))$. Consider the matrix equation $-PJ_{\tau_0}(z^*) - J_{\tau_0}^\top(z^*)P = Q_{\tau_0}$ for $-J_{\tau_0}(z^*)$ where

$$Q_{\tau_0} = \begin{bmatrix} I & L_0^\top \\ 0 & I \end{bmatrix} B_{\tau_0} \begin{bmatrix} I & 0 \\ L_0 & I \end{bmatrix} \tag{4.10}$$

with

$$B_{\tau_0} = \begin{bmatrix} Q_1 & P_1 D_{12} f(z^*) - \mathtt{S}(J_1(z^*)) L_0^\top P_2 \\ (P_1 D_{12} f(z^*) - \mathtt{S}(J_1(z^*)) L_0^\top P_2)^\top & P_2 L_0 D_{12} f(z^*) + (P_2 L_0 D_{12} f(z^*))^\top + \tau_0 Q_2 \end{bmatrix}$$

which can be verified by straightforward calculations. The matrix B_{τ_0} is a symmetric matrix, and it is positive definite. Indeed, first observe that $Q_1 \succ 0$ and $Q_2 \succ 0$. Then showing $B_{\tau_0} \succ 0$ reduces to showing $P_2 L_0 D_{12} f(z^*) + (P_2 L_0 D_{12} f(z^*))^\top \succeq 0$, which is the case because $D_2^2 f(z^*) \prec 0$ implies that $P_2 \prec 0$ so that

$$P_2 (D_2^2 f(z^*))^{-1} D_{12}^\top f(z^*) D_{12} f(z^*) + (P_2 (D_2^2 f(z^*))^{-1} D_{12}^\top f(z^*) D_{12} f(z^*))^\top \succeq 0$$

since the produce of negative definite matrices is positive definite and so is the product of positive definite matrices. Now, since $B_{\tau_0} \succ 0$ so is Q_{τ_0} since they are congruent. Since $\operatorname{spec}(-J_{\tau_0}(z^*)) \subset$

$\mathbb{C}_-^\circ$, $Q_{\tau_0} \succ 0$ implies that $P = P^\top \prec 0$ (by Lyapunov's theorem). Hence, P_1 and P_2 must be negative definite since P is congruent to $\text{diag}(P_1, P_2)$, but this gives us a contradiction with the fact that P_1 has the same inertia as $-\mathtt{S}_1(J_1(z^*))$ which we assumed to have at least one positive eigenvalue. Hence, if $\text{spec}(-J_{\tau_0}(z^*)) \subset \mathbb{C}_-^\circ$ for some $\tau_0 \in (0, \infty)$, then it must be the case that $\mathtt{S}_1(J_1(z^*)) \succ 0$ which means z^* is a differential Stackelberg equilibrium since we also have $-D_2^2 f(z^*) \succ 0$ by Lemma 4.5.

Thus, we can finish the proof of part 1 as follows. Consider any critical point $\tilde{z}$ that is not a differential Stackelberg equilibrium and is unstable for the nominal τ_0. Then, we claim that $\text{spec}(-J_\tau(\tilde{z})) \not\subset \mathbb{C}_-^\circ$ for all $\tau \geq \tau_0$. Suppose not. That is, there is some $\tau_1 \geq \tau_0$ such that $\text{spec}(-J_{\tau_1}(\tilde{z})) \subset \mathbb{C}_-^\circ$. But by our argument above, since $-D_2^2 f(\tilde{x}) \succ 0$, this implies that $\mathtt{S}_1(J_1(\tilde{z})) \succ 0$ which contradicts that $\tilde{z}$ is not a differential Stackelberg equilibrium. Hence, any critical point z^* that is not a differential Stackelberg equilibrium is unstable for all $\tau \in (0, \infty)$.

Proof of 2. We note that the fact there exists a finite $\tau^* \in (0, \infty)$ such that a differential Stackelberg is stable is known (Fiez and Ratliff, 2020). Let τ^* denote the minimum τ^* such that a differential Stackelberg equilibrium z^* is stable. To see that $\text{spec}(-J_\tau(z^*)) \subset \mathbb{C}_-^\circ$ for all $\tau \geq \tau^*$ given that $\text{spec}(-J_{\tau^*}(z^*)) \subset \mathbb{C}_-^\circ$, we can again examine the Lyapunov equation under the congruent transformation. We define the matrix P as above in equation (4.9) where $P_1 \prec 0$ and $P_2 \prec 0$ since $-\mathtt{S}(J_1(z^*)) \prec 0$ and $D_2^2 f(z^*) \prec 0$ with $Q_1, Q_2 \succ 0$. With

$$Q_\tau = \begin{bmatrix} I & L_0^\top \\ 0 & I \end{bmatrix} B_\tau \begin{bmatrix} I & 0 \\ L_0 & I \end{bmatrix}$$

and

$$B_\tau = \begin{bmatrix} Q_1 & P_1 D_{12} f(z^*) - \mathtt{S}(J_1(z^*)) L_0^\top P_2 \\ (P_1 D_{12} f(z^*) - \mathtt{S}(J_1(z^*)) L_0^\top P_2)^\top & P_2 L_0 D_{12} f(z^*) + (P_2 L_0 D_{12} f(z^*))^\top + \tau Q_2 \end{bmatrix}$$

we again can see that $Q_\tau \succ 0$ for the same reason as above for any $\tau \geq \tau^*$. This, in turn, implies that $\text{spec}(-J_\tau(z^*)) \subset \mathbb{C}_-^\circ$ for all $\tau \geq \tau^*$ since we constructed a Lyapunov function for z^*. Thus, we conclude that if z^* is a differential Stackelberg equilibrium, then $\text{spec}(-J_\tau(z^*)) \subset \mathbb{C}_-^\circ$ for all $\tau \in [\tau_*, \infty)$ where τ_* is the minimum $\tau \in (0, \infty)$ such that $\text{spec}(-J_\tau(z^*)) \subset \mathbb{C}_-^\circ$ and a finite τ_* is guaranteed to exist.

4.C Global Asymptotic Convergence Analysis

We now provide the proofs pertaining to the results presented in Section 4.5.

4.C.1 Proof of Lemma 4.1

Let the best response be denoted by

$$y_*(x) \in \underset{y \in \mathcal{X}_2}{\arg\max}\, f(x, y).$$

Since the function $f(x, \cdot)$ is PŁ, there set maximizers may not be a singleton. Hence, $y_*(x)$ is an element of the set of maximizers.

We claim that for any $\Gamma \in (0, 1/7]$,

$$\Phi(x_{k+1}, y_{k+1}) - \Phi(x_k, y_k) = \Gamma \underbrace{(f(x_k, y_k) - f(x_{k+1}, y_{k+1}))}_{(i)} + \underbrace{f(x_{k+1}, y_*(x_{k+1})) - f(x_k, y_*(x_k))}_{(ii)} < 0$$

To show this, we need to bound each of the two terms (i) and (ii).

Bounding term (i): $f(x_k, y_k) - f(x_{k+1}, y_{k+1})$. To begin, we add and subtract $f(x_k, y_{k+1})$ to (i) and get

$$f(x_k, y_k) - f(x_{k+1}, y_{k+1}) = f(x_k, y_k) - f(x_k, y_{k+1}) + f(x_k, y_{k+1}) - f(x_{k+1}, y_{k+1}). \tag{4.11}$$

From a Taylor expansion of $-f(x_k, y_{k+1})$ with respect to y_{k+1} we obtain

$$\begin{aligned} -f(x_k, y_{k+1}) &\leq -f(x_k, y_k) - \langle D_2 f(x_k, y_k), y_{k+1} - y_k \rangle + \frac{L_2}{2} \|y_{k+1} - y_k\|_2^2 \\ &= -f(x_k, y_k) - \tau\gamma \langle D_2 f(x_k, y_k), D_2 f(x_k, y_k) \rangle + \frac{\tau^2 \gamma^2 L_2}{2} \|D_2 f(x_k, y_k)\|_2^2 \\ &= -f(x_k, y_k) - \Big(\tau\gamma - \frac{\tau^2 \gamma^2 L_2}{2}\Big) \|D_2 f(x_k, y_k)\|_2^2. \end{aligned}$$

Hence, rearranging this bound and combining with (4.11), we get

$$f(x_k, y_k) - f(x_{k+1}, y_{k+1}) \leq -\Big(\tau\gamma - \frac{\tau^2 \gamma^2 L_2}{2}\Big) \|D_2 f(x_k, y_k)\|_2^2 + f(x_k, y_{k+1}) - f(x_{k+1}, y_{k+1}). \tag{4.12}$$

Now, from a Taylor expansion of $-f(x_{k+1}, y_{k+1})$ with respect to x_{k+1}, we get

$$\begin{aligned} -f(x_{k+1}, y_{k+1}) &\leq -f(x_k, y_{k+1}) - \langle D_1 f(x_k, y_{k+1}), x_{k+1} - x_k \rangle + \frac{L_1 \gamma^2}{2} \|x_{k+1} - x_k\|_2^2 \\ &= -f(x_k, y_{k+1}) + \gamma \langle D_1 f(x_k, y_{k+1}), D_1 f(x_k, y_k) \rangle + \frac{L_1 \gamma^2}{2} \|D_1 f(x_k, y_k)\|_2^2. \end{aligned}$$

We now add and subtract $\gamma \|D_1 f(x_k, y_k)\|_2^2$ to obtain

$$\begin{aligned} f(x_k, y_{k+1}) - f(x_{k+1}, y_{k+1}) &\leq \gamma D_1 f(x_k, y_{k+1})^\top D_1 f(x_k, y_k) + \frac{L_1 \gamma^2}{2} \|D_1 f(x_k, y_k)\|_2^2 \\ &\quad + \gamma D_1 f(x_k, y_k)^\top D_1 f(x_k, y_k) - \gamma \|D_1 f(x_k, y_k)\|_2^2 \\ &= \gamma \langle D_1 f(x_k, y_{k+1}) - D_1 f(x_k, y_k), D_1 f(x_k, y_k) \rangle \\ &\quad + \frac{L_1 \gamma^2 + 2\gamma}{2} \|D_1 f(x_k, y_k)\|_2^2. \end{aligned} \tag{4.13}$$

Next, we combine this with (4.12) to get

$$\begin{aligned} f(x_k, y_k) - f(x_{k+1}, y_{k+1}) &\leq -\Big(\tau\gamma - \frac{\tau^2\gamma^2 L_2}{2}\Big)\|D_2 f(x_k, y_k)\|^2 + \frac{L_1\gamma^2 + 2\gamma}{2}\|D_1 f(x_k, y_k)\|_2^2 \\ &\quad + \gamma\langle D_1 f(x_k, y_{k+1}) - D_1 f(x_k, y_k), D_1 f(x_k, y_k)\rangle \\ &\leq -\Big(\tau\gamma - \frac{\tau^2\gamma^2 L_2}{2}\Big)\|D_2 f(x_k, y_k)\|^2 + \frac{L_1\gamma^2 + 2\gamma}{2}\|D_1 f(x_k, y_k)\|_2^2 \\ &\quad + \gamma\|D_1 f(x_k, y_{k+1}) - D_1 f(x_k, y_k)\|\|D_1 f(x_k, y_k)\| \end{aligned}$$

where we used Cauchy-Schwartz in the last inequality. Applying Young's inequality on the last term, we have that

$$\begin{aligned} f(x_k, y_k) - f(x_{k+1}, y_{k+1}) &\leq -\left(\tau\gamma - \frac{\tau^2\gamma^2 L_2}{2}\right)\|D_2 f(x_k, y_k)\|^2 + \frac{L_1\gamma^2 + 2\gamma + \gamma}{2}\|D_1 f(x_k, y_k)\|_2^2 \\ &\quad + \frac{\gamma}{2}\|D_1 f(x_k, y_{k+1}) - D_1 f(x_k, y_k)\|^2 \\ &\leq -\left(\tau\gamma - \frac{\tau^2\gamma^2 L_2}{2} - \frac{\gamma(\tau\gamma)^2 L_2^2}{2}\right)\|D_2 f(x_k, y_k)\|^2 \\ &\quad + \frac{L_1\gamma^2 + 2\gamma + \gamma}{2}\|D_1 f(x_k, y_k)\|_2^2. \end{aligned} \tag{4.14}$$

Bounding (ii): $f(x_{k+1}, y_*(x_{k+1})) - f(x_k, y_*(x_k))$. To bound (ii), we take a Taylor expansion of $f(x_{k+1}, y_*(x_{k+1}))$ to get that

$$f(x_{k+1}, y_*(x_{k+1})) - f(x_k, y_*(x_k)) \leq \langle Df(x_k, y_*(x_k)), x_{k+1} - x_k\rangle + \frac{L_4\gamma^2}{2}\|D_1 f(x_k, y_k)\|^2 \tag{4.15}$$

$$= -\gamma\langle Df(x_k, y_*(x_k)), D_1 f(x_k, y_k)\rangle + \frac{L_4\gamma^2}{2}\|D_1 f(x_k, y_k)\|^2 \tag{4.16}$$

where $L_4 \leq L_1 + L_2 L_3$ is the Lipschitz bound on the total derivative of $f(x, y_*(x))$.[6]

Combining bounds. We now combine the bounds on (i) and (ii). Combining the (4.14) and (4.16), we have that

$$\begin{aligned} \Phi(x_{k+1}, y_{k+1}) - \Phi(x_k, y_k) &\leq -\gamma\langle Df(x_k, y_*(x_k)), D_1 f(x_k, y_k)\rangle \\ &\quad + \frac{\gamma}{2}\left(\frac{L_4\gamma + \Gamma L_1\gamma + 3\Gamma}{2}\right)\|D_1 f(x_k, y_k)\|^2 \\ &\quad - \Gamma\left(\tau\gamma - \frac{\tau^2\gamma^2 L_2}{2} - \frac{\gamma(\tau\gamma)^2 L_2^2}{2}\right)\|D_2 f(x_k, y_k)\|^2. \end{aligned} \tag{4.17}$$

To further bound the above expression, we start by bounding the first two terms. Towards this

[6] See Lemma 4.7 for the derivation of L_3.

end, define

$$V = -\gamma\langle Df(x_k, y_*(x_k)), D_1 f(x_k, y_k)\rangle + \frac{\gamma}{2}(L_4\gamma + \Gamma L_1\gamma + 3\Gamma)\|D_1 f(x_k, y_k)\|_2^2. \tag{4.18}$$

Recall that $\gamma < 1/(2L_4)$ and $\gamma < 1/(2\max\{L_1, L_2\})$. Since $\Gamma \leq 1/7$,

$$\gamma(L_4 + \Gamma L_1) + \Gamma 3 \leq \frac{1}{2} + \frac{7\Gamma}{2} \leq 1.$$

Completing the square in (4.18), we have that

$$\begin{aligned} V &\leq -\gamma\langle Df(x_k, y_*(x_k)), D_1 f(x_k, y_k)\rangle + \frac{\gamma}{2}\|D_1 f(x_k, y_k)\|^2 \\ &\leq -\frac{\gamma}{2}\|Df(x_k, y_*(x_k))\|^2 + \frac{\gamma}{2}\|Df(x_k, y_*(x_k)) - D_1 f(x_k, y_k)\|^2 \\ &\leq -\frac{\gamma}{2}\|Df(x_k, y_*(x_k))\|^2 + \frac{\gamma L_2^2}{2}\|y_k - y_*(x_k)\|^2 \qquad (4.19) \\ &\leq -\frac{\gamma}{2}\|Df(x_k, y_*(x_k))\|^2 + \frac{\gamma\kappa^2}{2}\|D_2 f(x_k, y_k)\|^2 \qquad (4.20) \end{aligned}$$

where in (4.19) we used the fact that $D_1 f(x, y)$ is Lipschitz in y and $Df(x_k, y_*(x_k)) = D_1 f(x, y)|_{y=y_*(x)}$ by Lemma 4.6. Further, in (4.20), we used the quadratic growth property of PŁ functions.

Now, by combining the bound on V with the remaining terms in (4.17), we have

$$\begin{aligned} \Phi(x_{k+1}, y_{k+1}) - \Phi(x_k, y_k) \leq &-\frac{\gamma}{2}\|Df(x_k, y_*(x_k))\|^2 \\ &+ \tau\gamma\left(\frac{\kappa^2}{2\tau} - \Gamma + \frac{\Gamma\tau\gamma L_2}{2} + \frac{\Gamma\tau\gamma^2 L_2^2}{2}\right)\|D_2 f(x_k, y_k)\|^2. \end{aligned} \tag{4.21}$$

Let

$$C = -\Gamma + \frac{\Gamma\tau\gamma L_2}{2} + \frac{\Gamma\tau\gamma^2 L_2^2}{2} + \frac{\kappa^2}{2\tau}$$

As long as $C < 0$, as claimed, $\Phi(\cdot)$ will be a potential function. To see this, we upper bound C as follows:

$$C = \frac{\kappa^2}{2\tau} - \Gamma\left(1 - \frac{\tau\gamma}{2}\left(L_2 + \gamma L_2^2\right)\right) \leq \frac{\kappa^2}{2\tau} - \Gamma\left(1 - \tau\gamma\frac{3}{4}L_2\right) \leq \frac{\Gamma}{2}\left(\frac{\kappa^2}{\Gamma\tau} - 2 + \frac{3}{2}\tau\gamma L_2\right)$$

since $\gamma < 1/(2\max\{L_1, L_2\})$. Since $\tau > \Gamma^{-1}\kappa^2$ and $\gamma \leq 1/(3L_2\tau)$, we have

$$-C \leq \frac{\Gamma}{2}\left(\frac{\kappa^2}{\Gamma\tau} - 2 + \frac{3}{2}\tau\gamma L_2\right) \leq \frac{\Gamma}{2}\left(\frac{3}{2}\tau\gamma L_2 - 1\right) < -\frac{\Gamma}{4}.$$

Hence,

$$\Phi(x_{k+1}, y_{k+1}) - \Phi(x_k, y_k) \leq -\frac{\gamma}{2}\|Df(x_k, y_*(x_k))\|^2 - \frac{\tau\gamma\Gamma}{2}\|D_2 f(x_k, y_k)\|^2$$

which completes the proof of the claim that Φ is a potential function since it is decreasing along trajectories.

4.C.2 Proof of Theorem 4.2

This result follows nearly immediately from Theorem 4.1, Lemma 4.1, and Lemma 4.2. In particular, the potential function Φ from Lemma 4.1 guarantees that

$$\Phi(x_{k+1}, y_{k+1}) - \Phi(x_k, y_k) \leq -\frac{\gamma}{2}\|Df(x_k, y_*(x_k))\|^2 - \frac{\tau\gamma\Gamma}{2}\|D_2 f(x_k, y_k)\|^2.$$

Thus, the potential function only stops decreasing when we have both

$$\|Df(x_k, y_*(x_k))\|^2 = 0 \quad \text{and} \quad \|D_2 f(x_k, y_k)\|^2 = 0.$$

By the quadratic growth of μ-PŁ functions (see Lemma 4.4) and the PŁ property in nonconvex-PŁ zero-sum games, we have that

$$\|y_k - y_*(x_k)\|^2 \leq \frac{2}{\mu}(\max_y f(x_k, y) - f(x_k, y_k)) \leq \frac{2}{\mu^2}\|D_2 f(x_k, y_k)\|^2$$

Hence, if $\|D_2 f(x_k, y_k)\|^2 \to 0$ then $y_k \to y_*(x_k)$. In particular, when $\|D_2 f(x_k, y_k)\|^2 = 0$, we have that $y_k = y_*(x_k)$ so that $\|Df(x_k, y_*(x_k))\|^2 = 0$ if and only if $\|D_1 f(x_k, y_k)\|^2 = 0$. This implies that the potential function only stops decreasing at critical points and this is the only place that a limit cycle could exist. Since the only stable points of $\dot{z} = -\Lambda_\tau(z)$ are differential Stackelberg equilibrium by Theorem 4.1 and τ-GDA avoids strict saddle points of $\dot{z} = -\Lambda_\tau(z)$ almost surely by Lemma 4.2 and the assumption that all saddle points are strict saddles, we can conclude that τ-GDA almost surely only reaches a critical point if it is a differential Stackelberg equilibrium. Thus, these facts ensure that the τ-GDA dynamics almost surely converge to a differential Stackelberg equilibrium.

4.C.3 Proof of Corollary 4.1

For this proof, consider any $\varepsilon > 0$. Our approach will be to show that for

$$T \geq \frac{2(\Phi(x_0, y_0) - \Phi(x_T, y_T))}{\varepsilon^2 \gamma \min\{1, \tau\Gamma\}},$$

we have that

$$\min_{0 \leq k \leq T-1} \max\Big\{\|Df(x_k, y_*(x_k))\|, \|D_2 f(x_k, y_k)\|\Big\} := \max\Big\{\|Df(x_s, y_*(x_s))\|, \|D_2 f(x_s, y_s)\|\Big\} \leq \varepsilon.$$

Then, we prove that given this fact[7],

$$\|Df(x_s, y_s)\| \leq \Big(1 + \frac{L_5}{\mu}\Big)\varepsilon.$$

[7]See Assumption 4.7 for the derivation of L_5.

This will then allow us to conclude for

$$T \geq \frac{2\Big(1+\frac{L_5}{\mu}\Big)^2(\Phi(x_0,y_0)-\Phi(x_T,y_T))}{\varepsilon^2\gamma\min\{1,\tau\Gamma\}}, \tag{4.22}$$

we have both

$$\|Df(x_s,y_s)\| \leq \varepsilon \quad \text{and} \quad \|D_2 f(x_s,y_s)\| \leq \varepsilon.$$

Then, by selecting the parameters to minimize the right-hand side of (4.22), we are able to conclude that at least one iterate of the τ-GDA dynamics are an ε-differential Stackelberg equilibrium after

$$T \geq \frac{2\Big(1+\frac{L_5}{\mu}\Big)^2(\Phi(x_0,y_0)-\Phi(x_T,y_T))}{\varepsilon^2\gamma\min\{1,\tau\Gamma\}}$$

iterations.

We now formally prove this. Summing the bound on the potential function from Lemma 4.1, we get the following that is justified below:

$$\Phi(x_0,y_0)-\Phi(x_T,y_T) = \sum_{k=0}^{T-1}\Big(\Phi(x_k,y_k)-\Phi(x_{k+1},y_{k+1})\Big) \tag{4.23}$$

$$\geq \frac{\gamma}{2}\sum_{k=0}^{T-1}\|Df(x_k,y_*(x_k))\|^2 + \frac{\tau\gamma\Gamma}{2}\sum_{k=0}^{T-1}\|D_2 f(x_k,y_k)\|^2 \tag{4.24}$$

$$\geq \frac{\gamma}{2}\min\{1,\tau\Gamma\}\sum_{k=0}^{T-1}\Big(\|Df(x_k,y_*(x_k))\|^2 + \|D_2 f(x_k,y_k)\|^2\Big) \tag{4.25}$$

$$\geq \frac{\gamma}{2}\min\{1,\tau\Gamma\}\sum_{k=0}^{T-1}\max\Big\{\|Df(x_k,y_*(x_k))\|^2, \|D_2 f(x_k,y_k)\|^2\Big\} \tag{4.26}$$

$$\geq \frac{\gamma T}{2}\min\{1,\tau\Gamma\}\min_{0\leq k\leq T-1}\max\Big\{\|Df(x_k,y_*(x_k))\|^2, \|D_2 f(x_k,y_k)\|^2\Big\}. \tag{4.27}$$

Observe that (4.23) follows from telescoping of the sum, (4.24) is a result of applying the bound on the potential function, (4.25) holds since it is replacing a coefficient of a positive number with something smaller, (4.26) holds since the sum of positive numbers is greater than the max, and (4.27) is obtained from the fact that the sum of T positive numbers is greater than T times the minimum number.

From the previous steps, and also rearranging terms and then taking the square root, we have

$$\sqrt{\frac{2(\Phi(x_0,y_0)-\Phi(x_T,y_T))}{T\gamma\min\{1,\tau\Gamma\}}} \geq \min_{0\leq k\leq T-1}\max\Big\{\|Df(x_k,y_*(x_k))\|, \|D_2 f(x_k,y_k)\|\Big\}.$$

We now want to find T such that

$$\min_{0\leq k\leq T-1} \max\Big\{\|Df(x_k, y_*(x_k))\|, \|D_2 f(x_k, y_k)\|\Big\} \leq \sqrt{\frac{2(\Phi(x_0, y_0) - \Phi(x_T, y_T))}{T\gamma \min\{1, \tau\Gamma\}}} \leq \varepsilon. \tag{4.28}$$

By moving terms around, we find that the inequality in (4.28) holds for any T such that

$$T \geq T^* := \frac{2(\Phi(x_0, y_0) - \Phi(x_T, y_T))}{\varepsilon^2 \gamma \min\{1, \tau\Gamma\}}.$$

This proves that there exists some iterate $0 \leq s \leq T-1$ such that for $T \geq T^*$, we have both

$$\|Df(x_s, y_*(x_s))\| \leq \varepsilon \quad \text{and} \quad \|D_2 f(x_s, y_s)\| \leq \varepsilon.$$

We now show that this implies a bound on $\|Df(x_s, y_s)\|$. In particular, using the fact that $f(x, \cdot)$ is μ-PŁ, we have that

$$\|y_s - y_*(x_s)\|^2 \leq \frac{2}{\mu}\Big(\max_{y\in\mathcal{X}_2} f(x_s, y) - f(x_s, y_s)\Big) = \frac{2}{\mu}\Big(f(x_s, y_*(x_s)) - f(x_s, y_s)\Big) \leq \frac{2}{\mu^2}\|D_2 f(x_s, y_s)\|^2.$$

Since $\|D_2 f(x_s, y_s)\| \leq \varepsilon$, we know that $\|y_s - y_*(x_s)\| \leq \frac{\sqrt{2}\epsilon}{\mu}$. Then, observe that we have

$$\begin{aligned}
\|Df(x_s, y_s)\| &= \|Df(x_s, y_s) - Df(x_s, y_*(x_s)) + Df(x_s, y_*(x_s))\| \\
&\leq \|Df(x_s, y_s) - Df(x_s, y_*(x_s))\| + \|Df(x_s, y_*(x_s))\| \\
&\leq L_5\|y_s - y_*(x_s)\| + \|Df(x_s, y_*(x_s))\| \\
&\leq \frac{\sqrt{2}L_5\varepsilon}{\mu} + \varepsilon = \Big(1 + \frac{\sqrt{2}L_5}{\mu}\Big)\varepsilon,
\end{aligned}$$

where we obtain the first inequality using the triangle inequality, the second inequality using the Lipschitz bound, and the final inequality using that $\|D_2 f(x_s, y_s)\| \leq \varepsilon$ and $\|y_s - y_*(x_s)\| \leq \frac{\sqrt{2}\epsilon}{\mu}$.

Thus, in order to determine the iteration complexity T needed to get that $\|Df(x_s, y_s)\| \leq \varepsilon$ we redefine the given ε to be $\varepsilon = \varepsilon'(1 + \frac{\sqrt{2}L_5}{\mu})$ and get that for

$$T \geq T^* := \frac{2\Big(1 + \frac{\sqrt{2}L_5}{\mu}\Big)^2(\Phi(x_0, y_0) - \Phi(x_T, y_T))}{\varepsilon^2 \gamma \min\{1, \tau\Gamma\}},$$

we have both

$$\|Df(x_s, y_s))\| \leq \varepsilon \quad \text{and} \quad \|D_2 f(x_s, y_s)\| \leq \varepsilon.$$

To finish, we select $\tau = 8\kappa^2$, $\Gamma = 1/8$, and $\gamma = \frac{1}{2}\min\{\frac{1}{3L_2\tau}, \frac{1}{2(L_1 + L_3 L_2)}\}$ to get an iteration complexity of.

$$T^* = \frac{4\Big(1 + \frac{\sqrt{2}L_4}{\mu}\Big)^2 \max\{24L_2\kappa^2, 2(L_1 + L_3 L_2)\}(\Phi(x_0, y_0) - \Phi(x_T, y_T)}{\varepsilon^2}.$$

Thus the iteration complexity is $\widetilde{\mathcal{O}}(\varepsilon^{-2})$ as claimed to reach an ε-differential Stackelberg equilibrium. We note that the assumption of initializing in the region of attraction of a differential Stackelberg equilibrium is based on the the stability result from Theorem 4.1 that guarantees there exists a local neighborhood on which τ–`GDA` converges toward the equilibrium that is stable.

Part II

Sequential Decision-Making under Uncertainty

Chapter 5

Sequential Decision-Making in Peer Review Bidding Systems

This chapter being our study of sequential decision-making problems. A number of applications involve sequential arrival of users, and require showing each user an ordering of items. A prime example (which forms the focus of this chapter) is the bidding process in conference peer review where reviewers enter the system sequentially, each reviewer needs to be shown the list of submitted papers, and the reviewer then "bids" to review some papers. The order of the papers shown has a significant impact on the bids due to primacy effects. In deciding on the ordering of papers to show, there are two competing goals: (i) obtaining sufficiently many bids for each paper, and (ii) satisfying reviewers by showing them relevant items. In this chapter, we begin by developing a framework to study this problem in a principled manner. We present an algorithm called SUPER*, inspired by the A* algorithm, for this goal. Theoretically, we show a local optimality guarantee of our algorithm and prove that popular baselines are considerably suboptimal. Moreover, under a community model for the similarities, we prove that SUPER* is near-optimal whereas the popular baselines are considerably suboptimal. In experiments on real data from ICLR 2018 and synthetic data, we find that SUPER* considerably outperforms baselines deployed in existing systems, consistently reducing the number of papers with fewer than requisite bids by 50-75% or more, and is also robust to various real world complexities.

5.1 Introduction

It is well known that peer review is essential for ensuring the quality and scientific value of research (Bianchi and Squazzoni, 2015; Black et al., 1998; Thurner and Hanel, 2011). A fundamental challenge in peer review is matching or assigning papers to qualified and willing reviewers. Common methods to deal with this problem often rely on access to a *similarity matrix* containing scores for each paper-reviewer pair expressing the estimated match quality between them. The similarity matrix is often obtained using feature-based or profile-based matching mechanisms that leverage keywords and available reviewer publications (Charlin and Zemel, 2013; Price and Flach, 2017). A number of automated methods to match papers with reviewers using similarity scores have been proposed that optimize objectives such as cumulative similarity or fairness notions (Garg et al., 2010; Karimzadehgan et al., 2008; Long et al., 2013; Stelmakh et al., 2019b; Tang et al., 2010).

A shortcoming of automating the paper matching process stems from the failure to actively incorporate reviewers within the paper assignment phase of the review process. The outright dependence on the similarity scores can be problematic since the preferences of reviewers can change frequently and the similarity scores themselves can be noisy. Bidding has emerged as a mechanism for aiding in and improving the peer review process under the guise that active engagement of the

reviewer leads to assignments more aligned with their preferences and higher review quality (Di Mauro et al., 2005).

In typical peer review process, when the bidding process opens, reviewers enter the system in an arbitrary sequential order. Upon entering, a list of papers is shown to them and they are asked to place bids on papers they would prefer to review. Following the bidding process, bids can be incorporated into the reviewer-paper assignment mechanism. It is known that the order of papers presented to reviewers in the bidding stage can greatly impact the number of bids that a paper receives (Cabanac and Preuss, 2013). From the perspective of the platform, there are two competing goals: (i) ensure that each paper has a sufficient number of bids, and (ii) ensure individual reviewer satisfaction by showing relevant papers.

With regard to goal (i), the platform aims to select a display order for each reviewer such that at the end of the bidding process, each paper has at least a certain number of bids. The main objective of ensuring a minimum number of bids on each paper is to improve review quality for all papers (Shah et al., 2018). The well-documented primacy effect (Murphy et al., 2006) suggests that papers shown on top of the ordering are the ones on which reviewers are more likely to bid. Consequently, this objective strongly suggests that papers with few bids should be placed higher in the list. Indeed, Cabanac and Preuss (2013) make the following remark:

> *"It is advised to counterbalance order effects during the bidding phase of peer review by promoting the submissions with fewer bids to potential referees. This manipulation intends to better share bids out among submissions in order to attract qualified referees for all submissions."*

With regard to goal (ii), the platform aims to display 'well-matched' papers to each reviewer. That is, the set of papers to be displayed is composed of papers on which the reviewer is most likely to bid. There are several reasons to select well-matched papers. It is generally assumed that reviewers are more likely to place bids on papers they are qualified to review (Rodriguez et al., 2007). Furthermore, reviewers that place positive bids on papers are more likely to give a review with high confidence and voice sharp opinions of acceptance or rejection that help guide final decisions on papers (Cabanac and Preuss, 2013). A number of comprehensive surveys also indicate that a primary motivation of reviewers is the ability to help and contribute to the work of colleagues (Mulligan et al., 2013; Ware, 2008). Failing to display relevant papers to reviewers can result in several unintended negative consequences. If irrelevant papers are shown early in the order to a reviewer, it may cause the reviewer to opt-out and disengage with the system even if further down the list there was an option that they would have happily bid on. Similarly, a poorly selected ordering may result in significantly fewer bids from a reviewer.

Competing objectives of this form are not unique to peer review systems and they appear in a number of applications. A fitting example is an intermediary between distinct user groups that seeks to facilitate interactions and satisfy each party. For example, in online labor markets, the platform must ensure each job obtains a sufficient number of applicants and that workers are presented with tasks they are qualified enough for to be considered. Similarly, in online e-commerce marketplaces, as the platform decides how to show products to users, there is a definite trade-off between satisfying merchants offering products that need to be sold and users that want to be shown relevant items. We maintain peer review as a running example and comment further on relevant applications at the end of this chapter.

In peer review, it is recognized that actively engaging reviewers in the paper assignment process via bidding can greatly improve the review process. If administered inadequately, bidding can in fact have a significant negative impact on the quality of the review process. In the words of Rodriguez et al. (2007),

> *"Since bidding is the preliminary component of the manuscript-to-referee matching algorithm, sloppy bidding can have dramatic effects on which referees actually review which submissions."*

A study on the 2016 Neural Information Processing Systems (NeurIPS) conference revealed the distribution of bids arising from a typical bidding process leaves significant challenges to match papers with reviewers (Shah et al., 2018). It was observed that a considerable number of reviewers do not place a sufficient number of bids and papers commonly fail to obtain as many bids as the number of reviewers needed. This phenomenon is detailed in (Shah et al., 2018) amongst the 3,200 reviewers and 2,400 papers.

> *"Moreover, there are 148 reviewers with no (positive or negative) bids and 1201 reviewers with at most 2 positive bids... We thus observe that a large number of reviewers do not even provide positive bids amounting to the number of papers they would review. As a consequence of the low number of bids by reviewers, we are left with 278 papers with at most 2 positive bids and 816 papers with at most 5 positive bids... There is thus a significant fraction of papers with fewer positive bids than the number of requisite reviewers."*

The study also found that there were 1090 papers with no positive bids from the area chairs. The inability to elicit meaningful bidding information in NeurIPS is far from an aberration. In a study of the 2005 Joint Conference on Digital Libraries, 146 out of the 264 submissions did not obtain any bids (Rodriguez et al., 2007). The shortfalls of existing bidding systems shift the onus of the reviewing assignments away from the participants and to the paper matching mechanisms.

Despite the importance of the bidding process in peer review, there is not yet much fundamental research on the problem of optimizing the display order during the bidding process, and much less so in consideration of the two objectives identified thus far. In practice, the display order is typically determined via heuristics such as a fixed ordering (e.g., order of submission), or in decreasing order of the relevance of the papers to that reviewer, or in increasing order of the number of bids received by the paper until then.

A key reason that bidding can fail is that papers are suboptimally displayed to the reviewers. Consider a paper that is not an ideal match for any reviewer in the system. If papers are ranked for display simply by how well-matched they are to reviewers, this particular paper may be shown far down in the ranking for each reviewer and hence, not receive many, if any, bids. The risk of this scenario is elevated for interdisciplinary research, which is know to face significant impediments as a consequence of the lack of ideally matched peers (Porter and Rossini, 1985; Travis and Collins, 1991).

On the other hand, if papers are inversely ranked by the number of bids they have obtained, then papers with fewer bids are more likely to be shown higher on the list regardless of how well-matched they are to any particular reviewer. This display order may cause reviewer dissatisfaction,

which in the worst case could result in zero bids. Similarly, ordering heuristics that are based on a fixed baseline may lead to bias in the review process. Indeed, in the report of a study of 42 peer-reviewed conferences in Computer Science, it was observed that under a fixed ordering (based on the submission time), the number of bids on papers is heavily influenced by the order of submission times of the papers (Cabanac and Preuss, 2013). It was concluded that the later the paper is submitted, the fewer bids it will receive.

Given the flaws of existing peer review bidding systems, we study the important problem of selecting the ordering of papers to display to each arriving reviewer in a principled manner.

5.1.1 Contributions

The key contributions of this chapter are summarized as follows.

Problem identification and formulation (Section 5.2). The bidding process is highly consequential, yet one of the most understudied components of the conference peer-review process. We identify a key source of unfairness and inefficiency in the bidding process, and develop principled methods to address it. A key challenge is suitably formalizing the peer review bidding process, for which to the best of our knowledge there are no prior formulations. We formulate an objective function that captures the competing goals of the platform while reflecting the underlying decision-making process of reviewers. The framework developed in this chapter to analyze the problem is an important step toward future improvements on bidding systems.

Algorithm design (Section 5.3). We present a sequential decision-making algorithm called SUPER* to address this problem. The algorithm takes as input the "similarities" between each reviewer-paper pair and the bids made by all past reviewers, and outputs the ordering of papers to show to any current reviewer.

Theoretical results (Section 5.4). We show two sets of theoretical results. We first consider a notion of 'local' performance: the performance with respect to a single reviewer. We prove that SUPER* is locally optimal whereas all popular baselines are considerably suboptimal. Our second set of theoretical results are based on a community model, where we prove that SUPER* is near-optimal (globally) and all popular baselines are considerably suboptimal.

Experiments on real and synthetic data (Section 5.5). We run extensive experiments using similarity scores from ICLR 2018 and on synthetic data. The experiments reveal that the SUPER* algorithm outperforms all popular baselines. For instance, it consistently reduces the number of papers with fewer than requisite bids by 50-75% while maintaining individual reviewer satisfaction. In addition, we see that SUPER* is very robust to model mismatches and complexities of the real-world review process.

We remark that the proofs of theoretical results are left to an appendix that follows the chapter.

5.1.2 Related Work

The paper ordering problem for the bidding process in peer review bears a strong resemblance to the learning to rank problem (Aslanyan and Porwal, 2019; Cao et al., 2007; Momma et al., 2019; Singh and Joachims, 2019; Svore et al., 2011; Yadav et al., 2019). Typically, the goal of learning to rank is to learn an overall ranking of items via supervised methods or by querying users, where the

latter provides further information on the relative ranking of items. In peer review, the objective of finding a ranking most suitable for an arriving reviewer during the bidding process is analogous to learning to rank methods that consider the utility of rankings for users along with the impact on the items being ranked (Singh and Joachims, 2019; Yadav et al., 2019). Moreover, the bidding model considered in this work is motivated from that which is commonly adopted in learning to rank models (Aslanyan and Porwal, 2019).

As formulated in this chapter, the goal for the design of the bidding process in peer review is to optimize for multiple criteria reflecting the objectives of the reviewers and the papers, respectively. This is not unlike the methods of Singh and Joachims (2019) and Yadav et al. (2019), which consider a fairness objective in combination with a ranking quality objective, or the multi-objective learning to rank problems studied by Svore et al. (2011) and Momma et al. (2019). In the works of Singh and Joachims (2019) and Yadav et al. (2019), the objective of ensuring fairness is encoded as a constraint in the optimization problems. Similarly, Svore et al. (2011) optimize a linear combination of ranking measures referred to as a 'graded measure' and (Momma et al., 2019) convert a constrained optimization problem into an unconstrained problem by penalizing constraint violations in the objective. In each of the aforementioned works, the ranking measures are separable in the arriving users, meaning that the contribution of any individual user to the overall objective is independent of the other users.

The problem we study is also abstractly similar to online recommendation systems facing competing objectives (Agarwal et al., 2011; Jambor and Wang, 2010; Rodriguez et al., 2012). However, a prevailing approach is to convert the multi-objective problem to a constrained optimization problem (Jambor and Wang, 2010; Rodriguez et al., 2012). Both the approach of incorporating objectives as constraints in the optimization problem formulations and combining objectives in a linear fashion is considered by Agarwal et al. (2011). Analogous to the learning to rank problem, the objectives are separable in the users.

The objective in the peer review problem as formulated in this work presents unique challenges not addressed in the aforementioned works on learning to rank and recommendation systems. Notably, it is not separable between the reviewers since it depends on the number of bids on each paper after each reviewer has arrived and placed bids on the papers. Being applicable to more general multi-criteria settings, our approach to the design of the bidding processes in peer review may also be applied to the learning to rank problem. This is a direction worthy of further study.

Our work also contributes to a growing literature on improving various aspects of the peer review process such as reviewer assignment (Charlin and Zemel, 2013; Garg et al., 2010; Kobren et al., 2019; Lian et al., 2018; Stelmakh et al., 2019b), biases (Stelmakh et al., 2019a; Tomkins et al., 2017), subjectivity (Noothigattu et al., 2018), miscalibration (Roos et al., 2012; Wang and Shah, 2019), strategic behavior (Aziz et al., 2019; Xu et al., 2019), and others (Church, 2005; Ding et al., 2020; Jecmen et al., 2020; Kang et al., 2018; Lawrence and Cortes, 2014; Shah et al., 2018; Stelmakh et al., 2021a,b; Wing, 2011). This work addresses the bidding process in conference peer review, which has largely been unexplored in past literature. The concurrent work of (Meir et al., 2020), which appeared after an initial workshop version of our work (Fiez et al., 2019d), is the only work besides our own that we are aware of to focus on methods for improving bidding in peer review. However, their approach is to design a market for bidding, which is entirely different from ours.

5.2 Problem Formulation

Consider $d \geq 2$ papers and $n \geq 2$ reviewers indexed as $\{1, \ldots, d\}$ and $\{1, \ldots, n\}$ respectively.[1] For each reviewer-paper pair, we have access to a *similarity score* that captures the similarity between the reviewer and the paper. We use the notation $S_{i,j} \in [0, 1]$ to denote the given similarity between any reviewer $i \in [n]$ and paper $j \in [d]$. A higher similarity score indicates a greater relevance of the paper to that reviewer. There are several systems in use today that compute similarities (Charlin and Zemel, 2013; Price et al., 2010), and in our work, we treat them as being given.

In the bidding period, reviewers sequentially arrive into the system and place bids on the papers. In our work, for any reviewer and paper, we only consider the existence of a bid or not, and do not consider the possibility of multiple bidding options. We assume for simplicity that all n reviewers arrive exactly once, and that a reviewer arrives after the previous reviewer has completed their bidding.[2] We do not make any assumptions on the arrival order of the reviewers. The problem is to determine the ordering of papers to show each reviewer on arrival in the interest of influencing the papers they decide to bid on while ensuring individual satisfaction. When deciding the paper ordering for any reviewer, the bids made by all reviewers who arrived in the past along with the paper orderings presented to them are known, but the bids made by the current or future reviewers are unknown. Let Π_d denote the set of all possible $d!$ permutations of the d papers. In what follows, for any reviewer $i \in [n]$, we let $\pi_i \in \Pi_d$ denote the ordering (permutation) of the papers shown to reviewer i. We also use the notation $\pi_i(j)$ to denote the position of paper $j \in [d]$ in the ordering π_i.

Gain function (objective). Any algorithm to determine the ordering of papers must trade-off between two competing objectives: ensuring each paper receives a sufficient number of bids and ensuring each reviewer gets to see relevant papers early in the ordering. A combination of the objectives comprise our "gain function," which is the objective we aim to optimize. We begin by discussing each objective component.

Paper-side gain: The paper-side gain is associated with a given function $\gamma_p : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$. At the end of the entire bidding process, the paper-side gain $\mathcal{G}_p$ is

$$\mathcal{G}_p = \sum_{j \in [d]} \gamma_p(g_j),$$

where g_j is the number of bids received by paper $j \in [d]$. We assume the function γ_p is non-decreasing and concave. The non-decreasing property represents an improved gain if there are more bids, and the concavity property captures diminishing returns.[3] An example of a choice for the paper-side gain is the square-root function $\gamma_p(x) = \sqrt{x}$. This function is increasing, smooth, and captures the diminishing returns property. The reader may keep this function in mind as a running example for concreteness. A second example is $\gamma_p(x) = \min\{x, r\}$ for a given parameter

[1] Henceforth, for any positive integer κ, we will use the standard shorthand $[\kappa]$ to denote the set $\{1, \ldots, \kappa\}$.

[2] However, in Section 5.5.1, we show that our algorithm is empirically robust to violations of these assumptions.

[3] Our algorithm easily adapts to paper-side gains that may also be a function of the similarity scores of the reviewers who bid; for example, a higher gain for bids from expert reviewers. We omit this detail for sake of brevity.

$r \geq 1$, which emphasizes having at least r bids per paper.

Reviewer-side gain: This objective captures the desideratum that the reviewers should be shown papers with high relevance early in the paper ordering. The reviewer-side gain is associated with some predetermined function $\gamma_r : [d] \times [0,1] \to \mathbb{R}_{\geq 0}$. Given this function, the reviewer-side gain $\mathcal{G}_r$ is defined as:

$$\mathcal{G}_r = \sum_{i \in [n]} \sum_{j \in [d]} \gamma_r(\pi_i(j), S_{i,j}).$$

The function γ_r is assumed to be non-increasing in the position (its first argument) and non-decreasing in the similarity (its second argument). One example choice of this function, which the reader may choose to keep in mind as a running example, is the Discounted Cumulative Gain or DCG used commonly in data mining (Järvelin and Kekäläinen, 2000). In our setting, the function is given by

$$\gamma_r(\pi_i(j), S_{i,j}) = \frac{2^{S_{i,j}} - 1}{\log_2(\pi_i(j) + 1)}, \tag{5.1}$$

where we have set the "relevance" parameter in DCG to be the similarity $S_{i,j}$.

Overall gain function: Finally, we assume there is a trade-off parameter $\lambda \geq 0$, chosen by the program chairs, which trades off between these two objectives so that the overall gain function is given by

$$\mathcal{G} = \mathcal{G}_p + \lambda \mathcal{G}_r. \tag{5.2}$$

The goal is to determine the orderings of papers to show each reviewer to maximize the expected overall gain, $\mathbb{E}\big[\mathcal{G}\big]$, where the expectation is taken over the randomness in the bids made by the reviewers (see reviewer bidding model below) and any randomness in the algorithm.

Reviewer bidding model. An important aspect of any system that displays a list to users is the presence of primacy effects. In the context of our problem, the primacy effect means a reviewer is more likely to bid on a paper shown at the top of the list rather than later (Murphy et al., 2006). A second aspect of bidding is that a reviewer is more likely to bid on papers with greater similarity, although the reviewer may not bid on exactly the papers with the highest similarity since the similarities are noisy representations of their reviewing interests.

Thus in order to model reviewer bidding, we revert to literature on position-based click models that have a nearly identical setting (where clicks are analogous to our bids). We model the bidding via a given function $f : [d] \times [0,1] \to [0,1]$, where $f(\pi_i(j), S_{i,j})$ is non-increasing in the position that a paper is shown (the first argument) and non-decreasing in the similarity score (the second argument). Any reviewer $i \in [n]$ bids on paper $j \in [d]$ independently with probability

$$p_{i,j} = f(\pi_i(j), S_{i,j}).$$

As a running example throughout, note that in position-based click models, the click probability decomposes into a product of relevance and position bias (Chuklin et al., 2015). Moreover, the literature considers the click probability to decay logarithmically as a function of the position (Aslanyan and Porwal, 2019). The translation of these models into our setting gives rise to

the example bidding function

$$f(\pi_i(j), S_{i,j}) = \frac{S_{i,j}}{\log_2(\pi_i(j)+1)}. \tag{5.3}$$

Baselines. We consider the following three methods of ordering papers as baselines.

Random baseline (RAND): A commonplace practice (Cabanac and Preuss, 2013) is to show papers to reviewers in some fixed order, such as in order of submission of the papers. As a baseline, we consider a better variant of this practice, in which each reviewer is shown an independently and randomly selected paper ordering.

Similarity baseline (SIM): A second common practice, followed in several conference management systems today, is to order the papers according to their similarities. In other words, any reviewer $i \in [n]$ is shown the papers in order of the values in $\{S_{i,j}\}_{j\in[d]}$ (where the paper with maximum similarity is shown at the top, and so on). Any ties are broken by showing papers with fewer bids higher, and further ties are broken uniformly at random.

Bid baseline (BID): A third baseline shows papers to greedily optimize the minimum bid count. Each reviewer is shown papers in increasing order of the number of bids received so far (from the reviewers who arrived previously). Any ties are broken in favor of the paper with a higher similarity, and further ties are broken uniformly at random.

5.3 Algorithm

The key challenge in designing a suitable algorithm for the problem at hand stems from the fact that the paper-side gain is coupled (non-separable) across the orderings of papers presented to all reviewers so the impact of each individual paper ordering cannot be fully realized until the entire bidding process is complete. Conversely, the reviewer-side gain is decoupled (separable) across reviewers. This means the reviewer-side gain that can be obtained from any given reviewer is independent of the ordering of papers presented to any other reviewer. Thus, an algorithm for this problem is required to make local decisions, where the effect of the decision on the global gain (or cost) is only partially known. This perspective is reminiscent of the classical A* algorithm (Hart et al., 1968), and using A* as an inspiration, we now present an algorithm which we call SUPER* for our problem[4].

The A* algorithm operates with a goal of finding the minimum cost path between a pair of vertices in a cost-weighted graph. For any node in consideration, it considers two functions: a function which captures the cost so far and a second function—called the "heuristic"—which captures some estimate of the cost from the current node to the destination. The A* algorithm then finds a path based on these two functions. Before moving to a description of SUPER*, we discuss such a heuristic in the context of the problem at hand.

5.3.1 Heuristic for Future Bids

In a manner analogous to the A* algorithm, at any point in time SUPER* keeps track of the gains so far and also takes as input a heuristic that captures the "unseen" events. The heuristic in

[4]The name SUPER* stands for SUperior PERmutations and also indicates the inspiration from A*.

A* provides, for every vertex in the given graph, an estimate of the cost incurred in the future. Analogously, the heuristic in SUPER* provides, for every arrival of a reviewer, an estimate of the number of bids each paper will receive in the future. Formally, let us index the reviewers as $i \in [n]$ in the order of arrival (note that this order is unknown a priori). The heuristic comprises a collection of vectors $\{h_1, \ldots, h_n\}$, where each $h_i \in [0, n-i]^d$ represents an estimate of the number of bids each of the d papers will receive from all future reviewers $\{i+1, \ldots, n\}$. The vector h_i is provided to the SUPER* algorithm on arrival of the i^{th} reviewer. Two examples of heuristic functions that we consider in the subsequent narrative are described as follows.

- *Zero heuristic:* $h_i = 0$ for every $i \in [n]$.
- *Mean heuristic:* This function computes the expected number of bids each paper will receive if the permutations shown to all future reviewers are chosen independently and uniformly at random. Formally: $h_{i,j} = \frac{1}{d} \sum_{i'=i+1}^{n} \sum_{j' \in [d]} f(j', S_{i',j}) \ \forall \ i \in [n-1], j \in [d]$.

We set $h_n = 0$ for any heuristic, implying there are no bids placed after the last reviewer. This is analogous to setting the heuristic value to zero for the target vertex in the A* algorithm.

5.3.2 Intuition Behind the Algorithm

We first provide some intuition about the SUPER* algorithm, and subsequently present a formal description. Since a primary impediment to designing an algorithm is the inability to fully realize the impact of a paper ordering on the paper-side gain until the end of the bidding process, we begin by considering the scenario where $(n-1)$ reviewers have already departed, and the problem is to determine the ordering of papers to show the final reviewer. In this scenario, we have access to the bids of all $(n-1)$ reviewers that have already arrived and the orderings of papers presented to them. We use the notation $g_{n-1,j} \in \{0, \ldots, n-1\}$ to denote the number of bids received by any paper $j \in [d]$ at the time of arrival of the last reviewer. The values $\{g_{n-1,1}, \ldots, g_{n-1,d}\}$ are thus known at the time when the final reviewer arrives. As a result, we can formulate an optimization problem for the final reviewer n to maximize the expected gain from (5.2) in the following manner. For every $j \in [d]$, let $\mathcal{B}_{n,j}$ denote a Bernoulli random variable with mean $p_{i,j} = f(\pi_n(j), S_{n,j})$, independent of all else. The random variable $\mathcal{B}_{n,j}$ represents the bid of the final reviewer on paper $j \in [d]$. The optimization problem can be written as

$$\max_{\pi_n \in \Pi_d} \sum_{j \in [d]} \mathbb{E}[\gamma_p(g_{n-1,j} + \mathcal{B}_{n,j})] + \lambda \sum_{j \in [d]} \gamma_r(\pi_n(j), S_{n,j}), \tag{5.4}$$

where the expectation is taken over the distribution of the random variables $\mathcal{B}_{n,1}, \ldots, \mathcal{B}_{n,d}$.

Observe that the constraint set for the optimization problem in (5.4) is the set Π_d of all permutations. This set is, in general, not very well behaved (Ailon et al., 2008; Shah et al., 2016), which makes even this one-step optimization a challenge. As we discuss later in the formal algorithm description along with Theorem 5.1, for the final reviewer SUPER* optimally solves (5.4) and it is computationally efficient. The aforementioned subproblem forms the starting point for the SUPER* algorithm. Now that we know to handle a single (last) reviewer in an optimal fashion, we now describe the SUPER* algorithm for a general reviewer, say, $i \in [n]$. When reviewer i arrives, we have access to the number of bids made by all past reviewers on any paper $j \in [d]$, which we denote by

Algorithm 5.1: SUPER*

Input: $\gamma_p : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, paper-side gain
$\gamma_r : [d] \times [0,1] \to \mathbb{R}_{\geq 0}$, reviewer-side gain
$f : [d] \times [0,1] \to [0,1]$, bidding model
$\lambda \geq 0$, trade-off parameter
$S \in [0,1]^{n \times d}$, similarity matrix.

Algorithm:

1. Initialize bids on papers to zero: $g_0 \leftarrow 0_d$
2. For each reviewer arrival $i \in [n]$
   (a) Compute or input heuristic $h_i \in [0, n-i]^d$
   (b) $\pi_i \leftarrow$ FindPaperOrder
   (c) Present paper order π_i
   (d) Observe bids $b_i \in \{0,1\}^d$
   (e) Update paper bid counts: $g_i = g_{i-1} + b_i$

Algorithm 5.2: FindPaperOrderEfficient

1. Compute weights for each $j \in [d]$:

$$\chi_{i,j} = \gamma_p(g_{i-1,j} + h_{i,j} + 1) - \gamma_p(g_{i-1,j} + h_{i,j})$$
$$\alpha_{i,j} = \lambda \gamma_r^S(S_{i,j}) + f^S(S_{i,j})\chi_{i,j}$$

2. $\pi_i = \sigma(\alpha_i)$, where $\sigma : \mathbb{R}^d \to [d]^d$ returns the rank from maximum to minimum of each input in place and breaks ties arbitrarily.

Output: π_i

Algorithm 5.3: FindPaperOrder

1. Compute weight matrix $w \in \mathbb{R}^{d \times d}$ such that $\forall\ j \in [d], k \in [d]$:

$$w_{j,k} = \lambda \gamma_r(k, S_{i,j}) + f(k, S_{i,j})(\gamma_p(g_{i-1,j} + h_{i,j} + 1) - \gamma_p(g_{i-1,j} + h_{i,j}))$$

2. Solve linear program to obtain $x^* \in \mathbb{R}^{d \times d}$:

$$x^* \in \operatorname*{arg\,max}_{x \in [0,1]^{d \times d}} \sum_{j \in [d]} \sum_{k \in [d]} w_{j,k} x_{j,k}$$
$$\text{s.t.} \sum_{k \in [d]} x_{j,k} = 1\ \forall\ j \in [d],\ \sum_{j \in [d]} x_{j,k} = 1\ \forall\ k \in [d]$$

with ties broken arbitrarily between the set of maximizing solutions

3. $\pi_i(j) = k$ such that $x^*_{j,k} = 1$ for each $j \in [d]$

Output: π_i

$g_{i-1,j} \in \{0, \ldots, i-1\}$.

We now recall the A* algorithm: for any vertex, A* considers the cost "g" so far and a heuristic estimate "h" of the subsequent cost. Then, considering the cost of any vertex as "$g + h$", the A* algorithm takes the one-step optimal action given by selecting the neighboring vertex with the smallest value of "$g + h$". In an analogous fashion, SUPER* considers the number of bids so far (g_{i-1}) and takes as input a heuristic (h_i) for the number of bids in the future. Then, considering the number of bids from all other reviewers as "$g_{i-1} + h_i$", the SUPER* algorithm takes the action which is the one-step optimal action. In other words, SUPER* solves for each paper ordering using:

$$\max_{\pi_i \in \Pi_d} \sum_{j \in [d]} \mathbb{E}[\gamma_p(g_{i-1,j} + h_{i,j} + \mathcal{B}_{i,j})] + \lambda \sum_{j \in [d]} \gamma_r(\pi_i(j), S_{i,j}), \tag{5.5}$$

where $\mathcal{B}_{i,j}$ is a Bernoulli random variable with mean $p_{i,j} = f(\pi_i(j), S_{i,j})$ and is independent of all

else. As for the final reviewer, SUPER* solves this problem in an efficient manner for any arbitrary reviewer.

5.3.3 Formal Algorithm Description

The SUPER* algorithm is presented in Algorithm 5.1. To determine a paper ordering to show any reviewer, SUPER* calls a procedure to efficiently solve (5.5). We give a general method in Algorithm 5.3 and a faster method in Algorithm 5.2 that is applicable for a special class of reviewer-side gain and bidding functions.

General version. In the general version of SUPER*, Algorithm 5.3 is called to return a paper ordering that is a solution to (5.5) each time a reviewer arrives. In the proof of Theorem 5.1, we show that the optimization problem over the set of permutations given in (5.4) to find the optimal paper ordering for the final reviewer can be reformulated as an integer linear programming problem with a totally unimodular constraint set. The totally unimodular property of the constraint set guarantees that the solution of a relaxed linear program is in fact the integer optimal solution. The application of this reduction from an optimization problem over permutations to a linear programming problem for any given reviewer forms the technique given in Algorithm 5.3 to efficiently obtain a solution to (5.5). Finally, the per-reviewer time complexity of the general version of SUPER* given the evaluations of the heuristic is $\mathcal{O}(d^3)$ as a consequence of the call to solve a linear assignment problem in Algorithm 5.3.

Faster specialized version. Given a bidding model that can be decomposed into the form $f(\pi_i(j), S_{i,j}) = f^S(S_{i,j})f^\pi(\pi_i(j))$ where $f^S : [0,1] \to [0,1]$ is non-decreasing and $f^\pi : [d] \to [0,1]$ is non-increasing, along with a reviewer-side gain function that can be decomposed as $\gamma_r(\pi_i(j), S_{i,j}) = \gamma_r^S(S_{i,j})f^\pi(\pi_i(j))$ where $\gamma_r^S : [0,1] \to \mathbb{R}_{\geq 0}$ is non-decreasing, SUPER* calls Algorithm 5.2 to return a paper ordering that is a solution to (5.5) each time a reviewer arrives. For this model class that the problem from (5.4) to find the optimal paper ordering for the final reviewer after evaluating the expectation can be reformulated as

$$\max_{\pi_n \in \Pi_d} \quad \sum_{j \in [d]} \alpha_{n,j} f^\pi(\pi_n(j)) \tag{5.6}$$

for some non-negative weights $\{\alpha_{n,j}\}_{j\in[d]}$. The problem in (5.6) admits a simple solution: f^π is non-increasing on the domain, so the objective is maximized by presenting papers in decreasing order of the weights $\{\alpha_{n,j}\}_{j\in[d]}$. Obtaining this solution only requires sorting the weights, which has a time complexity of $\mathcal{O}(d\log(d))$. The application of this problem reformulation for the given model class and any reviewer forms the technique given in Algorithm 5.2 to obtain a solution to (5.5).

Before moving on to present our theoretical results, we comment on the relevance of this model class. Importantly, the DCG reviewer-side gain function and bidding model $f(S_{i,j}, \pi_i(j)) = S_{i,j}/\log_2(\pi_i(j)+1)$, which we have mentioned as running examples that can be kept in mind, satisfy the decomposition for which SUPER* is computationally efficient. This choice of functions is standard in the past literature on ranking models and click behavior (Aslanyan and Porwal, 2019; Järvelin and Kekäläinen, 2000), meaning that the time complexity result for this model class is quite relevant.

5.4 Theoretical Results

We now present the main theoretical results of this chapter. Complete proofs of all results are in Section 5.A of the appendix that follows this chapter.

5.4.1 Local Optimality

The property of local optimality, as the name suggests, means that the algorithm is optimal with respect to the reviewer under consideration. Achieving even a good local performance in a computationally efficient manner is challenging due to the optimization over permutations in (5.4). The following results show that SUPER*, which is computationally efficient, is locally optimal.

The result is first presented in terms of the final reviewer for simplicity and extended to a general reviewer subsequently. In the following theorem, since we consider only the final reviewer, note that the heuristic for SUPER* is irrelevant because the heuristic value for the final reviewer is always set to zero.

Theorem 5.1. *Given any history of paper orderings and bids from reviewers that arrived previously, the paper ordering given by SUPER* to the final reviewer maximizes the expected gain conditioned on the history.*

In other words, the expected amount by which the gain is increased from the final reviewer is maximized. This result follows from the fact that SUPER* exactly solves a problem that maximizes the expected gain increase from a reviewer. To generalize the previous result to a local optimality result for any reviewer, let the immediate gain from a reviewer be defined as the difference between the gain after and before the reviewer arrived.

Corollary 5.1. *Given any history of paper orderings and bids from reviewers that arrived previously, the paper ordering given to any reviewer by SUPER* with zero heuristic maximizes the expected immediate gain from that reviewer conditioned on the history.*

The property of local optimality also implies optimality of SUPER* (with any heuristic) when the paper-side gain function is linear (See Section 5.B.2). The proof of Theorem 5.1 is in Section 5.A.1 and the proof of Corollary 5.1 is in Section 5.A.2.

We now show that an analogous statement cannot be made regarding the other baseline methods. In fact, in contrast to SUPER*, all the popular baselines are considerably suboptimal.

Theorem 5.2. *Consider a model with the paper-side gain function $\gamma_p(g_j) = \sqrt{g_j}$, the reviewer-side gain function $\gamma_r(\pi_i(j), S_{i,j}) = (2^{S_{i,j}} - 1)/\log_2(\pi_i(j) + 1)$, and the bidding function $f(\pi_i(j), S_{i,j}) = S_{i,j}/\log_2(\pi_i(j) + 1)$. There exists a constant $c > 0$ such that for every $d \geq 2$ and $\lambda \geq 0$, in the worst case for the final reviewer:*

(a) SIM is suboptimal by an additive factor of at least $cd/\log_2^2(d)$;

(b) BID is suboptimal by an additive factor of at least $cd \max\{1, \lambda\}/\log_2^2(d)$;

(c) RAND is suboptimal by an additive factor of at least $cd \max\{1, \lambda\}/\log_2^2(d)$.

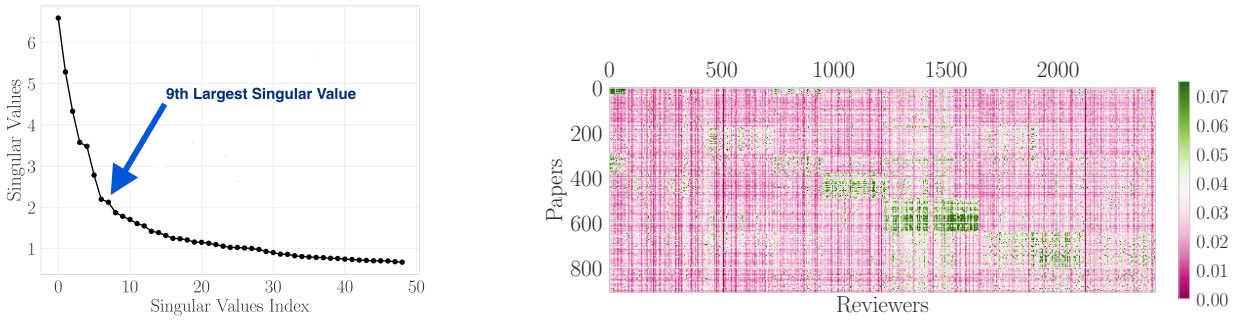
Theorems 5.1 and 5.2 in tandem show that SUPER* not only is locally optimal but can outperform currently popular algorithms by a wide margin. In the proof of this result, we construct instances where naively optimizing only a piece of the objective is highly suboptimal. The proof of Theorem 5.2 is in Section 5.A.3.

5.4.2 Global Optimality Under a Community Model

We now transition to consider the global performance of the algorithms. Given our focus on the application of peer review, we are motivated to give guarantees on the performance of SUPER* for similarity matrix classes that would be encountered in a real conference.

A common characteristic of networks is community structure (Newman and Girvan, 2004; Porter et al., 2009), where nodes can be grouped into clusters and links between groups are not as common. This phenomena has been documented in social and biological networks among others (Girvan and Newman, 2002). Pertinent to this work, empirical investigations have revealed community structures in scientific collaboration networks (Newman, 2001). Given this close connection, and the fact that scientific research is highly specialized, it is intuitive that communities exist in major conferences pertaining to different subtopics.

We explore the possible existence of such structure in the ICLR 2018 similarity matrix that was reconstructed by Xu et al. (2019) and is of size $n = 2435$ and $d = 935$. Recall that the ICLR similarity matrix is of size (2435×935). To begin, we investigate the spectral properties of the similarity matrix from ICLR 2018, and in particular, whether it is low rank. We plot the singular values of the similarity matrix in Figure 5.1a, where the (heuristic) elbow method suggests a low rank (≈ 10). In Figure 5.1b we plot the entries of the similarity matrix after permuting its rows and columns according to the spectral co-clustering algorithm (Dhillon, 2001). The result suggests the ICLR 2018 similarity matrix exhibits some characteristics of a noisy block diagonal structure.



(a) (b)

Figure 5.1: (a) The 50 singular values (excluding the maximum singular value) of ICLR 2018 similarity matrix, which shows low-rank structure. (b) Similarity scores of the permuted ICLR matrix as a heatmap indicating the block diagonal structure.

In what follows, we now perform an associated theoretical analysis of the algorithms under such community structures of the similarity matrix. We begin by proposing a simple model which we call the 'noiseless community model'.

Noiseless community model. Informally, the noiseless community model we study is a set of similarity matrices that up to a permutation of rows and columns belong to a subclass of block diagonal matrices. Formally, let $\mathbf{0}_{q\times q}$ and $\mathbf{1}_{q\times q}$ denote $q \times q$ matrices of all zeros and all ones

respectively. Define an $mq \times mq$ block diagonal matrix B as:

$$B = \begin{bmatrix} \mathbf{1}_{q\times q} & \mathbf{0}_{q\times q} & \cdots & \mathbf{0}_{q\times q} \\ \mathbf{0}_{q\times q} & \mathbf{1}_{q\times q} & \cdots & \mathbf{0}_{q\times q} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{q\times q} & \mathbf{0}_{q\times q} & \cdots & \mathbf{1}_{q\times q} \end{bmatrix}.$$

Finally, denote by $\mathcal{P}_{mq\times mq}$ the set of all $mq \times mq$ permutation matrices. Recall that a permutation matrix is a matrix obtained by permuting the rows of an identity matrix. Also recall that left multiplying a matrix by a permutation matrix permutes the rows of the matrix and right multiplying a matrix by a permutation matrix permutes the columns of the matrix. With this background, the noiseless community model is defined as the following set of similarity matrices for $m \geq 2$ and $q \geq 2$:

$$\text{Noiseless Community Model} = \{S \in \mathbb{R}^{mq\times mq} : S = P(sB)\widetilde{P}, \quad s \in [0.01, 1], \quad P, \widetilde{P} \in \mathcal{P}_{mq\times mq}\}. \tag{5.7}$$

The number of reviewers is given by $n = mq$ and the number of papers is by $d = mq$. In words, this is the set of all similarity matrices obtained via a permutation of the rows and columns of the block matrix B.

We begin our theoretical results for this chapter by showing that under the noiseless community formulation, both SUPER* and SIM are optimal, whereas BID and RAND fare poorly. Recall that $d = n = mq$ in the noiseless community model.

Theorem 5.3. *Consider a model with a paper-side gain function $\gamma_p(g_j) = \sqrt{g_j}$, the reviewer-side gain function $\gamma_r(\pi_i(j), S_{i,j}) = (2^{S_{i,j}} - 1)/\log_2(\pi_i(j) + 1)$, and the bidding function $f(\pi_i(j), S_{i,j}) = \mathbb{1}\{\pi_i(j) = 1\}\mathbb{1}\{S_{i,j} > s/2\}$. Then, under the noiseless community model from (5.7), for all $m \geq 2$, $q \geq 2$ and $\lambda \geq 0$:*

(a) SUPER with zero heuristic is optimal;*

(b) SIM is optimal.

In contrast, there exists a constant $c > 0$ such that for every $m \geq 2$, $q \geq 2$ and $\lambda \geq 0$:

(c) BID is suboptimal by an additive factor of at least $c\lambda d/\log_2^2(d)$;

(d) RAND is suboptimal by an additive factor of at least cd.

Although SIM is optimal in the noiseless community model, this optimality turns out to be quite brittle. As we show below, even an infinitesimally small amount of noise makes SIM considerably suboptimal. In contrast, SUPER* is robust enough and suffers by only a small amount.

Noisy community model. More formally, we first define a 'noisy community model'. Under this model, we assume that the similarity matrix is generated by first selecting any similarity matrix S' from the noiseless community model defined in (5.7), and then adding noise to its entries as:

$$S_{i,j} = \begin{cases} s - \nu_{i,j} & \text{if } S'_{i,j} = s \\ \nu_{i,j} & \text{if } S'_{i,j} = 0, \end{cases} \tag{5.8}$$

where $\nu_{i,j}$ is drawn independently and uniformly from $(0, \xi)$ for each reviewer-paper pair, for some small value ξ to be defined subsequently.

The next result shows that even under an arbitrarily small perturbation ξ from a noiseless community model, the baselines become significantly suboptimal. In contrast, SUPER* is robust to the noise and is near-optimal. Recall that $d = n = mq$ in the noisy community model.

Theorem 5.4. *Consider the gain and bidding functions from Theorem 5.3 and the noisy community model given in* (5.8) *with any noise bound satisfying* $\xi \leq (1+\lambda)^{-1}e^{-emq}$. *Then, for all* $m \geq 2$, $q \geq 2$, *and* $\lambda \geq 0$*:*

(a) SUPER* *with zero heuristic is within at least an additive factor of* 0.0001 *of the optimal. Moreover, there exists a constant* $c > 0$ *such that for every* $m \geq 2$, $q \geq 2$, *and* $\lambda \geq 0$, *with respect to* SUPER* *with zero heuristic:*

(b) SIM *is suboptimal by an additive factor of at least* cd*;*

(c) BID *is suboptimal by an additive factor of at least* $c\lambda d/\log_2^2(d)$*;*

(d) RAND *is suboptimal by an additive factor of at least* cd.

This result thus establishes the global optimality of the proposed SUPER* algorithm under the community model, while in contrast all popular baselines are considerably suboptimal. The intuition behind the proofs in this subsection is that to optimize the objective, each paper should only be shown in the highest position exactly once when it has a high similarity score. Doing so maximizes the expected paper-side gain, while incurring minimal loss in the expected reviewer-side gain. The proofs of Theorem 5.3 and Theorem 5.4 can be found in Section 5.A.4 and Section 5.A.5, respectively.

5.5 Experimental Results

We now empirically evaluate SUPER* (with zero and mean heuristics) and compare it with the baselines SIM, BID, and RAND (discussed earlier in Section 5.2). The experimentation methodology is as follows for any chosen set of model parameters including the gain functions, bidding probability, trade-off parameter, and number of reviewers and papers. Given a fixed similarity matrix, we shuffle the rows of the matrix to randomize the sequence of reviewer arrivals and simulate each of the algorithms. Then, for each algorithm, we record the gain along with the number of papers that end up with bid counts in the intervals $\{0,1,2\}$, $\{3,4,5\}$, $\{6,7,8\}$, and $\{9+\}$. We repeat this process 20 times for a given similarity matrix if it is fixed and draw a similarity matrix at random for each run if the score structure being simulated is a distribution. To evaluate performance, we show the means of the relative gains (additive gains relative to the gain of a baseline) across the runs and include error bars representing the standard error of the mean. Moreover, we present the mean number of papers across the repeated simulations that finish with bid counts in each of the previously given bid count intervals.

5.5.1 ICLR Similarity Matrix

To begin our experiments, we perform evaluations on a similarity matrix from the ICLR 2018 conference discussed earlier in Section 5.4. Recall that the similarity matrix consists of $n = 2435$ reviewers and $d = 935$ papers. In the following experiments, we run a simulation using a default model configuration, then we explore the impact of changing components of the model, and we finish by exploring the robustness of the algorithm to various real-world complexities.

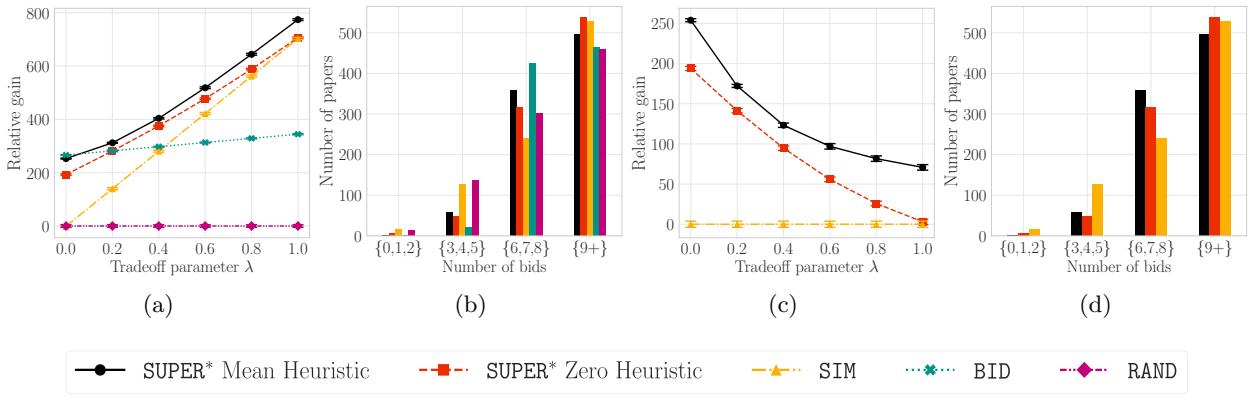


Figure 5.2: ICLR experiment with the default model configuration. Legend for each bid distribution plot (b, d): within each bid-count interval, the bars correspond to the algorithms given in the legend and are presented in an order consistent with the legend itself when read from left to right and (d) only includes SUPER* with mean heuristic, SUPER* with zero heuristic, and SIM.

Default model configuration. We begin by evaluating a default model configuration that is considered throughout the the remainder of the experiments unless otherwise specified. The model consists of the paper-side gain function $\gamma_p(g_j) = \min\{g_j, 6\}$, the reviewer-side gain function $\gamma_r(\pi_i(j), S_{i,j}) = (2^{S_{i,j}} - 1)/\log_2(\pi_i(j) + 1)$, and the bidding probability model $f(\pi_i(j), S_{i,j}) = S_{i,j}/\log_2(\pi_i(j) + 1)$. We remark that the paper-side gain function is a natural choice given that conferences often assign three reviewers to each paper and as such they may seek twice the number of bids per paper. Moreover, recall that for this pair of reviewer-side gain and bidding functions, the efficient routine in Algorithm 5.2 can be called in place of Algorithm 5.3 in SUPER* to retrieve a paper ordering, which is what we implement.

The results of the experiment are presented in Figure 5.2. In Figures 5.2a–5.2b we compare SUPER* to each baseline and in Figures 5.2c–5.2d we zoom in and only show the results for SUPER* and SIM. In terms of the gain results shown in Figures 5.2a and 5.2c, each version of SUPER* outperforms the baseline algorithms, while BID outperforms SIM when minimal weight λ is given to the reviewer-side gain and vice versa when a significant amount of weight λ is given to the reviewer-side gain. In Figures 5.2b and 5.2d, the distribution of the bid counts obtained for the algorithms are shown with $\lambda = 0.8$, which was chosen since this parameter choice gave nearly equal paper-side and weighted reviewer-side gain for RAND. While BID has a similar number of papers with fewer than the minimum number of desired bids as each version of SUPER*, the gain demonstrates why it is not a generally adopted method. As a result of showing papers of limited relevance early in the paper orderings to elicit bids on papers with few bids, the overall gain is significantly smaller than SIM and SUPER* since the reviewer-side gain is worse. The distributions also illustrate that both versions of SUPER* end the bidding process with approximately a 60% reduction of the number of papers with fewer than the desired minimum number of bids compared to SIM and RAND.

Varying model parameters. We now consider variations of the default model consideration. In Figures 5.3a–5.3d, results are shown when the paper-side gain function is changed from $\gamma_p(g_j) = \min\{g_j, 6\}$ to $\gamma_p(g_j) = \sqrt{g_j}$. In Figures 5.3a–5.3b we compare SUPER* to each baseline and in Figures 5.3c–5.3d we zoom in and only show the results for SUPER* and SIM. In Figures 5.3b

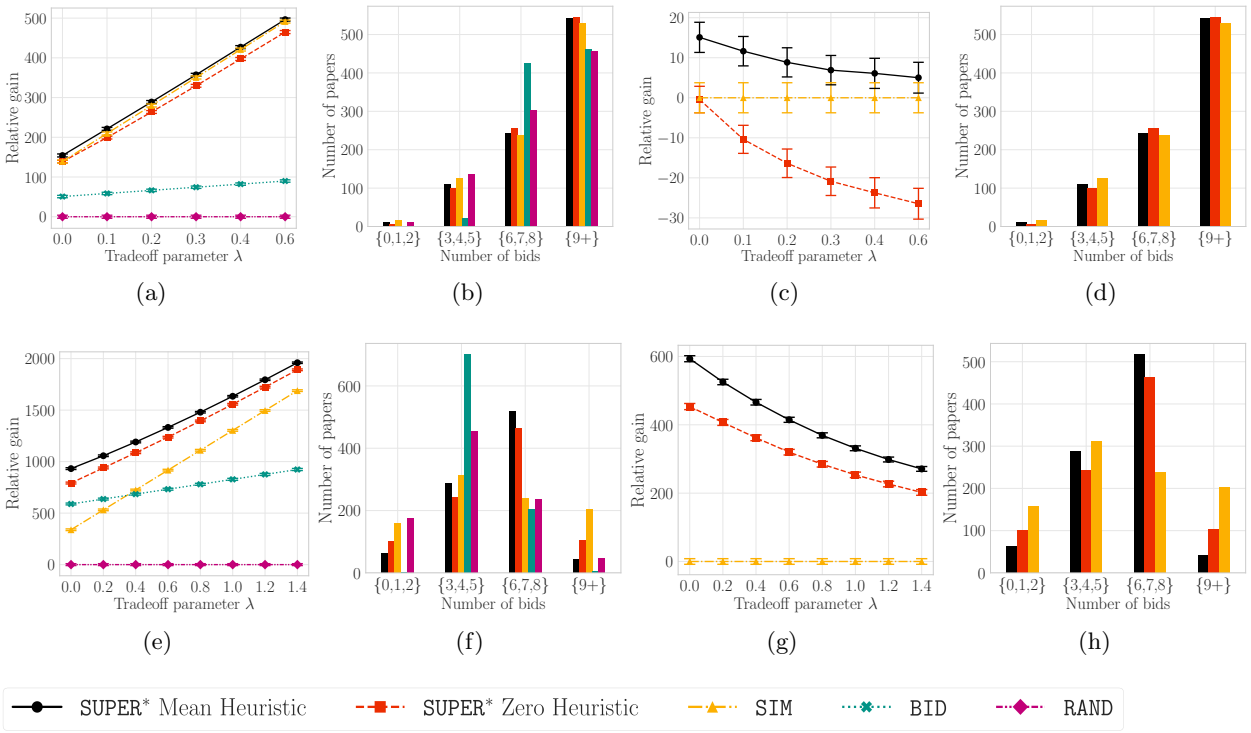


Figure 5.3: ICLR 2018 experiment with variations of the default model configuration. In Figures 5.3a–5.3d, the paper-side gain function is changed. In Figures 5.3e–5.3h, the reviewer-side gain and bidding function are changed. Legend for each bid distribution plot (b, d, f, h): within each bid-count interval, the bars correspond to the algorithms given in the legend and are presented in an order consistent with the legend itself when read from left to right and (d, h) only includes SUPER* with mean heuristic, SUPER* with zero heuristic, and SIM.

and 5.3d, the distribution of the bid counts obtained for the algorithms are shown with $\lambda = 0.4$, which was chosen since this parameter choice gave nearly equal paper-side and weighted reviewer-side gain for RAND. For this model configuration, BID and RAND are significantly suboptimal in terms of the gain. SUPER* with the mean heuristic outperforms SIM by a marginal amount in terms of the gain, while SIM outperforms SUPER* with the zero heuristic by a marginal amount in terms of the gain. The bid distributions show that each version of SUPER* ends the bidding process with a smaller number of papers obtaining fewer than six bids compared to SIM. Compared to the default paper-side gain function, the discrepancy is not as significant since SUPER* is optimizing an objective that rewards getting more than five bids per paper even though the returns are diminishing.

In Figures 5.3a–5.3d, the default paper-side gain function is considered, but the reviewer-side gain function is changed to $\gamma_r(\pi_i(j), S_{i,j}) = (2^{S_{i,j}} - 1)/\sqrt{\pi_i(j)}$ and the bidding function is changed to $f(\pi_i(j), S_{i,j}) = S_{i,j}/\sqrt{\pi_i(j)}$. The effect of this change is that the probability of obtaining a bid on a paper from a reviewer decays faster with the position the paper is shown and similarly the reviewer-side gain from a paper diminishes faster as a function of the position the paper is shown to a reviewer. In Figures 5.3e–5.3f we compare SUPER* to each baseline and in Figures 5.3g–5.3h we zoom in and only show the results for SUPER* and SIM. In Figures 5.3f and 5.3h, the distribution

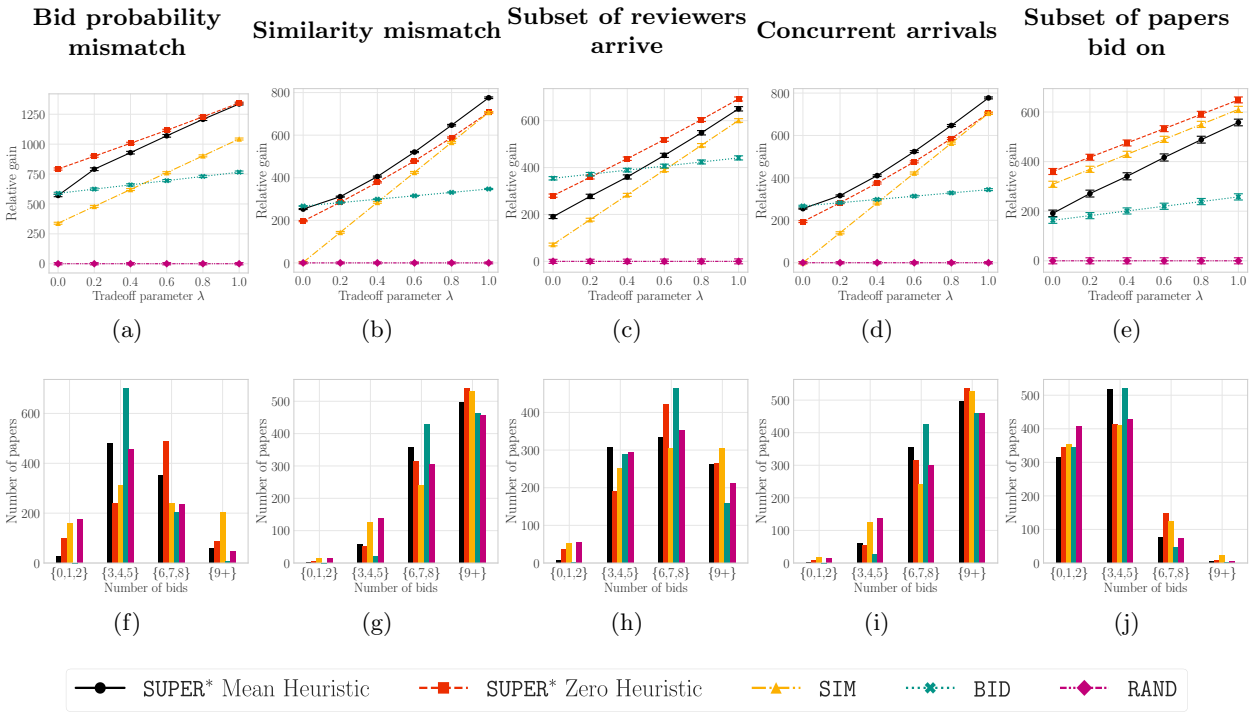


Figure 5.4: ICLR 2018 experiments of robustness to real-world complexities. Legend for each bid distribution plot (f)–(j): within each bid-count interval, the bars correspond to the algorithms given in the legend and are presented in an order consistent with the legend itself when read from left to right.

of the bid counts obtained for the algorithms are shown with $\lambda = 1.2$, which was chosen since this parameter choice gave nearly equal paper-side and weighted reviewer-side gain for RAND. For this model configuration, each version of SUPER* significantly outperforms each of the baselines in terms of the gain. The bid count distributions show SUPER* with zero and mean heuristic reduce the number of papers with fewer than three bids by 35% and 60% compared to SIM, respectively. Moreover, both versions of SUPER* end up with half as many papers obtaining fewer than six bids compared to BID.

Robustness to real-world complexities. The previous experiments were performed under settings faithful to the model described earlier in Section 5.2. We now evaluate the robustness of SUPER* to the models by evaluating the performance under various vagaries and complexities of real-world peer review. For this set of experiments, we consider that SUPER* is optimizing the default model configuration described previously in this section. The results of the following simulations that consider deviations from the model being optimized are given in Figure 5.4. For each experiment we show the gains of each of the algorithms relative to RAND across a sweep of the parameter λ and the bid count distributions for each of algorithms with $\lambda = 0.8$ as selected for the default model previously.

We begin looking at model mismatch for the bidding function. In Figures 5.4a and 5.4f, the situation is considered in which the actual bids are performed under the bidding function $f(\pi_i(j), S_{i,j}) = S_{i,j}/\sqrt{\pi_i(j)}$, whereas SUPER* assumes $f(\pi_i(j), S_{i,j}) = S_{i,j}/\log_2(\pi_i(j)+1)$. The results show each version of SUPER* is robust to this deviation and outperforms the baselines in terms

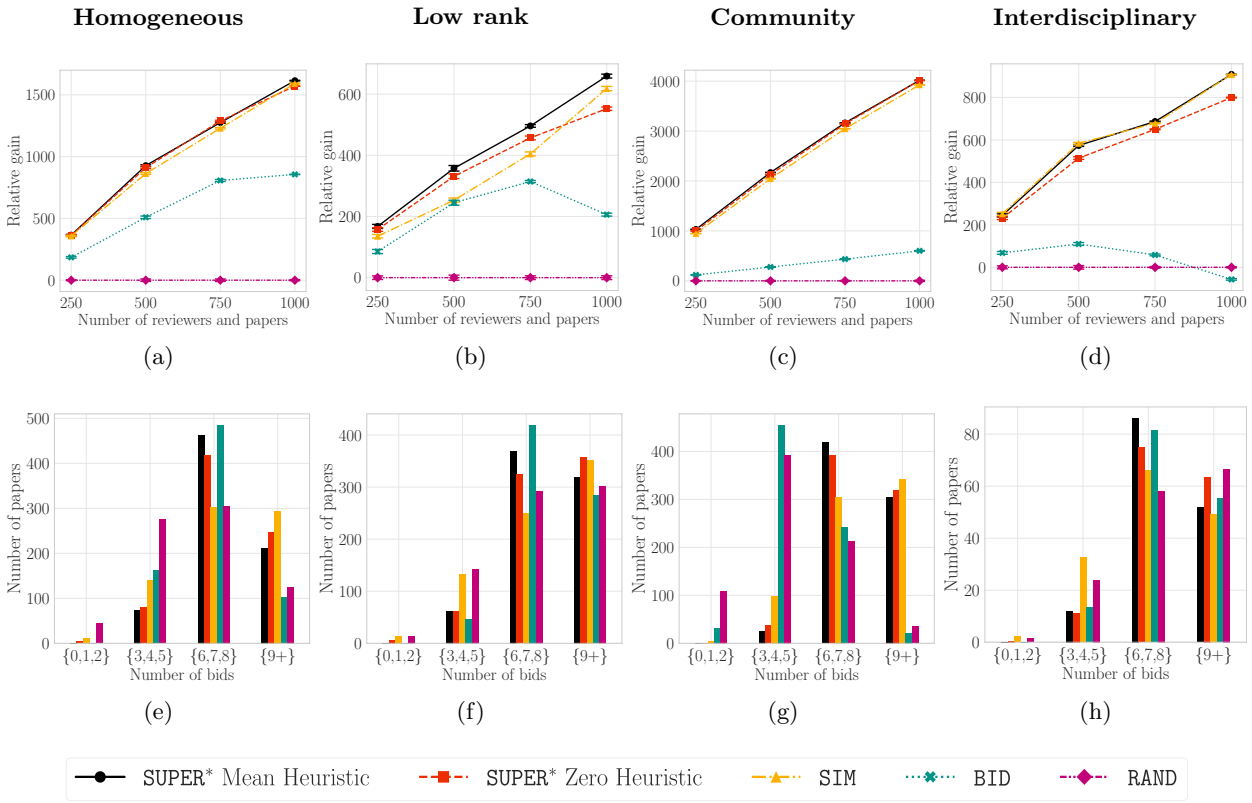


Figure 5.5: Experiments on synthetic similarity scores under the default model configuration. Legend for each bid distribution plot (e)–(h): within each bid-count interval, the bars correspond to the algorithms given in the legend and are presented in an order consistent with the legend itself when read from left to right.

of the gain. Moreover, SUPER* with zero heuristic is especially robust since it is not as dependent on the model of bids as SUPER* with mean heuristic. The bid count distributions show that SUPER* with mean heuristic reduces the number of papers with fewer than three bids by 85% relative to SIM, and SUPER* with zero heuristic reduces the number of papers with fewer than six bids by 50% compared to BID. In Figures 5.4b and 5.4g, we consider that the probability of a reviewer bidding on a paper is actually given by $f(\pi_i(j), S_{i,j}) = (S_{i,j} + \mathcal{N}(0, \sigma^2))/\log_2(\pi_i(j)+1)$ where $\sigma = 0.01$ and we remark that the mean of the similarity scores is approximately 0.03 so the noise magnitude is not insignificant. We clip the noisy bid probabilities to guarantee that they remain in the interval $[0, 1]$. We observe that SUPER* is again robust to this model mismatch and outperforms the baselines in terms of the gain and each version ends up reducing the fewer the number of papers having below the minimum desired number of bids by approximately 60% compared to SIM and RAND.

In practice, not all reviewers may participate in the bidding process. We consider that only three quarters of the reviewers arrive, but this is unknown a priori to the algorithms. This proportion is roughly based on the number of reviewers that were found to not place any positive bids during the NeurIPS 2016 review process (Shah et al., 2018). The results under this real-world complexity are shown in Figures 5.4c and 5.4h. Moreover, reviewers may not actually arrive sequentially.

Figures 5.4d and 5.4i consider the setting where Poisson(1) reviewers arrive at each time, and the algorithm must present paper orderings to all these reviewers simultaneously. For this pair of real-world complexities, SUPER* remains quite robust and generally performs favorably compared to the baselines in terms of both the gain and the bid count distributions. It is worth noting that when not all reviewers arrive, SUPER* with zero heuristic outperforms SUPER* with zero heuristic since it does attempt to account for bids that may come from reviewers that have not arrived.

A common feature in peer review bidding systems is the ability for a reviewer to search papers by subject area or keyword and then only bid within the resulting subset of papers. We now evaluate the robustness of SUPER* to this type of reviewer behavior in the following manner. In our simulations, on arrival of any reviewer, one quarter of the papers are randomly selected and required to be shown to the reviewer (these are the papers that are assumed to meet the search query). The remaining papers are not presented to the reviewer, are bid on with probability zero, and do not factor into the reviewer-side gain. The algorithms compute the paper orderings over only the selected subset of papers. The results of this experiment are shown in Figures 5.4e and 5.4j. We observe that SUPER* with zero heuristic outperforms the rest of the algorithms in terms of the gain while obtaining a favorable bid distribution. SUPER* with mean heuristic is not quite as robust since fewer bids come from future reviewers than anticipated when computing the mean heuristic – if one has estimates of the amount of selection done by reviewers via search, then this issue may be mitigated by appropriately scaling down the heuristic value.

5.5.2 Synthetic Similarities

We perform several simulations on synthetic similarity scores comparing the algorithms as presented in Figure 5.5. For this set of simulations, we consider the default model configuration from the previous section. Moreover, for each similarity structure we let the number of reviewers and the number of papers (n, d) be among the set of pairs $\{(250, 250), (500, 500), (750, 750), (1000, 1000)\}$ and fix the trade-off parameter to be $\lambda = 0.8$ since this gave roughly equal paper-side and weighted reviewer-side gains for RAND with $n = d = 750$.

Homogeneous similarity scores. We consider a synthetic similarity matrix structure where each entry is drawn from at random from a Beta distribution with parameters $\alpha = 1$ and $\beta = 15$. This distribution is highly skewed and the expected value of a draw from it is 0.0625. The results of this experiment are given in Figures 5.5a and 5.5e. The gain of SUPER* with each heuristic exceeds that of each of the baselines and similarly the bid count distributions show SUPER* with each heuristic ends up with at least 50% fewer papers obtaining under the minimum desired by the paper-side gain function compared to the baselines. We tried other homogeneous similarity structures and observed similar trends throughout.

Low rank structure. We experimented with the following low rank structure to generate the similarity scores. The similarity matrix is split into 10 groups of reviewers which correspond to a block of rows in the matrix. The similarity scores for block $\ell \in \{1, \ldots, 10\}$ are given by the rank-1 matrix $u(v^\ell)^\top$ where u is a $n/10$-dimensional vector of ones and v^ℓ is a d-dimensional vector with each entry drawn at random from a Beta distribution with parameters $\alpha = \ell$ and $\beta = 60$. This set of parameter choices for the distribution was made so that the scores were near the scale of the ICLR similarity matrix and to create a disparity in similarity scores between the blocks. Combining the blocks forms a similarity matrix of rank 10 where within each block the reviewers are identical

and between the blocks the similarity score distribution changes. The structure can be viewed as if there are 10 types of reviewers with varying levels of relevance to the papers. The result of the experiment with this similarity matrix structure is shown in Figures 5.5b and 5.5f. We again see that SUPER* with each heuristic outperforms the baselines in terms of the gain and they obtain a favorable distribution of the bid counts with a 50% reduction in the number of papers with fewer than six bids compared to SIM and RAND.

Community model structure. We now consider a community model type block structure as motivated in Section 5.4. To form this similarity matrix, we create a block matrix where each submatrix is of dimension 25×25. The similarity score of each entry in the submatrix (ℓ, k) is 0 if $\ell \neq k$ and 0.7 if $\ell = k$. In other words the matrix is block diagonal. We then add noise to each similarity score drawn uniformly at random from the interval $[0, 0.05]$. The results are given in Figures 5.5c and 5.5g. BID and RAND are highly suboptimal in terms of the gain and bid count distribution since they end up showing papers with similarity scores near zero early in the paper orderings even if later on there are reviewers to arrive which are closely matched to the papers. Remarkably, SUPER* with each heuristic reduces the number of papers with fewer than the minimum number of desired bids by at least 90% compared to BID and RAND. Moreover, SUPER* with each heuristic outperforms SIM in terms of the gain and reduces the number of papers with fewer than six bids by over 60%. This happens since for papers with close similarity scores, SUPER* shows the paper with fewer bids ahead if the similarity score is only marginally smaller.

Interdisciplinary papers. To conclude our simulations, we consider the impact our algorithm could have on interdisciplinary papers. As mentioned in Section 5.1, such papers face additional challenges in the peer review process owing to the lack of ideally matched peers. To simulate this phenomena, we consider a similarity matrix structure where there are two groups of reviewers of equal size and then three groups of papers which make up 40%, 40%, and 20% of the papers, respectively. Each reviewer in group one has similarity scores of 0.17, 0.005, and 0.085 with the respective paper groups and each reviewer in group two has similarity scores of 0.005, 0.17, and 0.085 with the respective paper groups. This reflects a scenario where the reviewer pool has two distinct areas of expertise and there is a set of interdisciplinary papers (paper group with similarity score of 0.075 with all reviewers). We show the results of the experiment in Figures 5.5d and 5.5h. In terms of the gain, SUPER* and SIM perform significantly outperform BID and RAND. For the bid count distribution in Figure 5.5h, we only consider the interdisciplinary papers. SUPER* with each heuristic mitigates negative impacts on the interdisciplinary papers as the number with an insufficient number of bids is curtailed by 65% and 50% compared to SIM and RAND. This is a result of the fact that SUPER* works to trade-off the paper-side and reviewer-side objectives so the interdisciplinary papers end up not always being shown after the papers matching the reviewers expertise as occurs for SIM.

5.6 Conclusion

We develop a sequential decision-making algorithm called SUPER* to optimize the bidding process and show that it empirically outperforms baseline methods on real conference data and has several compelling theoretical guarantees. An obvious open problem is that of developing an online algorithm that is globally optimal for every similarity matrix. Conversely, showing possible computational hardness of the problem as a negative result could be a path of future work. In several

automated reviewer-paper assignment methods, bids and and similarities are combined to form the scores used to compute the assignment (Shah et al., 2018). Given the tight connection between the bidding and matching systems, it is natural to design methods for jointly optimizing the components that govern the assignment process. Finally, given the online nature of reviewer arrivals and the need to immediately show the paper ordering to an arriving reviewer, it is of interest to solve the passive problem and this could also serve as an effective heuristic function for SUPER*.

While our work focuses on peer review, there are several revelant applications. An example is crowdsourcing, where a common goal of the platform is to ensure that each task gets sufficiently many qualified workers. From the perspective of the worker, it has been documented that workers put a non-trivial emphasis on a task if it is of interest to them (Hossain, 2012; Kaufmann et al., 2011). This means there is a trade-off of ensuring each worker is satisfied while ensuring a minimum number of workers for each job. As such, it is reasonable to formulate a multi-objective optimization problem and approach it as in this work. Indeed, the crowdsourcing platform in this formulation seeks to optimize both for a task-side objective, which ensures each task gets a sufficient number of workers selecting it, and for a user-side objective, which is to present relevant tasks to users.

Another potentially viable application is crowdfunding and microlending platforms such as Kiva or KickStarter. In crowdfunding, users pledge toward funding a project, and the project is only funded if the cumulative contributions of the crowd meet a known target threshold. The platforms seek to maximize the number of funded projects by deciding, when, how often, and to which users the projects are displayed. Past work (Jain and Jamieson, 2018) has modeled the optimization as a multi-armed bandit problem where the goal is to maximize the number of funded projects with a minimum amount of user impressions. This problem has a fundamental trade-off between showing relevant projects to users while ensuring that the projects themselves are given a fair shot to be funded. As such, a model-dependent approach such of ours could be of potential interest. Conversely, a more model-free approach based on multi-armed bandits (a problem class we explore in the chapter that follows) would be interest for the ranking problem in bidding systems.

A final potential application outside of peer review for our work is in recommender systems and online advertising where there is the common trade-off between exploration (gaining sufficient feedback on all items) and exploitation (showing the most relevant items to users). In recommender systems, the cold start problem refers to the situation where the system is just beginning to interact with users or items are freshly included in the catalogue and no past user interaction information is available. The common trade-off arises again where there is a need to show relevant items to users, while the system benefits from gaining feedback on items for which the utility is unknown. Our framework easily adapts to a changing action set and could be relevant to this task.

CHAPTER 5 APPENDIX

5.A Proofs

In this appendix section, we present proofs of the theoretical results presented in the main text. We begin by formally defining the history of information available to an algorithm when a reviewer arrives and restating important notation that will be found throughout the proofs.

Definition 5.1 (History)**.** *The history of information available to an algorithm when reviewer $i \in [n]$ arrives is defined as $\mathcal{H}_{i-1} = \{\pi_\ell, b_\ell : \ell = 1, \ldots, i-1\}$, where $\pi_\ell \in \Pi_d$ is the paper ordering that was presented to reviewer ℓ and $b_\ell \in \{0,1\}^d$ contains the bid realizations on each paper from reviewer ℓ.*

Notation. We let $d \geq 2$ denote the number of papers and $n \geq 2$ denote the number of reviewers. We use the notation $S_{i,j} \in [0,1]$ to denote the given similarity score between any reviewer $i \in [n]$ and paper $j \in [d]$. In general, we let i index a reviewer and j index a paper. We commonly use the set notation $[\kappa] = \{1, 2, \ldots, \kappa\}$ for any positive integer κ. Π_d denotes the set of $d!$ permutations of d papers. For any reviewer $i \in [n]$, we let $\pi_i \in \Pi_d$ denote the ordering (permutation) of the papers shown to reviewer i. We use the notation $\pi_i(j)$ to denote the position of paper $j \in [d]$ in the ordering π_i. The notation $\mathcal{B}_{i,j}$ represents the random variable of reviewer $i \in [n]$ biding on paper $j \in [d]$ which follows a Bernoulli distribution with parameter $p_{i,j} = f(\pi_i(j), S_{i,j})$. We use the notation $g_{i-1,j} \in \{0, \ldots, i-1\}$ to denote the number of bids received by any paper $j \in [d]$ at the time of arrival of reviewer $i \in [n]$ and g_j to denote the number of bids at the end of the bidding process on paper $j \in [d]$. The heuristic estimating the number of bids paper $j \in [d]$ will obtain from reviewers $\{i+1, \ldots, n\}$ is denoted as $h_{i,j}$ and it is provided to the SUPER* algorithm on arrival of reviewer $i \in [n]$. Finally, we abbreviate 'with probability' by w.p and use the terminology almost surely to mean with probability one and almost never to mean with probability zero.

5.A.1 Proof of Theorem 5.1: Local Optimality of SUPER* for Final Reviewer

In this proof, we show that selecting the optimal ordering to present to the final reviewer can be simplified to a tractable optimization problem. The SUPER* algorithm solves exactly this problem to determine an ordering of papers to present the final reviewer, and is hence an optimal algorithm for the final reviewer.

The optimization problem for the final reviewer n to maximize the expected gain conditioned on the history is

$$\max_{\pi_n \in \Pi_d} \quad \sum_{j \in [d]} \mathbb{E}[\gamma_p(g_j) | \mathcal{H}_{n-1}] + \lambda \sum_{i \in [n]} \sum_{j \in [d]} \mathbb{E}[\gamma_r(\pi_i(j), S_{i,j}) | \mathcal{H}_{n-1}] \tag{5.9}$$

where the expectation is taken over the randomness in the bid to be placed by the reviewer. Conditioned on the history $\mathcal{H}_{n-1}$, the final number of bids on any paper $j \in [d]$ given by g_j is

the sum of the deterministic number of bids prior to the final reviewer denoted as $g_{n-1,j}$ and a Bernoulli random variable $\mathcal{B}_{n,j}$ with parameter $p_{n,j} = f(\pi_n(j), S_{n,j})$ representing the random bid of the final reviewer. We incorporate this fact and remove terms independent of the optimization variable from the problem in (5.9) to equivalently obtain

$$\max_{\pi_n \in \Pi_d} \quad \sum_{j \in [d]} \mathbb{E}[\gamma_p(g_{n-1,j} + \mathcal{B}_{n,j})] + \lambda \sum_{j \in [d]} \gamma_r(\pi_n(j), S_{n,j}) \tag{5.10}$$

where the expectation on the reviewer-side gain is removed as it is independent of the random reviewer bids.

We now simplify the paper-side gain term by expanding the expectation for each $j \in [d]$. Observe that

$$\gamma_p(g_{n-1,j} + \mathcal{B}_{n,j}) = \begin{cases} \gamma_p(g_{n-1,j} + 1) & \text{w.p.} \quad f(\pi_n(j), S_{n,j}) \\ \gamma_p(g_{n-1,j}) & \text{w.p.} \quad 1 - f(\pi_n(j), S_{n,j}). \end{cases}$$

Therefore,

$$\mathbb{E}[\gamma_p(g_{n-1,j} + \mathcal{B}_{n,j})] = f(\pi_n(j), S_{n,j})(\gamma_p(g_{n-1,j} + 1) - \gamma_p(g_{n-1,j})) + \gamma_p(g_{n-1,j}). \tag{5.11}$$

Substituting (5.11) into (5.10) and removing the term independent of the optimization variable gives the problem

$$\max_{\pi_n \in \Pi_d} \quad \sum_{j \in [d]} f(\pi_n(j), S_{n,j})(\gamma_p(g_{n-1,j} + 1) - \gamma_p(g_{n-1,j})) + \lambda \sum_{j \in [d]} \gamma_r(\pi_n(j), S_{n,j}). \tag{5.12}$$

We now reformulate (5.12) into the following equivalent integer linear programming problem:

$$\begin{aligned} \max_{x \in \mathbb{R}^{d \times d}} \quad & \sum_{j \in [d]} \sum_{k \in [d]} f(k, S_{n,j})(\gamma_p(g_{n-1,j} + 1) - \gamma_p(g_{n-1,j}))x_{j,k} + \lambda \sum_{j \in [d]} \sum_{k \in [d]} \gamma_r(k, S_{n,j})x_{j,k} \\ \text{s.t.} \quad & \sum_{k \in [d]} x_{j,k} = 1 \ \forall \ j \in [d], \quad \sum_{j \in [d]} x_{j,k} = 1 \ \forall \ k \in [d], \quad x_{j,k} \in \{0, 1\} \ \forall \ j, k \in [d]. \end{aligned}$$

In this formulation, x is a $d \times d$ matrix for which $x_{j,k}$ is an indicator of paper $j \in [d]$ being shown at position $k \in [d]$. The constraint $\sum_{k \in [d]} x_{j,k} = 1 \ \forall \ j \in [d]$ ensures each paper is included strictly once in the ordering shown to the reviewer. The constraint $\sum_{j \in [d]} x_{j,k} = 1 \ \forall \ k \in [d]$ ensures strictly one paper is selected to be shown at each position. The final constraint ensures that each index of x is integer valued. This integer linear programming problem is known as a linear sum assignment problem.

The key step of this proof is to recall that the linear sum assignment problem can be solved as a linear program. Indeed, the final constraint ensuring an integer solution can be relaxed to $0 \leq x_{j,k} \leq 1 \ \forall \ j, k \in [d]$ and the optimal solution of the relaxed linear program will be the integer optimal solution. This property of the linear sum assignment problem is a consequence of the relaxed linear program containing a totally unimodular constraint set which guarantees the optimal solution to be the integral solution (see, e.g., Chapter 4 in Burkard et al., 2012).

The optimization problem arising from recognizing that the integer constraint can be relaxed is

given by

$$\begin{aligned} \max_{x \in \mathbb{R}^{d \times d}} \quad & \sum_{j \in [d]} \sum_{k \in [d]} w_{j,k} x_{j,k} \\ \text{s.t.} \quad & \sum_{k \in [d]} x_{j,k} = 1 \ \forall \ j \in [d], \quad \sum_{j \in [d]} x_{j,k} = 1 \ \forall \ k \in [d], \quad 0 \leq x_{j,k} \leq 1 \ \forall \ j,k \in [d], \end{aligned} \tag{5.13}$$

where

$$w_{j,k} = f(k, S_{n,j})(\gamma_p(g_{n-1,j} + 1) - \gamma_p(g_{n-1,j})) + \lambda \gamma_r(k, S_{n,j}) \ \forall \ j,k \in [d]. \tag{5.14}$$

This formulation shows that the paper ordering for the final reviewer that maximizes the expected gain conditioned on the history can be obtained efficiently by solving a linear program.

Local optimality of SUPER* for final reviewer. To determine the ordering of papers to show any reviewer $i \in [n]$ for the general class of assumed gain and bidding functions, SUPER* calls Algorithm 5.3. Algorithm 5.3 solves the optimization problem in (5.13) using the weights

$$w_{j,k} = f(k, S_{i,j})(\gamma_p(g_{i-1,j} + h_{i,j} + 1) - \gamma_p(g_{i-1,j} + h_{i,j})) + \lambda \gamma_r(k, S_{i,j}) \ \forall \ j,k \in [d]. \tag{5.15}$$

Given any heuristic function, the heuristic for the final reviewer is such that $h_n = 0$. This means SUPER* selects the paper ordering for the final reviewer by solving the optimization problem from (5.13–5.14) to maximize the expected gain conditioned on the history. As a result, we conclude SUPER* is locally optimal for the final reviewer with any heuristic function.

5.A.2 Proof of Corollary 5.1: Local Optimality of SUPER* for Any Reviewer

The immediate gain from any reviewer $i \in [n]$ is the difference between the gain after and before the reviewer arrived. Formally, the immediate gain from any reviewer $i \in [n]$ is given by the quantity

$$\mathcal{G}_i = \sum_{j \in [d]} \Big(\gamma_p\Big(\sum_{\ell \in [i]} \mathcal{B}_{\ell,j} \Big) - \gamma_p\Big(\sum_{\ell \in [i-1]} \mathcal{B}_{\ell,j} \Big) \Big) + \lambda \sum_{j \in [d]} \gamma_r(\pi_i(j), S_{i,j}),$$

where $\mathcal{B}_{\ell,j}$ is a Bernoulli random variable representing the random bid of a reviewer $\ell \in [n]$ on a paper $j \in [d]$ that depends on the position the paper was shown to the reviewer.

The optimization problem to maximize the immediate expected gain from reviewer $i \in [n]$ conditioned on the history (see Definition 5.1) is thus given by

$$\max_{\pi_i \in \Pi_d} \quad \sum_{j \in [d]} \mathbb{E}\Big[\gamma_p\Big(\sum_{\ell \in [i]} \mathcal{B}_{\ell,j} \Big) - \gamma_p\Big(\sum_{\ell \in [i-1]} \mathcal{B}_{\ell,j} \Big) \big| \mathcal{H}_{i-1} \Big] + \lambda \sum_{j \in [d]} \mathbb{E}[\gamma_r(\pi_i(j), S_{i,j})].$$

Since $\sum_{\ell \in [i-1]} \mathcal{B}_{\ell,j}$ is the deterministic bid count $g_{i-1,j}$ conditioned on the history $\mathcal{H}_{i-1}$ for each paper $j \in [d]$ and the reviewer-side gain from reviewer $i \in [n]$ is deterministic given a fixed paper ordering $\pi_i \in \Pi_d$, the previous optimization problem is equivalently given by

$$\max_{\pi_i \in \Pi_d} \quad \sum_{j \in [d]} \mathbb{E}[\gamma_p(g_{i-1,j} + \mathcal{B}_{i,j}) - \gamma_p(g_{i-1,j})] + \lambda \sum_{j \in [d]} \gamma_r(\pi_i(j), S_{i,j}). \tag{5.16}$$

We now evaluate the expectation in (5.16) using (5.11) and then simplify to obtain the optimization problem

$$\max_{\pi_i \in \Pi_d} \quad \sum_{j \in [d]} f(\pi_i(j), S_{i,j})(\gamma_p(g_{i-1,j}+1) - \gamma_p(g_{i-1,j})) + \lambda \sum_{j \in [d]} \gamma_r(\pi_i(j), S_{i,j}). \tag{5.17}$$

The problem in (5.17) is equivalent to that given in (5.12) from the proof of Theorem 5.1 up to the reviewer index. Consequently, we follow the steps after (5.12) in the proof of Theorem 5.1 to simplify (5.17) into a tractable representation. In doing so, we get that the problem in (5.17) is equivalent to the linear program in (5.13) with the weights

$$w_{j,k} = f(k, S_{i,j})(\gamma_p(g_{i-1,j}+1) - \gamma_p(g_{i-1,j})) + \lambda \gamma_r(k, S_{i,j}) \ \forall \ j, k \in [d].$$

Local optimality of SUPER* with zero heuristic for any reviewer. To determine the ordering of papers to show any reviewer $i \in [n]$ for the general class of assumed gain and bidding functions, SUPER* calls Algorithm 5.3. Algorithm 5.3 solves the optimization problem in (5.13) using the weights

$$w_{j,k} = f(k, S_{i,j})(\gamma_p(g_{i-1,j}+h_{i,j}+1) - \gamma_p(g_{i-1,j}+h_{i,j})) + \lambda \gamma_r(k, S_{i,j}) \ \forall \ j, k \in [d].$$

For any reviewer $i \in [n]$, given the zero heuristic function, $h_i = 0$ by definition. This means SUPER* with zero heuristic selects the paper ordering for any reviewer by solving the optimization problem to maximize the immediate expected gain conditioned on the history, so we conclude it is locally optimal for any reviewer.

5.A.3 Proof of Theorem 5.2: Suboptimality of Baselines for Final Reviewer

The organization of this proof is as follows. In Section 5.A.3.1, we present notation and preliminary information common to the analysis for each of the baselines. We prove the suboptimality bounds for the SIM, BID, and RAND baselines separately in Sections 5.A.3.2, 5.A.3.3, and 5.A.3.4, respectively. Combining the results for each of the baselines proves the theorem statement. We conclude in Section 5.A.3.5 with proofs of technical lemmas invoked in the analysis of the baselines.

5.A.3.1 Notation and Preliminaries

We denote the gain from the final reviewer n of an arbitrary algorithm ALG presenting a potentially random paper ordering $\pi_n^{\texttt{ALG}}$ to the reviewer as

$$\mathcal{G}_n^{\texttt{ALG}} = \mathcal{G}_{p,n}^{\texttt{ALG}} + \lambda \mathcal{G}_{r,n}^{\texttt{ALG}}.$$

The paper-side gain from the final reviewer $\mathcal{G}_{p,n}^{\texttt{ALG}}$ is given by

$$\mathcal{G}_{p,n}^{\texttt{ALG}} = \sum_{j \in [d]} \Big(\gamma_p\Big(\sum_{i \in [n]} \mathcal{B}_{i,j} \Big) - \gamma_p\Big(\sum_{i \in [n-1]} \mathcal{B}_{i,j} \Big) \Big),$$

where again $\mathcal{B}_{i,j}$ is a Bernoulli random variable representing the random bid of a reviewer $i \in [n]$ on a paper $j \in [d]$ that depends on the position the paper was shown to the reviewer. The reviewer-side

gain from the final reviewer $\mathcal{G}_{r,n}^{\texttt{ALG}}$ is given by

$$\mathcal{G}_{r,n}^{\texttt{ALG}} = \sum_{j\in[d]} \gamma_r(\pi_n^{\texttt{ALG}}(j), S_{n,j}).$$

Accordingly,

$$\mathcal{G}_{n}^{\texttt{ALG}} = \sum_{j\in[d]} \Big(\gamma_p\Big(\sum_{i\in[n]} \mathcal{B}_{i,j}\Big) - \gamma_p\Big(\sum_{i\in[n-1]} \mathcal{B}_{i,j}\Big)\Big) + \lambda \sum_{j\in[d]} \gamma_r(\pi_n^{\texttt{ALG}}(j), S_{n,j}).$$

The expected gain from the final reviewer conditioned on the history of bids and paper orderings $\mathcal{H}_{n-1}$ (see Definition 5.1) is given by

$$\mathbb{E}[\mathcal{G}_{n}^{\texttt{ALG}}|\mathcal{H}_{n-1}] = \mathbb{E}[\mathcal{G}_{p,n}^{\texttt{ALG}}|\mathcal{H}_{n-1}] + \lambda\mathbb{E}[\mathcal{G}_{r,n}^{\texttt{ALG}}|\mathcal{H}_{n-1}]$$

where the expectation is with respect to the randomness in the algorithm and the bids from the final reviewer. Observe that

$$\mathbb{E}[\mathcal{G}_{p,n}^{\texttt{ALG}}|\mathcal{H}_{n-1}] = \mathbb{E}_{\pi_n^{\texttt{ALG}}}\Big[\sum_{j\in[d]} f(\pi_n^{\texttt{ALG}}(j), S_{n,j})(\gamma_p(g_{n-1,j}+1) - \gamma_p(g_{n-1,j}))\Big]$$

since $\sum_{i\in[n-1]} \mathcal{B}_{i,j}$ is the deterministic quantity $g_{n-1,j}$ for each $j \in [d]$ conditioned on $\mathcal{H}_{n-1}$ and $\mathcal{B}_{n,j}$ is a Bernoulli random variable with parameter $p_{n,j} = f(\pi_n^{\texttt{ALG}}(j), S_{n,j})$ for each $j \in [d]$ given the fixed paper ordering $\pi_n^{\texttt{ALG}}$. Moreover,

$$\mathbb{E}[\mathcal{G}_{r,n}^{\texttt{ALG}}|\mathcal{H}_{n-1}] = \mathbb{E}_{\pi_n^{\texttt{ALG}}}\Big[\sum_{j\in[d]} \gamma_r(\pi_n^{\texttt{ALG}}(j), S_{n,j})\Big].$$

It follows that

$$\mathbb{E}[\mathcal{G}_{n}^{\texttt{ALG}}|\mathcal{H}_{n-1}] = \mathbb{E}_{\pi_n^{\texttt{ALG}}}\Big[\sum_{j\in[d]} f(\pi_n^{\texttt{ALG}}(j), S_{n,j})(\gamma_p(g_{n-1,j}+1) - \gamma_p(g_{n-1,j})) + \lambda\sum_{j\in[d]} \gamma_r(\pi_n^{\texttt{ALG}}(j), S_{n,j})\Big].$$

The given biding function can be decomposed into the form

$$f(\pi_i(j), S_{i,j}) = \frac{S_{i,j}}{\log_2(\pi_i(j)+1)} = S_{i,j} f^{\pi}(\pi_i(j)) \tag{5.18}$$

where

$$f^{\pi}(\pi_i(j)) = \frac{1}{\log_2(\pi_i(j)+1)}$$

denotes the component of the bidding function f that only depends on the paper ordering and is independent of the similarity score. The reviewer-side gain function can similarly be decomposed into the form

$$\gamma_r(\pi_i(j), S_{i,j}) = (2^{S_{i,j}} - 1) f^{\pi}(\pi_i(j)). \tag{5.19}$$

Using the decomposed forms of the bidding function and the reviewer-side gain function

from (5.18) and (5.19), the expected gain from the final reviewer n of an arbitrary algorithm $\texttt{ALG}$ is given by

$$\begin{aligned}\mathbb{E}[\mathcal{G}_n^{\texttt{ALG}}|\mathcal{H}_{n-1}] &= \mathbb{E}[\mathcal{G}_{p,n}^{\texttt{ALG}}|\mathcal{H}_{n-1}] + \lambda\mathbb{E}[\mathcal{G}_{r,n}^{\texttt{ALG}}|\mathcal{H}_{n-1}] \\ &= \mathbb{E}_{\pi_n^{\texttt{ALG}}}\Big[\sum_{j\in[d]} S_{n,j}(\gamma_p(g_{n-1,j}+1) - \gamma_p(g_{n-1,j}))f^\pi(\pi_n^{\texttt{ALG}}(j)) + \lambda\sum_{j\in[d]}(2^{S_{n,j}}-1)f^\pi(\pi_n^{\texttt{ALG}}(j))\Big]. \quad (5.20)\end{aligned}$$

The optimal paper ordering for the final reviewer is thus given by the solution to the following optimization problem

$$\pi_n^* = \arg\max_{\pi_n\in\Pi_d} \quad \sum_{j\in[d]} \alpha_{n,j} f^\pi(\pi_n(j)) \quad (5.21)$$

where

$$\alpha_{n,j} = S_{n,j}(\gamma_p(g_{n-1,j}+1) - \gamma_p(g_{n-1,j})) + \lambda(2^{S_{n,j}}-1) \;\forall\; j\in[d]. \quad (5.22)$$

The optimal solution to (5.21) ranks papers in a decreasing order of their corresponding values in $\{\alpha_{n,j}\}_{j\in[d]}$ since the function f^π is decreasing in the decision variable. This observation will be used to obtain an explicit form of π_n^* for each problem we subsequently construct to show the suboptimality of the baselines. From Theorem 5.1, $\texttt{SUPER}^*$ with any heuristic is optimal for the final reviewer, which means $\pi_n^{\texttt{SUPER}^*} = \pi_n^*$. See that $\pi_n^{\texttt{SUPER}^*}$ is a non-random quantity given a deterministic tie-breaking mechanism and the expected gain is independent of the tie-breaking mechanism. Thus, without loss of generality, we assume ties are broken by the paper indexes in favor of $j < j'$ for $\texttt{SUPER}^*$.

We need to compare the expected gain obtained from the final reviewer using the $\texttt{SUPER}^*$ algorithm presenting the optimal paper ordering $\pi_n^{\texttt{SUPER}^*} = \pi_n^*$ with any baseline $\texttt{ALG} \in \{\texttt{SIM}, \texttt{BID}, \texttt{RAND}\}$ presenting a potentially random ordering $\pi_n^{\texttt{ALG}}$. Therefore, we analyze the quantity

$$\begin{aligned}\mathbb{E}[\mathcal{G}_n^{\texttt{SUPER}^*} - \mathcal{G}_n^{\texttt{ALG}}|\mathcal{H}_{n-1}] &= \mathbb{E}[\mathcal{G}_{p,n}^{\texttt{SUPER}^*} - \mathcal{G}_{p,n}^{\texttt{ALG}}|\mathcal{H}_{n-1}] + \lambda\mathbb{E}[\mathcal{G}_{r,n}^{\texttt{SUPER}^*} - \mathcal{G}_{r,n}^{\texttt{ALG}}|\mathcal{H}_{n-1}] \\ &= \mathbb{E}_{\pi_n^{\texttt{ALG}}}\Big[\sum_{j\in[d]} S_{n,j}(\gamma_p(g_{n-1,j}+1) - \gamma_p(g_{n-1,j}))(f^\pi(\pi_n^{\texttt{SUPER}^*}(j)) - f^\pi(\pi_n^{\texttt{ALG}}(j))) \\ &\qquad + \lambda\sum_{j\in[d]}(2^{S_{n,j}}-1)(f^\pi(\pi_n^{\texttt{SUPER}^*}(j)) - f^\pi(\pi_n^{\texttt{ALG}}(j)))\Big] \quad (5.23)\end{aligned}$$

for each baseline $\texttt{ALG} \in \{\texttt{SIM}, \texttt{BID}, \texttt{RAND}\}$. The $\texttt{SIM}$ and $\texttt{BID}$ algorithms are deterministic for the final reviewer conditioned on the history, up to the tie-breaking mechanism. The $\texttt{RAND}$ algorithm is random, so the expectation over the paper ordering in (5.23) is necessary when analyzing $\texttt{RAND}$. In the remainder of the proof, we analyze $\texttt{SIM}$, then $\texttt{BID}$, and finish with $\texttt{RAND}$.

5.A.3.2 Suboptimality of SIM for Final Reviewer

In this section, we prove the worst case performance of the $\texttt{SIM}$ baseline for the final reviewer.

Intuition. The $\texttt{SIM}$ algorithm directly optimizes the expected reviewer-side gain since it shows papers in a decreasing order of the similarity scores. Consequently, it obtains the maximum expected reviewer-side gain that can be achieved. However, the algorithm gives no attention to the number of bids on each paper, which play an important role in the expected paper-side gain that can be

obtained upon the arrival of the final reviewer. This point suggests that SIM may be suboptimal for the combined objective.

To build intuition for when this can occur, consider that there is only a pair of papers j and j'. Moreover, suppose paper j has only marginally higher similarity score than paper j', but paper j has significantly more bids than paper j'. In this scenario, the expected reviewer-side gain of any paper ordering is nearly equal. However, showing paper j' ahead of paper j results in significantly higher expected paper-side gain owing to the diminishing returns of bids from the paper-side gain function. Since SIM instead shows paper j ahead of paper j', it would be suboptimal. The following construction now generalizes this observation.

Construction. We now construct a problem instance that will be used to prove SIM is significantly suboptimal for the final reviewer in the worst case. Consider the similarity scores for the final reviewer to be $S_{n,j} = 1 - 1/(j\epsilon)$ for each paper $j \in [d]$, where $\epsilon = (1+\lambda)e^{e^e}$ and $\lambda \geq 0$ is the fixed and given trade-off parameter. In this construction, the similarity scores for the final reviewer are nearly equal, but they are increasing in the paper index. For the time being, assume the number of papers d is even. At the end of this section, we handle when the number of papers d is odd. Let the number of bids on the papers from previous reviewers be

$$g_{n-1,j} = \begin{cases} 0, & \text{if } j \in \{1, \ldots, d/2\} \\ 1, & \text{if } j \in \{d/2+1, \ldots, d\}. \end{cases}$$

The bid counts are such that papers among the top half of the similarity scores obtained a bid in the past, and papers among the bottom half of the similarity scores did not obtain any bids from previous reviewers.

We now derive the explicit form of the optimal paper ordering for the final reviewer. Recall from (5.22) that the weights of the optimization problem for the final reviewer given in (5.21) are defined by

$$\alpha_{n,j} = S_{n,j}(\gamma_p(g_{n-1,j}+1) - \gamma_p(g_{n-1,j})) + \lambda(2^{S_{n,j}} - 1) \;\forall\; j \in [d]. \tag{5.24}$$

Moreover, from the structure of the optimization problem in (5.21), if $\alpha_{n,j} > \alpha_{n,j'}$, then $\pi_n^{\texttt{SUPER}^*}(j) < \pi_n^{\texttt{SUPER}^*}(j')$ so that paper j is shown ahead of paper j' in the ranking. Observe that $\alpha_{n,j}$ is increasing in the similarity score $S_{n,j}$ and decreasing in the number of bids $g_{n-1,j}$ for each $j \in [d]$. Consequently, if a pair of papers $j, j' \in [d]$ are such that $g_{n-1,j} = g_{n-1,j'}$ and $S_{n,j} > S_{n,j'}$, then $\alpha_{n,j} > \alpha_{n,j'}$ and in turn, $\pi_n^{\texttt{SUPER}^*}(j) < \pi_n^{\texttt{SUPER}^*}(j')$.

We now show that in the optimal ordering for this instance, each paper with zero bids is shown ahead of each paper with a non-zero number of bids. Among the set of papers with zero bids, namely those indexed by $\{1, \ldots, d/2\}$, the minimum similarity score $S_{n,j}$ and minimum weight $\alpha_{n,j}$ occur at $j = 1$. Among the set of papers with a non-zero number of bids, namely those indexed by $\{d/2+1, \ldots, d\}$, the maximum similarity score $S_{i,j'}$ and maximum weight $\alpha_{n,j'}$ occur at $j' = d$. Thus, if $\alpha_{n,1} - \alpha_{n,d} > 0$, then we can conclude each paper with zero bids is shown ahead of each paper with a non-zero number of bids. To prove this, we need the following lemma, the proof of which can be found in Section 5.A.3.5.1.

Lemma 5.1. *For $\gamma_p(x) = \sqrt{x}$, any $\lambda \geq 0$, $d \geq 2$, and $\epsilon = (1+\lambda)e^{e^e}$, it must be that*

$$(\gamma_p(1) - \gamma_p(0))(1 - 1/\epsilon) - (\gamma_p(2) - \gamma_p(1))(1 - 1/(d\epsilon)) + \lambda(2^{(1-1/\epsilon)} - 2^{(1-1/(d\epsilon))}) \geq 1/2. \tag{5.25}$$

From the given similarity scores and bid counts, and then applying Lemma 5.1, we obtain

$$\alpha_{n,1} - \alpha_{n,d} = (\gamma_p(1) - \gamma_p(0))(1 - 1/\epsilon) - (\gamma_p(2) - \gamma_p(1))(1 - 1/(d\epsilon)) + \lambda(2^{(1-1/\epsilon)} - 2^{(1-1/(d\epsilon))}) \geq 1/2.$$

Therefore, the optimal paper ordering shows all papers with zero bids ahead of every paper with a non-zero number of bids. Moreover, recall if a pair of papers $j, j' \in [d]$ are such that $g_{n-1,j} = g_{n-1,j'}$ and $S_{n,j} > S_{n,j'}$, then $\alpha_{n,j} > \alpha_{n,j'}$. This fact allows us to determine that among each group of papers (zero and non-zero bids), the optimal paper ordering presents the papers in decreasing order of the similarity scores.

Combining the previous conclusions, $\texttt{SUPER}^*$ shows the optimal paper ordering

$$\pi_n^{\texttt{SUPER}^*}(j) = \begin{cases} d/2 - j + 1, & \text{if } j \in \{1, \ldots, d/2\} \\ d + d/2 + 1 - j, & \text{if } j \in \{d/2 + 1, \ldots, d\}. \end{cases} \tag{5.26}$$

The SIM algorithm shows papers in a decreasing order of the similarity scores so that $\pi_n^{\texttt{SIM}}(j) = d - j + 1$ for $j \in [d]$. Observe that for this problem, $\texttt{SUPER}^*$ shows the papers with zero bids much earlier in the paper ordering than SIM. We now move on to lower bounding (5.23) for this construction and begin by considering the expected paper-side gain.

Bounding the expected paper-side gain. Substituting the similarity scores, the number of bids on each paper, and the (deterministic) paper orderings presented by each algorithm for this construction into the paper-side component of (5.23), we obtain

$$\begin{aligned} \mathbb{E}[\mathcal{G}_{p,n}^{\texttt{SUPER}^*} - \mathcal{G}_{p,n}^{\texttt{SIM}} | \mathcal{H}_{n-1}] &= (\gamma_p(1) - \gamma_p(0)) \sum_{j=1}^{d/2} (1 - 1/(j\epsilon))(f^\pi(d/2 - j + 1) - f^\pi(d - j + 1)) \\ &\quad + (\gamma_p(2) - \gamma_p(1)) \sum_{j=d/2+1}^{d} (1 - 1/(j\epsilon))(f^\pi(d + d/2 + 1 - j) - f^\pi(d - j + 1)). \end{aligned}$$

Manipulating the indexing of the sum over the last half of the papers gives

$$\begin{aligned} \mathbb{E}[\mathcal{G}_{p,n}^{\texttt{SUPER}^*} - \mathcal{G}_{p,n}^{\texttt{SIM}} | \mathcal{H}_{n-1}] &= (\gamma_p(1) - \gamma_p(0)) \sum_{j=1}^{d/2} (1 - 1/(j\epsilon))(f^\pi(d/2 - j + 1) - f^\pi(d - j + 1)) \\ &\quad + (\gamma_p(2) - \gamma_p(1)) \sum_{j=1}^{d/2} (1 - 1/((j + d/2)\epsilon))(f^\pi(d - j + 1) - f^\pi(d/2 - j + 1)). \end{aligned}$$

Noting that $(f^\pi(d/2-j+1)-f^\pi(d-j+1)) = -(f^\pi(d-j+1)-f^\pi(d/2-j+1))$, we obtain

$$\mathbb{E}[\mathcal{G}_{p,n}^{\texttt{SUPER}^*} - \mathcal{G}_{p,n}^{\texttt{SIM}}|\mathcal{H}_{n-1}] = (\gamma_p(1)-\gamma_p(0))\sum_{j=1}^{d/2}(1-1/(j\epsilon))(f^\pi(d/2-j+1)-f^\pi(d-j+1)) \\ -(\gamma_p(2)-\gamma_p(1))\sum_{j=1}^{d/2}(1-1/((j+d/2)\epsilon))(f^\pi(d/2-j+1)-f^\pi(d-j+1)).$$

For every $j \in [d/2]$, we have $f^\pi(d/2-j+1)-f^\pi(d-j+1) > 0$ since $d/2-j+1 < d-j+1$ and $f^\pi(\pi_i(j)) = 1/\log_2(\pi_i(j)+1)$ is a decreasing function on the domain $\mathbb{R}_{>0}$. Moreover, for every $j \in [d/2]$, $1-1/(j\epsilon) \geq 1-1/\epsilon$ and $1-1/((j+d/2)\epsilon) \leq 1-1/(d\epsilon)$, and the given paper-side gain function γ_p is increasing on the domain $\mathbb{R}_{\geq 0}$. Thus,

$$\mathbb{E}[\mathcal{G}_{p,n}^{\texttt{SUPER}^*} - \mathcal{G}_{p,n}^{\texttt{SIM}}|\mathcal{H}_{n-1}] \geq ((\gamma_p(1)-\gamma_p(0))(1-1/\epsilon)-(\gamma_p(2)-\gamma_p(1))(1-1/(d\epsilon))) \\ \sum_{j=1}^{d/2}(f^\pi(d/2-j+1)-f^\pi(d-j+1)).$$

Finally, manipulating the indexing of the sum gives

$$\mathbb{E}[\mathcal{G}_{p,n}^{\texttt{SUPER}^*} - \mathcal{G}_{p,n}^{\texttt{SIM}}|\mathcal{H}_{n-1}] \geq ((\gamma_p(1)-\gamma_p(0))(1-1/\epsilon)-(\gamma_p(2)-\gamma_p(1))(1-1/(d\epsilon))) \\ \sum_{j=1}^{d/2}(f^\pi(j)-f^\pi(j+d/2)). \tag{5.27}$$

We now bound the expected reviewer-side gain including the trade-off parameter λ from (5.23). The steps that follow are analogous to those exercised in bounding the expected paper-side gain.

Bounding the expected reviewer-side gain. Substituting the values of the similarity scores, the number of bids on each paper, and the (deterministic) paper orderings presented by each algorithm into the reviewer-side component of (5.23), we obtain

$$\lambda\mathbb{E}[\mathcal{G}_{r,n}^{\texttt{SUPER}^*} - \mathcal{G}_{r,n}^{\texttt{SIM}}|\mathcal{H}_{n-1}] = \lambda\sum_{j=1}^{d/2}(2^{(1-1/(j\epsilon))}-1)(f^\pi(d/2-j+1)-f^\pi(d-j+1)) \\ +\lambda\sum_{j=d/2+1}^{d}(2^{(1-1/(j\epsilon))}-1)(f^\pi(d+d/2+1-j)-f^\pi(d-j+1)).$$

Manipulating the indexing of the sum over the last half of the papers results in

$$\lambda\mathbb{E}[\mathcal{G}_{r,n}^{\texttt{SUPER}^*} - \mathcal{G}_{r,n}^{\texttt{SIM}}|\mathcal{H}_{n-1}] = \lambda\sum_{j=1}^{d/2}(2^{(1-1/(j\epsilon))} - 1)(f^\pi(d/2-j+1) - f^\pi(d-j+1)) \\ + \lambda\sum_{j=1}^{d/2}(2^{(1-1/((j+d/2)\epsilon))} - 1)(f^\pi(d-j+1) - f^\pi(d/2-j+1)).$$

Noting that $f^\pi(d/2-j+1) - f^\pi(d-j+1) = -(f^\pi(d-j+1) - f^\pi(d/2-j+1))$, we obtain

$$\lambda\mathbb{E}[\mathcal{G}_{r,n}^{\texttt{SUPER}^*} - \mathcal{G}_{r,n}^{\texttt{ALG}}|\mathcal{H}_{n-1}] = \lambda\sum_{j=1}^{d/2}(2^{(1-1/(j\epsilon))} - 1)(f^\pi(d/2-j+1) - f^\pi(d-j+1)) \\ - \lambda\sum_{j=1}^{d/2}(2^{(1-1/((j+d/2)\epsilon))} - 1)(f^\pi(d/2-j+1) - f^\pi(d-j+1)).$$

For every $j \in [d/2]$, we have $f^\pi(d/2-j+1) - f^\pi(d-j+1) > 0$ since $d/2-j+1 < d-j+1$ and $f^\pi(\pi_i(j)) = 1/\log_2(\pi_i(j)+1)$ is a decreasing function on the domain $\mathbb{R}_{>0}$. Furthermore, observe that for every $j \in [d/2]$, $1-1/(j\epsilon) \geq 1-1/(\epsilon)$ and $1-1/((j+d/2)\epsilon) \leq 1-1/(d\epsilon)$. This set of facts leads to the bound

$$\lambda\mathbb{E}[\mathcal{G}_{r,n}^{\texttt{SUPER}^*} - \mathcal{G}_{r,n}^{\texttt{SIM}}|\mathcal{H}_{n-1}] \geq \lambda\big(2^{(1-1/\epsilon)} - 2^{(1-1/(d\epsilon))}\big)\sum_{j=1}^{d/2}(f^\pi(d/2-j+1) - f^\pi(d-j+1)).$$

To finish this sequence of steps, we manipulate the indexing of the sum to conclude

$$\mathbb{E}[\mathcal{G}_{r,n}^{\texttt{SUPER}^*} - \mathcal{G}_{r,n}^{\texttt{SIM}}|\mathcal{H}_{n-1}] \geq \lambda\big(2^{(1-1/\epsilon)} - 2^{(1-1/(d\epsilon))}\big)\sum_{j=1}^{d/2}(f^\pi(j) - f^\pi(j+d/2)). \tag{5.28}$$

Completing the lower bound. Combining the bounds on the expected paper-side and reviewer-side gain terms from (5.27) and (5.28), we obtain an initial lower bound given by

$$\mathbb{E}[\mathcal{G}_{n}^{\texttt{SUPER}^*} - \mathcal{G}_{n}^{\texttt{SIM}}|\mathcal{H}_{n-1}] \geq \mathcal{C}\sum_{j=1}^{d/2}(f^\pi(j) - f^\pi(j+d/2)), \tag{5.29}$$

where for ease of notation we define

$$\mathcal{C} = (\gamma_p(1)-\gamma_p(0))(1-1/\epsilon) - (\gamma_p(2)-\gamma_p(1))(1-1/(d\epsilon)) + \lambda(2^{(1-1/\epsilon)} - 2^{(1-1/(d\epsilon))}). \tag{5.30}$$

We apply Lemma 5.1 to get that $\mathcal{C} \geq 1/2$. The following lemma, the proof of which can be found in Section 5.A.3.5.2, provides a bound on the sum in (5.29).

Lemma 5.2. *Let $f^\pi(x) = 1/\log_2(x+1)$. Fix d to be an even integer such that $d \geq 2$. Then,*

$$\sum_{j=1}^{d/2}(f^\pi(j) - f^\pi(j+d/2)) \geq \frac{d}{16\log_2^2(d)}.$$

From (5.29) along with the fact that $\mathcal{C} \geq 1/2$ and Lemma 5.2, we obtain

$$\mathbb{E}[\mathcal{G}_n^{\texttt{SUPER}^*} - \mathcal{G}_n^{\texttt{SIM}}|\mathcal{H}_{n-1}] \geq \frac{d}{32\log_2^2(d)}. \tag{5.31}$$

Lemma 5.2, and consequently the bound in (5.31), are applicable when d is even. The following lemma shows that an equivalent result (up to constants) holds when the number of papers d is odd. The bound is obtained by looking at an identical problem construction for $d' = d-1$ papers, and then including an additional paper that has a similarity score of zero with the final reviewer and one previous bid. This change is such that both $\texttt{SUPER}^*$ and SIM show the additional paper last, and moreover, the expected gain from paper d is zero since the similarity score is zero.

Lemma 5.3. *If d is odd, then in the worst case for the final reviewer under the assumptions of Theorem 5.2,*

$$\mathbb{E}[\mathcal{G}_n^{\textit{SUPER}^*} - \mathcal{G}_n^{\textit{SIM}}|\mathcal{H}_{n-1}] \geq \frac{d}{64\log_2^2(d)}.$$

The proof of Lemma 5.3 is in Section 5.A.3.5.3.

Combining the bound from (5.31) which holds for d even with the bound from Lemma 5.3 which holds for d odd, we find that for every $d \geq 2$ and $\lambda \geq 0$,

$$\mathbb{E}[\mathcal{G}_n^{\texttt{SUPER}^*} - \mathcal{G}_n^{\texttt{SIM}}|\mathcal{H}_{n-1}] \geq \frac{d}{64\log_2^2(d)}.$$

This proves the claim in Theorem 5.2 stating that there exists a constant $c > 0$ such that for every $d \geq 2$ and $\lambda \geq 0$, SIM is suboptimal by an additive factor of at least $cd/\log_2^2(d)$ in the worst case for the final reviewer.

5.A.3.3 Suboptimality of BID for Final Reviewer

In this section, we prove the worst case performance of the BID baseline for the final reviewer.

Intuition. The BID algorithm greedily optimizes the minimum bid count since it shows papers in an increasing order of the number of bids received previously. The underlying problem with this method is that the paper ordering selected for any reviewer is independent of the similarity scores for that reviewer (up to serving as a tie-breaking mechanism). Since the paper-side gain function and the reviewer-side gain function both depend on the similarity scores, this property of BID leads to suboptimality in terms of both the expected paper-side gain and the reviewer-side gain.

To build some intuition for when BID is suboptimal, consider there is only a pair of papers j and j'. Moreover, suppose paper j has a much higher similarity score than paper j' and paper j has only one more bid than j'. In this scenario, the expected reviewer-side gain from showing paper j

ahead of paper j' is significantly higher than showing paper j' ahead of paper j. Moreover, since the probability of obtaining a bid on paper j is significantly higher at a given position in the paper ordering than for paper j' and the number of bids on the papers are nearly equal, the expected paper-side gain is also maximized if paper j is shown ahead of paper j'. Since `BID` instead shows paper j' ahead of paper j, it would be suboptimal for both paper-side and reviewer-side gain. The following construction now generalizes this observation.

Construction. We now construct a problem instance that will be used to prove `BID` is significantly suboptimal for the final reviewer in the worst case. Consider the similarity scores for the final reviewer as

$$S_{n,j} = \begin{cases} 1, & \text{if } j \in \{1, \dots, d/2\} \\ 0, & \text{if } j \in \{d/2+1, \dots, d\}. \end{cases}$$

For now, assume the number of papers d is even. We handle when the number of papers d is odd at the end of this proof. Let the number of bids on the papers from previous reviewers be such that

$$g_{n-1,j} = \begin{cases} 1, & \text{if } j \in \{1, \dots, d/2\} \\ 0, & \text{if } j \in \{d/2+1, \dots, d\}. \end{cases}$$

In words, half of the papers have a similarity score of one and have recieved bids, and the other half of the papers have a similarity score of zero and have obtained no bids.

We now derive the optimal paper ordering for the final reviewer. Recall from (5.22) that the weights of the optimization problem for the final reviewer given in (5.21) are defined by

$$\alpha_{n,j} = S_{n,j}(\gamma_p(g_{n-1,j}+1) - \gamma_p(g_{n-1,j})) + \lambda(2^{S_{n,j}} - 1) \;\forall\; j \in [d].$$

Moreover, from the structure of the optimization problem in (5.21), if $\alpha_{n,j} > \alpha_{n,j'}$, then $\pi_n^{\texttt{SUPER}^*}(j) < \pi_n^{\texttt{SUPER}^*}(j')$ so that paper j is shown ahead of paper j' in the ranking. Observe that for each $j \in \{1, \dots, d/2\}$, $\alpha_{n,j}$ is a fixed number. Similarly, for each $j' \in \{d/2+1, \dots, d\}$, $\alpha_{n,j'}$ is a fixed number. If $\alpha_{n,j} - \alpha_{n,j'} > 0$ for any $j \in \{1, \dots, d/2\}$ and $j' \in \{d/2+1, \dots, d\}$, then we can conclude each paper with a bid is shown ahead of each paper without a bid. We consider $j = 1$ and $j' = d$. Since $\gamma_p(x) = \sqrt{x}$ and $\alpha_{n,d} = 0$,

$$\alpha_{n,1} - \alpha_{n,d} = \gamma_p(2) - \gamma_p(1) + \lambda = \sqrt{2} - 1 + \lambda \geq 1/3 + \lambda. \tag{5.32}$$

Consequent of the fact $\lambda \geq 0$, we conclude $\alpha_{n,1} - \alpha_{n,d} > 0$, which means $\texttt{SUPER}^*$ shows each paper with a bid ahead of each paper without a bid. Finally, since $\alpha_{n,j}$ is a fixed number for each $j \in \{1, \dots, d/2\}$ and $\alpha_{n,j'}$ is a fixed number for each $j' \in \{d/2+1, \dots, d\}$, as long as $\pi_n^{\texttt{SUPER}^*}(j) < \pi_n^{\texttt{SUPER}^*}(j')$ for every j, j' pair, then the paper ordering is optimal. In other words, any paper ordering which shows the papers with a bid in an arbitrary order followed by the papers without a bid in an arbitrary order is optimal.

We conclude $\pi_n^{\texttt{SUPER}^*}(j) = j$ for each $j \in [d]$, where without loss of generality, to simplify the analysis, we assume if a pair of papers have equal weights in the optimization problem, then ties are broken in order of the paper indexes since the tie-breaking mechanism will not change the expected gain the paper ordering obtains from the final reviewer.

The BID baseline will show papers in an increasing order of the number of bids so that

$$\pi_n^{\texttt{BID}}(j) = \begin{cases} j + d/2, & \text{if } j \in \{1, \ldots, d/2\} \\ j - d/2, & \text{if } j \in \{d/2+1, \ldots, d\}. \end{cases}$$

This paper ordering is derived from recalling that BID breaks ties by the similarity scores and further ties are broken uniformly at random. However, without loss of generality, to simplify the analysis, we assume if a pair of papers have equal similarity scores and bid counts, then ties are broken in order of the paper indexes since the tie-breaking mechanism among this set of papers will not impact the expected gain. We now move on to lower bounding (5.23) for this construction.

Bounding the expected gain. Substituting the similarity scores, the number of bids on each paper, and the (deterministic) paper orderings presented by each algorithm for this construction into (5.23), we obtain

$$\mathbb{E}[\mathcal{G}_n^{\texttt{SUPER}^*} - \mathcal{G}_n^{\texttt{BID}} | \mathcal{H}_{n-1}] = \sum_{j=1}^{d/2} (\gamma_p(2) - \gamma_p(1))(f^\pi(j) - f^\pi(j + d/2)) + \lambda \sum_{j=1}^{d/2} (f^\pi(j) - f^\pi(j + d/2))$$

where the terms for papers in the set $\{d/2+1, \ldots, d\}$ dropped out since the similarity scores are zero. Simplifying the expression, we obtain

$$\mathbb{E}[\mathcal{G}_n^{\texttt{SUPER}^*} - \mathcal{G}_n^{\texttt{BID}} | \mathcal{H}_{n-1}] = (\gamma_p(2) - \gamma_p(1) + \lambda) \sum_{j=1}^{d/2} (f^\pi(j) - f^\pi(j + d/2)). \tag{5.33}$$

From (5.32), we get

$$\gamma_p(2) - \gamma_p(1) + \lambda \geq 1/3 + \lambda. \tag{5.34}$$

Moreover, from Lemma 5.2,

$$\sum_{j=1}^{d/2} (f^\pi(j) - f^\pi(j + d/2)) \geq \frac{d}{16 \log_2^2(d)}. \tag{5.35}$$

Combining (5.33), (5.34), and (5.35), we obtain

$$\mathbb{E}[\mathcal{G}_n^{\texttt{SUPER}^*} - \mathcal{G}_n^{\texttt{BID}} | \mathcal{H}_{n-1}] \geq \Big(\frac{1}{48} + \frac{\lambda}{16}\Big)\Big(\frac{d}{\log_2^2(d)}\Big) \tag{5.36}$$

whenever the number of papers d is even.

Lemma 5.2 and the bound in (5.36) are applicable when d is even. The following lemma shows that an equivalent result (up to constants) holds when the number of papers d is odd. The approach to obtain the result is similar to that for deriving Lemma 5.3. We obtain the bound for an odd number of papers d by looking at an identical problem construction for $d' = d - 1$ papers, and then include an additional paper that has a similarity score of zero with the final reviewer and one previous bid. This change is such that both SUPER* and BID show the additional paper last, and moreover, the expected gain from paper d is zero since the similarity score is zero.

Lemma 5.4. *If d is odd, then in the worst case for the final reviewer under the assumptions of Theorem 5.2,*

$$\mathbb{E}[\mathcal{G}_n^{\textit{SUPER}^*} - \mathcal{G}_n^{\textit{BID}} | \mathcal{H}_{n-1}] \geq \Big(\frac{1}{96} + \frac{\lambda}{32}\Big)\Big(\frac{d}{\log_2^2(d)}\Big).$$

The proof of Lemma 5.4 is in Section 5.A.3.5.4.

Combining the bound from (5.36) which holds for d even with the bound from Lemma 5.4 which holds for d odd, we find that for every $d \geq 2$ and $\lambda \geq 0$,

$$\mathbb{E}[\mathcal{G}_n^{\texttt{SUPER}^*} - \mathcal{G}_n^{\texttt{BID}} | \mathcal{H}_{n-1}] \geq \Big(\frac{1}{96} + \frac{\lambda}{32}\Big)\Big(\frac{d}{\log_2^2(d)}\Big). \tag{5.37}$$

This proves the claim in Theorem 5.2 that there exists a constant $c > 0$ such that for all $d \geq 2$ and $\lambda \geq 0$, `BID` is suboptimal by an additive factor of at least $cd \max\{1, \lambda\}/\log_2^2(d)$ in the worst case for the final reviewer.

5.A.3.4 Suboptimality of `RAND` for Final Reviewer

In this section, we prove the the worst case performance of the `RAND` baseline for the final reviewer.

Intuition. The `RAND` algorithm selects an ordering of papers to show a reviewer uniformly at random from the set of permutations. Since this method is agnostic to the similarity scores and the number of bids, `RAND` can select highly suboptimal paper orderings with some non-zero probability.

To see when this can occur, consider the example that provided intuition for the suboptimality of `BID` in Section 5.A.3.3 that consisted of only a pair of papers j and j'. In this example, paper j has a much higher similarity score than paper j' and paper j has only one more bid than j'. The expected reviewer-side gain from showing paper j ahead of paper j' is significantly higher than showing paper j' ahead of paper j. Moreover, since the probability of obtaining a bid on paper j is significantly higher at a given position in the paper ordering than for paper j' and the number of bids on the papers are nearly equal, the expected paper-side gain is also maximized if paper j is shown ahead of paper j'. Since there are only two permutations of the papers that can be selected, with probability $1/2$, `RAND` would show paper j' ahead of paper j and be suboptimal for both paper-side and reviewer-side gain. The problem construction from Section 5.A.3.3 is sufficient to generalize this observation. For completeness, we repeat the construction below.

Construction. In the remainder of the proof, we show the problem construction from Section 5.A.3.3 can be used to prove `RAND` is significantly suboptimal for the final reviewer in the worst case. In this construction, the similarity scores for the final reviewer are

$$S_{n,j} = \begin{cases} 1, & \text{if } j \in \{1, \ldots, d/2\} \\ 0, & \text{if } j \in \{d/2+1, \ldots, d\}. \end{cases}$$

For now, assume the number of papers d is divisible by four. We deal with a number of papers d that is not divisible by four at the end of this section. The number of bids on the papers from

previous reviewers are

$$g_{n-1,j} = \begin{cases} 1, & \text{if } j \in \{1, \dots, d/2\} \\ 0, & \text{if } j \in \{d/2+1, \dots, d\}. \end{cases}$$

In Section 5.A.3.3, we showed that $\texttt{SUPER}^*$ selects the optimal ordering $\pi_n^{\texttt{SUPER}^*}(j) = j$ for each $j \in [d]$. The $\texttt{RAND}$ baseline will select a paper ordering $\pi_n^{\texttt{RAND}}$ uniformly at random from the set of permutations Π_d.

Bounding the expected gain. For this construction, we need to lower bound (5.23). As an initial step, we simplify the quantity by substituting the similarity scores, the number of bids on each paper, and the paper ordering presented by $\texttt{SUPER}^*$ to obtain

$$\mathbb{E}[\mathcal{G}_n^{\texttt{SUPER}^*} - \mathcal{G}_n^{\texttt{RAND}} | \mathcal{H}_{n-1}] = \mathbb{E}_{\pi_n^{\texttt{RAND}}}\Big[\sum_{j=1}^{d/2} (\gamma_p(2) - \gamma_p(1))(f^\pi(j) - f^\pi(\pi_n^{\texttt{RAND}}(j))) + \lambda \sum_{j=1}^{d/2} (f^\pi(j) - f^\pi(\pi_n^{\texttt{RAND}}(j)))\Big],$$

where the terms for papers in the set $\{d/2+1, \dots, d\}$ dropped out since the similarity scores are zero. Combining the sums and using the fact that $\gamma_p(2) - \gamma_p(1) = \sqrt{2} - 1 \geq 1/3$ gives

$$\mathbb{E}[\mathcal{G}_n^{\texttt{SUPER}^*} - \mathcal{G}_n^{\texttt{RAND}} | \mathcal{H}_{n-1}] \geq \mathbb{E}_{\pi_n^{\texttt{RAND}}}\Big[(1/3 + \lambda) \sum_{j=1}^{d/2} (f^\pi(j) - f^\pi(\pi_n^{\texttt{RAND}}(j)))\Big]. \tag{5.38}$$

Before proceeding, we provide some intuition that guides the remainder of the proof. The expression in (5.38) only depends on the positions $\texttt{RAND}$ shows the papers in the set $\{1, \dots, d/2\}$ to the final reviewer. Recalling that the given function f^π is decreasing on the domain $\mathbb{R}_{>0}$, we can observe that the number of positive summand in $\sum_{j=1}^{d/2}(f^\pi(j) - f^\pi(\pi_n^{\texttt{RAND}}(j)))$ increases with the number of papers from the set $\{1, \dots, d/2\}$ that are not presented in the set of positions $\{1, \dots, d/2\}$ from a selection $\pi_n^{\texttt{RAND}}$ and the remaining summand are zero. This point suggests if with probability bounded away from zero, sufficiently many papers from the set $\{1, \dots, d/2\}$ are not presented in the set of positions $\{1, \dots, d/2\}$ in the ordering selected by $\texttt{RAND}$, then it should be suboptimal in expectation.

Toward formalizing this line of reasoning, the following lemma provides a lower bound on the probability that $\texttt{RAND}$ selects a paper ordering that shows fewer than $d/4$ papers from the set $\{1, \dots, d/2\}$ in the set of positions $\{1, \dots, d/2\}$. The proof is given in Section 5.A.3.5.5.

Lemma 5.5. *Assume d is divisible by four and consider a set of papers $[d]$. Let $\mathcal{E}$ be the event that a permutation π of the paper set $[d]$ drawn uniformly at random from Π_d has fewer than $d/4$ of the papers $\{1, \dots, d/2\}$ in the positions $\{1, \dots, d/2\}$. Then, $\mathbb{P}(\mathcal{E}) \geq 1/6$.*

Define $T_1 \subset \Pi_d$ as the set of paper orderings with fewer than $d/4$ of the papers from the set $\{1, \dots, d/2\}$ in the set of positions $\{1, \dots, d/2\}$ and $T_2 \subset \Pi_d$ as the set containing the remaining paper orderings so that $T_1 \cup T_2 = \Pi_d$. Now, beginning from (5.38), we evaluate and bound the

expectation as follows:

$$\begin{aligned}\mathbb{E}[\mathcal{G}_n^{\texttt{SUPER}^*} - \mathcal{G}_n^{\texttt{RAND}} | \mathcal{H}_{n-1}] &= (1/3+\lambda) \sum_{\pi_n \in T_1} \mathbb{P}(\pi_n^{\texttt{RAND}} = \pi_n) \sum_{j=1}^{d/2} (f^\pi(j) - f^\pi(\pi_n(j))) \\ &\quad + (1/3+\lambda) \sum_{\pi_n \in T_2} \mathbb{P}(\pi_n^{\texttt{RAND}} = \pi_n) \sum_{j=1}^{d/2} (f^\pi(j) - f^\pi(\pi_n(j))) \\ &\geq (1/3+\lambda) \mathbb{P}(\pi_n^{\texttt{RAND}} \in T_1) \min_{\pi_n \in T_1} \sum_{j=1}^{d/2} (f^\pi(j) - f^\pi(\pi_n(j))) \\ &\quad + (1/3+\lambda) \mathbb{P}(\pi_n^{\texttt{RAND}} \in T_2) \min_{\pi_n \in T_2} \sum_{j=1}^{d/2} (f^\pi(j) - f^\pi(\pi_n(j))).\end{aligned}$$

Observe that $\min_{\pi_n \in T_2} \sum_{j=1}^{d/2} (f^\pi(j) - f^\pi(\pi_n(j))) = 0$ since the optimal paper ordering that shows each paper in the set $\{1, \ldots, d/2\}$ in the set of positions $\{1, \ldots, d/2\}$ is contained in T_2. Moreover, from Lemma 5.5, $\mathbb{P}(\pi_n^{\texttt{RAND}} \in T_1) \geq 1/6$. This results in the bound

$$\mathbb{E}[\mathcal{G}_n^{\texttt{SUPER}^*} - \mathcal{G}_n^{\texttt{RAND}} | \mathcal{H}_{n-1}] \geq \left(\frac{1}{18} + \frac{\lambda}{6}\right) \min_{\pi_n \in T_1} \sum_{j=1}^{d/2} (f^\pi(j) - f^\pi(\pi_n(j))). \tag{5.39}$$

We now need to reason about the minimizer of $\min_{\pi_n \in T_1} \sum_{j=1}^{d/2} (f^\pi(j) - f^\pi(\pi_n(j)))$. Equivalently, we can find the maximizer of $\max_{\pi_n \in T_1} \sum_{j=1}^{d/2} f^\pi(\pi_n(j))$. Since the given function f^π is decreasing on the domain $\mathbb{R}_{>0}$, the quantity $\sum_{j=1}^{d/2} f^\pi(\pi_n(j))$ is maximized when the papers in the set $\{1, \ldots, d/2\}$ are shown the earliest in the ordering π_n that is feasible subject to the constraint that fewer than $d/4$ papers from the set $\{1, \ldots, d/2\}$ are presented in the set of positions $\{1, \ldots, d/2\}$. This means

$$\pi_n(j) = \begin{cases} j, & \text{if } j \in \{1, \ldots, d/4 - 1\} \\ j + d/4 + 1, & \text{if } j \in \{d/4, \ldots, d/2\} \\ j - d/4 - 1, & \text{if } j \in \{d/2 + 1, \ldots, 3d/4 + 1\} \\ j, & \text{if } j \in \{3d/4 + 2, \ldots, d\} \end{cases} \tag{5.40}$$

is a minimizer of $\sum_{j=1}^{d/2} (f^\pi(j) - f^\pi(\pi_n(j)))$ among the set T_1. Substituting the paper ordering from (5.40) as the minimizer into (5.39), we obtain

$$\mathbb{E}[\mathcal{G}_n^{\texttt{SUPER}^*} - \mathcal{G}_n^{\texttt{RAND}} | \mathcal{H}_{n-1}] \geq \left(\frac{1}{18} + \frac{\lambda}{6}\right) \sum_{j=d/4}^{d/2} (f^\pi(j) - f^\pi(j + d/4 + 1)). \tag{5.41}$$

The following lemma provides a bound on the sum in (5.41).

Lemma 5.6. *Let $f^\pi(x) = 1/\log_2(x+1)$ and fix $d \geq 4$ and divisible by four. Then,*

$$\sum_{j=d/4}^{d/2} (f^\pi(j) - f^\pi(j + d/4 + 1)) \geq \frac{d}{32\log_2^2(d)}.$$

The proof of Lemma 5.6 can be found in Section 5.A.3.5.6.

Combining (5.41) and Lemma 5.6 results in the following bound whenever d is divisible by four:

$$\mathbb{E}[\mathcal{G}_n^{\texttt{SUPER}^*} - \mathcal{G}_n^{\texttt{RAND}} | \mathcal{H}_{n-1}] \geq \Big(\frac{1}{576} + \frac{\lambda}{192}\Big)\Big(\frac{d}{\log_2^2(d)}\Big). \tag{5.42}$$

The next lemma shows that if the number of papers d is not divisible by four, an equivalent result (up to constants) holds. For $d \in \{2, 3\}$, the bound is rather immediate since we can compute the probability that `RAND` selects the paper ordering `BID` shows for this construction and then apply the bound from (5.37) on the suboptimality of `BID` that holds for any d. For $d > 3$ and not divisible by four, the bound is obtained by looking at an identical problem construction for the maximum $d' < d$ divisible by four and then including $d - d'$ papers with a similarity score of zero and one previous bid. This change is such that the bound from (5.42) applies as a function of d', so the result then follows immediately.

Lemma 5.7. *If d is not divisible by four, then in the worst case for the final reviewer under the assumptions of Theorem 5.2,*

$$\mathbb{E}[\mathcal{G}_n^{\textit{SUPER}^*} - \mathcal{G}_n^{\textit{RAND}} | \mathcal{H}_{n-1}] \geq \Big(\frac{1}{1728} + \frac{\lambda}{576}\Big)\Big(\frac{d}{\log_2^2(d)}\Big).$$

The proof of Lemma 5.7 can be found in Section 5.A.3.5.7.

Combining the bound from (5.42) which holds for d divisible by four with the bound from Lemma 5.7 which holds for d not divisible by four, we find that for every $d \geq 2$ and $\lambda \geq 0$,

$$\mathbb{E}[\mathcal{G}_n^{\texttt{SUPER}^*} - \mathcal{G}_n^{\texttt{RAND}} | \mathcal{H}_{n-1}] \geq \Big(\frac{1}{1728} + \frac{\lambda}{576}\Big)\Big(\frac{d}{\log_2^2(d)}\Big).$$

This proves there is a constant $c > 0$ such that for all $d \geq 2$ and $\lambda \geq 0$, `RAND` is suboptimal by an additive factor of at least $cd\max\{1, \lambda\}/\log_2^2(d)$ in the worst case for the final reviewer as claimed in Theorem 5.2.

5.A.3.5 Proofs of Lemmas 5.1–5.7

In this section, we present the proofs of technical lemmas stated in the primary proof of Theorem 5.2.

5.A.3.5.1 Proof of Lemma 5.1. Recall from the lemma statement, $\gamma_p(x) = \sqrt{x}$. Moreover, the fixed and given quantities are $\lambda \geq 0$, $d \geq 2$, and $\epsilon = (1+\lambda)e^{e^e}$. We derive the following bound

justified below:

$$
\begin{aligned}
&(\gamma_p(1)-\gamma_p(0))(1-1/\epsilon)-(\gamma_p(2)-\gamma_p(1))(1-1/(d\epsilon))+\lambda(2^{(1-1/\epsilon)}-2^{(1-1/(d\epsilon))}) \\
&\quad\geq (\gamma_p(1)-\gamma_p(0))(1-1/\epsilon)-(\gamma_p(2)-\gamma_p(1))+\lambda(2^{(1-1/\epsilon)}-2) && (5.43)\\
&\quad= (\gamma_p(1)-\gamma_p(0))(1-((1+\lambda)e^{e^e})^{-1})-(\gamma_p(2)-\gamma_p(1))+\lambda(2^{(1-1/((1+\lambda)e^{e^{e^e}})}-2) && (5.44)\\
&\quad\geq (0.99)(\gamma_p(1)-\gamma_p(0))-(\gamma_p(2)-\gamma_p(1))-0.01 && (5.45)\\
&\quad\geq 1/2. && (5.46)
\end{aligned}
$$

We obtain (5.43) using the fact that $(1-1/(d\epsilon)) \leq 1$ for any given d. Equation (5.44) follows from plugging in the explicit form of $\epsilon = (1+\lambda)e^{e^e}$. To see the inequality in (5.45), observe that $(1-((1+\lambda)e^{e^e})^{-1})$ is an increasing function of λ. Consequently, $(1-((1+\lambda)e^{e^e})^{-1}) \geq (1-(e^{e^e})^{-1}) \geq 0.99$. Furthermore, the quantity $\lambda(2^{(1-1/((1+\lambda)e^{e^{e^e}}))}-2)$ is a decreasing function of λ, from which we determine

$$\lambda(2^{(1-1/((1+\lambda)e^{e^{e^e}}))}-2) \geq \lim_{\lambda'\to\infty} \lambda'(2^{(1-1/((1+\lambda')e^{e^{e^e}}))}-2) = -e^{-e^e}\log(4) \geq -0.01.$$

Then, we obtain the final bound in (5.46) as follows:

$$(0.99)(\gamma_p(1)-\gamma_p(0))-(\gamma_p(2)-\gamma_p(1))-0.01 = 0.99-(\sqrt{2}-1)-0.01 \geq 1/2.$$

5.A.3.5.2 Proof of Lemma 5.2. Recall that $f^\pi(x) = 1/\log_2(x+1)$. Fixing $d=2$, we obtain

$$
\begin{aligned}
\sum_{j=1}^{d/2}(f^\pi(j)-f^\pi(j+d/2)) &= f^\pi(1)-f^\pi(2) \\
&= \frac{1}{\log_2(2)}-\frac{1}{\log_2(3)} \\
&\geq \Big(\frac{1}{4}\Big)\Big(\frac{1}{\log_2^2(2)}\Big) && (5.47)\\
&= \Big(\frac{1}{8}\Big)\Big(\frac{d}{\log_2^2(d)}\Big). && (5.48)
\end{aligned}
$$

The inequality in (5.47) follows from the fact that $\log_2(3)-1 \geq 1/2$ and $\log_2(3) \leq 2\log_2(2)$.

Now consider $d \geq 4$ and d even. We derive a bound as follows that is justified below:

$$\sum_{j=1}^{d/2}(f^\pi(j) - f^\pi(j+d/2)) \geq \sum_{j=1}^{\lfloor d/4 \rfloor}(f^\pi(j) - f^\pi(j+d/2)) \tag{5.49}$$

$$\geq \Big(\Big\lfloor \frac{d}{4} \Big\rfloor\Big)(f^\pi(\lfloor d/4 \rfloor) - f^\pi(d/2)) \tag{5.50}$$

$$= \Big(\Big\lfloor \frac{d}{4} \Big\rfloor\Big)\Big(\frac{1}{\log_2(\lfloor d/4 \rfloor + 1)} - \frac{1}{\log_2(d/2+1)}\Big)$$

$$= \Big(\Big\lfloor \frac{d}{4} \Big\rfloor\Big)\Big(\frac{\log_2(d/2+1) - \log_2(\lfloor d/4 \rfloor + 1)}{\log_2(d/2+1)\log_2(\lfloor d/4 \rfloor + 1)}\Big)$$

$$\geq \Big(\Big\lfloor \frac{d}{4} \Big\rfloor\Big)\Big(\frac{\log_2(3) - 1}{\log_2^2(d)}\Big) \tag{5.51}$$

$$\geq \Big(\frac{1}{16}\Big)\Big(\frac{d}{\log_2^2(d)}\Big). \tag{5.52}$$

The inequality in (5.49) follows from the fact that $f^\pi(j) \geq f^\pi(j+d/2)$ for every $j \in [d/2]$ since f^π is a decreasing function on the domain $\mathbb{R}_{>0}$. Similarly, we obtain (5.50) using the observation that $f^\pi(j) - f^\pi(j+d/2) \geq f^\pi(\lfloor d/4 \rfloor) - f^\pi(d/2)$ for every $j \in [\lfloor d/4 \rfloor]$ since f^π is a decreasing function on the domain $\mathbb{R}_{>0}$. To see how (5.51) is derived, observe that

$$\log_2(d/2+1) - \log_2(\lfloor d/4 \rfloor + 1) \geq \log_2(d/2+1) - \log_2(d/4+1) \geq \log_2(3) - 1 \tag{5.53}$$

where the final inequality in (5.53) follows since $\log_2(d/2+1) - \log_2(d/4+1)$ is increasing in d and $d \geq 4$ by assumption. Moreover, $\log_2(d/2+1)\log_2(\lfloor d/4 \rfloor + 1) \leq \log_2^2(d)$ for $d \geq 4$. The final inequality in (5.52) holds since $\log_2(3) - 1 \geq 1/2$ and $\lfloor d/4 \rfloor = d/4 - (d \bmod 4)/4 \geq d/8$ for $d \geq 4$.

Combining the bound for $d = 2$ from (5.48) and the bound for $d \geq 4$ and even from (5.52), we conclude

$$\sum_{j=1}^{d/2}(f^\pi(j) - f^\pi(j+d/2)) \geq \Big(\frac{1}{16}\Big)\Big(\frac{d}{\log_2^2(d)}\Big),$$

whenever $d \geq 2$ and even.

5.A.3.5.3 Proof of Lemma 5.3. In this proof, we show a simple adaptation of the problem construction from Section 5.A.3.2 results in a suboptimality bound on `SIM` for the final reviewer when the number of papers d is odd that matches (up to constants) the bound given in (5.31) that holds whenever the paper count d is even.

Fix d odd and let $d' = d-1$ (and hence d' is an even number). We consider an identical problem construction for the papers $j \in [d']$ as from Section 5.A.3.2 and then include a paper d that has a similarity score of zero and one bid. This change is such that for the given class of functions in the model, `SUPER`* and `SIM` show the paper d after the papers in the set $[d']$ and the expected gain from paper d is deterministically zero since the similarity score is zero.

In particular, let the similarity scores for the final reviewer be $S_{n,j} = 1 - 1/(j\epsilon)$ for each paper $j \in [d']$, where $\epsilon = (1+\lambda)e^{e^e}$ and $\lambda \geq 0$ is the fixed and given trade-off parameter. Moreover, let

$S_{n,d} = 0$. Set the number of bids on the papers from previous reviewers to be

$$g_{n-1,j} = \begin{cases} 0, & \text{if } j \in \{1, \ldots, d'/2\} \\ 1, & \text{if } j \in \{d'/2+1, \ldots, d\}. \end{cases}$$

In Section 5.A.3.2, we showed if a pair of papers j, j' are such that $g_{n-1,j} = g_{n-1,j'}$ and $S_{n,j} > S_{n,j'}$, then $\pi_n^{\texttt{SUPER}^*}(j) < \pi_n^{\texttt{SUPER}^*}(j')$. Since paper $j' = d$ is such that for every paper $j \in \{d'/2 + 1, \ldots, d'\}$, $g_{n-1,j} = g_{n-1,j'}$ and $S_{n,j} > S_{n,j'}$, we find $\pi_n^{\texttt{SUPER}^*}(j') < \pi_n^{\texttt{SUPER}^*}(j')$. From (5.26), this means for this construction $\pi_n^{\texttt{SUPER}^*}(d) = d$ and that $\texttt{SUPER}^*$ shows the paper ordering

$$\pi_n^{\texttt{SUPER}^*}(j) = \begin{cases} d'/2 - j + 1, & \text{if } j \in \{1, \ldots, d'/2\} \\ d' + d'/2 + 1 - j, & \text{if } j \in \{d'/2+1, \ldots, d'\} \\ j, & \text{if } j \in \{d\}. \end{cases}$$

The SIM algorithm shows papers in a decreasing order of the similarity scores so that

$$\pi_n^{\texttt{SIM}}(j) = \begin{cases} d' - j + 1, & \text{if } j \in \{1, \ldots, d'\} \\ j, & \text{if } j \in \{d\}. \end{cases}$$

Observe that this construction is identical to the problem construction from Section 5.A.3.2 for papers in the set $[d']$. Moreover, from (5.20) it is clear that the expected gain from papers with zero similarity score is zero independent of the paper ordering. This allows us to conclude that the bound given in (5.31) for d even applies to this construction as a function of d'. In other words, given d odd, we obtain

$$\mathbb{E}[\mathcal{G}_n^{\texttt{SUPER}^*} - \mathcal{G}_n^{\texttt{SIM}} | \mathcal{H}_{n-1}] \geq \Big(\frac{1}{32}\Big)\Big(\frac{d'}{\log_2^2(d')}\Big).$$

Moreover, since $d' = d - 1$, whenever d is odd,

$$\mathbb{E}[\mathcal{G}_n^{\texttt{SUPER}^*} - \mathcal{G}_n^{\texttt{SIM}} | \mathcal{H}_{n-1}] \geq \Big(\frac{1}{32}\Big)\Big(\frac{d-1}{\log_2^2(d-1)}\Big) \geq \Big(\frac{1}{64}\Big)\Big(\frac{d}{\log_2^2(d)}\Big),$$

where the final inequality follows from the facts that $\log_2^2(d-1) \leq \log_2^2(d)$ and $d - 1 \geq d/2$ for $d \geq 2$.

5.A.3.5.4 Proof of Lemma 5.4. In this proof, we show a simple adaptation of the problem construction from Section 5.A.3.3 leads to a suboptimality bound on BID for the final reviewer when the number of papers is odd that matches (up to constants) the bound given in (5.36) that holds whenever the number of papers d is even. This approach is analogous to the method to obtain a suboptimality bound for SIM with an odd number of papers using the bound that held for an even number of papers from Lemma 5.3.

Fix d odd and let $d' = d - 1$ (and hence d' is an even number). We consider an identical problem construction for the papers $j \in [d']$ as from Section 5.A.3.3 and then include a paper d that has a similarity score of zero and one bid. This change is such that for the given class of functions in the model, $\texttt{SUPER}^*$ and BID show the paper d after the papers in the set $[d']$ and the expected gain from paper d is deterministically zero since the similarity score is zero.

Let the similarity scores for the final reviewer be

$$S_{n,j} = \begin{cases} 1, & \text{if } j \in \{1, \ldots, d'/2\} \\ 0, & \text{if } j \in \{d'/2+1, \ldots, d\} \end{cases}$$

and the number of bids on the papers from previous reviewers be

$$g_{n-1,j} = \begin{cases} 1, & \text{if } j \in \{1, \ldots, d'/2\} \\ 0, & \text{if } j \in \{d'/2+1, \ldots, d'\} \\ 1, & \text{if } j = d. \end{cases}$$

We now derive the optimal paper ordering for the final reviewer. Recall from (5.22) that the weights of the optimization problem for the final reviewer given in (5.21) are defined by

$$\alpha_{n,j} = S_{n,j}(\gamma_p(g_{n-1,j}+1) - \gamma_p(g_{n-1,j})) + \lambda(2^{S_{n,j}} - 1) \; \forall \; j \in [d].$$

Moreover, from the structure of the optimization problem in (5.21), if $\alpha_{n,j} > \alpha_{n,j'}$, then $\pi_n^{\texttt{SUPER}^*}(j) < \pi_n^{\texttt{SUPER}^*}(j')$ so that paper j is shown ahead of paper j' in the ranking. Observe that for each paper $j \in \{1, \ldots, d'/2\}$, $\alpha_{n,j}$ is a fixed number. Similarly, for each $j' \in \{d'/2+1, \ldots, d'\} \cup \{d\}$, $\alpha_{n,j'}$ is a fixed number. In Section 5.A.3.3, we showed if a pair of papers j, j' are such that $g_{n-1,j} = 1$ and $g_{n-1,j'} = 0$ along with $S_{n,j} = 1$ and $S_{n,j'} = 0$, then $\alpha_{n,j} > \alpha_{n,j'}$ so that $\pi_n^{\texttt{SUPER}^*}(j) < \pi_n^{\texttt{SUPER}^*}(j')$. This immediately guarantees $\pi_n^{\texttt{SUPER}^*}(j) < \pi_n^{\texttt{SUPER}^*}(j')$ for each pair of papers $j \in \{1, \ldots, d'/2\}$, $j' \in \{d'/2+1, \ldots, d'\} \cup \{d\}$. We conclude $\pi_n^{\texttt{SUPER}^*}(j) = j$ for each $j \in [d]$, where without loss of generality, to simplify the analysis, we assume if a pair of papers have equal weights in the optimization problem, then ties are broken in order of the paper indexes since the tie-breaking mechanism will not change the expected gain the paper ordering obtains from the final reviewer.

The `BID` baseline will show papers in an increasing order of the number of bids so that

$$\pi_n^{\texttt{BID}}(j) = \begin{cases} j + d'/2, & \text{if } j \in \{1, \ldots, d'/2\} \\ j - d'/2, & \text{if } j \in \{d'/2+1, \ldots, d'\} \\ j & \text{if } j = d. \end{cases}$$

This paper ordering is derived from recalling that `BID` breaks ties by the similarity scores and further ties are broken uniformly at random. However, without loss of generality, to simplify the analysis, we assume if a pair of papers have equal similarity scores and bid counts, then ties are broken in order of the paper indexes since the tie-breaking mechanism among this set of papers will not impact the expected gain.

The construction in this proof and the paper orderings selected by `SUPER`* and `BID` are identical to the problem construction and paper orderings selected by `SUPER`* and `BID` from Section 5.A.3.3 for papers in the set $[d']$. From (5.20) it is clear that the expected gain from papers with zero similarity score is zero independent of the paper ordering. This allows us to conclude that the bound given in (5.36) for d even applies to this construction as a function of d'. In other words,

given d odd, we obtain

$$\mathbb{E}[\mathcal{G}_n^{\texttt{SUPER}^*} - \mathcal{G}_n^{\texttt{BID}} | \mathcal{H}_{n-1}] \geq \Big(\frac{1}{48} + \frac{\lambda}{16}\Big)\Big(\frac{d'}{\log_2^2(d')}\Big).$$

Since $d' = d - 1$, whenever d is odd,

$$\mathbb{E}[\mathcal{G}_n^{\texttt{SUPER}^*} - \mathcal{G}_n^{\texttt{BID}} | \mathcal{H}_{n-1}] \geq \Big(\frac{1}{48} + \frac{\lambda}{16}\Big)\Big(\frac{d-1}{\log_2^2(d-1)}\Big) \geq \Big(\frac{1}{96} + \frac{\lambda}{32}\Big)\Big(\frac{d}{\log_2^2(d)}\Big),$$

where the final inequality follows from the facts that $\log_2^2(d-1) \leq \log_2^2(d)$ and $d - 1 \geq d/2$ for $d \geq 2$.

5.A.3.5.5 Proof of Lemma 5.5. Recall that from the lemma statement, the number of papers d is assumed to be divisible by four. Let $\mathcal{E}$ denote the event that a permutation π of the paper set $[d]$ drawn uniformly at random from Π_d has fewer than $d/4$ of the papers from the set $[d/2]$ in the position set $[d/2]$.

The probability of the event $\mathcal{E}$ can be explained in the following manner. The number of outcomes presenting j papers from the paper set $[d/2]$ in the position set $[d/2]$ consists of $\binom{d/2}{j}$ combinations of potential papers that can be selected from the paper set $[d/2]$ and $\binom{d/2}{j}$ combinations of potential positions in the position set $[d/2]$. Moreover, there are $j!$ permutations of the selected papers in the chosen positions. Given that there are j papers selected from the paper set $[d/2]$ placed in the position set $[d/2]$, there are $\binom{d/2}{d/2-j}$ combinations of papers from the paper set $\{d/2+1, \ldots, d\}$ that can be placed in the remaining spots in the position set $[d/2]$. This set of papers can be permuted $(d/2-j)!$ ways in the given set of positions, and the remaining $d/2$ papers can be permuted $(d/2)!$ ways in the position set $\{d/2+1, \ldots, d\}$. To obtain the final probability of the event $\mathcal{E}$, we sum the number of outcomes for each $j < d/4$ and then normalize by the total number of outcomes $d!$. Accordingly,

$$\begin{aligned}
\mathbb{P}(\mathcal{E}) &= \frac{1}{d!} \sum_{j=0}^{d/4-1} \binom{d/2}{j}\binom{d/2}{j}(j!)\binom{d/2}{d/2-j}(d/2-j)!(d/2)! \\
&= \frac{1}{d!} \sum_{j=0}^{d/4-1} \binom{d/2}{j}\binom{d/2}{j}(j!)\Big(\frac{(d/2)!}{(d/2-j)!j!}\Big)(d/2-j)!(d/2)! \\
&= \frac{(d/2)!(d/2)!}{d!} \sum_{j=0}^{d/4-1} \binom{d/2}{j}^2. \qquad (5.54)
\end{aligned}$$

We now recall some facts about the binomial coefficients. The symmetry property of the binomial coefficients implies $\binom{n}{k} = \binom{n}{n-k}$ for $0 \leq k \leq n$ and Vandermonde's identity says that $\binom{m+n}{r} = \sum_{k=0}^{r} \binom{m}{k}\binom{n}{r-k}$ and as a corollary $\binom{m}{n} = \sum_{k=0}^{m} \binom{2m}{n}$. Using this set of facts, we work toward a lower bound on $\mathbb{P}(\mathcal{E})$ by obtaining a simplified form of the sum $\sum_{j=0}^{d/4-1} \binom{d/2}{j}^2$. Observe

that

$$\sum_{j=0}^{d/4-1}\binom{d/2}{j}^2 = \sum_{j=0}^{d/4}\binom{d/2}{j}^2 - \binom{d/2}{d/4}^2$$
$$= \frac{1}{2}\sum_{j=0}^{d/4}\binom{d/2}{j}^2 + \frac{1}{2}\sum_{j=0}^{d/4}\binom{d/2}{j}^2 - \binom{d/2}{d/4}^2.$$

From the symmetry property, $\binom{d/2}{j}^2 = \binom{d/2}{d/2-j}^2$, so we get

$$\sum_{j=0}^{d/4-1}\binom{d/2}{j}^2 = \frac{1}{2}\sum_{j=0}^{d/4}\binom{d/2}{j}^2 + \frac{1}{2}\sum_{j=0}^{d/4}\binom{d/2}{d/2-j}^2 - \binom{d/2}{d/4}^2.$$

Manipulating the indexing of the sum $\sum_{j=0}^{d/4}\binom{d/2}{d/2-j}^2$, we obtain

$$\sum_{j=0}^{d/4-1}\binom{d/2}{j}^2 = \frac{1}{2}\sum_{j=0}^{d/4}\binom{d/2}{j}^2 + \frac{1}{2}\sum_{j=d/4}^{d/2}\binom{d/2}{j}^2 - \binom{d/2}{d/4}^2.$$

Now, moving the term $\frac{1}{2}\binom{d/2}{d/4}^2$ out of the sum $\frac{1}{2}\sum_{j=d/4}^{d/2}\binom{d/2}{j}^2$ results in

$$\sum_{j=0}^{d/4-1}\binom{d/2}{j}^2 = \frac{1}{2}\sum_{j=0}^{d/4}\binom{d/2}{j}^2 + \frac{1}{2}\sum_{j=d/4+1}^{d/2}\binom{d/2}{j}^2 - \frac{1}{2}\binom{d/2}{d/4}^2.$$

Furthermore,

$$\sum_{j=0}^{d/4-1}\binom{d/2}{j}^2 = \frac{1}{2}\sum_{j=0}^{d/2}\binom{d/2}{j}^2 - \frac{1}{2}\binom{d/2}{d/4}^2.$$

Finally, applying Vandermonde's identity as given above, we get

$$\sum_{j=0}^{d/4-1}\binom{d/2}{j}^2 = \frac{1}{2}\binom{d}{d/2}^2 - \frac{1}{2}\binom{d/2}{d/4}^2. \tag{5.55}$$

Combing (5.54) with (5.55) and then simplifying, we get

$$\mathbb{P}(\mathcal{E}) = \Big(\frac{(d/2)!(d/2)!}{2d!}\Big)\Big(\binom{d}{d/2} - \binom{d/2}{d/4}^2\Big)$$
$$= \Big(\frac{(d/2)!(d/2)!}{2d!}\Big)\Big(\frac{d!}{(d/2)!(d/2)!}\Big) - \Big(\frac{(d/2)!(d/2)!}{2d!}\Big)\Big(\frac{(d/2)!}{(d/4)!(d/4)!}\Big)^2$$
$$= \frac{1}{2} - \Big(\frac{(d/2)!(d/2)!}{2d!}\Big)\Big(\frac{(d/2)!}{(d/4)!(d/4)!}\Big)^2.$$

The quantity $\Big(\frac{(d/2)!(d/2)!}{2d!}\Big)\Big(\frac{(d/2)!}{(d/4)!(d/4)!}\Big)^2$ is decreasing in d. Consequently, for every $d \geq 4$,

$$\Big(\frac{(d/2)!(d/2)!}{2d!}\Big)\Big(\frac{(d/2)!}{(d/4)!(d/4)!}\Big)^2 \leq \frac{1}{3}.$$

This allows us to conclude

$$\mathbb{P}(\mathcal{E}) \geq 1/2 - 1/3 = 1/6.$$

5.A.3.5.6 Proof of Lemma 5.6. This proof follows in a similar manner to the proof of Lemma 5.2. Recall that $f^\pi(x) = 1/\log_2(x+1)$. Fixing $d \geq 4$ and divisible by four, we obtain the following bound justified below:

$$\begin{aligned}
\sum_{j=d/4}^{d/2} (f^\pi(j) - f^\pi(j+d/4+1)) &\geq \sum_{j=d/4}^{\lfloor 3d/8 \rfloor} (f^\pi(j) - f^\pi(j+d/4+1)) && (5.56)\\
&\geq \Big(\Big\lfloor \frac{3d}{8} \Big\rfloor + 1 - \frac{d}{4}\Big)(f^\pi(\lfloor 3d/8 \rfloor) - f^\pi(d/2)) && (5.57)\\
&= \Big(\Big\lfloor \frac{3d}{8} \Big\rfloor + 1 - \frac{d}{4}\Big)\Big(\frac{1}{\log_2(\lfloor 3d/8 \rfloor + 1)} - \frac{1}{\log_2(d/2+1)}\Big)\\
&= \Big(\Big\lfloor \frac{3d}{8} \Big\rfloor + 1 - \frac{d}{4}\Big)\Big(\frac{\log_2(d/2+1) - \log_2(\lfloor 3d/8 \rfloor + 1)}{\log_2(d/2+1)\log_2(\lfloor 3d/8 \rfloor + 1)}\Big)\\
&\geq \Big(\Big\lfloor \frac{3d}{8} \Big\rfloor + 1 - \frac{d}{4}\Big)\Big(\frac{\log_2(3) - \log_2(12/8+1)}{\log_2^2(d)}\Big) && (5.58)\\
&= \Big(\frac{1}{32}\Big)\Big(\frac{d}{\log_2^2(d)}\Big). && (5.59)
\end{aligned}$$

The inequality in (5.56) follows from the fact that $f^\pi(j) \geq f^\pi(j + d/4 + 1)$ for every $j \in \{d/4, \ldots, \lfloor 3d/8 \rfloor\}$ since f^π is a decreasing function on the domain $\mathbb{R}_{>0}$. Similarly, we obtain (5.57) using the observation that $f^\pi(j) - f^\pi(j+d/4+1) \geq f^\pi(\lfloor 3d/8 \rfloor) - f^\pi(d/2)$ for every $j \in \{d/4, \ldots, \lfloor 3d/8 \rfloor\}$ since f^π is a decreasing function on the domain $\mathbb{R}_{>0}$. To see how (5.58) is derived, observe that

$$\log_2(d/2+1) - \log_2(\lfloor 3d/8 \rfloor + 1) \geq \log_2(d/2+1) - \log_2(3d/8+1) \geq \log_2(3) - \log_2(12/8+1), \quad (5.60)$$

where the final inequality in (5.60) follows since $\log_2(d/2+1) - \log_2(3d/8+1)$ is increasing in d and $d \geq 4$ by assumption. Moreover, $\log_2(d/2+1)\log_2(\lfloor 3d/8 \rfloor + 1) \leq \log_2^2(d)$ for $d \geq 4$. The final inequality in (5.59) holds since $\log_2(3) - \log_2(12/8+1) \geq 1/4$ and

$$\lfloor 3d/8 \rfloor + 1 - d/4 \geq 3d/8 - d/4 \geq d/8.$$

5.A.3.5.7 Proof of Lemma 5.7. Let us begin with $d \in \{2, 3\}$. It is immediate that `RAND` selects the paper ordering of `BID` with probability at least $1/6$ since $|\Pi_d| \leq 6$. Moreover, any paper ordering selected by `RAND` cannot obtain higher expected gain from the final reviewer than `SUPER`* since it is optimal for the final reviewer. Accordingly, combined with the bound on `BID` from (5.37)

which holds for any $d \geq 2$, we obtain for $d \in \{2, 3\}$,

$$\mathbb{E}[\mathcal{G}_n^{\texttt{SUPER}^*} - \mathcal{G}_n^{\texttt{RAND}} | \mathcal{H}_{n-1}] \geq \Big(\frac{1}{488} + \frac{\lambda}{192}\Big)\Big(\frac{d}{\log_2^2(d)}\Big). \tag{5.61}$$

We now focus on $d > 3$ and not divisible by four. The problem construction we consider is similar to that from Lemma 5.4 where the number of papers was odd and it was derived from including a paper with zero similarity score and a bid as an extra paper to the problem construction from Section 5.A.3.3. We follow the same approach, but let d' be the maximum number divisible by four such that $d' < d$ and consider an identical problem construction for the papers in the set $[d']$, but then include $d - d'$ extra papers with zero similarity and a bid. This change is such that the papers in the set $\{d' + 1, \ldots, d\}$ are shown after the papers in the set $[d']$ by $\texttt{SUPER}^*$ and the expected gain from them is deterministically zero since the similarity scores are zero.

In particular, let the similarity scores for the final reviewer be

$$S_{n,j} = \begin{cases} 1, & \text{if } j \in \{1, \ldots, d'/2\} \\ 0, & \text{if } j \in \{d'/2 + 1, \ldots, d\} \end{cases}$$

and the number of bids on the papers from previous reviewers be

$$g_{n-1,j} = \begin{cases} 1, & \text{if } j \in \{1, \ldots, d'/2\} \\ 0, & \text{if } j \in \{d'/2 + 1, \ldots, d'\} \\ 1, & \text{if } j \in \{d' + 1, \ldots, d\}. \end{cases}$$

Following the exact reasoning from the proof of Lemma 5.4, we conclude $\pi_n^{\texttt{SUPER}^*}(j) = j$ for each $j \in [d]$.

For this construction and $\texttt{RAND}$, we need to lower bound (5.23). We simplify the expression by substituting the similarity scores and the number of bids on each paper, and the paper ordering presented by $\texttt{SUPER}^*$ for this construction to obtain

$$\mathbb{E}[\mathcal{G}_n^{\texttt{SUPER}^*} - \mathcal{G}_n^{\texttt{RAND}} | \mathcal{H}_{n-1}] = \mathbb{E}_{\pi_n^{\texttt{RAND}}}\big[(\gamma_p(2) - \gamma_p(1) + \lambda) \sum_{j=1}^{d'/2} (f^{\pi}(j) - f^{\pi}(\pi_n^{\texttt{RAND}}(j)))\big].$$

where the terms for papers in the set $\{d'/2 + 1, \ldots, d\}$ dropped out since the similarity scores are zero. From this point, it is clear that the analysis beginning from (5.38) in Section 5.A.3.4 can be repeated as a function of d' to obtain the bound

$$\mathbb{E}[\mathcal{G}_n^{\texttt{SUPER}^*} - \mathcal{G}_n^{\texttt{RAND}} | \mathcal{H}_{n-1}] \geq \Big(\frac{1}{576} + \frac{\lambda}{192}\Big)\Big(\frac{d'}{\log_2^2(d')}\Big).$$

Since $d' \geq d - 3$, we obtain for $d > 3$ and not divisible by four,

$$\mathbb{E}[\mathcal{G}_n^{\texttt{SUPER}^*} - \mathcal{G}_n^{\texttt{RAND}} | \mathcal{H}_{n-1}] \geq \Big(\frac{1}{576} + \frac{\lambda}{192}\Big)\Big(\frac{d-3}{\log_2^2(d-3)}\Big) \geq \Big(\frac{1}{1728} + \frac{\lambda}{576}\Big)\Big(\frac{d}{\log_2^2(d)}\Big) \tag{5.62}$$

where the final inequality follows from the facts that $\log_2^2(d-3) \leq \log_2^2(d)$ and $d - 3 \geq d/3$ for $d \geq 5$.

Combining the bound from (5.61) for $d \in \{2, 3\}$ with the bound from (5.62) for $d > 3$ and not divisible by four, we conclude for every $d \geq 2$ and not divisible by four,

$$\mathbb{E}[\mathcal{G}_n^{\texttt{SUPER}^*} - \mathcal{G}_n^{\texttt{RAND}} | \mathcal{H}_{n-1}] \geq \Big(\frac{1}{1728} + \frac{\lambda}{576}\Big)\Big(\frac{d}{\log_2^2(d)}\Big).$$

5.A.4 Proof of Theorem 5.3: Noiseless Community Model Result

In this proof, we show for the noiseless community model defined in Section 5.4.2 that SUPER* with zero heuristic and SIM are optimal. Moreover, we prove that BID and RAND are significantly suboptimal. The organization of this proof is as follows. In Section 5.A.4.1, we present additional notation that is needed in the proof. Section 5.A.4.2 presents simplifying preliminary analysis that is needed throughout the proof to analyze the expected paper-side and reviewer-side gains of the algorithms. In Section 5.A.4.3, we characterize the optimal policy for the noiseless community model. We show in Sections 5.A.4.4 and 5.A.4.5 that SUPER* with zero heuristic and SIM are equivalent to the optimal policy, respectively. We prove the suboptimality bounds for BID and RAND in Sections 5.A.4.6 and 5.A.4.7, respectively. Combining the results from the sections of this proof gives the stated result of Theorem 5.3. We relegate the proofs of technical lemmas needed for this result to Section 5.A.4.8.

5.A.4.1 Notation

Theorem 5.3 holds for any similarity matrix S belonging to the noiseless community model defined in Section 5.4.2 and formally in (5.7). From this point on in the proof, any reference to a similarity matrix S is such that it belongs to the noiseless community model. Recall that the number of reviewers is given by $n = mq$ and the number of papers is given by $d = mq$ where $m \geq 2$ and $q \geq 2$.

We now state some additional notation for the proof and recall the class of gain and bidding functions assumed in this claim. Let us define for each reviewer $i \in [n]$ the set

$$\mathcal{D}_i = \{j \in [d] : S_{i,j} = s\}, \tag{5.63}$$

which comprises the papers on the block diagonal of the noiseless community model similarity matrix for the reviewer up to a permutation of rows and columns. Similarly, define for each paper $j \in [d]$ the set

$$\mathcal{D}_j = \{i \in [n] : S_{i,j} = s\}, \tag{5.64}$$

which comprises the reviewers on the block diagonal of the noiseless community model similarity matrix for the paper up to a permutation of rows and columns. Observe that $|\mathcal{D}_i| = q$ for each reviewer $i \in [n]$ and $|\mathcal{D}_j| = q$ for each paper $j \in [d]$. In the remainder of the proof, we simply refer to the set $\mathcal{D}_i$ as the papers on the block diagonal for a reviewer $i \in [n]$ and the set $\mathcal{D}_j$ as the reviewers on the block diagonal for a paper $j \in [d]$, and omit the wording of up to a permutation of rows and columns for brevity. Similarly, if a reviewer-paper pair (i, j) is such that $S_{i,j} = s$ so that $i \in \mathcal{D}_i$ and $j \in \mathcal{D}_i$, we say the reviewer-paper pair is on the block diagonal and omit that this is up to a permutation of rows and columns. Moreover, we denote by $\mathcal{D}_i^c$ the complement of the set

$\mathcal{D}_i$ for any reviewer $i \in [n]$, which contains each paper $j \in [d]$ not in the set $\mathcal{D}_i$ and corresponds to the papers not on the block diagonal for the reviewer. Similarly, we let $\mathcal{D}_j^c$ denote the complement of the set $\mathcal{D}_j$ for any paper $j \in [d]$, which contains each reviewer $i \in [n]$ not in the set $\mathcal{D}_j$ and corresponds to the reviewers not on the block diagonal for the paper. Finally, if a reviewer-paper pair (i, j) is such that $S_{i,j} = s$ so that $i \in \mathcal{D}_i^c$ and $j \in \mathcal{D}_i^c$, we say the reviewer-paper pair is off the block diagonal. For each complement set, we again omit the wording of up to a permutation of rows and columns.

We denote the expected gain of any algorithm $\texttt{ALG}$ presenting a potentially random sequence of paper orderings $\pi_1^{\texttt{ALG}}, \ldots, \pi_n^{\texttt{ALG}}$ as

$$\mathbb{E}[\mathcal{G}^{\texttt{ALG}}] = \mathbb{E}[\mathcal{G}_p^{\texttt{ALG}}] + \lambda \mathbb{E}[\mathcal{G}_r^{\texttt{ALG}}],$$

where the expectation is with respect to the randomness in the bids placed by reviewers and any randomness in the algorithm and $\lambda \geq 0$ is the trade-off parameter. The expected paper-side gain is given by the quantity

$$\mathbb{E}[\mathcal{G}_p^{\texttt{ALG}}] = \mathbb{E}\Big[\sum_{j \in [d]} \gamma_p(g_j)\Big], \tag{5.65}$$

where $g_j = \sum_{i \in [n]} \mathcal{B}_{i,j}$ is random the number of bids on paper $j \in [d]$ at the end of the bidding process and the paper-side gain function for this result is $\gamma_p(x) = \sqrt{x}$. Recall that $\mathcal{B}_{i,j}$ is a Bernoulli random variable denoting the random bid of reviewer $i \in [n]$ on paper $j \in [d]$. From the assumptions of Theorem 5.3, the success probability of $\mathcal{B}_{i,j}$ for any reviewer $i \in [n]$ and paper $j \in [d]$ is given by the bidding function

$$f(\pi_i^{\texttt{ALG}}(j), S_{i,j}) = \mathbb{1}\{\pi_i^{\texttt{ALG}}(j) = 1\}\mathbb{1}\{S_{i,j} > s/2\}. \tag{5.66}$$

In the remainder of the proof, if $\pi_i^{\texttt{ALG}}(j) = 1$, we often say the paper is shown in the highest or top position of the paper ordering. The expected reviewer-side gain is given by

$$\mathbb{E}[\mathcal{G}_r^{\texttt{ALG}}] = \mathbb{E}\Big[\sum_{i \in [n]} \sum_{j \in [d]} \gamma_r(\pi_i^{\texttt{ALG}}(j), S_{i,j})\Big], \tag{5.67}$$

where for this result the reviewer-side gain function is

$$\gamma_r(\pi^{\texttt{ALG}}(j), S_{i,j}) = \frac{2^{S_{i,j}} - 1}{\log_2(\pi_i^{\texttt{ALG}}(j) + 1)} = (2^{S_{i,j}} - 1)\gamma_r^{\pi}(\pi_i^{\texttt{ALG}}(j)). \tag{5.68}$$

We let

$$\gamma_r^{\pi}(\pi_i^{\texttt{ALG}}(j)) = \frac{1}{\log_2(\pi_i^{\texttt{ALG}}(j) + 1)} \tag{5.69}$$

denote the component of the reviewer-side gain function that only depends on the position the paper is in.

5.A.4.2 Preliminaries

The focus of this section is to simplify the expressions for the expected paper-side gain and expected reviewer-side gain from (5.65) and (5.67) respectively, using the similarity matrix structure and the given class of gain and bidding functions. There are several immediate characteristics of the reviewer bidding behavior and the relation between the similarity scores as a result of the given bidding function from (5.66) and the noiseless community model similarity score structure from (5.7) that we reference throughout the proof:

- If the reviewer-paper pair (i, j) is on the block diagonal of the noiseless community model similarity matrix S so that $i \in \mathcal{D}_j$ and $j \in \mathcal{D}_i$, then paper $j \in [d]$ is bid on almost surely by reviewer $i \in [n]$ when $\pi_i^{\texttt{ALG}}(j) = 1$ and almost never when $\pi_i^{\texttt{ALG}}(j) \neq 1$.

- If the reviewer-paper pair (i, j) is not on the block diagonal of the noiseless community model similarity matrix S so that $i \in \mathcal{D}_j^c$ and $j \in \mathcal{D}_i^c$, then paper $j \in [d]$ is bid on almost never by reviewer $i \in [n]$ independent of $\pi_i^{\texttt{ALG}}(j)$.

- If the reviewer-paper pair (i, j) is on the block diagonal of the noiseless community model matrix S so that $i \in \mathcal{D}_j$ and $j \in \mathcal{D}_i$, and the reviewer-paper pair (i, j') is not on the block diagonal of the noiseless community model matrix S so that $i \in \mathcal{D}_{j'}^c$ and $j' \in \mathcal{D}_i^c$, then $S_{i,j} > S_{i,j'}$.

Observe that the statements above further imply that each reviewer bids on at most one paper almost surely.

We now show that the expected paper-side gain from any paper $j \in [d]$ only depends on the positions it is shown to reviewers $i \in [n]$ by some algorithm $\texttt{ALG}$ for which the reviewer-paper pair (i, j) is on the block diagonal of the similarity matrix. Indeed, the expected paper-side gain for any paper $j \in [d]$ simplifies to be

$$\mathbb{E}[\gamma_p(g_j)] = \mathbb{E}\Big[\gamma_p\Big(\sum_{i\in[n]} \mathcal{B}_{i,j}\Big)\Big] = \mathbb{E}\Big[\gamma_p\Big(\sum_{i\in\mathcal{D}_j} \mathcal{B}_{i,j}\Big)\Big] = \sum_{\ell=0}^{q} \mathbb{P}\Big(\ell = \sum_{i\in\mathcal{D}_j} \mathbb{1}\{\pi_i^{\texttt{ALG}}(j) = 1\}\Big)\gamma_p(\ell). \quad (5.70)$$

The preceding equation follows from the fact that the the bid from any reviewer $i \in \mathcal{D}_j^c$ on paper $j \in [d]$ is zero almost surely independent of the position the paper is shown, and since any reviewer $i \in \mathcal{D}_j$ bids on the paper $j \in [d]$ almost surely if $\pi_i^{\texttt{ALG}}(j) = 1$ and almost never if $\pi_i^{\texttt{ALG}}(j) \neq 1$.

To obtain a final simplified version of the expected paper-side gain given in (5.65), we sum (5.70) over the paper set $[d]$ and get that

$$\mathbb{E}[\mathcal{G}_p^{\texttt{ALG}}] = \sum_{j\in[d]} \mathbb{E}[\gamma_p(g_j)] = \sum_{j\in[d]} \sum_{\ell=0}^{q} \mathbb{P}\Big(\ell = \sum_{i\in\mathcal{D}_j} \mathbb{1}\{\pi_i^{\texttt{ALG}}(j) = 1\}\Big)\gamma_p(\ell). \quad (5.71)$$

It is now clear from (5.71) that to analyze the expected paper-side gain of any algorithm $\texttt{ALG}$, we only need to determine the distribution on the number of times each paper is shown in the highest position to reviewers for which the reviewer-paper pair is on the block diagonal of the similarity matrix.

We now turn to deriving a simplified form of the expected reviewer-side gain given in (5.67). Beginning from (5.67), we substitute in the form of the reviewer-side gain function from (5.68) and then plug in the similarity scores of the noiseless community model matrix to obtain

$$\mathbb{E}[\mathcal{G}_r^{\texttt{ALG}}] = \mathbb{E}\Big[\sum_{i\in[n]}\sum_{j\in[d]}(2^{S_{i,j}}-1)\gamma_r^{\pi}(\pi_i^{\texttt{ALG}}(j))\Big] = \mathbb{E}\Big[(2^s-1)\sum_{i\in[n]}\sum_{j\in\mathcal{D}_i}\gamma_r^{\pi}(\pi_i^{\texttt{ALG}}(j))\Big]. \tag{5.72}$$

To be clear, the final equality above follows from the facts that $\mathcal{D}_i \cup \mathcal{D}_i^c = [d]$ for each reviewer $i \in [n]$ and $S_{i,j} = s$ for $j \in \mathcal{D}_i$ and $S_{i,j'} = 0$ for $j' \in \mathcal{D}_i^c$. It is now evident that the expected reviewer-side gain only depends on the positions the papers on the block diagonal for each individual reviewer are presented.

5.A.4.3 Optimal Policy

In this section, we characterize the optimal policy for the noiseless community model and the given class of gain and bidding functions. To do so, we independently explain how the expected paper-side and reviewer-side gain are maximized. Then, we show that they can be simultaneously maximized to obtain the optimal policy.

Policy to maximize the expected paper-side gain. The expected paper-side gain is maximized by any policy that shows a paper among the set with the minimum number of bids within $\mathcal{D}_i$ in the highest position to each reviewer $i \in [n]$. We now characterize the maximum expected paper-side gain that can be obtained and then show that the aforementioned policy achieves it.

From the characteristics of the reviewer bidding behavior given in Section 5.A.4.2, each reviewer bids on at most one paper almost surely. This means that the maximum number of bids that can be obtained by any policy is equal to the number of reviewers $n = mq$ almost surely. The expected paper-side gain from (5.65) for the given paper-side gain function is the sum of the expected value of a strictly concave function of the number of bids on a paper over each the $d = mq$ papers. Consequently, since the maximum number of bids that be obtained by any algorithm is equal to the number of papers almost surely, the expected paper-side gain is maximized if the bids are evenly distributed among the papers so that each paper has exactly one bid almost surely. It then immediately follows that the maximum expected paper-side gain that can be obtained from any algorithm ALG is

$$\mathbb{E}[\mathcal{G}_p^{\texttt{ALG}}] = \sum_{j\in[d]}\mathbb{E}[\gamma_p(g_j)] = \sum_{j\in[d]}\gamma_p(1) = mq. \tag{5.73}$$

We now show that any policy presenting a paper among the set with a minimum number of bids within $\mathcal{D}_i$ in the highest position to each reviewer $i \in [n]$ maximizes the expected paper-side gain. For any given reviewer $i \in [n]$, the q papers in $\mathcal{D}_i$ are each in $\mathcal{D}_{i'}$ for $q-1$ other reviewers $i' \in [n]$ and also in $\mathcal{D}_{i''}^c$ for each of the remaining reviewers $i'' \in [n]$. If a paper from $\mathcal{D}_i$ is shown in the highest position to reviewer $i \in [n]$, then it is bid on almost surely. Moreover, any paper that is not shown in the highest position to the reviewer is bid on with probability zero. Together, this means that upon the arrival of each reviewer $i \in [n]$, there is a paper in $\mathcal{D}_i$ with zero bids that has not been shown in the highest position to any reviewer previously almost surely. Consequently, each paper $j \in [d]$ is shown exactly once almost surely in the highest position to some reviewer $i \in \mathcal{D}_j$. It then follows from the decomposition in (5.70) that the expected paper-side gain of this

policy ALG is

$$\mathbb{E}[\mathcal{G}_p^{\texttt{ALG}}] = \sum_{j\in[d]} \mathbb{E}[\gamma_p(g_j)] = \sum_{j\in[d]} \gamma_p(1) = mq. \tag{5.74}$$

We conclude that the policy maximizes the expected paper-side gain since it was shown in (5.73) that the maximum expected paper-side gain that can be obtained is mq.

Policy to maximize the expected reviewer-side gain. The expected reviewer-side gain as given in (5.72) is decoupled between each of the reviewers. Moreover, the expected reviewer-side gain from any reviewer $i \in [n]$ only depends on the positions that papers in the set $\mathcal{D}_i$ are shown. Since the function γ_r^π as given in (5.69) is decreasing on the domain $\mathbb{R}_{>0}$, as long as each paper in the set $\mathcal{D}_i$ is shown before the papers in the set $\mathcal{D}_i^c$, then the expected reviewer-side gain from any reviewer $i \in [n]$ is maximized. This means that if a policy shows papers this way for each reviewer $i \in [n]$, then the expected reviewer-side gain is maximized.

Overall optimal policy. The expected paper-side and reviewer-side gains can be simultaneously maximized. Indeed, if a paper among the set with the minimum number of bids from $\mathcal{D}_i$ is shown in the highest position to each reviewer $i \in [n]$, then the expected paper-side gain is maximized. Furthermore, if the remaining papers in $\mathcal{D}_i$ are shown ahead of each paper in $\mathcal{D}_i^c$ for each reviewer $i \in [n]$, then the expected reviewer-side gain is maximized. It then follows that this is the optimal policy. We refer to such a policy as OPT in the remainder of the proof.

5.A.4.4 Optimality of SUPER* with Zero Heuristic

We show in this section that SUPER* with zero heuristic is equivalent to the optimal policy under the noiseless community model for the given class of gain and bidding functions.

Informal description of SUPER* with zero heuristic policy. Recall that as explained in Section 5.3 and formally characterized in Section 5.4, SUPER* with zero heuristic is designed to maximize the immediate expected gain from each reviewer conditioned on the history. We show that the immediate expected paper-side and reviewer-side gain from any reviewer $i \in [n]$ are both maximized by showing a paper with the minimum number of bids among $\mathcal{D}_i$ in the highest position, followed by the remaining papers in $\mathcal{D}_i$ in any order, and then the papers from $\mathcal{D}_i^c$ in any order.

The immediate expected paper-side gain from any reviewer $i \in [n]$ is maximized by showing a paper with the minimum number of bids among $\mathcal{D}_i$ in the highest position in the paper ordering. To see why, observe that the immediate expected paper-side gain from any paper that is not shown in the highest position is zero since the probability of it being bid on is zero. Moreover, the probability of a paper being bid on that is shown in the highest position is only non-zero if it is in the set of papers $\mathcal{D}_i$. Then, since the given paper-side gain function is strictly concave so the returns of bids are diminishing, we determine that the immediate expected paper-side gain from the paper shown in the highest position of the ordering is maximized if it is a paper with the minimum number of bids among $\mathcal{D}_i$.

The expected reviewer-side gain from any reviewer $i \in [n]$ is maximized as long as papers in $\mathcal{D}_i$ are shown ahead of $\mathcal{D}_i^c$. This follows from the fact that the expected reviewer-side gain as given in (5.72) is decoupled between the reviewers. Furthermore, the expected reviewer-side gain from any reviewer $i \in [n]$ only depends on the positions that papers in the set $\mathcal{D}_i$ are shown. Since the function γ_r^π as given in (5.69) is decreasing on the domain $\mathbb{R}_{>0}$, as long as each paper in the set $\mathcal{D}_i$

is shown before the papers in the set $\mathcal{D}_i^c$, then the expected reviewer-side gain from the reviewer is maximized.

Formal description of SUPER* with zero heuristic policy. We now formally state the policy of SUPER* with zero heuristic policy for the noiseless community model and the given gain and bidding functions. The proof of Lemma 5.8 is given in Section 5.A.4.8.

Lemma 5.8. *Under the assumptions of Theorem 5.3, SUPER* with zero heuristic shows a paper among the set with the minimum number of bids from $\mathcal{D}_i$ in the highest position to each reviewer $i \in [n]$. Moreover, the remaining papers in $\mathcal{D}_i$ are shown in an arbitrary order ahead of the papers in $\mathcal{D}_i^c$ which are also shown in arbitrary order to each reviewer $i \in [n]$.*

The policy of SUPER* with zero heuristic given in Lemma 5.8 is equivalent to the optimal policy derived in Section 5.A.4.3. We conclude SUPER* with zero heuristic is optimal for the noiseless community model with the given class of gain and bidding functions.

5.A.4.5 Optimality of SIM

The SIM policy shows papers to each reviewer in decreasing order of the similarity scores with ties between a pair of papers broken in favor of the paper with fewer bids and any remaining ties are broken uniformly at random. The similarity score of each paper $j \in \mathcal{D}_i$ is greater than the similarity score of each paper $j' \in \mathcal{D}_i^c$ for each reviewer. By definition of the policy, the previous fact immediately implies that for each reviewer $i \in [n]$, SIM shows each paper in $\mathcal{D}_i$ ahead of each paper in $\mathcal{D}_i^c$. Moreover, the tie-breaking mechanism of SIM guarantees that a paper with the minimum number of bids among $\mathcal{D}_i$ is shown in the highest position of the paper ordering to each reviewer $i \in [n]$. This policy is equivalent to the optimal policy given in Section 5.A.4.3, and hence SIM is optimal for the noiseless community model with the given class of gain and bidding functions.

5.A.4.6 Suboptimality of BID

We now prove the suboptimality of BID for the noiseless community model.

Intuition and BID policy. The BID algorithm presents papers in an increasing order of the number of bids and ties between papers are broken in favor of the paper with the higher similarity score. In this section, we go on to show that this policy maximizes the expected paper-side gain. This follows from the fact that almost surely a paper with zero bids and a similarity score exceeding the threshold necessary for a reviewer to bid on a paper is shown in the highest position to each reviewer and bid on. However, for the noiseless community model similarity class, the algorithm is suboptimal for the combined objective since the expected reviewer-side gain obtained is suboptimal. The fundamental problem with BID is that, except for as a tie-breaking mechanism, the similarity scores are ignored by the algorithm. For the given bidding model, papers which are not shown in the highest position are bid on with probability zero. Consequently, showing papers with fewer bids closer, but not in the highest position, cannot improve the expected paper-side gain and reduces the expected reviewer-side gain.

Bounding the expected paper-side gain. Recall from Section 5.A.4.3 that any policy presenting a paper among the set with a minimum number of bids within $\mathcal{D}_i$ in the highest position

to each reviewer $i \in [n]$ maximizes the expected paper-side gain. We now follow similar arguments from Section 5.A.4.3 to determine that BID shows a paper among the set with a minimum number of bids among $\mathcal{D}_i$ in the highest position to each reviewer $i \in [n]$ almost surely so that is maximizes the expected paper-side gain.

For any given reviewer $i \in [n]$, the q papers in $\mathcal{D}_i$ are in $\mathcal{D}_{i'}$ for the same $q-1$ other reviewers $i' \in [n]$ and also in $\mathcal{D}_{i''}^c$ for each of the remaining reviewers $i'' \in [n]$. For any reviewer $i \in [n]$, any paper $j \in \mathcal{D}_i^c$ is bid on almost never and any paper $j \in \mathcal{D}_i$ is only bid on with non-zero probability if shown in the highest position to the reviewer. This means that upon the arrival of each reviewer $i \in [n]$, there is a paper in $\mathcal{D}_i$ with zero bids almost surely. Furthermore, the similarity score of any paper in $\mathcal{D}_i$ is greater than the similarity score of any paper in $\mathcal{D}_i^c$ for each reviewer $i \in [n]$. Hence, BID shows a paper in $\mathcal{D}_i$ with zero bids in the highest position to each reviewer $i \in [n]$ almost surely since papers are shown in increasing order of the number of bids and ties are broken in favor of the paper with the higher similarity score. The structure of the bidding function guarantees that if a paper in $\mathcal{D}_i$ is shown in the highest position of the paper ordering to reviewer $i \in [n]$, then it is bid on by the reviewer almost surely. Consequently, each paper $j \in [d]$ is shown exactly once almost surely in the highest position to some reviewer $i \in \mathcal{D}_j$. It then follows from the decomposition in (5.70) that the expected paper-side gain of BID for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$, is given by

$$\mathbb{E}[\mathcal{G}_p^{\texttt{BID}}] = \sum_{j \in [d]} \mathbb{E}[\gamma_p(g_j)] = \sum_{j \in [d]} \gamma_p(1) = mq.$$

From the expected paper-side gain of OPT given in (5.74), we conclude that for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$,

$$\mathbb{E}[\mathcal{G}_p^{\texttt{OPT}}] - \mathbb{E}[\mathcal{G}_p^{\texttt{BID}}] = 0. \tag{5.75}$$

Bounding the expected reviewer-side gain. We now show that the optimal policy OPT obtains significantly more expected reviewer-side gain than BID. This requires deriving a suitable lower bound on the following expression based on (5.72):

$$\lambda \mathbb{E}[\mathcal{G}_r^{\texttt{OPT}} - \mathcal{G}_r^{\texttt{BID}}] = \lambda \mathbb{E}\Big[\sum_{i \in [n]} (2^s - 1) \sum_{j \in \mathcal{D}_i} (\gamma_r^{\pi}(\pi_i^{\texttt{OPT}}(j)) - \gamma_r(\pi_i^{\texttt{BID}}(j)))\Big]. \tag{5.76}$$

Let us begin by defining a "good event" for any reviewer and paper under which if the paper has probability zero of being bid on then it is not bid on and if the paper has probability one of being bid on then it is bid on. Formally, for any reviewer $k \in [n]$, paper $j \in [d]$, and paper ordering $\pi_k^{\texttt{ALG}}$ given by an algorithm ALG, we define

$$\mathcal{E}_{k,j}^{\texttt{ALG}} = \{\pi_k^{\texttt{ALG}}(j) = 1, S_{k,j} > s/2, \mathcal{B}_{k,j} = 1\} \cup \{\pi_k^{\texttt{ALG}}(j) \neq 1, \mathcal{B}_{k,j} = 0\} \cup \{S_{k,j} < s/2, \mathcal{B}_{k,j} = 0\}.$$

Moreover, for each reviewer $i \in [n]$, define the following event $\mathcal{E}_i = \cup_{k=1}^{i-1} \cup_{j=1}^{d} \{\mathcal{E}_{k,j}^{\texttt{OPT}} \cup \mathcal{E}_{k,j}^{\texttt{BID}}\}$ which says the good event held for each reviewer that arrived previously for every paper and observe that the complement of this event occurs on a measure zero space by the structure of the bidding function given in (5.66). Consequently, from the law of total expectation, an equivalent form of (5.76) is

given by

$$\lambda\mathbb{E}[\mathcal{G}_r^{\texttt{OPT}} - \mathcal{G}_r^{\texttt{BID}}] = \lambda\mathbb{E}\Big[\sum_{i\in[n]}(2^s-1)\mathbb{E}\Big[\sum_{j\in\mathcal{D}_i}(\gamma_r^\pi(\pi_i^{\texttt{OPT}}(j)) - \gamma_r(\pi_i^{\texttt{BID}}(j)))\Big|\mathcal{E}_i\Big]\Big]. \tag{5.77}$$

Recall from the derivation of the expected paper-side gain of OPT in Section 5.A.4.3 and BID in this section that each algorithm obtains exactly one bid almost surely from each reviewer and on each paper. Define $\mathcal{F}$ as the set of initial $\lfloor mq/4 \rfloor$ reviewers for which upon arrival of such a reviewer $i \in \mathcal{F}$ at least one paper on the block diagonal for the reviewer given by $\mathcal{D}_i$ has received a bid previously. Observe that OPT obtains at least as much expected reviewer-side gain as BID from each reviewer since it was shown in Section 5.A.4.3 that the policy maximizes the expected reviewer-side gain from each individual reviewer. As a result, we get the following lower bound on (5.77):

$$\lambda\mathbb{E}[\mathcal{G}_r^{\texttt{OPT}} - \mathcal{G}_r^{\texttt{BID}}] \geq \lambda\mathbb{E}\Big[\sum_{i\in\mathcal{F}}(2^s-1)\mathbb{E}\Big[\sum_{j\in\mathcal{D}_i}(\gamma_r^\pi(\pi_i^{\texttt{OPT}}(j)) - \gamma_r(\pi_i^{\texttt{BID}}(j)))\Big|\mathcal{E}_i\Big]\Big]. \tag{5.78}$$

We now separate papers into relevant groups defined upon arrival for each reviewer $i \in \mathcal{F}$ given the event $\mathcal{E}_i$. Let $T_{i,1}$ be the set of papers in $\mathcal{D}_i$ with zero bids and $T_{i,2}$ be the set of papers in $\mathcal{D}_i$ with one bid. Denote by $T_{i,3}$ the set of papers in $\mathcal{D}_i^c$ with zero bids and $T_{i,4}$ as the papers in $\mathcal{D}_i^c$ with one bid. Moreover, we let $N_{i,k} = |T_{i,k}|$ for $k \in \{1,2,3,4\}$ denote the number of papers in each set and define $\ell_i = N_{i,1} + 1$. Using this notation, (5.78) is equivalently

$$\lambda\mathbb{E}[\mathcal{G}_r^{\texttt{OPT}} - \mathcal{G}_r^{\texttt{BID}}] \geq \lambda\mathbb{E}\Big[\sum_{i\in\mathcal{F}}(2^s-1)\mathbb{E}\Big[\sum_{j\in T_{i,1}\cup T_{i,2}}(\gamma_r^\pi(\pi_i^{\texttt{OPT}}(j)) - \gamma_r(\pi_i^{\texttt{BID}}(j)))\Big|\mathcal{E}_i\Big]\Big]. \tag{5.79}$$

As shown in Section 5.A.4.3, OPT shows a paper with the minimum number of bids among $\mathcal{D}_i$ in the highest position of the paper ordering to each reviewer $i \in \mathcal{F}$. Again, this paper corresponds to a paper in the set $T_{i,1}$ with zero bids. After this paper, the remaining papers in $\mathcal{D}_i$ are shown in any arbitrary order. This group of papers contains papers among $T_{i,1} \cup T_{i,2}$. Since it has no impact on the expected gain in the analysis that follows, without loss of generality, consider that OPT shows the papers in $T_{i,1}$ ahead of the papers in $T_{i,2}$.

The BID policy shows a paper with the minimum number of bids among $\mathcal{D}_i$ in the highest position of the paper ordering to each reviewer $i \in \mathcal{F}$ almost surely as proved earlier. See that such a paper corresponds to a paper in the set $T_{i,1}$ with zero bids. After this paper, the remaining papers with zero bids, which by definition belong to $T_{i,1} \cup T_{i,3}$, are shown with ties broken in favor of the paper with the higher similarity score. Since each paper in $\mathcal{D}_i$ has a higher similarity score than each paper in $\mathcal{D}_i^c$, we conclude that BID shows the remaining papers in $T_{i,1}$ after the paper shown in the highest position.

Consequently, the papers in $T_{i,1}$ are shown among the positions $\{1,\dots,N_{i,1}\}$ by both OPT and BID conditioned on the event $\mathcal{E}_i$. This allows us to simplify (5.79) and get that

$$\lambda\mathbb{E}[\mathcal{G}_r^{\texttt{OPT}} - \mathcal{G}_r^{\texttt{BID}}] \geq \lambda\mathbb{E}\Big[\sum_{i\in\mathcal{F}}(2^s-1)\mathbb{E}\Big[\sum_{j\in T_{i,2}}(\gamma_r^\pi(\pi_i^{\texttt{OPT}}(j)) - \gamma_r(\pi_i^{\texttt{BID}}(j)))\Big|\mathcal{E}_i\Big]\Big]. \tag{5.80}$$

As we just showed, conditioned on the event $\mathcal{E}_i$, OPT shows the papers in $T_{i,2}$ in an arbitrary order immediately after the papers in $T_{i,1}$ to each reviewer $i \in \mathcal{F}$. This means that the papers in $T_{i,2}$ are shown among the position set $\{\ell_i, \ldots, \ell_i + N_{i,2} - 1\}$ by OPT to each reviewer $i \in \mathcal{F}$ conditioned on $\mathcal{E}_i$.

In contrast, BID shows the papers in $T_{i,2}$ after the papers in $T_{i,1} \cup T_{i,3}$, but before the papers in $T_{i,4}$. Indeed, the papers in $T_{i,3}$ each have zero bids and the papers in $T_{i,2}$ each have one bid, so by definition of the policy, BID shows the papers in $T_{i,3}$ ahead of the papers in $T_{i,2}$. Furthermore, by definition of the sets, the similarity score of each paper in $T_{i,2}$ is greater than the similarity score of each paper in $T_{i,4}$, which combined with the tie-breaking mechanism of BID ensures that papers in $T_{i,2}$ are shown ahead of the papers in $T_{i,4}$ even though the number of bids are equal. This means that the papers in $T_{i,2}$ are shown among the position set $\{\ell_i + N_{i,3}, \ldots, \ell_i + N_{i,2} + N_{i,3} - 1\}$ by BID to each reviewer $i \in \mathcal{F}$ conditioned on $\mathcal{E}_i$.

From this set of facts and continuing from (5.80), we obtain

$$\lambda \mathbb{E}[\mathcal{G}_r^{\texttt{OPT}} - \mathcal{G}_r^{\texttt{BID}}] \geq \lambda \mathbb{E}\Big[(2^s - 1) \sum_{i \in \mathcal{F}} \mathbb{E}\Big[\sum_{j=\ell_i}^{\ell_i + N_{i,2} - 1} (\gamma_r^\pi(j) - \gamma_r^\pi(j + N_{i,3})) \Big| \mathcal{E}_i \Big]\Big]. \tag{5.81}$$

Minimizing over $i \in \mathcal{F}$ in (5.81) and using the definition $|\mathcal{F}| = \lfloor mq/4 \rfloor$, we get the bound

$$\lambda \mathbb{E}[\mathcal{G}_r^{\texttt{OPT}} - \mathcal{G}_r^{\texttt{BID}}] \geq \lambda \mathbb{E}\Big[(2^s - 1)(\lfloor mq/4 \rfloor) \min_{i \in \mathcal{F}} \mathbb{E}\Big[\sum_{j=\ell_i}^{\ell_i + N_{i,2} - 1} (\gamma_r^\pi(j) - \gamma_r^\pi(j + N_{i,3})) \Big| \mathcal{E}_i \Big]\Big]. \tag{5.82}$$

Moreover, for every $m \geq 2$ and $q \geq 2$, it holds that

$$\lfloor mq/4 \rfloor = mq/4 - (mq \bmod 4)/4 \geq mq/8, \tag{5.83}$$

and by definition of the noiseless community model

$$2^s - 1 \geq 2^{0.01} - 1 \geq 1/150. \tag{5.84}$$

Combining (5.82), (5.83), and (5.84), we have

$$\lambda \mathbb{E}[\mathcal{G}_r^{\texttt{OPT}} - \mathcal{G}_r^{\texttt{BID}}] \geq \Big(\frac{\lambda}{1200}\Big) \mathbb{E}\Big[\min_{i \in \mathcal{F}} \mathbb{E}\Big[\sum_{j=\ell_i}^{\ell_i + N_{i,2} - 1} (\gamma_r^\pi(j) - \gamma_r^\pi(j + N_{i,3})) \Big| \mathcal{E}_i \Big]\Big]. \tag{5.85}$$

Toward the goal of bounding the right-hand side of (5.85), we now work on verifying the following claim.

Claim 1. For each reviewer $i \in \mathcal{F}$ and conditioned on the event $\mathcal{E}_i$,

$$N_{i,3} \geq mq - q - m - \lfloor mq/4 \rfloor + N_{i,2} + 1 \geq N_{i,2} \geq 1. \tag{5.86}$$

Recall that $N_{i,3}$ denotes the number of papers in $\mathcal{D}_i^c$ with zero bids upon the arrival of reviewer $i \in \mathcal{F}$. By definition, the number of papers in $\mathcal{D}_i^c$ is $mq - q$. To bound $N_{i,3}$, we need to bound the maximum number of papers $\mathcal{D}_i^c$ that could have been bid on previously upon the arrival of the reviewer.

Observe that upon the arrival of the reviewer, there could be at most $(\lfloor mq/4 \rfloor - 1) - (N_{i,2} - 1)$ reviewers from $\mathcal{F}$ that previously arrived and bid on a paper in $\mathcal{D}_i^c$. This follows from the fact that $|\mathcal{F}| = \lfloor mq/4 \rfloor$ and each reviewer bids on at most one paper almost surely from the structure of the bidding function given in (5.66), so the total number of bids from this set of reviewers previously is at most $(\lfloor mq/4 \rfloor - 1)$. Furthermore, of the $(\lfloor mq/4 \rfloor - 1)$ bids from the reviewer set $\mathcal{F}$, the number of bids on papers which are in $\mathcal{D}_i$ instead of $\mathcal{D}_i^c$ is given by $(N_{i,2} - 1)$ since prior to the arrival of the reviewer a paper in $\mathcal{D}_i$ had to be bid on by definition of the reviewer set $\mathcal{F}$. Finally, at most $(m-1)$ papers in $\mathcal{D}_i^c$ are bid on before the arrival of the reviewer from previous reviewers which do not belong to $\mathcal{F}$ since there are m blocks in the similarity matrix. Accordingly, the number of papers with a bid in $\mathcal{D}_i^c$ is at most $(m-1) + (\lfloor mq/4 \rfloor - 1) - (N_{i,2} - 1)$. We conclude that the number of papers in $\mathcal{D}_i^c$ without a bid given by $N_{i,3}$ for any reviewer $i \in \mathcal{F}$ conditioned on $\mathcal{E}_i$ is bounded below as follows

$$N_{i,3} \geq mq - q - m - \lfloor mq/4 \rfloor + N_{i,2} + 1. \tag{5.87}$$

We now show

$$mq - q - m - \lfloor mq/4 \rfloor + N_{i,2} + 1 \geq N_{i,2}. \tag{5.88}$$

To see (5.88), observe that

$$mq - q - m - \lfloor mq/4 \rfloor + 1 \geq mq - q - m - mq/4 + 1 = 3mq/4 - q - m + 1. \tag{5.89}$$

The quantity $(3mq/4 - q - m + 1)$ is increasing in m and q for $m \geq 2$, $q \geq 2$. Using this fact, we get that for every $m \geq 2$ and $q \geq 2$,

$$3mq/4 - q - m + 1 \geq 0. \tag{5.90}$$

Combining (5.89) and (5.90) immediately implies that (5.88) holds. Finally, $N_{i,2} \geq 1$ for each reviewer $i \in \mathcal{F}$ conditioned on the event $\mathcal{E}_i$ by definition of the reviewer set, which proves the final inequality of (5.86).

Using the result from (5.86), we now prove the following claim to bound the right-hand side of (5.85).

Claim 2. Conditioned on the event $\mathcal{E}_i$, for each reviewer $i \in \mathcal{F}$ it must be that

$$\sum_{j=\ell_i}^{\ell_i + N_{i,2} - 1} (\gamma_r^\pi(j) - \gamma_r^\pi(j + N_{i,3})) \geq \Big(\frac{2}{5}\Big)\Big(\frac{1}{\log_2^2(mq)}\Big). \tag{5.91}$$

To begin, for any $i \in \mathcal{F}$ conditioned on the event $\mathcal{E}_i$ we get that

$$\sum_{j=\ell_i}^{\ell_i + N_{i,2} - 1} (\gamma_r^\pi(j) - \gamma_r^\pi(j + N_{i,3})) \geq \sum_{j=\ell_i}^{\ell_i + N_{i,2} - 1} (\gamma_r^\pi(j) - \gamma_r^\pi(j + mq - q - m - \lfloor mq/4 \rfloor + N_{i,2} + 1)). \tag{5.92}$$

The inequality in (5.92) relies upon the facts that $\ell_i \geq 1$ for each reviewer $i \in \mathcal{F}$ by definition and the function γ_r^π as given in (5.69) is decreasing on the domain $\mathbb{R}_{>0}$. As a result of each property, we can invoke the lower bound on $N_{i,3}$ from (5.86) to get the stated bound in (5.92). Moreover,

$\ell_i + N_{i,2} - 1 = |\mathcal{D}_i| = q$ by definition for any $i \in \mathcal{F}$ given the event $\mathcal{E}_i$, so an equivalent form of the bound in (5.92) is

$$\sum_{j=\ell_i}^{\ell_i+N_{i,2}-1} (\gamma_r^\pi(j) - \gamma_r^\pi(j+N_{i,3})) \geq \sum_{j=\ell_i}^{q} (\gamma_r^\pi(j) - \gamma_r^\pi(j+mq-q-m-\lfloor mq/4 \rfloor + N_{i,2}+1)). \quad (5.93)$$

Now, since $\ell_i \geq 1$ for each reviewer $i \in \mathcal{F}$ by definition, the function γ_r^π as given in (5.69) is decreasing on the domain $\mathbb{R}_{>0}$, and $mq - q - m - \lfloor mq/4 \rfloor + N_{i,2} + 1 \geq 1$ from (5.86), we determine that each summand in (5.93) is positive. Hence, to obtain a lower bound on (5.93), we take the maximum ℓ_i over each reviewer $i \in \mathcal{F}$. Recall that $\ell_i - 1 = N_{i,1}$, which gives the total number of papers in $\mathcal{D}_i$ without a bid by definition. For each reviewer $i \in \mathcal{F}$, there must be at least one paper with a bid in $\mathcal{D}_i$ by definition of the reviewer set given the event $\mathcal{E}_i$. Then, using the fact that $|\mathcal{D}_i| = q$, we get $\ell_i - 1 = N_{i,1} \leq q - 1$ so that $\ell_i \leq q$ for any reviewer $i \in \mathcal{F}$ given the event $\mathcal{E}_i$. Hence, (5.93) is lower bounded as follows:

$$\sum_{j=\ell_i}^{\ell_i+N_{i,2}-1} (\gamma_r^\pi(j) - \gamma_r^\pi(j+N_{i,3})) \geq \gamma_r^\pi(q) - \gamma_r^\pi(mq-m-\lfloor mq/4 \rfloor + N_{i,2}+1). \quad (5.94)$$

Combining the fact that $N_{i,2} \geq 1$ for each reviewer $i \in \mathcal{F}$ conditioned on the event $\mathcal{E}_i$ by definition of the reviewer set with (5.86), we obtain

$$mq - m - \lfloor mq/4 \rfloor + N_{i,2} + 1 \geq mq - m - \lfloor mq/4 \rfloor + 2 \geq q + 1. \quad (5.95)$$

Since γ_r^π as given in (5.69) is decreasing on the domain $\mathbb{R}_{>0}$, the inequality in (5.95) immediately implies

$$\gamma_r^\pi(mq - m - \lfloor mq/4 \rfloor + N_{i,2} + 1) \leq \gamma_r^\pi(mq - m - \lfloor mq/4 \rfloor + 2). \quad (5.96)$$

Then, combining (5.94) and (5.96) results in the bound

$$\sum_{j=\ell_i}^{\ell_i+N_{i,2}-1} (\gamma_r^\pi(j) - \gamma_r^\pi(j+N_{i,3})) \geq \gamma_r^\pi(q) - \gamma_r^\pi(mq-m-\lfloor mq/4 \rfloor + 2). \quad (5.97)$$

To bound $(\gamma_r^\pi(q) - \gamma_r^\pi(mq-m-\lfloor mq/4 \rfloor + 2))$, we need the following result proved in Section 5.A.4.8.2.

Lemma 5.9. *Fix $m \geq 2$, $q \geq 2$, and let $\gamma_r^\pi(x) = 1/\log_2(x+1)$. Then,*

$$\gamma_r^\pi(q) - \gamma_r^\pi(mq - m - \lfloor mq/4 \rfloor + 2) \geq \Big(\frac{2}{5}\Big)\Big(\frac{1}{\log_2^2(mq)}\Big).$$

Applying Lemma 5.9 to (5.97), we arrive at the lower bound claimed in (5.91). Then, relating (5.91) back to (5.82), for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$, the following bound holds

$$\lambda \mathbb{E}[\mathcal{G}_r^{\texttt{OPT}} - \mathcal{G}_r^{\texttt{BID}}] \geq \Big(\frac{1}{3000}\Big)\Big(\frac{\lambda mq}{\log_2^2(mq)}\Big). \quad (5.98)$$

Observe that the expectation in the right-hand side of (5.98) is dropped since it is not a random variable.

Completing the bound. Combining the bounds on the expected paper-side and reviewer-side gain between OPT and BID given in (5.77) and (5.98), we find for every $m \geq 2, q \geq 2, \lambda \geq 0$,

$$\begin{aligned}\mathbb{E}[\mathcal{G}^{\texttt{OPT}} - \mathcal{G}^{\texttt{BID}}] &= \mathbb{E}[\mathcal{G}_p^{\texttt{OPT}} - \mathcal{G}_p^{\texttt{BID}}] + \lambda \mathbb{E}[\mathcal{G}_r^{\texttt{OPT}} - \mathcal{G}_r^{\texttt{BID}}] \\ &\geq \Big(\frac{1}{3000}\Big)\Big(\frac{\lambda m q}{\log_2^2(q)}\Big).\end{aligned}$$

We conclude that there exists a constant $c > 0$ such that for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$, BID is suboptimal by an additive factor of at least $c\lambda mq/\log_2^2(mq)$ for the noiseless community model.

5.A.4.7 Suboptimality of RAND

In this section, we show the suboptimality of RAND for the noiseless community model.

Intuition and RAND policy. The RAND algorithm selects a paper ordering uniformly at random from the set of permutations of papers. For the given class of gain and bidding functions, this is problematic since to obtain a bid from a reviewer, a paper from the block diagonal for the reviewer must be shown in the highest position. Since at least half of the papers are not on the block diagonal of the similarity matrix for any reviewer, there is a significant probability that RAND fails to induce a bid from each reviewer. This causes the algorithm to be suboptimal for the expected paper-side gain.

Bounding the expected paper-side gain. Recall from (5.70) that the expected paper-side gain from any paper $j \in [d]$ is given by

$$\mathbb{E}[\gamma_p(g_j)] = \sum_{\ell=0}^{q} \mathbb{P}\Big(\ell = \sum_{i \in \mathcal{D}_j} \mathbb{1}\{\pi_i^{\texttt{RAND}}(j) = 1\}\Big)\gamma_p(\ell). \tag{5.99}$$

To bound this quantity for a given paper, we need to characterize the distribution of the number of times the paper is shown in the highest position to reviewers for which it is on the block diagonal.

The RAND algorithm selects a paper ordering uniformly at random from the set of paper permutations. This means the probability of paper any paper $j \in [d]$ being shown in the highest position to any reviewer $i \in [n]$ is $1/mq$ since there are $d = mq$ papers. Consequently, the number of times paper $j \in [d]$ is shown in the highest position to reviewers in the set $\mathcal{D}_j$ follows a binomial distribution with q trials, since the cardinality of $\mathcal{D}_j$ is q, and a success probability of $1/mq$. This means the expected paper-side gain from any paper $j \in [d]$ given in (5.99) for RAND is equivalently

$$\mathbb{E}[\gamma_p(g_j)] = \sum_{\ell=0}^{q} \binom{q}{\ell}\Big(\frac{1}{mq}\Big)^{\ell}\Big(1 - \frac{1}{mq}\Big)^{q-\ell}\gamma_p(\ell). \tag{5.100}$$

To bound (5.100), we need the following lemma that bounds the expectation of the square root of a binomial random variable with n trials and success probability p.

Lemma 5.10. *Fix $n \geq 2$ and $p \in [0,1]$. Then,*

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \sqrt{k} \leq np(1-p)^{n-1}\Big(1 - \frac{\sqrt{2}}{2}\Big) + np\Big(\frac{\sqrt{2}}{2}\Big).$$

The proof of Lemma 5.10 is provided in Section 5.A.4.8.3.

We can directly apply Lemma 5.10 to (5.100) since the given paper-side gain function is the square root function. The number of trials is $q \geq 2$ and the success probability is $1/mq$, so for any paper $j \in [d]$, we obtain

$$\mathbb{E}[\gamma_p(g_j)] = \sum_{\ell=0}^{q} \binom{q}{\ell} \Big(\frac{1}{mq}\Big)^{\ell} \Big(1 - \frac{1}{mq}\Big)^{q-\ell} \gamma_p(\ell) \leq \Big(\frac{1}{m}\Big)\Big(1 - \frac{1}{mq}\Big)^{q-1}\Big(1 - \frac{\sqrt{2}}{2}\Big) + \frac{\sqrt{2}}{2m}. \tag{5.101}$$

The bound in the right-hand side of (5.101) is decreasing in m and q for $m \geq 2$ and $q \geq 2$. This means for every $m \geq 2$, $q \geq 2$, and any paper $j \in [d]$,

$$\mathbb{E}[\gamma_p(g_j)] \leq \frac{6 + \sqrt{2}}{16}.$$

To get a final bound on the expected paper-side gain of the algorithm, we sum the previous bound over the number of papers and obtain

$$\mathbb{E}[\mathcal{G}_p^{\texttt{RAND}}] = \sum_{j \in [d]} \mathbb{E}[\gamma_p(g_j)] \leq \Big(\frac{6 + \sqrt{2}}{16}\Big) mq. \tag{5.102}$$

Combining (5.102) with the expected paper-side gain of the optimal policy which was shown to be mq in Section 5.A.4.3, this implies for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$,

$$\mathbb{E}[\mathcal{G}_p^{\texttt{OPT}} - \mathcal{G}_p^{\texttt{RAND}}] \geq mq - \Big(\frac{6 + \sqrt{2}}{16}\Big) mq. \tag{5.103}$$

Bounding the expected reviewer-side gain. We compare the expected reviewer-side gain of the optimal algorithm `OPT` and `RAND`. Previously in Section 5.A.4.3 we showed that the optimal algorithm maximizes the expected reviewer-side gain. This means that the expected reviewer-side gain of `RAND` cannot exceed that from the optimal policy `OPT`. Consequently, for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$ we get the bound

$$\lambda \mathbb{E}[\mathcal{G}_r^{\texttt{OPT}} - \mathcal{G}_r^{\texttt{RAND}}] \geq 0. \tag{5.104}$$

Completing the bound. Combining the bounds on the expected paper-side and reviewer-side gain between the optimal algorithm `OPT` and `RAND` given in (5.103) and (5.104), for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$, we get that

$$\begin{aligned}\mathbb{E}[\mathcal{G}^{\texttt{OPT}} - \mathcal{G}^{\texttt{RAND}}] &= \mathbb{E}[\mathcal{G}_p^{\texttt{OPT}} - \mathcal{G}_p^{\texttt{RAND}}] + \lambda \mathbb{E}[\mathcal{G}_r^{\texttt{OPT}} - \mathcal{G}_r^{\texttt{RAND}}] \\ &\geq mq - \Big(\frac{6 + \sqrt{2}}{16}\Big) mq \geq mq/2.\end{aligned}$$

We conclude that there exists a constant $c > 0$ such that for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$, RAND is suboptimal by an additive factor of at least cmq for the noiseless community model.

5.A.4.8 Proofs of Lemmas 5.8–5.10

In this section, we present the proofs of technical lemmas stated in the primary proof of Theorem 5.3.

5.A.4.8.1 Proof of Lemma 5.8. In the proof of Corollary 5.A.2 given in Section 5.A.2, we showed in (5.17) that SUPER* with zero heuristic solves the problem

$$\pi_i^{\texttt{SUPER}^*} = \operatorname*{arg\,max}_{\pi_i \in \Pi_d} \quad \sum_{j \in [d]} f(\pi_i(j), S_{i,j})(\gamma_p(g_{i-1,j}+1) - \gamma_p(g_{i-1,j})) + \lambda \sum_{j \in [d]} \gamma_r(\pi_i(j), S_{i,j}) \tag{5.105}$$

in order to determine the ordering of papers $\pi_i^{\texttt{SUPER}^*}$ to present to reviewer $i \in [n]$ so that the immediate expected gain is maximized conditioned on the history of bids from reviewers that arrived previously. Recalling that the bidding function is $f(\pi_i(j), S_{i,j}) = \mathbb{1}\{\pi_i(j) = 1\}\mathbb{1}\{S_{i,j} > s/2\}$, the optimization problem in (5.105) is equivalent to

$$\pi_i^{\texttt{SUPER}^*} = \operatorname*{arg\,max}_{\pi_i \in \Pi_d} \sum_{j \in [d]} \mathbb{1}\{\pi_i(j) = 1\}\mathbb{1}\{S_{i,j} > s/2\}(\gamma_p(g_{i-1,j}+1) - \gamma_p(g_{i-1,j})) + \lambda \sum_{j \in [d]} \gamma_r(\pi_i(j), S_{i,j}). \tag{5.106}$$

Observe that $\mathcal{D}_i \cup \mathcal{D}_i^c = [d]$. Moreover, if $j \in \mathcal{D}_i$, then $S_{i,j} > s/2$ since $S_{i,j} = s$ by definition of the noiseless community model similarity matrix and $s \in [0.01, 1]$. Analogously, if $j \in \mathcal{D}_i^c$, then $S_{i,j} < s/2$ since $S_{i,j} = 0$ by definition of the noiseless community model similarity matrix and $s \in [0.01, 1]$. This allows us to simplify (5.106) to the following problem:

$$\pi_i^{\texttt{SUPER}^*} = \operatorname*{arg\,max}_{\pi_i \in \Pi_d} \quad \sum_{j \in \mathcal{D}_i} \mathbb{1}\{\pi_i(j) = 1\}(\gamma_p(g_{i-1,j}+1) - \gamma_p(g_{i-1,j})) + \lambda \sum_{j \in [d]} \gamma_r(\pi_i(j), S_{i,j}). \tag{5.107}$$

The given paper-side gain function γ_p is such that $\gamma_p(g_{i-1,j}+1) - \gamma_p(g_{i-1,j})$ is decreasing as a function of the number of bids $g_{i-1,j}$. As a result, the expected paper-side gain term from (5.107), which is given by

$$\sum_{j \in \mathcal{D}_i} \mathbb{1}\{\pi_i(j) = 1\}(\gamma_p(g_{i-1,j}+1) - \gamma_p(g_{i-1,j})), \tag{5.108}$$

is maximized by showing a paper $j \in \mathcal{D}_i$ with the minimum number of bids in the highest position of the paper ordering. Moreover, the given reviewer-side gain function γ_r from (5.68) is decreasing in the position $\pi_i(j)$ in which a paper is shown and increasing in the similarity score $S_{i,j}$. Consequently, the expected reviewer-side gain term from (5.107), which is given by

$$\sum_{j \in [d]} \gamma_r(\pi_i(j), S_{i,j}), \tag{5.109}$$

is maximized by showing papers in decreasing order of the similarity scores. The similarity score is $S_{i,j} = s \in [0.01, 1]$ for papers in $\mathcal{D}_i$ and the similarity score is $S_{i,j} = 0$ for papers in $\mathcal{D}_i^c$ by definition of the noiseless community model. Accordingly, the expected reviewer-side gain term in (5.109) is maximized as long as each paper in $\mathcal{D}_i$ is shown earlier in the paper ordering than each paper in

$\mathcal{D}_i^c$.

Since the expected paper-side and reviewer-side gain terms of (5.107) given by (5.108) and (5.109) respectively can be simultaneously maximized by showing any of the papers with the minimum number of bids among $\mathcal{D}_i$ in the highest position, followed by the remaining papers in $\mathcal{D}_i$ in any arbitrary order, and then the papers in $\mathcal{D}_i^c$ in any arbitrary order, we conclude this is the policy of $\texttt{SUPER}^*$ with zero heuristic.

5.A.4.8.2 Proof of Lemma 5.9. Recall that $\gamma_r^\pi = 1/\log_2(x+1)$. Moreover, fix $m \geq 2$ and $q \geq 2$. Simplifying the expression we seek to bound, we obtain

$$\begin{aligned}\gamma_r^\pi(q) - \gamma_r^\pi(mq - m - \lfloor mq/4 \rfloor + 2) &= \frac{1}{\log_2(q+1)} - \frac{1}{\log_2(mq - m - \lfloor mq/4 \rfloor + 2 + 1)} \\ &= \frac{\log_2(mq - m - \lfloor mq/4 \rfloor + 2 + 1) - \log_2(q+1)}{\log_2(mq - m - \lfloor mq/4 \rfloor + 2 + 1)\log_2(q+1)}. \end{aligned} \tag{5.110}$$

We now lower bound the numerator of the right-hand side of (5.110). For any fixed $m \geq 2$ and $q \geq 2$,

$$\log_2(mq - m - \lfloor mq/4 \rfloor + 2 + 1) - \log_2(q+1) \geq \log_2(2q - \lfloor q/2 \rfloor + 1) - \log_2(q+1). \tag{5.111}$$

To see why, observe that $mq - m - \lfloor mq/4 \rfloor$ is non-decreasing as a function of m for $m \geq 2$ and $q \geq 2$. The non-decreasing property follows from the fact that

$$\begin{aligned}(m+1)q - (m+1) - \lfloor (m+1)q/4 \rfloor &= mq - m + q - 1 - \lfloor mq/4 + q/4 \rfloor \\ &\geq mq - m + q - 1 - (\lfloor mq/4 \rfloor + \lfloor q/4 \rfloor + 1) \\ &= mq - m - \lfloor mq/4 \rfloor + q - \lfloor q/4 \rfloor - 2 \\ &\geq mq - m - \lfloor mq/4 \rfloor. \end{aligned}$$

To get the final inequality, consider

$$q - \lfloor q/4 \rfloor - 2 = q - q/4 + (q \bmod 4)/4 - 2 = 3q/4 + (q \bmod 4)/4 - 2 \tag{5.112}$$

and notice that for $q = 2$, (5.112) is zero, and for $q > 2$, (5.112) is positive.

Now, see that $\log_2(2q - \lfloor q/2 \rfloor + 1) - \log_2(q+1)$ is increasing as a function of q since $2q - \lfloor q/2 \rfloor > q$ for $q \geq 2$. Accordingly, for every $m \geq 2$ and $q \geq 2$,

$$\log_2(2q - \lfloor q/2 \rfloor + 1) - \log_2(q+1) \geq \log_2(4) - \log_2(3) \geq 2/5. \tag{5.113}$$

To finish, we obtain a lower bound on (5.110) by finding an upper bound on the denominator in the right-hand side. Observe that $\log_2(mq - m - \lfloor mq/4 \rfloor + 2 + 1) \leq \log_2(mq)$ since $m + \lfloor mq/4 \rfloor \geq 3$ and $\log_2(q+1) \leq \log_2(mq)$ for every $m \geq 2$ and $q \geq 2$. Then, combined with (5.110), (5.111), and (5.113), we obtain the stated result of

$$\gamma_r^\pi(q) - \gamma_r^\pi(mq - m - \lfloor mq/4 \rfloor + 2) \geq \Big(\frac{2}{5}\Big)\Big(\frac{1}{\log_2^2(mq)}\Big).$$

5.A.4.8.3 Proof of Lemma 5.10. Given $n \geq 2$ and $p \in [0,1]$, we need to prove the bound

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \sqrt{k} \leq np(1-p)^{n-1}\Big(1 - \frac{\sqrt{2}}{2}\Big) + np\Big(\frac{\sqrt{2}}{2}\Big).$$

To begin, observe that

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \sqrt{k} = \sum_{k=1}^{n} \binom{n}{k} p^k (1-p)^{n-k} \sqrt{k}. \tag{5.114}$$

Then, since

$$\binom{n}{k} p^k = \Big(\frac{n!}{k!(n-k)!}\Big) p^k = \Big(\frac{np}{k}\Big)\Big(\frac{(n-1)!}{(k-1)!(n-k)!}\Big) p^{k-1} = \Big(\frac{np}{k}\Big)\binom{n-1}{k-1} p^{k-1},$$

we can simplify (5.114) to obtain

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \sqrt{k} = np \sum_{k=1}^{n} \Big(\frac{\sqrt{k}}{k}\Big)\binom{n-1}{k-1} p^{k-1}(1-p)^{n-k}. \tag{5.115}$$

From the fact that $\frac{\sqrt{k}}{k} \leq \frac{\sqrt{2}}{2}$ for $k \geq 2$, we bound (5.115) as follows:

$$\begin{aligned} np \sum_{k=1}^{n} \Big(\frac{\sqrt{k}}{k}\Big)\binom{n-1}{k-1} p^{k-1}(1-p)^{n-k} &= np(1-p)^{n-1} + np \sum_{k=2}^{n} \Big(\frac{\sqrt{k}}{k}\Big)\binom{n-1}{k-1} p^{k-1}(1-p)^{n-k} \\ &\leq np(1-p)^{n-1} + np\Big(\frac{\sqrt{2}}{2}\Big) \sum_{k=2}^{n} \binom{n-1}{k-1} p^{k-1}(1-p)^{n-k}. \end{aligned} \tag{5.116}$$

Relating (5.116) back to (5.115), we get

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \sqrt{k} \leq np(1-p)^{n-1} + np\Big(\frac{\sqrt{2}}{2}\Big) \sum_{k=2}^{n} \binom{n-1}{k-1} p^{k-1}(1-p)^{n-k}. \tag{5.117}$$

From addition and subtraction of $np(1-p)^{n-1}\big(\frac{\sqrt{2}}{2}\big)$ into the right-hand side of (5.117), we obtain

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \sqrt{k} \leq np(1-p)^{n-1}\Big(1 - \frac{\sqrt{2}}{2}\Big) + np\Big(\frac{\sqrt{2}}{2}\Big) \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1}(1-p)^{n-k}. \tag{5.118}$$

Now, see that from an indexing manipulation

$$\sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1}(1-p)^{n-k} = \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k}. \tag{5.119}$$

Moreover, from the Binomial theorem,

$$\sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} = (p + (1-p))^{n-1} = 1. \tag{5.120}$$

Combining (5.118), (5.119), and (5.120) gives the final result of

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \sqrt{k} \leq np(1-p)^{n-1}\Big(1 - \frac{\sqrt{2}}{2}\Big) + np\Big(\frac{\sqrt{2}}{2}\Big).$$

5.A.5 Proof of Theorem 5.4: Noisy Community Model Result

In this proof, we show for the noisy community model that SUPER* with zero heuristic is near optimal and each of the baselines is significantly suboptimal with respect to SUPER* with zero heuristic. The organization of this proof is as follows. In Section 5.A.5.1, we present notation and preliminary analysis that is needed throughout the proof. In Section 5.A.5.2, we analyze SUPER* with zero heuristic and compute the expected paper-side gain for the similarity matrix class. We prove the suboptimality bounds for the SIM, BID, and RAND baselines with respect to SUPER* with zero heuristic separately in Sections 5.A.5.3, 5.A.5.4, and 5.A.5.5 respectively. We finish the proof in Section 5.A.5.6 by showing that SUPER* with zero heuristic is near optimal. Combining the results in each section of this proof gives the stated result of the theorem. Proofs of technical lemmas needed only for this proof can be found in Section 5.A.5.7. The proofs of technical lemmas used in this proof, but introduced in the proof of Theorem 5.3, are given in Section 5.A.4.8. Finally, we remark that a number of methods for proving this result are similar to that from the proof of Theorem 5.3 and we point out in several places where this is the case as well as where the techniques differ.

5.A.5.1 Notation and Preliminaries

The notation and terminology in this proof follow that from the proof of Theorem 5.3 in Section 5.A.4.1 since the gain and bidding functions are shared between the results and the noisy community model is based on the noiseless community model. The primary adjustment is that any reference to a similarity matrix S refers to that from the noisy community model, which is generated by selecting some similarity matrix S' from the noiseless community model as given in in (5.7), and then adding noise in the manner described in (5.8). Recall that the noise in the similarity score for each reviewer-paper pair (i, j) denoted by $\nu_{i,j}$ is drawn independently and uniformly from $(0, \xi)$ where $\xi \leq (1+\lambda)^{-1} e^{-emq}$ for the given trade-off parameter $\lambda \geq 0$. We also follow the notation from the proof of Theorem 5.3 in Section 5.A.4.1, in terms of terminology of reviewers and papers on the block diagonal and keep the sets $\mathcal{D}_i$ for all $i \in [n]$ and $\mathcal{D}_j$ for all $j \in [d]$ from (5.63) and (5.64) defined in terms of the noiseless community model similarity matrix now given by S'.

In an analogous manner to the preliminaries section of the proof of Theorem 5.3, we present several characteristics of the reviewer bidding behavior and the similarity scores that are needed throughout the proof. This set of rather immediate results also enable a decomposition of the expected paper-side gain equivalent to that for the noiseless community model from the proof of Theorem 5.3 given in (5.70).

We begin by showing if a paper is on the block diagonal for a reviewer, then it is bid on almost surely when shown in the highest position of the paper ordering to the reviewer and almost never when it is not.

Lemma 5.11. *Under the assumptions of Theorem 5.4, if the reviewer-paper pair (i, j) is on the block diagonal of the noiseless community model matrix S' so that $i \in \mathcal{D}_j$ and $j \in \mathcal{D}_i$, then in the noisy community model matrix $S_{i,j} > s/2$. Moreover, the paper $j \in [d]$ is bid on by reviewer $i \in [n]$ almost surely when $\pi_i^{ALG}(j) = 1$ and almost never when $\pi_i^{ALG}(j) \neq 1$.*

Proof of Lemma 5.11. If the reviewer-paper pair (i, j) is on the block diagonal of the noiseless community model matrix S', then by definition $S'_{i,j} = s$ and $S_{i,j} = s - \nu_{i,j}$ where the noise $\nu_{i,j}$ is drawn uniformly at random from the interval $(0, \xi)$. Moreover, recall that $s \in [0.01, 1]$ and $\xi \leq (1 + \lambda)^{-1} e^{-emq}$. Accordingly, for $\lambda \geq 0, m \geq 2$, and $q \geq 2$, we obtain

$$\xi \leq (1 + \lambda)^{-1} e^{-emq} \leq e^{-4e} < 0.01/2 \leq s/2. \tag{5.121}$$

Since $\nu_{i,j} \in (0, \xi)$, we immediately get $S_{i,j} > s - \xi$. Then applying (5.121), we conclude that $S_{i,j} > s/2$. Finally, since the probability of reviewer i bidding on paper j is given by the quantity $f(\pi_i^{\texttt{ALG}}(j), S_{i,j}) = \mathbb{1}\{\pi_i^{\texttt{ALG}}(j) = 1\}\mathbb{1}\{S_{i,j} > s/2\}$, the reviewer bids on the paper almost surely when $\pi_i^{\texttt{ALG}}(j) = 1$ and almost never when $\pi_i^{\texttt{ALG}}(j) \neq 1$. □

We now show that if a paper is not on the block diagonal for a given reviewer, then the reviewer bids on the paper almost never independent of the position the paper is shown.

Lemma 5.12. *Under the assumptions of Theorem 5.4, if the reviewer-paper pair (i, j) is not on the block diagonal of the noiseless community model matrix S' so that $i \in \mathcal{D}_j^c$ and $j \in \mathcal{D}_i^c$, then in the noisy community model matrix $S_{i,j} < s/2$. Moreover, paper $j \in [d]$ is bid on almost never by reviewer $i \in [n]$ independent of $\pi_i^{ALG}(j)$.*

Proof of Lemma 5.12. If the reviewer-paper pair (i, j) is not on the block diagonal of the noiseless community model matrix S', then by definition $S'_{i,j} = 0$ and $S_{i,j} = \nu_{i,j}$ where the noise $\nu_{i,j}$ is drawn uniformly at random from the interval $(0, \xi)$. Since $\nu_{i,j} \in (0, \xi)$, we immediately get $S_{i,j} < \xi$. Then applying (5.121), we conclude that $S_{i,j} < s/2$. Finally, since the probability of reviewer i bidding on paper j is given by the quantity $f(\pi_i^{\texttt{ALG}}(j), S_{i,j}) = \mathbb{1}\{\pi_i^{\texttt{ALG}}(j) = 1\}\mathbb{1}\{S_{i,j} > s/2\}$, the reviewer bids on the paper almost never independent of the position the paper is shown to the reviewer given from $\pi_i^{\texttt{ALG}}(j)$. □

Observe that Lemmas 5.11 and 5.12 imply that each reviewer bids on at most one paper almost surely. Moreover, they can also be combined to determine that any paper on the block diagonal for a reviewer is guaranteed to have a higher similarity score than any paper not on the block diagonal for the reviewer.

Lemma 5.13. *Under the assumptions of Theorem 5.4, if the reviewer-paper pair (i, j) is on the block diagonal of the noiseless community model matrix S' so that $i \in \mathcal{D}_j$ and $j \in \mathcal{D}_i$, and the reviewer-paper pair (i, j') is not on the block diagonal of the noiseless community model matrix S' so that $i \in \mathcal{D}_{j'}^c$ and $j' \in \mathcal{D}_i^c$, then in the noisy community model matrix $S_{i,j} > S_{i,j'}$.*

We now apply the preceding results to show that the expected paper-side gain from a given paper $j \in [d]$ only depends on the positions it is shown to reviewers $i \in [n]$ by some algorithm ALG for which the reviewer-paper pair (i, j) is on the block diagonal of the similarity matrix. The expected paper-side gain for any paper $j \in [d]$ simplifies to be

$$\mathbb{E}[\gamma_p(g_j)] = \mathbb{E}\Big[\gamma_p\Big(\sum_{i\in[n]} \mathcal{B}_{i,j}\Big)\Big] = \mathbb{E}\Big[\gamma_p\Big(\sum_{i\in\mathcal{D}_j} \mathcal{B}_{i,j}\Big)\Big] = \sum_{\ell=0}^{q} \mathbb{P}\Big(\ell = \sum_{i\in\mathcal{D}_j} \mathbb{1}\{\pi_i^{\texttt{ALG}}(j) = 1\}\Big)\gamma_p(\ell). \quad (5.122)$$

The above equation follows from Lemma 5.12, which indicates that the the bid from any reviewer $i \in \mathcal{D}_j^c$ is zero almost surely independent of the position the paper is shown, and from Lemma 5.11, which guarantees any reviewer $i \in \mathcal{D}_j$ bids on the paper $j \in [d]$ almost surely if $\pi_i^{\texttt{ALG}}(j) = 1$ and almost never if $\pi_i^{\texttt{ALG}}(j) \neq 1$. As mentioned at the beginning of this section, this decomposition of the expected paper-side gain for the noisy community model in (5.122) is equivalent to that for the noiseless community model given in (5.70).

5.A.5.2 Analyzing SUPER* with Zero Heuristic

In this section, we present a preliminary analysis of SUPER* with zero heuristic for the noisy community model similarity matrix class with the given gain and bidding functions. We begin by characterizing the behavior of SUPER* with zero heuristic. Following deriving the policy, the expected paper-side gain of the algorithm is computed. The analysis of the expected reviewer-side gain is deferred to the sections showing the suboptimality of the baselines with respect to SUPER* with zero heuristic (specifically, see Sections 5.A.5.3 and 5.A.5.4). However, we do provide intuition in this section for why the expected reviewer-side gain is nearly optimal.

Intuition and SUPER* with zero heuristic policy. The given bidding function is such that only the paper shown in the highest position to a reviewer has a non-zero probability of being bid on. Intuitively Lemma 5.11 and Lemma 5.12 suggest that to optimize the expected paper-side gain, the algorithm should seek to show a paper on the block diagonal for the reviewer in the highest position. Moreover, since the given paper-side gain function exhibits diminishing returns in the number of bids, showing the paper on the block diagonal with fewest bids maximizes the immediate expected paper-side gain.

The given reviewer-side gain function is decreasing in the position a paper is shown and increasing in the similarity score of the paper. This indicates that to maximize the immediate expected reviewer-side gain, papers should be shown in a decreasing order of the similarity scores to the reviewer. For the given noisy community model similarity class, the similarity scores of papers on the block diagonal for a given reviewer are significantly higher than the similarity scores of papers off the block diagonal for the given reviewer as was formalized in Lemma 5.13. Furthermore, the noise is bounded in a small interval. Consequently, the similarity scores for papers on the block diagonal for a reviewer are nearly identical and the similarity scores for papers off the block diagonal for a reviewer are also nearly identical. This suggests that as long as papers on the block diagonal are shown ahead of papers off the block diagonal for a reviewer, then the expected reviewer-side gain from a given reviewer should be close to the maximum that can be obtained.

The high-level view of the objective the algorithm is optimizing indicates that the immediate expected paper-side gain can be maximized with minimal cost to the immediate expected reviewer-

side gain. This can be achieved by showing the paper on the block diagonal for the reviewer with the minimum number of bids in the highest position and the remaining papers in a decreasing order of the similarity scores. The following lemma formalizes the intuition that has been given.

Lemma 5.14. *Under the assumptions of Theorem 5.4, when reviewer $i \in [n]$ arrives, if there is a paper in $\mathcal{D}_i$ with zero bids and each paper in $\mathcal{D}_i$ has at most one bid, then* SUPER* *with zero heuristic shows the reviewer the paper with the maximum similarity score among the papers without a bid in $\mathcal{D}_i$ at the highest position followed by the remaining papers in a decreasing order of the similarity scores.*

The proof of Lemma 5.14 is provided in Section 5.A.5.7.1. It turns out that the conditions of Lemma 5.14, namely the existence of a paper in $\mathcal{D}_i$ with zero bids and each paper in $\mathcal{D}_i$ having at most one bid upon the arrival of reviewer $i \in [n]$, are met using SUPER* with zero heuristic almost surely. We now formally characterize this statement and then compute the expected paper-side gain of the algorithm.

Computing the expected paper-side gain. Consider a group of q reviewers denoted by $\mathcal{R}$ for which $\mathcal{D}_i = \mathcal{D}_{i'}$ for every pair of reviewers $i, i' \in \mathcal{R}$, meaning that the papers on the block diagonal for the reviewers are equivalent. Observe that from the structure of the noisy community model, for every reviewer $i \in \mathcal{R}$, it also holds that $\mathcal{D}_i \subset \mathcal{D}^c_{i''}$ for all $i'' \in \mathcal{R}^c$, meaning that the papers on the block diagonal for each reviewer in $\mathcal{R}$ are off the block diagonal for all reviewers in $\mathcal{R}^c$. Moreover, there are m such blocks of reviewers analogous to the given group $\mathcal{R}$.

Upon the initial arrival of a reviewer i from $\mathcal{R}$, from Lemma 5.12 each paper in $\mathcal{D}_i$ has zero bids almost surely since they are off the block diagonal for all reviewers that arrived previously. From Lemma 5.14, SUPER* with zero heuristic shows this reviewer the paper in $\mathcal{D}_i$ with the maximum similarity score in the highest position of the paper ordering. Lemma 5.11 guarantees that this paper is bid on by the reviewer almost surely and the rest of the papers in $\mathcal{D}_i$ are bid on almost never by the reviewer.

We now consider the next arrival of a reviewer i' from $\mathcal{R}$ and note that $\mathcal{D}_{i'} = \mathcal{D}_i$ by definition of this set of reviewers. Between the arrivals of reviewers i and i', none of the papers in $\mathcal{D}_{i'}$ obtain any more bids almost surely since again from Lemma 5.12 any paper that is off the block diagonal for a reviewer is bid on almost never independent of the position the paper is shown. This means each paper in $\mathcal{D}_{i'}$ has zero bids almost surely except for the paper that has a bid from reviewer i almost surely. Accordingly, we apply Lemma 5.14 to determine that SUPER* with zero heuristic shows this reviewer the paper with the maximum similarity score among the papers without a bid within $\mathcal{D}_{i'}$ in the highest position of the paper ordering. Then, Lemma 5.11 guarantees that this paper is bid on by the reviewer almost surely and the rest of the papers in $\mathcal{D}_{i'}$ are bid on almost never by the reviewer.

Repeatedly applying this argument, upon the final arrival of a reviewer i'' from $\mathcal{R}$, each paper in $\mathcal{D}_{i''}$ has exactly one bid except for one paper that remains without a bid almost surely. We note again that by definition of this set of reviewers $\mathcal{D}_{i''} = \mathcal{D}_i$. From Lemma 5.14, SUPER* with zero heuristic shows the final paper without a bid within $\mathcal{D}_{i''}$ in the highest position of the paper ordering and Lemma 5.11 ensures that this paper is bid on by the reviewer almost surely and the rest of the papers in $\mathcal{D}_{i'}$ are bid on almost never by the reviewer.

Following the arrival of reviewer i'', the papers in $\mathcal{D}_i$ never appear on the block diagonal for a reviewer again and finish with exactly one bid almost surely after each being shown in the highest

position of the paper ordering exactly once to some reviewer for which they are on the block diagonal almost surely. The line of reasoning applied to the group of reviewers $\mathcal{R}$ can be duplicated for each of the m blocks of reviewers which share papers on the block diagonal. In doing so, it immediately follows from the decomposition in (5.122) that the expected paper-side gain of SUPER* with zero heuristic for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$, is given by

$$\mathbb{E}[\mathcal{G}_p^{\texttt{SUPER}^*}] = \sum_{j \in [d]} \mathbb{E}[\gamma_p(g_j)] = \sum_{j \in [d]} \gamma_p(1) = mq. \tag{5.123}$$

Identically as in the derivation of the optimal expected paper-side gain for the noiseless community model given in Section 5.A.4.3, this is the optimal expected paper-side gain that can be obtained since each reviewer bids on at most one paper almost surely and the given paper-side gain function is strictly concave so evenly distributing the bids over the papers maximizes the expected paper-side gain.

Properties of SUPER* with zero heuristic for the noisy community model similarity matrix. Since the conditions of Lemma 5.14 are satisfied for each reviewer almost surely, SUPER* with zero heuristic shows each reviewer $i \in [n]$ the paper with the maximum similarity score among the papers without a bid in $\mathcal{D}_i$ followed by the remaining papers in a decreasing order of the similarity scores. This fact leads to several properties of the algorithm for this similarity matrix class. From Lemma 5.13, $S_{i,j} > S_{i,j'}$ for $j \in \mathcal{D}_i, j' \in \mathcal{D}_i^c$. This means the paper with the maximum similarity score among the papers without a bid in $\mathcal{D}_i$ is equivalently the paper with the maximum similarity score among the papers without a bid. This results in the following property of the algorithm.

Property 5.1. *SUPER* with zero heuristic presents the paper with the maximum similarity score among the papers without a bid in the highest position of the paper ordering and the remaining papers in a decreasing order of the similarity scores to each reviewer $i \in [n]$ almost surely.*

Similarly, using Lemma 5.13, we can determine that the algorithm shows each paper that is on the block diagonal of the similarity matrix for the reviewer ahead of each paper off the block diagonal.

Property 5.2. *SUPER* with zero heuristic shows every paper in $\mathcal{D}_i$ ahead of every paper in $\mathcal{D}_i^c$ to each reviewer $i \in [n]$ almost surely.*

The final property that again follows from Lemma 5.13 is that the algorithm shows the papers off the block diagonal in a decreasing order of the similarity scores.

Property 5.3. *SUPER* with zero heuristic shows papers among $\mathcal{D}_i^c$ in a decreasing order of the similarity scores to each reviewer $i \in [n]$ almost surely.*

The properties of SUPER* with zero heuristic provided are going to assist the comparison of the expected reviewer-side gain with that from the baselines and the optimal policy. As discussed previously, intuitively Property 5.2 should guarantee that the algorithm obtains near-optimal expected reviewer-side gain since the noise in the similarity scores is bounded in a small interval and the similarity scores of papers on the block diagonal are much higher than that for papers off the block diagonal for a reviewer.

5.A.5.3 Suboptimality of SIM

In this section, we analyze SIM for the noisy community model similarity matrix class with the given gain and bidding functions.

Intuition and SIM policy. The SIM algorithm presents papers in a decreasing order of the similarity scores. This approach maximizes the expected reviewer-side gain since the given reviewer-side gain function is decreasing in the position a paper is shown and increasing in the similarity score of the paper. However, for the noisy community model similarity matrix class, the algorithm is suboptimal for the combined objective since the expected paper-side gain is far from optimal. As shown in the analysis of SUPER* with zero heuristic, to maximize the expected paper-side gain, each paper should only be shown in the highest position of the paper ordering to a reviewer once almost surely. Moreover, this must be when the paper is on the block diagonal for a reviewer so that it obtains a bid almost surely. The problem with SIM for this similarity matrix class is that it is oblivious to the number of bids on papers. As a result, the algorithm may show a paper in the highest position of the paper ordering to a reviewer that has only marginally higher similarity score, but many more bids, than another option. While this may result in a scarce amount more gain from the reviewer-side objective, we show it is costly in terms of the paper-side objective. We formalize this by showing SIM is significantly suboptimal for the expected paper-side gain, and that it only achieves a marginal amount more expected reviewer-side gain than SUPER* with zero heuristic.

Bounding the expected paper-side gain. Recall from (5.122) that the expected paper-side gain from any paper $j \in [d]$ is given by

$$\mathbb{E}[\gamma_p(g_j)] = \sum_{\ell=0}^{q} \mathbb{P}\Big(\ell = \sum_{i \in \mathcal{D}_j} \mathbb{1}\{\pi_i^{\texttt{SIM}}(j) = 1\}\Big)\gamma_p(\ell). \tag{5.124}$$

To bound this quantity for a given paper, we need to characterize the distribution of the number of times the paper is shown in the highest position of the paper ordering to reviewers for which it is on the block diagonal.

Toward this goal, let us consider any reviewer $i \in [n]$ and the probability of each paper being shown in the highest position of the paper ordering for the reviewer. For the given reviewer, the set of papers on the block diagonal is given by $\mathcal{D}_i$ and this set has cardinality q. Moreover, from Lemma 5.13, $S_{i,j} > S_{i,j'}$ for $j \in \mathcal{D}_i, j' \in \mathcal{D}_i^c$. This result says the similarity score of any paper on the block diagonal for the reviewer is greater than the similarity score of any paper off the block diagonal for the reviewer.

The SIM algorithm shows papers in a decreasing order of the similarity scores, which combined with Lemma 5.13, guarantees that the probability of any paper $j \in \mathcal{D}_i^c$ being shown in the highest position of the paper ordering is zero. For any paper $j \in \mathcal{D}_i$, the similarity score is given by $S_{i,j} = s - \nu_{i,j}$. The noise $\nu_{i,j}$ for each reviewer-paper pair (i, j) is drawn independently and uniformly at random from a bounded interval. This implies that the probability of any paper $j \in \mathcal{D}_i$ being shown in the highest position to the reviewer is $1/q$ since there are q papers in the set.

Recall that $\mathcal{D}_j$ for any paper $j \in [d]$ denotes the reviewers $i \in [n]$ for which the reviewer-paper pair (i, j) is on the block diagonal of the similarity matrix. From the preceding reasoning, the

probability of paper $j \in [d]$ being shown in the highest position to each reviewer $i \in \mathcal{D}_j$ is $1/q$. Consequently, the number of times paper $j \in [d]$ is shown in the highest position to reviewers in the set $\mathcal{D}_j$ follows a Binomial distribution with q trials since the cardinality of $\mathcal{D}_j$ is q and a success probability of $1/q$. This means the expected paper-side gain from any paper $j \in [d]$ given in (5.124) for SIM, is equivalently expressed as

$$\mathbb{E}[\gamma_p(g_j)] = \sum_{\ell=0}^{q} \binom{q}{\ell} \Big(\frac{1}{q}\Big)^{\ell} \Big(1 - \frac{1}{q}\Big)^{q-\ell} \gamma_p(\ell). \tag{5.125}$$

To bound (5.125), we can directly apply Lemma 5.10, which bounds the expectation of the square root of a binomial random variable, and obtain

$$\mathbb{E}[\gamma_p(g_j)] = \sum_{\ell=0}^{q} \binom{q}{\ell} \Big(\frac{1}{q}\Big)^{\ell} \Big(1 - \frac{1}{q}\Big)^{q-\ell} \gamma_p(\ell) \leq \Big(1 - \frac{1}{q}\Big)^{q-1} \Big(1 - \frac{\sqrt{2}}{2}\Big) + \frac{\sqrt{2}}{2}. \tag{5.126}$$

The bound in (5.126) is decreasing in q for $q \geq 2$. This means for every $q \geq 2$ and any paper $j \in [d]$,

$$\mathbb{E}[\gamma_p(g_j)] \leq \frac{2 + \sqrt{2}}{4}.$$

To get the expected paper-side gain of the algorithm, we sum this bound over the number of papers and obtain

$$\mathbb{E}[\mathcal{G}_p^{\texttt{SIM}}] = \sum_{j \in [d]} \mathbb{E}[\gamma_p(g_j)] \leq \Big(\frac{2 + \sqrt{2}}{4}\Big) mq. \tag{5.127}$$

Combining (5.127) with the expected paper-side gain of SUPER* with zero heuristic from (5.123), this implies for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$,

$$\mathbb{E}[\mathcal{G}_p^{\texttt{SUPER}^*} - \mathcal{G}_p^{\texttt{SIM}}] \geq mq - \Big(\frac{2 + \sqrt{2}}{4}\Big) mq. \tag{5.128}$$

Bounding the expected reviewer-side gain. We now turn our attention to comparing the expected reviewer-side gain of SUPER* with zero heuristic and SIM. We need to bound

$$\lambda \mathbb{E}[\mathcal{G}_r^{\texttt{SUPER}^*} - \mathcal{G}_r^{\texttt{SIM}}] = \lambda \mathbb{E}\Big[\sum_{i \in [n]} \sum_{j \in [d]} (\gamma_r(\pi_i^{\texttt{SUPER}^*}(j), S_{i,j}) - \gamma_r(\pi_i^{\texttt{SIM}}(j), S_{i,j}))\Big]. \tag{5.129}$$

Toward doing so, recall Property 5.2, which says SUPER* with zero heuristic shows every paper in $\mathcal{D}_i$ ahead of every paper in $\mathcal{D}_i^c$ to each reviewer $i \in [n]$ almost surely. Moreover, Property 5.3 says the papers among $\mathcal{D}_i^c$ are shown in decreasing order of the similarity scores to each reviewer $i \in [n]$ almost surely. In comparison, SIM shows papers in decreasing order of the similarity scores to each reviewer $i \in [n]$. From Lemma 5.13, the similarity scores of papers in $\mathcal{D}_i$ are greater than the similarity scores of papers in $\mathcal{D}_i^c$ for each reviewer $i \in [n]$. Combining this fact with the policy of SUPER* with zero heuristic and SIM, we can see that the algorithms show the papers among $\mathcal{D}_i^c$ in identical positions almost surely. This means the reviewer-side gain from this set of papers is equivalent for each of the algorithms almost surely.

Since the noise is bounded in a small interval, we expect that the ordering among papers in $\mathcal{D}_i$ would not impact the expected reviewer-side gain significantly as long as they are shown before the papers in $\mathcal{D}_i^c$. The following result formalizes this intuition and provides a bound.

Lemma 5.15. *Let π_i^ℓ denote the paper ordering that presents papers in decreasing order of the similarity scores. Moreover, denote by $\pi_i^{\ell'}$ any paper ordering that shows each paper in $\mathcal{D}_i$ ahead of each paper in $\mathcal{D}_i^c$ and papers among $\mathcal{D}_i^c$ in a decreasing order of the similarity scores. Then, under the assumptions of Theorem 5.4, the following bound holds for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$:*

$$\lambda \sum_{j\in[d]} \big(\gamma_r(\pi_i^{\ell'}(j), S_{i,j}) - \gamma_r(\pi_i^{\ell}(j), S_{i,j})\big) \geq -qe^{-emq}\log(4).$$

The proof of Lemma 5.15 is provided in Section 5.A.5.7.2.

The result of Lemma 5.15 immediately applies to (5.129) for each reviewer conditioned on the almost sure events given the characteristics of SUPER* with zero heuristic and SIM mentioned earlier and since it holds for any realization of the noise in the similarity scores. Combining (5.129) with Lemma 5.15 and noting that there are $n = mq$ reviewers, we obtain for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$,

$$\lambda\mathbb{E}[\mathcal{G}_r^{\texttt{SUPER}^*} - \mathcal{G}_r^{\texttt{SIM}}] = \lambda\mathbb{E}\Big[\sum_{i\in[n]}\sum_{j\in[d]} (\gamma_r(\pi_i^{\texttt{SUPER}^*}(j), S_{i,j}) - \gamma_r(\pi_i^{\texttt{SIM}}(j), S_{i,j}))\Big] \geq -mq^2e^{-emq}\log(4). \tag{5.130}$$

Finally, see that for every $m \geq 2$ and $q \geq 2$,

$$-mq^2e^{-emq}\log(4) \geq -8e^{-4e}\log(4) \geq -0.0001 \tag{5.131}$$

since $-mq^2e^{-emq}\log(4)$ is negative and increasing as a function of m and q for $m \geq 2$ and $q \geq 2$. From (5.130) and (5.131), we get that for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$,

$$\lambda\mathbb{E}[\mathcal{G}_r^{\texttt{SUPER}^*} - \mathcal{G}_r^{\texttt{SIM}}] \geq -0.0001. \tag{5.132}$$

Completing the bound. Combining the bounds on the expected paper-side and reviewer-side gain between SUPER* with zero heuristic and SIM given in (5.128) and (5.132), we get that for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$,

$$\begin{aligned}\mathbb{E}[\mathcal{G}^{\texttt{SUPER}^*} - \mathcal{G}^{\texttt{SIM}}] &= \mathbb{E}[\mathcal{G}_p^{\texttt{SUPER}^*} - \mathcal{G}_p^{\texttt{SIM}}] + \lambda\mathbb{E}[\mathcal{G}_r^{\texttt{SUPER}^*} - \mathcal{G}_r^{\texttt{SIM}}] \\ &\geq mq - \Big(\frac{2+\sqrt{2}}{4}\Big)mq - 0.0001 \\ &\geq mq/10.\end{aligned}$$

We conclude that there is a constant $c > 0$ such that for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$, SUPER* with zero heuristic obtains an additive factor of at least cmq more expected gain than SIM in the noisy community model.

5.A.5.4 Suboptimality of BID

We now analyze BID for the noisy community model with the given gain and bidding functions. We remark that much of the analysis in this section follows very similarly to that for BID given in Section 5.A.4.6 from the proof of Theorem 5.3 regarding the noiseless community model since the reviewer bidding behavior is identical in the noisy community model as characterized by Lemmas 5.11, 5.12, and 5.13. The primary adjustments are in the analysis of the expected reviewer-side gain with respect to SUPER* with zero heuristic.

Bounding the expected paper-side gain. For any given reviewer $i \in [n]$, the q papers in $\mathcal{D}_i$ are each in $\mathcal{D}_{i'}$ for $q-1$ other reviewers $i' \in [n]$ and also in $\mathcal{D}_{i''}^c$ for each of the remaining reviewers $i'' \in [n]$. For any reviewer $i \in [n]$, any paper $j \in \mathcal{D}_i^c$ is bid on almost never from Lemma 5.12 and any paper $j \in \mathcal{D}_i$ is only bid on with non-zero probability if shown in the highest position to the reviewer from Lemma 5.11. This means that upon the arrival of each reviewer $i \in [n]$, there is a paper in $\mathcal{D}_i$ with zero bids that has not been shown in the highest position to any reviewer previously almost surely. Furthermore, from Lemma 5.13, the similarity score of any paper in $\mathcal{D}_i$ is greater than the similarity score of any paper in $\mathcal{D}_i^c$ for each reviewer $i \in [n]$. Therefore, BID shows a paper in $\mathcal{D}_i$ with zero bids in the highest position to each reviewer $i \in [n]$ almost surely since papers are shown in increasing order of the number of bids and ties are broken in favor of the paper with the higher similarity score. Lemma 5.11 guarantees that if a paper in $\mathcal{D}_i$ is shown in the highest position of the paper ordering to reviewer $i \in [n]$, then it is bid on by the reviewer almost surely. Consequently, each paper $j \in [d]$ is shown exactly once almost surely in the highest position to some reviewer to some reviewer $i \in \mathcal{D}_j$. It then follows from the decomposition in (5.122) that the expected paper-side gain of BID for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$, is given by

$$\mathbb{E}[\mathcal{G}_p^{\texttt{BID}}] = \sum_{j \in [d]} \mathbb{E}[\gamma_p(g_j)] = \sum_{j \in [d]} \gamma_p(1) = mq.$$

From the expected paper-side gain of SUPER* with zero heuristic given in (5.123), we conclude that for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$,

$$\mathbb{E}[\mathcal{G}_p^{\texttt{SUPER}^*}] - \mathbb{E}[\mathcal{G}_p^{\texttt{BID}}] = 0. \tag{5.133}$$

Before moving on to the expected reviewer-side gain, we point out that SUPER* with zero heuristic and BID show the same paper in the highest position to each reviewer almost surely as direct result of Lemma 5.14, the definition of the BID policy, and the derivations of the expected paper-side gain for each algorithm.

Property 5.4. *SUPER* with zero heuristic and BID show the same paper in the highest position of the paper ordering to each reviewer almost surely.*

As we now show, the suboptimality of BID stems from the fact that after the paper shown in the highest position, the remaining papers are presented in increasing order of the number of bids, whereas SUPER* with zero heuristic shows the remaining papers in decreasing order of the similarity scores.

Bounding the expected reviewer-side gain. We now focus on showing SUPER* with zero heuristic obtains significantly more expected reviewer-side gain than BID. This requires deriving a

suitable lower bound on the following expression:

$$\lambda\mathbb{E}[\mathcal{G}_r^{\texttt{SUPER}^*} - \mathcal{G}_r^{\texttt{BID}}] = \lambda\mathbb{E}\Big[\sum_{i\in[n]}\sum_{j\in[d]}(\gamma_r(\pi_i^{\texttt{SUPER}^*}(j), S_{i,j}) - \gamma_r(\pi_i^{\texttt{BID}}(j), S_{i,j}))\Big]. \tag{5.134}$$

Let us begin by defining a "good event" for any reviewer and paper under which if the paper has probability zero of being bid on then it is not bid on and if the paper has probability one of being bid on then it is bid on. Formally, for any reviewer $k \in [n]$, paper $j \in [d]$, and paper ordering $\pi_k^{\texttt{ALG}}$ given by an algorithm $\texttt{ALG}$, we define

$$\mathcal{E}_{k,j}^{\texttt{ALG}} = \{\pi_k^{\texttt{ALG}}(j) = 1, S_{k,j} > s/2, \mathcal{B}_{k,j} = 1\} \cup \{\pi_k^{\texttt{ALG}}(j) \neq 1, \mathcal{B}_{k,j} = 0\} \cup \{S_{k,j} < s/2, \mathcal{B}_{k,j} = 0\}.$$

Moreover, for each reviewer $i \in [n]$, define the following event $\mathcal{E}_i = \cup_{k=1}^{i-1}\cup_{j=1}^{d}\{\mathcal{E}_{k,j}^{\texttt{SUPER}^*} \cup \mathcal{E}_{k,j}^{\texttt{BID}}\}$ which says the good event held for each reviewer that arrived previously for every paper and observe that the complement of this event occurs on a measure zero space by the structure of the bidding function given in (5.66). Consequently, from the law of total expectation, an equivalent form of (5.134) is given by

$$\lambda\mathbb{E}[\mathcal{G}_r^{\texttt{SUPER}^*} - \mathcal{G}_r^{\texttt{BID}}] = \lambda\mathbb{E}\Big[\sum_{i\in[n]}\mathbb{E}\Big[\sum_{j\in[d]}(\gamma_r(\pi_i^{\texttt{SUPER}^*}(j), S_{i,j}) - \gamma_r(\pi_i^{\texttt{BID}}(j), S_{i,j}))\Big|\mathcal{E}_i\Big]\Big]. \tag{5.135}$$

From Property 5.4, $\texttt{SUPER}^*$ with zero heuristic and $\texttt{BID}$ show the same paper in the highest position of the paper ordering to any reviewer $i \in [n]$ given the event $\mathcal{E}_i$. Moreover, from Property 5.1, $\texttt{SUPER}^*$ with zero heuristic presents the remainder of the papers in a decreasing order of the similarity scores given the event $\mathcal{E}_i$. This implies that for each reviewer $i \in [n]$, $\texttt{SUPER}^*$ with zero heuristic obtains at least as much reviewer-side gain as $\texttt{BID}$ given the event $\mathcal{E}_i$ since the reviewer-side gain function is increasing in the similarity score and decreasing in the position a paper is shown. Define $\mathcal{F}$ as the initial set of $\lfloor mq/4 \rfloor$ reviewers for which upon arrival of such a reviewer $i \in \mathcal{F}$ at least one paper on the block diagonal for the reviewer given by $\mathcal{D}_i$ has received a bid previously, where we recall that $\texttt{SUPER}^*$ with zero heuristic and $\texttt{BID}$ each obtain exactly one bid from each reviewer and on each paper almost surely as proved in the analysis of the expected paper-side gains. From (5.135) and the fact that $\texttt{SUPER}^*$ with zero heuristic obtains at least as much expected reviewer-side gain as $\texttt{BID}$ from each reviewer given the event $\mathcal{E}$, we obtain

$$\lambda\mathbb{E}[\mathcal{G}_r^{\texttt{SUPER}^*} - \mathcal{G}_r^{\texttt{BID}}] \geq \lambda\mathbb{E}\Big[\sum_{i\in\mathcal{F}}\mathbb{E}\Big[\sum_{j\in[d]}(\gamma_r(\pi_i^{\texttt{SUPER}^*}(j), S_{i,j}) - \gamma_r(\pi_i^{\texttt{BID}}(j), S_{i,j}))\Big|\mathcal{E}_i\Big]\Big]. \tag{5.136}$$

We now separate papers into relevant groups defined upon arrival for each reviewer $i \in \mathcal{F}$. Let $T_{i,1}$ be the set of papers containing only the paper that each algorithm shows in the highest position that has zero bids and belongs to the set $\mathcal{D}_i$. Let $T_{i,2}$ denote the remaining set of papers in $\mathcal{D}_i$ with zero bids and $T_{i,3}$ be the set of papers in $\mathcal{D}_i$ with one bid. Denote by $T_{i,4}$ the set of papers in $\mathcal{D}_i^c$ with zero bids and $T_{i,5}$ as the papers in $\mathcal{D}_i^c$ with one bid. Moreover, we let $N_{i,k} = |T_{i,k}|$ for $k \in \{1,2,3,4,5\}$ denote the number of papers in each set and define $\ell_i = N_{i,1} + N_{i,2} + 1$.

Accordingly, (5.136) is equivalently

$$\lambda\mathbb{E}[\mathcal{G}_r^{\texttt{SUPER}^*} - \mathcal{G}_r^{\texttt{BID}}] \geq \lambda\mathbb{E}\Big[\sum_{i\in\mathcal{F}}\mathbb{E}\Big[\sum_{j\in T_{i,1}\cup T_{i,2}\cup T_{i,3}\cup T_{i,4}\cup T_{i,5}}(\gamma_r(\pi_i^{\texttt{SUPER}^*}(j), S_{i,j}) - \gamma_r(\pi_i^{\texttt{BID}}(j), S_{i,j}))\Big|\mathcal{E}_i\Big]\Big]. \tag{5.137}$$

From Property 5.4, $\texttt{SUPER}^*$ with zero heuristic and $\texttt{BID}$ show the same paper in the highest position of the paper ordering to each reviewer $i \in [n]$ given $\mathcal{E}_i$. This allows us to simplify (5.137) and get that

$$\lambda\mathbb{E}[\mathcal{G}_r^{\texttt{SUPER}^*} - \mathcal{G}_r^{\texttt{BID}}] \geq \lambda\mathbb{E}\Big[\sum_{i\in\mathcal{F}}\mathbb{E}\Big[\sum_{j\in T_{i,2}\cup T_{i,3}\cup T_{i,4}\cup T_{i,5}}(\gamma_r(\pi_i^{\texttt{SUPER}^*}(j), S_{i,j}) - \gamma_r(\pi_i^{\texttt{BID}}(j), S_{i,j}))\Big|\mathcal{E}_i\Big]\Big]. \tag{5.138}$$

Given event $\mathcal{E}_i$, $\texttt{SUPER}^*$ with zero heuristic shows papers among $T_{i,2}\cup T_{i,3}$ in decreasing order of the similarity scores followed by papers among $T_{i,4}\cup T_{i,5}$ in decreasing order of the similarity scores consequent of Properties 5.1, 5.2, and 5.3.

Now consider an algorithm $\texttt{ALG}$ that shows papers from $T_{i,k}$ ahead of papers from $T_{i,k+1}$ for each $k \in \{1,2,3,4\}$. Moreover, let this algorithm present papers among each group $T_{i,k}$ for $k \in \{1,2,3,4,5\}$ in decreasing order of the similarity scores. The given reviewer-side gain function is decreasing in the position a paper is shown and increasing in the similarity score. This means the expected reviewer-side gain from any reviewer is maximized by showing papers in decreasing order of the similarity scores. Consequently, the expected reviewer-side gain of $\texttt{SUPER}^*$ with zero heuristic from each reviewer is at least as much as that from $\texttt{ALG}$. This fact leads to a lower bound on (5.138) of

$$\lambda\mathbb{E}[\mathcal{G}_r^{\texttt{SUPER}^*} - \mathcal{G}_r^{\texttt{BID}}] \geq \lambda\mathbb{E}\Big[\sum_{i\in\mathcal{F}}\mathbb{E}\Big[\sum_{j\in T_{i,2}\cup T_{i,3}\cup T_{i,4}\cup T_{i,5}}(\gamma_r(\pi_i^{\texttt{ALG}}(j), S_{i,j}) - \gamma_r(\pi_i^{\texttt{BID}}(j), S_{i,j}))\Big|\mathcal{E}_i\Big]\Big] \tag{5.139}$$

where now $\mathcal{E}_i = \cup_{k=1}^{i-1}\cup_{j=1}^{d}\{\mathcal{E}_{k,j}^{\texttt{ALG}}\cup\mathcal{E}_{k,j}^{\texttt{BID}}\}$.

The $\texttt{BID}$ policy shows papers in $T_{i,2}\cup T_{i,4}$ ahead of papers in $T_{i,3}\cup T_{i,5}$. Moreover, papers in $T_{i,2}$ are shown ahead of papers in $T_{i,4}$ and papers in $T_{i,3}$ ahead of papers in $T_{i,5}$. Papers among each group $T_{i,k}$ for $k \in \{2,3,4,5\}$ are shown in decreasing order of the similarity scores. This characterization of the $\texttt{BID}$ policy follows from definition, since papers are shown in increasing order of the number of bids with ties broken by the similarity scores. Recall that the similarity scores of papers in $T_{i,2}\cup T_{i,3}$ are greater than the similarity scores of papers in $T_{i,4}\cup T_{i,5}$ from Lemma 5.13. It is now clear that $\texttt{ALG}$ and $\texttt{BID}$ show papers among $T_{i,2}\cup T_{i,5}$ in identical positions given the event $\mathcal{E}_i$. Combining this fact with (5.139), we get that

$$\lambda\mathbb{E}[\mathcal{G}_r^{\texttt{SUPER}^*} - \mathcal{G}_r^{\texttt{BID}}] \geq \lambda\mathbb{E}\Big[\sum_{i\in\mathcal{F}}\mathbb{E}\Big[\sum_{j\in T_{i,3}\cup T_{i,4}}(\gamma_r(\pi_i^{\texttt{ALG}}(j), S_{i,j}) - \gamma_r(\pi_i^{\texttt{BID}}(j), S_{i,j}))\Big|\mathcal{E}_i\Big]\Big]. \tag{5.140}$$

We now separate the sum over papers in $T_{i,3}$ from the sums over papers in $T_{i,4}$ in (5.140) to

obtain

$$\lambda\mathbb{E}[\mathcal{G}_r^{\texttt{SUPER}^*} - \mathcal{G}_r^{\texttt{BID}}] \geq \lambda\mathbb{E}\Big[\sum_{i\in\mathcal{F}}\mathbb{E}\Big[\sum_{j\in T_{i,3}}(\gamma_r(\pi_i^{\texttt{ALG}}(j), S_{i,j}) - \gamma_r(\pi_i^{\texttt{BID}}(j), S_{i,j}))\Big|\mathcal{E}_i\Big] + \sum_{i\in\mathcal{F}}\mathbb{E}\Big[\sum_{j\in T_{i,4}}(\gamma_r(\pi_i^{\texttt{ALG}}(j), S_{i,j}) - \gamma_r(\pi_i^{\texttt{BID}}(j), S_{i,j}))\Big|\mathcal{E}_i\Big]\Big]. \tag{5.141}$$

The ALG policy shows papers in $T_{i,3}$ followed by papers in $T_{i,4}$, with each group of papers being presented in decreasing order of the similarity scores to each reviewer given the event $\mathcal{E}_i$. In contrast, the BID policy shows papers in $T_{i,4}$ followed by papers in $T_{i,3}$, with each group of papers being presented in decreasing order of the similarity scores. This means that BID shows each paper in $T_{i,3}$ later in the paper ordering by $N_{i,4}$ positions compared to ALG to each reviewer given $\mathcal{E}_i$. Analogously, ALG shows each paper in $T_{i,4}$ later in the paper ordering by $N_{i,3}$ positions compared to BID to each reviewer given $\mathcal{E}_i$. This set of facts and continuing from (5.137), leads to the bound

$$\lambda\mathbb{E}[\mathcal{G}_r^{\texttt{SUPER}^*} - \mathcal{G}_r^{\texttt{BID}}] \geq \lambda\mathbb{E}\Big[\sum_{i\in\mathcal{F}}\mathbb{E}\Big[\sum_{j\in T_{i,3}}(\gamma_r(\pi_i^{\texttt{ALG}}(j), S_{i,j}) - \gamma_r(\pi_i^{\texttt{ALG}}(j) + N_{i,4}, S_{i,j}))\Big|\mathcal{E}_i\Big] - \sum_{i\in\mathcal{F}}\mathbb{E}\Big[\sum_{j\in T_{i,4}}(\gamma_r(\pi_i^{\texttt{BID}}(j), S_{i,j}) - \gamma_r(\pi_i^{\texttt{BID}}(j) + N_{i,3}, S_{i,j}))\Big|\mathcal{E}_i\Big]\Big]. \tag{5.142}$$

From the decomposed form of the reviewer-side gain function in (5.69), an equivalent form of (5.142) is

$$\lambda\mathbb{E}[\mathcal{G}_r^{\texttt{SUPER}^*} - \mathcal{G}_r^{\texttt{BID}}] \geq \lambda\mathbb{E}\Big[\sum_{i\in\mathcal{F}}\mathbb{E}\Big[\sum_{j\in T_{i,3}}(2^{S_{i,j}} - 1)(\gamma_r^{\pi}(\pi_i^{\texttt{ALG}}(j)) - \gamma_r^{\pi}(\pi_i^{\texttt{ALG}}(j) + N_{i,4}))\Big|\mathcal{E}_i\Big] - \sum_{i\in\mathcal{F}}\mathbb{E}\Big[\sum_{j\in T_{i,4}}(2^{S_{i,j}} - 1)(\gamma_r^{\pi}(\pi_i^{\texttt{BID}}(j)) - \gamma_r^{\pi}(\pi_i^{\texttt{BID}}(j) + N_{i,3}))\Big|\mathcal{E}_i\Big]\Big]. \tag{5.143}$$

The similarity scores of papers in $T_{i,3}$ are given by $S_{i,j} = s - \nu_{i,j}$ and the similarity score of papers in $T_{i,4}$ are $S_{i,j} = \nu_{i,j}$. Recall that the noise is bounded in the interval $(0, \xi)$. Combining this with the fact that the function γ_r^{π} from (5.69) is decreasing on the domain $\mathbb{R}_{>0}$, we bound (5.143) as follows

$$\lambda\mathbb{E}[\mathcal{G}_r^{\texttt{SUPER}^*} - \mathcal{G}_r^{\texttt{BID}}] \geq \lambda\mathbb{E}\Big[(2^{s-\xi} - 1)\sum_{i\in\mathcal{F}}\mathbb{E}\Big[\sum_{j\in T_{i,3}}(\gamma_r^{\pi}(\pi_i^{\texttt{ALG}}(j)) - \gamma_r^{\pi}(\pi_i^{\texttt{ALG}}(j) + N_{i,4}))\Big|\mathcal{E}_i\Big] - (2^{\xi} - 1)\sum_{i\in\mathcal{F}}\mathbb{E}\Big[\sum_{j\in T_{i,4}}(\gamma_r^{\pi}(\pi_i^{\texttt{BID}}(j)) - \gamma_r^{\pi}(\pi_i^{\texttt{BID}}(j) + N_{i,3}))\Big|\mathcal{E}_i\Big]\Big]. \tag{5.144}$$

Now recall that ALG shows the papers in $T_{i,3}$ after the papers in $T_{i,1} \cup T_{i,2}$. Similarly, BID shows the papers in $T_{i,4}$ after the papers in $T_{i,1} \cup T_{i,2}$. Recall that $\ell_i = N_{i,1} + N_{i,2} + 1$. From this notation

and (5.144), we obtain

$$\begin{aligned}\lambda\mathbb{E}[\mathcal{G}_r^{\texttt{SUPER}^*} - \mathcal{G}_r^{\texttt{BID}}] \geq \lambda\mathbb{E}\Big[(2^{s-\xi}-1)\sum_{i\in\mathcal{F}}\mathbb{E}\Big[\sum_{j=\ell_i}^{\ell_i+N_{i,3}-1}(\gamma_r^\pi(j)-\gamma_r^\pi(j+N_{i,4}))\Big|\mathcal{E}_i\Big] \\ - (2^{\xi}-1)\sum_{i\in\mathcal{F}}\mathbb{E}\Big[\sum_{j=\ell_i}^{\ell_i+N_{i,4}-1}(\gamma_r^\pi(j)-\gamma_r^\pi(j+N_{i,3}))\Big|\mathcal{E}_i\Big]\Big].\end{aligned} \tag{5.145}$$

Following the exact techniques to prove Claim 1 in Section 5.A.4.6 for the expected reviewer-side gain analysis of BID in the noiseless community model, we get that for each reviewer $i \in \mathcal{F}$ and conditioned on the event $\mathcal{E}_i$,

$$N_{i,4} \geq mq - q - m - \lfloor mq/4 \rfloor + N_{i,3} + 1 \geq N_{i,3} \geq 1. \tag{5.146}$$

Now, toward the goal of bounding (5.145), we perform the following indexing manipulations:

$$\begin{aligned}\sum_{j=\ell_i}^{\ell_i+N_{i,4}-1}(\gamma_r^\pi(j)-\gamma_r^\pi(j+N_{i,3})) &= \sum_{j=\ell_i}^{\ell_i+N_{i,4}-1}\gamma_r^\pi(j) - \sum_{j=\ell_i+N_{i,3}}^{\ell_i+N_{i,3}+N_{i,4}-1}\gamma_r^\pi(j) \\ &= \sum_{j=\ell_i}^{\ell_i+N_{i,3}-1}\gamma_r^\pi(j) - \sum_{j=\ell_i+N_{i,4}}^{\ell_i+N_{i,3}+N_{i,4}-1}\gamma_r^\pi(j) \qquad (5.147) \\ &= \sum_{j=\ell_i}^{\ell_i+N_{i,3}-1}(\gamma_r^\pi(j)-\gamma_r^\pi(j+N_{i,4})). \qquad (5.148)\end{aligned}$$

To obtain (5.147), we used the fact from (5.146) that $N_{i,4} \geq N_{i,3}$ for any reviewer $i \in \mathcal{F}$ given the event $\mathcal{E}_i$.

Then from (5.144) and (5.148), we get

$$\lambda\mathbb{E}[\mathcal{G}_r^{\texttt{SUPER}^*} - \mathcal{G}_r^{\texttt{BID}}] \geq \lambda\mathbb{E}\Big[(2^{s-\xi}-2^s)\sum_{i\in\mathcal{F}}\mathbb{E}\Big[\sum_{j=\ell_i}^{\ell_i+N_{i,3}-1}(\gamma_r^\pi(j)-\gamma_r^\pi(j+N_{i,4}))\Big|\mathcal{E}_i\Big]\Big]. \tag{5.149}$$

Minimizing over $i \in \mathcal{F}$ in (5.149) and using the definition $|\mathcal{F}| = \lfloor mq/4 \rfloor$, we have

$$\lambda\mathbb{E}[\mathcal{G}_r^{\texttt{OPT}} - \mathcal{G}_r^{\texttt{BID}}] \geq \lambda\mathbb{E}\Big[(2^s-1)(\lfloor mq/4 \rfloor)\min_{i\in\mathcal{F}}\mathbb{E}\Big[\sum_{j=\ell_i}^{\ell_i+N_{i,3}-1}(\gamma_r^\pi(j)-\gamma_r^\pi(j+N_{i,4}))\Big|\mathcal{E}_i\Big]\Big]. \tag{5.150}$$

Moreover, for every $m \geq 2$, $q \geq 2$, it holds that

$$\lfloor mq/4 \rfloor = mq/4 - (mq \bmod 4)/4 \geq mq/8. \tag{5.151}$$

By definition of the noisy community model and the given bound on ξ, for every $m \geq 2$, $q \geq 2$, and

$\lambda \geq 0$, we get

$$2^{s-\xi} - 2^{\xi} = 2^{s-(1+\lambda)^{-1}e^{-emq}} - 2^{(1+\lambda)^{-1}e^{-emq}} \geq 2^{0.01-e^{-4e}} - 2^{e^{-4e}} \geq 1/150. \tag{5.152}$$

Combining (5.150), (5.151), and (5.152), we have

$$\lambda \mathbb{E}[\mathcal{G}_r^{\texttt{SUPER}^*} - \mathcal{G}_r^{\texttt{BID}}] \geq \Big(\frac{\lambda}{1200}\Big)\mathbb{E}\Big[\min_{i\in\mathcal{F}} \mathbb{E}\Big[\sum_{j=\ell_i}^{\ell_i+N_{i,3}-1} (\gamma_r^{\pi}(j) - \gamma_r^{\pi}(j+N_{i,4}))\Big|\mathcal{E}_i\Big]\Big]. \tag{5.153}$$

Then, applying exactly the same techniques to prove Claim 2 in Section 5.A.4.6 for the expected reviewer-side gain analysis of BID in the noiseless community model, for every $i \in \mathcal{F}$ conditioned on the event $\mathcal{E}_i$, we have

$$\min_{i\in\mathcal{F}} \sum_{j=\ell_i}^{\ell_i+N_{i,3}-1} (\gamma_r^{\pi}(j) - \gamma_r^{\pi}(j+N_{i,4})) \geq \Big(\frac{2}{5}\Big)\Big(\frac{1}{\log_2^2(mq)}\Big). \tag{5.154}$$

Finally, combining (5.154) with (5.153), for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$, the following bound holds

$$\lambda \mathbb{E}[\mathcal{G}_r^{\texttt{OPT}} - \mathcal{G}_r^{\texttt{BID}}] \geq \Big(\frac{1}{3000}\Big)\Big(\frac{\lambda mq}{\log_2^2(mq)}\Big). \tag{5.155}$$

Observe that the expectation in the right-hand side of (5.155) is dropped since it is not a random variable.

Completing the bound. Combining the bounds on the expected paper-side and reviewer-side gain between $\texttt{SUPER}^*$ with zero heuristic and BID given in (5.133) and (5.155), we find for every $m \geq 2, q \geq 2, \lambda \geq 0$,

$$\begin{aligned}\mathbb{E}[\mathcal{G}^{\texttt{SUPER}^*} - \mathcal{G}^{\texttt{BID}}] &= \mathbb{E}[\mathcal{G}_p^{\texttt{SUPER}^*} - \mathcal{G}_p^{\texttt{BID}}] + \lambda \mathbb{E}[\mathcal{G}_r^{\texttt{SUPER}^*} - \mathcal{G}_r^{\texttt{BID}}] \\ &\geq \Big(\frac{1}{3000}\Big)\Big(\frac{\lambda mq}{\log_2^2(mq)}\Big).\end{aligned}$$

We conclude that there exists a constant $c > 0$ such that for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$, $\texttt{SUPER}^*$ with zero heuristic obtains an additive factor of at least $c\lambda mq/\log_2^2(mq)$ more expected gain than BID for the noisy community model.

5.A.5.5 Suboptimality of RAND

In this section, we analyze RAND for the noisy community model similarity class with the given gain and bidding functions. A significant amount of the analysis in this section follows identically to that from analyzing RAND in the noiseless community model from Section 5.A.4.7 and the reason for the suboptimal behavior is identical.

Bounding the expected paper-side gain. Recall from (5.122) that the expected paper-side gain from any paper $j \in [d]$ is given by

$$\mathbb{E}[\gamma_p(g_j)] = \sum_{\ell=0}^{q} \mathbb{P}\Big(\ell = \sum_{i \in \mathcal{D}_j} \mathbb{1}\{\pi_i^{\texttt{RAND}}(j) = 1\}\Big)\gamma_p(\ell). \tag{5.156}$$

We remark that the decomposition of the expected paper-side gain from any paper given in (5.156) for RAND in the noisy community model is identical to that given in (5.99) for RAND in the noiseless community model. Since the RAND policy is independent of the similarity scores and the reviewer bids, the distribution of the number of times a paper is shown in the highest position to reviewers for which it is on the block diagonal is identical in the noisy community model as it is in the noiseless community model. Accordingly, we directly bound the expected paper-side gain of RAND in the noisy community model using the bound from (5.102) derived in Section 5.A.4.7 for RAND in the noiseless community model. Then, combining with the expected paper-side gain of SUPER* with zero heuristic from (5.123), we get that for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$,

$$\mathbb{E}[\mathcal{G}_p^{\texttt{SUPER}^*} - \mathcal{G}_p^{\texttt{RAND}}] \geq mq - \Big(\frac{6+\sqrt{2}}{16}\Big)mq. \tag{5.157}$$

Bounding the expected reviewer-side gain. We now need to compare the expected reviewer-side gain of SUPER* with zero heuristic and RAND. The SIM algorithm obtains the maximum expected reviewer-side gain that can be achieved since the reviewer-side gain function is increasing in the similarity score and decreasing in the position a paper is shown. Consequently, the bound on the expected reviewer-side gain from (5.132) between SUPER* with zero heuristic and SIM applies to RAND. Using the bound from (5.132), we get that for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$,

$$\lambda\mathbb{E}[\mathcal{G}_r^{\texttt{SUPER}^*} - \mathcal{G}_r^{\texttt{RAND}}] \geq -0.0001. \tag{5.158}$$

Completing the bound. Combining the bounds on the expected paper-side and reviewer-side gain between SUPER* with zero heuristic and RAND given in (5.157) and (5.158), for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$,

$$\begin{aligned}\mathbb{E}[\mathcal{G}^{\texttt{SUPER}^*} - \mathcal{G}^{\texttt{RAND}}] &= \mathbb{E}[\mathcal{G}_p^{\texttt{SUPER}^*} - \mathcal{G}_p^{\texttt{RAND}}] + \lambda\mathbb{E}[\mathcal{G}_r^{\texttt{SUPER}^*} - \mathcal{G}_r^{\texttt{RAND}}] \\ &\geq mq - \Big(\frac{6+\sqrt{2}}{16}\Big)mq - 0.0001 \geq mq/2.\end{aligned}$$

We conclude that there exists a constant $c > 0$ such that for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$, SUPER* with zero heuristic obtains an additive factor of at least cmq more expected gain than RAND in the noisy community model.

5.A.5.6 Near-Optimality of SUPER* with Zero Heuristic

In this section, we show that SUPER* with zero heuristic is nearly optimal. We let OPT denote the optimal algorithm for the expected gain.

Bounding the expected paper-side gain. As explained in Section 5.A.5.2, the expected paper-side gain of SUPER* with zero heuristic is the maximum that can be achieved. Indeed, this is consequent of the facts that each reviewer bids on at most one paper almost surely and the given paper-side gain function is strictly concave so evenly distributing the bids over the papers maximizes the expected paper-side gain. We conclude that for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$,

$$\mathbb{E}[\mathcal{G}_p^{\texttt{SUPER}^*}] - \mathbb{E}[\mathcal{G}_p^{\texttt{OPT}}] \geq 0. \tag{5.159}$$

Bounding the expected reviewer-side gain. The SIM algorithm obtains the maximum expected reviewer-side gain that can be achieved since the reviewer-side gain function as given in (5.68) is increasing in the similarity score and decreasing in the position a paper is shown, which means showing papers in decreasing order of the similarity scores to each reviewer maximizes the expected reviewer-side gain. Consequently, the bound on the expected reviewer-side gain from (5.132) between SUPER* with zero heuristic and SIM applies to the optimal algorithm. Using the bound from (5.132), we get that for $m \geq 2, q \geq 2$, and $\lambda \geq 0$,

$$\lambda\mathbb{E}[\mathcal{G}_r^{\texttt{SUPER}^*} - \mathcal{G}_r^{\texttt{OPT}}] \geq -0.0001. \tag{5.160}$$

Completing the bound. Combining the bounds on the expected paper-side and reviewer-side gain between SUPER* with zero heuristic and OPT given in (5.159) and (5.160), we get that for every $m \geq 2, q \geq 2$, and $\lambda \geq 0$,

$$\mathbb{E}[\mathcal{G}^{\texttt{SUPER}^*} - \mathcal{G}^{\texttt{OPT}}] = \mathbb{E}[\mathcal{G}_p^{\texttt{SUPER}^*} - \mathcal{G}_p^{\texttt{OPT}}] + \lambda\mathbb{E}[\mathcal{G}_r^{\texttt{SUPER}^*} - \mathcal{G}_r^{\texttt{OPT}}] \geq -0.0001.$$

We conclude SUPER* with zero heuristic is always within at least an additive factor of 0.0001 of the optimal in the noisy community model.

5.A.5.7 Proofs of Lemmas 5.14–5.15

In this section, we present the proofs of technical lemmas invoked in the primary proof of Theorem 5.4.

5.A.5.7.1 Proof of Lemma 5.14. In the proof of Corollary 5.A.2 given in Section 5.A.2, we showed in (5.17) that SUPER* with zero heuristic solves the problem

$$\pi_i^{\texttt{SUPER}^*} = \underset{\pi_i \in \Pi_d}{\arg\max} \quad \sum_{j \in [d]} f(\pi_i(j), S_{i,j})(\gamma_p(g_{i-1,j}+1) - \gamma_p(g_{i-1,j})) + \lambda \sum_{j \in [d]} \gamma_r(\pi_i(j), S_{i,j}) \tag{5.161}$$

in order to determine the ordering of papers $\pi_i^{\texttt{SUPER}^*}$ to present to reviewer $i \in [n]$ so that the immediate expected gain is maximized conditioned on the history of bids from reviewers that arrived previously. Recalling that the bidding function is $f(\pi_i(j), S_{i,j}) = \mathbb{1}\{\pi_i(j) = 1\}\mathbb{1}\{S_{i,j} > s/2\}$, the optimization problem in (5.161) is equivalent to

$$\pi_i^{\texttt{SUPER}^*} = \underset{\pi_i \in \Pi_d}{\arg\max} \sum_{j \in [d]} \mathbb{1}\{\pi_i(j) = 1\}\mathbb{1}\{S_{i,j} > s/2\}(\gamma_p(g_{i-1,j}+1) - \gamma_p(g_{i-1,j})) + \lambda \sum_{j \in [d]} \gamma_r(\pi_i(j), S_{i,j}). \tag{5.162}$$

Observe that $\mathcal{D}_i \cup \mathcal{D}_i^c = [d]$. Moreover, if $j \in \mathcal{D}_i$, then $S_{i,j} > s/2$ from Lemma 5.11. Analogously, if $j \in \mathcal{D}_i^c$, then $S_{i,j} < s/2$ from Lemma 5.12. This allows us to simplify (5.162) to the following problem:

$$\pi_i^{\texttt{SUPER}^*} = \arg\max_{\pi_i \in \Pi_d} \quad \sum_{j \in \mathcal{D}_i} \mathbb{1}\{\pi_i(j) = 1\}(\gamma_p(g_{i-1,j} + 1) - \gamma_p(g_{i-1,j})) + \lambda \sum_{j \in [d]} \gamma_r(\pi_i(j), S_{i,j}). \quad (5.163)$$

Given the assumption that there is a paper in $\mathcal{D}_i$ with zero bids and each paper in $\mathcal{D}_i$ has at most one bid, we need to prove $\texttt{SUPER}^*$ with zero heuristic shows the paper with the maximum similarity score among the papers without a bid in $\mathcal{D}_i$ followed by the remaining papers in a decreasing order of the similarity scores. To do so, we analyze the solution to (5.163) when the paper with the maximum similarity score has zero bids and when the paper with the maximum similarity score has one bid. For each scenario, we show $\texttt{SUPER}^*$ with zero heuristic presents the paper with the maximum similarity score among the papers without a bid in the highest position and the remaining papers in a decreasing order of the similarity scores. This is equivalent to the stated result we seek to prove since from Lemma 5.13, $S_{i,j} > S_{i,j'}$ for $j \in \mathcal{D}_i, j' \in \mathcal{D}_i^c$, which guarantees the paper with the maximum similarity score belongs to the set $\mathcal{D}_i$ and the paper with the maximum similarity score among the papers without a bid belongs to the set $\mathcal{D}_i$.

Before analyzing each scenario, we recall some key properties of the functions in the optimization problem given in (5.163) under the assumptions. The given paper-side gain function γ_p is such that the quantity $\gamma_p(g_{i-1,j} + 1) - \gamma_p(g_{i-1,j})$ is decreasing as a function of the number of bids $g_{i-1,j}$. As a result, the expected paper-side gain term from (5.163), which is given by

$$\sum_{j \in \mathcal{D}_i} \mathbb{1}\{\pi_i(j) = 1\}(\gamma_p(g_{i-1,j} + 1) - \gamma_p(g_{i-1,j})), \quad (5.164)$$

is maximized by showing the paper $j \in \mathcal{D}_i$ with the minimum number of bids in the highest position of the paper ordering. Moreover, the given reviewer-side gain function γ_r from (5.68) is decreasing in the position $\pi_i(j)$ in which a paper is shown and increasing in the similarity score $S_{i,j}$. Consequently, the expected reviewer-side gain term from (5.163), which is given by

$$\sum_{j \in [d]} \gamma_r(\pi_i(j), S_{i,j}), \quad (5.165)$$

is maximized by showing papers in decreasing order of the similarity scores.

Solution when the paper with the maximum similarity score has zero bids. If the paper with the maximum similarity score has zero bids, then the solution to (5.163) is to present the papers in decreasing order of the similarity scores. To see why this solution is optimal, observe that it maximizes each component of (5.163) given in (5.164) and (5.165) since the paper with the maximum similarity score has the minimum number of bids among the set $\mathcal{D}_i$ and papers are in decreasing order of the similarity scores. This solution is equivalent to presenting the paper with the maximum similarity score among the papers without a bid in the highest position and the remaining papers in a decreasing order of the similarity scores since the paper with the maximum similarity score has zero bids.

Solution when the paper with the maximum similarity score has one bid. To determine the solution to (5.163) when the paper with the maximum similarity score has one bid, we consider groups of candidate solutions. We group potential solutions into the set of paper orderings that show a paper with at least one bid in the highest position (group 1) and the set of paper orderings that show a paper without a bid in the highest position (group 2). For each group of paper orderings, we find the solution that maximizes the objective of the optimization problem in (5.163). To resolve which is optimal, we compare the objective values of the solutions from each group.

Analyzing Group 1. For this group, the solution is constrained to the set of paper orderings that show a paper with at least one bid in the highest position. The solution among this group that maximizes the objective of (5.163) is to show papers in decreasing order of the similarity scores. We call this candidate solution π_i^ℓ.

The candidate solution π_i^ℓ can be seen to be optimal among the group since it maximizes (5.164) subject to the constraint of the group and it maximizes (5.165). Indeed, solution π_i^ℓ maximizes (5.164) subject to the constraint of the group since the paper with the maximum similarity score has the minimum number of bids among the papers in $\mathcal{D}_i$ with at least one bid. Moreover, solution π_i^ℓ maximizes (5.165) since papers are shown in decreasing order of the similarity scores.

Analyzing Group 2. For this group, the solution is constrained to the set of paper orderings that show a paper without a bid in the highest position. The solution among this group that maximizes the objective of (5.163) is to show the paper with the maximum similarity score among the papers with zero bids in the highest position and then present the remaining papers in decreasing order of the similarity scores. We call this candidate solution $\pi_i^{\ell'}$.

The candidate solution $\pi_i^{\ell'}$ can be seen to be optimal among the group since it maximizes (5.164) and it maximizes (5.165) subject to the constraint of the group as we now show. From assumption, there is at least one paper in $\mathcal{D}_i$ with zero bids. The similarity score of any paper in $\mathcal{D}_i$ is greater than the similarity score of any paper in $\mathcal{D}_i^c$ from Lemma 5.13. This implies that the paper with the maximum similarity score among the papers with zero bids is in $\mathcal{D}_i$. Therefore, we conclude solution $\pi_i^{\ell'}$ maximizes (5.164) since the paper with the maximum similarity score among the papers with zero bids is in $\mathcal{D}_i$ and it has the minimum number of bids among the papers in $\mathcal{D}_i$. Moreover, solution $\pi_i^{\ell'}$ maximizes (5.165) subject to the constraint of the group since the paper with maximum similarity among the set of papers with zero bids is shown in the highest position and the remaining papers are shown in decreasing order of the similarity scores.

Comparing candidate solutions π_i^ℓ and $\pi_i^{\ell'}$. We now compare the objective of (5.163) for the candidate solutions π_i^ℓ and $\pi_i^{\ell'}$. Our goal is to show the objective given $\pi_i^{\ell'}$ is greater than the objective given π_i^ℓ. This is to say, we wish to show the following quantity is positive

$$\sum_{j\in\mathcal{D}_i}\big(\mathbb{1}\{\pi_i^{\ell'}(j)=1\}-\mathbb{1}\{\pi_i^{\ell}(j)=1\}\big)(\gamma_p(g_{i-1,j}+1)-\gamma_p(g_{i-1,j}))+\lambda\sum_{j\in[d]}\big(\gamma_r(\pi_i^{\ell'}(j),S_{i,j})-\gamma_r(\pi_i^{\ell}(j),S_{i,j})\big). \tag{5.166}$$

To simplify notation, let the quantity in (5.166) be denoted by $\mathcal{C}$. Since π_i^ℓ shows a paper in $\mathcal{D}_i$ with one bid in the highest position and $\pi_i^{\ell'}$ shows a paper in $\mathcal{D}_i$ with zero bids in the highest position, we obtain

$$\mathcal{C}=(\gamma_p(1)-\gamma_p(0))-(\gamma_p(2)-\gamma_p(1))+\lambda\sum_{j\in[d]}\big(\gamma_r(\pi_i^{\ell'}(j),S_{i,j})-\gamma_r(\pi_i^{\ell}(j),S_{i,j})\big).$$

Since π_i^ℓ presents papers in decreasing order of the similarity scores and $\pi_i^{\ell'}$ shows the papers in $\mathcal{D}_i^c$ in a decreasing order of the similarity scores after the papers in $\mathcal{D}_i$, we can apply Lemma 5.15 to get

$$\mathcal{C} \geq (\gamma_p(1) - \gamma_p(0)) - (\gamma_p(2) - \gamma_p(1)) - qe^{-emq}\log(4). \tag{5.167}$$

Observe that $-qe^{-emq}\log(4)$ is negative and an increasing as a function of m and q on the domain $m \geq 2$ and $q \geq 2$. This means for every $m \geq 2$ and $q \geq 2$,

$$-qe^{-emq}\log(4) \geq -2e^{-4e}\log(4) \geq -0.01. \tag{5.168}$$

Moreover, for the given paper-side gain function,

$$(\gamma_p(1) - \gamma_p(0)) - (\gamma_p(2) - \gamma_p(1)) = 2 - \sqrt{2} \tag{5.169}$$

Combining (5.167), (5.168), and (5.169), we obtain

$$\mathcal{C} \geq 2 - \sqrt{2} - 0.01 > 0.$$

Since $\mathcal{C} > 0$, we can conclude the objective of (5.163) given $\pi_i^{\ell'}$ is greater than the objective of (5.163) given π_i^ℓ. This means that the solution when the paper with the maximum similarity score has one bid is to show the paper with the maximum similarity score among the papers with zero bids in the highest position and then present the remaining papers in decreasing order of the similarity scores.

Combining the solutions. We have now derived the solution to (5.163) when the paper with the maximum similarity score has not obtained a bid previously and when the paper with the maximum similarity score has obtained exactly one bid previously. For each scenario, we showed SUPER* with zero heuristic presents the paper with the maximum similarity score among the papers without a bid in the highest position and the remaining papers in a decreasing order of the similarity scores. This allows us to conclude that if there is a paper in $\mathcal{D}_i$ with zero bids and each paper in $\mathcal{D}_i$ has at most one bid, then SUPER* with zero heuristic shows the paper with the maximum similarity score among the papers without a bid in $\mathcal{D}_i$ followed by the remaining papers in a decreasing order of the similarity scores.

5.A.5.7.2 Proof of Lemma 5.15. From the stated result, π_i^ℓ denotes the paper ordering that presents papers in decreasing order of the similarity scores. Recall that $\mathcal{D}_i$ denotes the set of papers on the block diagonal of the similarity matrix for any reviewer $i \in [n]$. Moreover, $\pi_i^{\ell'}$ is any paper ordering that shows each paper in $\mathcal{D}_i$ ahead of each paper in $\mathcal{D}_i^c$ and papers among $\mathcal{D}_i^c$ in a decreasing order of the similarity scores. Given this information, we need to bound

$$\lambda \sum_{j \in [d]} (\gamma_r(\pi_i^{\ell'}(j), S_{i,j}) - \gamma_r(\pi_i^\ell(j), S_{i,j})).$$

Each paper ordering π_i^ℓ and $\pi_i^{\ell'}$ shows the papers in $\mathcal{D}_i^c$ in a decreasing order of the similarity scores after the papers in $\mathcal{D}_i$ since from Lemma 5.13, $S_{i,j} > S_{i,j'}$ for $j \in \mathcal{D}_i, j' \in \mathcal{D}_i^c$. This means

papers in $\mathcal{D}_i^c$ are shown in identical positions by each paper ordering, so we get

$$\lambda \sum_{j\in[d]} (\gamma_r(\pi_i^{\ell'}(j), S_{i,j}) - \gamma_r(\pi_i^{\ell}(j), S_{i,j}) = \lambda \sum_{j\in\mathcal{D}_i} (\gamma_r(\pi_i^{\ell'}(j), S_{i,j}) - \gamma_r(\pi_i^{\ell}(j), S_{i,j})).$$

Equivalently, from the decomposed form of the given reviewer-side gain function from (5.68),

$$\lambda \sum_{j\in[d]} (\gamma_r(\pi_i^{\ell'}(j), S_{i,j}) - \gamma_r(\pi_i^{\ell}(j), S_{i,j}) = \lambda \sum_{j\in\mathcal{D}_i} (2^{S_{i,j}} - 1)\gamma_r^{\pi}(\pi_i^{\ell'}(j)) - \lambda \sum_{j\in\mathcal{D}_i} (2^{S_{i,j}} - 1)\gamma_r^{\pi}(\pi_i^{\ell}(j)).$$

By definition, the similarity score of each paper $j \in \mathcal{D}_i$ is given by $S_{i,j} = s - \nu_{i,j}$. Moreover, $\nu_{i,j}$ is bounded in $(0, \xi)$, so $s - \xi < s - \nu_{i,j} < s$. This fact leads to the lower bound

$$\lambda \sum_{j\in[d]} (\gamma_r(\pi_i^{\ell'}(j), S_{i,j}) - \gamma_r(\pi_i^{\ell}(j), S_{i,j}) \geq \lambda(2^{s-\xi} - 1) \sum_{j\in\mathcal{D}_i} \gamma_r^{\pi}(\pi_i^{\ell'}(j)) - \lambda(2^{s} - 1) \sum_{j\in\mathcal{D}_i} \gamma_r^{\pi}(\pi_i^{\ell}(j)).$$

Each paper ordering π_i^{ℓ} and $\pi_i^{\ell'}$ shows the papers in $\mathcal{D}_i$ in some order among the set of positions $\{1, \ldots, q\}$ since there are q papers in $\mathcal{D}_i$ by definition and each paper in $\mathcal{D}_i$ is shown ahead of each paper in $\mathcal{D}_i^c$. From this observation, we obtain

$$\lambda \sum_{j\in[d]} (\gamma_r(\pi_i^{\ell'}(j), S_{i,j}) - \gamma_r(\pi_i^{\ell}(j), S_{i,j}) \geq \lambda(2^{s-\xi} - 2^{s}) \sum_{j\in[q]} \gamma_r^{\pi}(j),$$

which is equivalently

$$\lambda \sum_{j\in[d]} (\gamma_r(\pi_i^{\ell'}(j), S_{i,j}) - \gamma_r(\pi_i^{\ell}(j), S_{i,j}) \geq \lambda(2^{s-\xi} - 2^{s}) \sum_{j\in[q]} \frac{1}{\log_2(j+1)}$$

from the definition of γ_r^{π} given in (5.69). Since $\lambda(2^{s-\xi} - 2^{s}) \leq 0$ and $1/\log_2(j+1) \leq 1$ for each $j \in [q]$, we obtain

$$\lambda \sum_{j\in[d]} (\gamma_r(\pi_i^{\ell'}(j), S_{i,j}) - \gamma_r(\pi_i^{\ell}(j), S_{i,j}) \geq \lambda q(2^{s-\xi} - 2^{s}).$$

Recall that $\xi \leq (1+\lambda)^{-1}e^{-emq}$, which means

$$\lambda \sum_{j\in[d]} (\gamma_r(\pi_i^{\ell'}(j), S_{i,j}) - \gamma_r(\pi_i^{\ell}(j), S_{i,j}) \geq \lambda q(2^{s-(1+\lambda)^{-1}e^{-emq}} - 2^{s}).$$

Now, see that $\lambda(2^{s-(1+\lambda)^{-1}e^{-emq}} - 2^{s})$ is non-positive and a decreasing function of λ and s on the domain $\lambda \geq 0$ and $s \geq 0.01$. This means for every $\lambda \geq 0$ and $s \geq 0.01$, the following relation holds

$$\begin{aligned}
\lambda(2^{s-(1+\lambda)^{-1}e^{-emq}} - 2^{s}) &\geq \lambda(2^{1-(1+\lambda)^{-1}e^{-emq}} - 2) \\
&\geq \lim_{\lambda'\to\infty} \lambda'(2^{1-(1+\lambda')^{-1}e^{-emq}} - 2) \\
&= -e^{-emq}\log(4).
\end{aligned}$$

Consequently, we conclude

$$\lambda \sum_{j \in [d]} (\gamma_r(\pi_i^{\ell'}(j), S_{i,j}) - \gamma_r(\pi_i^{\ell}(j), S_{i,j}) \geq -qe^{-emq} \log(4).$$

5.B Additional Results

In this section, we formally state and prove a pair of results that were mentioned informally in the main part of the chapter. We characterize the time complexity per-reviewer of the SUPER* algorithm for the general model and for a selected set of gain and bidding functions that admit a computationally efficient solution. Moreover, we show that SUPER* with any heuristic is globally optimal given a linear paper-side gain. This result is a corollary of the fact that SUPER* is locally optimal as shown in Theorem 5.1.

5.B.1 Time Complexity of SUPER*

The following proposition characterizes the time complexity of the SUPER* algorithm for each reviewer given the evaluations of the heuristic for the general form and a relevant class of gain and bidding functions that admits a computational efficient solution.

Proposition 5.1. *SUPER* has a time complexity of $\mathcal{O}(d^3)$ per-reviewer given the evaluations of the heuristic function. The time complexity of SUPER* improves to $\mathcal{O}(d \log(d))$ given a bidding function that can be decomposed into the form $f(\pi_i(j), S_{i,j}) = f^{\pi}(\pi_i(j)) f^{S}(S_{i,j})$ where $f^{\pi} : [d] \to [0,1]$ is non-increasing and $f^{S} : [0,1] \to [0,1]$ is non-decreasing, along with a reviewer-side gain function that can be decomposed into the form $\gamma_r(\pi_i(j), S_{i,j}) = f^{\pi}(\pi_i(j)) \gamma_r^{S}(S_{i,j})$ where $\gamma_r^{S} : [0,1] \to \mathbb{R}_{\geq 0}$ is non-decreasing.*

Proof of Proposition 5.1. We partition this proof by first examining the time complexity under the general model and then after which we consider the time complexity for the special case.

General time complexity. We begin by showing the time complexity of SUPER* for a reviewer given the heuristic evaluations under the general class of gain and bidding functions. The general form of the SUPER* algorithm calls Algorithm 5.3 upon the arrival of a reviewer to determine the ordering of papers to show the reviewer. The optimization problem in Algorithm 5.3 is in the form of the linear assignment problem. It is well known that the Hungarian algorithm can solve for the optimal solution of a linear assignment problem with a time complexity of $\mathcal{O}(d^3)$ (see, e.g., Chapter 8 in Lawler, 1976). As a result, SUPER* has a time complexity of $\mathcal{O}(d^3)$ for the general class of gain and bidding functions under consideration for each reviewer given the evaluations of the heuristic function.

Special case time complexity. In the proof of Theorem 5.1, we showed that the optimal paper ordering to present the final reviewer could be obtained by solving the linear program given in (5.13) with the weights from (5.14). The general version of SUPER* determines the ordering of papers to present any reviewer by calling Algorithm 5.3, which solves the linear program given in (5.13) using the weights from (5.15). We now show an equivalence between that solution method and a sorting algorithm for the class of gain and bidding functions given in the claim.

Prior to deriving the linear program in (5.13) as a method to obtain the optimal solution for the final reviewer in the proof of Theorem 5.1, we showed in (5.12) that the optimization problem for the final reviewer was of the form

$$\max_{\pi_n \in \Pi_d} \quad \sum_{j \in [d]} f(\pi_n(j), S_{n,j})(\gamma_p(g_{n-1,j}+1) - \gamma_p(g_{n-1,j})) + \lambda \sum_{j \in [d]} \gamma_r(\pi_n(j), S_{n,j}).$$

Given the function forms $f(\pi_n(j), S_{n,j}) = f^\pi(\pi_n(j)) f^S(S_{n,j})$ where $f^\pi : [d] \to [0,1]$ is non-increasing and $f^S : [0,1] \to [0,1]$ is non-decreasing, and $\gamma_r(\pi_n(j), S_{n,j}) = f^\pi(\pi_n(j))\gamma_r^S(S_{n,j})$ where $\gamma_r^S : [0,1] \to \mathbb{R}_{\geq 0}$ is non-decreasing, the problem can be reformulated as

$$\max_{\pi_n \in \Pi_d} \quad \sum_{j \in [d]} \alpha_{n,j} f^\pi(\pi_n(j)) \tag{5.170}$$

where

$$\alpha_{n,j} = f^S(S_{n,j})(\gamma_p(g_{n-1,j}+1) - \gamma_p(g_{n-1,j})) + \lambda \gamma_r^S(S_{n,j}) \ \forall \ j \in [d].$$

Consequently, for the class of gain and bidding functions given in the claim, an equivalent form of the general SUPER* algorithm that calls Algorithm 5.3 to determine the ordering of papers to show any reviewer $i \in [n]$ instead solves the problem in (5.170) using weights

$$\alpha_{i,j} = f^S(S_{i,j})(\gamma_p(g_{i-1,j}+h_{i,j}+1) - \gamma_p(g_{i-1,j}+h_{i,j})) + \lambda \gamma_r^S(S_{i,j}) \ \forall \ j \in [d]. \tag{5.171}$$

The optimal solution to a problem of the form in (5.170) is simply to present the papers in decreasing order of their corresponding values of $\alpha_{i,j}$ since the function f^π is non-increasing. The sorting procedure requires a time complexity of just $\mathcal{O}(d \log(d))$. Since Algorithm 5.2 solves the problem in (5.170) using the weights in (5.171), we conclude that the per-reviewer time complexity of SUPER* for the given class of gain and bidding functions and given the evaluations of the heuristic is $\mathcal{O}(d \log(d))$. □

5.B.2 SUPER* Optimality for Linear Paper-Side Gain

In this section, we show that SUPER* with any heuristic is optimal when the paper-side gain function is linear. This property of the algorithm follows rather directly from the local optimality result in Theorem 5.1 since for this type of paper-side gain function, the global optimization problem is decoupled between each reviewer.

Proposition 5.2. *SUPER*, with any heuristic, is optimal when the paper-side gain function is linear.*

Proof of Proposition 5.2. The optimization objective over the set of reviewers is defined as

$$\max_{\pi_1, \ldots, \pi_n \in \Pi_d} \quad \sum_{j \in [d]} \mathbb{E}[\gamma_p(g_j)] + \lambda \sum_{i \in [n]} \sum_{j \in [d]} \mathbb{E}[\gamma_r(\pi_i(j), S_{i,j})],$$

where the expectation is taken over the randomness in the bids made by the reviewers. Under a

linear paper-side gain function, the problem is equivalently formulated as

$$\max_{\pi_1,\ldots,\pi_n\in\Pi_d} \quad \sum_{j\in[d]} \mathbb{E}\Big[\sum_{i\in[n]} \mathcal{B}_{i,j}\Big] + \lambda \sum_{i\in[n]}\sum_{j\in[d]} \gamma_r(\pi_i(j), S_{i,j}),$$

where $\mathcal{B}_{i,j}$ denotes the random bid of reviewer i on paper j and the expectation on the reviewer-side gain went away since it is deterministic given a paper ordering for any reviewer. Using the structure of the bidding model, we can simplify the expectation over the paper-side gain to obtain the objective function

$$\max_{\pi_1,\ldots,\pi_n\in\Pi_d} \quad \sum_{i\in[n]}\sum_{j\in[d]} f(\pi_i(j), S_{i,j}) + \lambda \sum_{i\in[n]}\sum_{j\in[d]} \gamma_r(\pi_i(j), S_{i,j}).$$

The paper-side and reviewer-side gains are now decoupled between the ordering presented to each reviewer. Consequently, the optimal paper-ordering to present to each reviewer $i \in [n]$ is given by the solution to the optimization problem

$$\max_{\pi_i\in\Pi_d.} \quad \sum_{j\in[d]} f(\pi_i(j), S_{i,j}) + \lambda \sum_{j\in[d]} \gamma_r(\pi_i(j), S_{i,j}). \tag{5.172}$$

The SUPER* algorithm solves the following problem to determine the ordering of papers to present each reviewer $i \in [n]$:

$$\max_{\pi_i\in\Pi_d} \quad \sum_{j\in[d]} f(\pi_i(j), S_{i,j})(\gamma_p(g_{i-1,j} + h_{i,j} + 1) - \gamma_p(g_{i-1,j} + h_{i,j})) + \lambda \sum_{j\in[d]} \gamma_r(\pi_i(j), S_{i,j}).$$

Under a linear paper-side gain, the optimization problem the SUPER* algorithm solves for each reviewer simplifies to the problem

$$\max_{\pi_i\in\Pi_d} \quad \sum_{j\in[d]} f(\pi_i(j), S_{i,j}) + \lambda \sum_{j\in[d]} \gamma_r(\pi_i(j), S_{i,j})$$

since the number of bids and the heuristic cancels. We showed in the proof of Proposition 5.1 that the SUPER* algorithm solves this problem efficiently, and exactly. Since the problem is equivalent to that in (5.172) which gives the optimal solution for each reviewer, the SUPER* algorithm is optimal with a linear paper-side gain function. □

Chapter 6

Sequential Experimental Design for Transductive Linear Bandits

In this chapter, we continue our study of sequential decision-making and introduce the pure exploration *transductive linear bandit problem*: given a set of measurement vectors $\mathcal{X} \subset \mathbb{R}^d$, a set of items $\mathcal{Z} \subset \mathbb{R}^d$, a fixed confidence δ, and an unknown vector $\theta^* \in \mathbb{R}^d$, the goal is to infer $\arg\max_{z \in \mathcal{Z}} z^\top \theta^*$ with probability $1-\delta$ by making as few sequentially chosen noisy measurements of the form $x^\top \theta^*$ as possible. When $\mathcal{X} = \mathcal{Z}$, this setting generalizes *linear bandits*, and when $\mathcal{X}$ is the standard basis vectors and $\mathcal{Z} \subset \{0,1\}^d$, *combinatorial bandits*. The transductive setting naturally arises when the set of measurement vectors is limited due to factors such as availability or cost. As an example, in drug discovery the compounds and dosages $\mathcal{X}$ a practitioner may be willing to evaluate in the lab in vitro due to cost or safety reasons may differ vastly from those compounds and dosages $\mathcal{Z}$ that can be safely administered to patients in vivo. Alternatively, in recommender systems for books, the set of books $\mathcal{X}$ a user is queried about may be restricted to known best-sellers even though the goal might be to recommend more esoteric titles $\mathcal{Z}$. This chapter provides an instance-dependent lower bound for the transductive setting, an algorithm that matches these up to logarithmic factors, and an evaluation. In particular, we present the first non-asymptotic algorithm for linear bandits that nearly achieves the information-theoretic lower bound.

6.1 Introduction

In content recommendation or property optimization in the physical sciences, often there is a set of items (e.g., products to purchase, drugs) described by a set of feature vectors $\mathcal{Z} \subset \mathbb{R}^d$, and the goal is to find the $z \in \mathcal{Z}$ that maximizes some response or property (e.g., affinity of user to the product, drug combating disease). A natural model for these settings is to assume that there is an unknown vector $\theta^* \in \mathbb{R}^d$ and the expected response to any item $z \in \mathcal{Z}$, if evaluated, is equal to $z^\top \theta^*$. However, we often cannot measure $z^\top \theta^*$ directly, but we may infer it transductively through some potentially noisy probes. That is, given a finite set of probes $\mathcal{X} \subset \mathbb{R}^d$ we observe $x^\top \theta^* + \eta$ for any $x \in \mathcal{X}$ where η is independent mean-zero, sub-Gaussian noise. Given a set of measurements $\{(x_i, r_i)\}_{i=1}^N$ one can construct the least squares estimator $\widehat{\theta} = \arg\min_\theta \sum_{i=1}^N (r_i - x_i^\top \theta)^2$ and then use $\widehat{\theta}$ as a plug-in estimate for θ^* to estimate the optimal $z_* := \arg\max_{z \in \mathcal{Z}} z^\top \theta^*$. However, the accuracy of such a plug-in estimator depends critically on the number and choice of probes used to construct $\widehat{\theta}$. Unfortunately, the optimal allocation of probes cannot be decided a priori: it must be chosen sequentially and adapt to the observations in real-time to optimize the accuracy of the prediction.

If the probing vectors (arms) $\mathcal{X}$ are *equal* to the item vectors $\mathcal{Z}$, this problem is known as *pure exploration for linear bandits* which is considered by Karnin (2016); Soare et al. (2014); Tao

et al. (2018); Xu et al. (2018). This naturally arises in content recommendation, for example, if $\mathcal{X} = \mathcal{Z}$ is a feature representation of songs, and θ^* represents a user's music preferences, a music recommendation system can elicit the preference for a particular song $z \in \mathcal{Z}$ directly by enqueuing it in the user's playlist. However, often times there are constraints on which items in $\mathcal{Z}$ can be shown to the user.

1. $\mathcal{X} \subset \mathcal{Z}$. Consider a whiskey bar with hundreds of whiskies ranging in price from dollars a shot to hundreds of dollars. The bar tender may have an implicit feature representation of each whiskey, the patron has an implicit preference vector θ^*, and the bar tender wants to select the affordable whiskeys $\mathcal{X} \subset \mathcal{Z}$ in a taste test to get an idea of the patron's preferences before recommending the expensive whiskies that optimize the patron's preferences in $\mathcal{Z}$.
2. $\mathcal{Z} \subset \mathcal{X}$. In drug discovery, thousands of compounds are evaluated in order to determine which ones are effective at combating a disease. However, it may be that while $\mathcal{Z}$ is the set of compounds and doses that are approved for medical use (e.g., safe), it may be advantageous to test even unsafe compounds or dosages $\mathcal{X}$ such that $\mathcal{X} \supset \mathcal{Z}$. Such unsafe $\mathcal{X}$ may aid in predicting the optimal $z_* \in \mathcal{Z}$ because they provide more information about θ^*.
3. $\mathcal{Z} \cap \mathcal{X} = \emptyset$. Consider a user shopping for a home among a set $\mathcal{Z}$ where each is parameterized by a number of factors like distance to work, school quality, crime rate, etc. so that each $z \in \mathcal{Z}$ can be described as a linear combination of the relevant factors described by $\mathcal{X}$: $z = \sum_{x \in \mathcal{X}} \alpha_{z,x} x$, where we may take each $x \in \mathcal{X}$ to simply be one-hot-encoded. The response $x^\top \theta^* + \eta$ reflects the user's preferences for the query x, a specific attribute of the house. Indeed, if all $\alpha_{z,x} \in \{0, 1\}$ this is known as *pure exploration for combinatorial bandits* (Cao and Krishnamurthy, 2017; Chen et al., 2017). That is, a house either has the attribute, or not.

Given items $\mathcal{Z}$, measurement probes $\mathcal{X}$, a confidence δ, and an unknown θ^*, this chapter develops algorithms to sequentially decide which measurements in $\mathcal{X}$ to take in order to minimize the number of measurements necessary in order to determine z_* with high probability.

6.1.1 Contributions

Our goals are broadly to first define the transductive bandit problem and then characterize the instance-optimal sample complexity for this problem. Our contributions include the following.

1. In Section 6.2 we provide instance dependent lower bounds for the transductive bandit problem that simultaneously generalize previous known lower bounds for linear bandits and combinatorial bandits using standard arguments.
2. In Section 6.3 we give an algorithm (Algorithm 6.1) for transductive linear bandits and prove an associated sample complexity result (Theorem 6.2). We show that the sample complexity we obtain matches the lower bound up to logarithmic factors. This is the primary contribution of the chapter. Along the way, we discuss how rounding procedures can be used to improve upon the computational complexity of this algorithm.
3. In Sections 6.4 and 6.5 we contrast our algorithm with previous work from a theoretical and empirical perspective, respectively. Our experiments show that our theoretically superior algorithm is empirically competitive with previous algorithms on a range of problem scenarios.

6.1.2 Notation

For each $z \in \mathcal{Z}$ define the *gap* of z, $\Delta(z) = (z_* - z)^\top \theta^*$ and furthermore, $\Delta_{\min} = \min_{z \neq z_*} \Delta(z)$. If $A \in \mathbb{R}^{d \times d}_{\geq 0}$ is a positive semidefinite matrix, and $y \in \mathbb{R}^d$ is a vector, let $\|y\|_A^2 := y^\top A y$ denote the induced semi-norm. Let $\triangle_{\mathcal{X}} := \{\lambda \in \mathbb{R}^{|\mathcal{X}|} : \lambda \geq 0, \sum_{x \in \mathcal{X}} \lambda_x = 1\}$ denote the set of probability distributions on $\mathcal{X}$. Taking $\mathcal{S} \subset \mathcal{Z}$ to a subset of the item set, we define $\mathcal{Y}(\mathcal{S}) = \{z - z' : \forall\ z, z' \in \mathcal{S}, z \neq z'\}$ as the directions obtained from the differences between each pair of arms and $\mathcal{Y}^*(\mathcal{S}) = \{z_* - z : \forall\ z \in \mathcal{S} \setminus z_*\}$ as the directions obtained from the differences between the optimal arm and each suboptimal arm. Finally, for an arbitrary set of vectors $\mathcal{V} \subset \mathbb{R}^d$, define $\rho(\mathcal{V}) = \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{v \in \mathcal{V}} \|v\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x x x^\top)^{-1}}$. This quantity will be crucial in the discussion of our sample complexity and it is motivated in Section 6.2.2

6.2 Transductive Linear Bandits Problem

Consider known finite collections of d-dimensional vectors $\mathcal{X} \subset \mathbb{R}^d$ and $\mathcal{Z} \subset \mathbb{R}^d$, known confidence $\delta \in (0, 1)$, and unknown $\theta^* \in \mathbb{R}^d$. The objective is to identify $z_* = \arg\max_{z \in \mathcal{Z}} z^\top \theta^*$ with probability at least $1 - \delta$ while taking as few measurements in $\mathcal{X}$ as possible. Formally, a transductive linear bandits algorithm is described by a **selection rule** $X_t \in \mathcal{X}$ at each time t given the history $(X_s, R_s)_{s<t}$, **stopping time** τ with respect to the filtration $\mathcal{F}_t = (X_s, R_s)_{s \leq t}$, and **recommendation rule** $\widehat{z} \in \mathcal{Z}$ invoked at time τ which is $\mathcal{F}_\tau$-measurable. We assume that X_t is $\mathcal{F}_{t-1}$-measurable and may use additional sources of randomness; in addition at each time t that $R_t = X_t^\top \theta^* + \eta_t$ where η_t is independent, zero-mean, and 1-sub-Gaussian. Let $\mathbb{P}_{\theta^*}, \mathbb{E}_{\theta^*}$ denote the probability law of $R_t | \mathcal{F}_{t-1}$ for all t.

Definition 6.1. *We say that an algorithm for a transductive bandit problem is δ-PAC for $\mathcal{X}, \mathcal{Z} \subset \mathbb{R}^d$ if for all $\theta^* \in \mathbb{R}^d$ we have $\mathbb{P}_{\theta^*}(\widehat{z} = z_*) \geq 1 - \delta$.*

6.2.1 Optimal allocations

In this section we discuss a number of ways we can allocate a measurement budget to the different arms. The following establishes a lower bound on the expected number of samples any δ-PAC algorithm must take.

Theorem 6.1. *Assume $\eta_t \overset{iid}{\sim} \mathcal{N}(0, 1)$ for all t. Then for any $\delta \in (0, 1)$, any δ-PAC algorithm must satisfy*

$$\mathbb{E}_{\theta^*}[\tau] \geq \log(1/2.4\delta) \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{z \in \mathcal{Z} \setminus \{z_*\}} \frac{\|z_* - z\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x x x^\top)^{-1}}}{((z_* - z)^\top \theta^*)^2}.$$

This lower bound is proved in Section 6.A. It is derived using typical lower bound techniques and employs the transportation inequality of Kaufmann et al. (2016). Moreover, the lower bound generalizes a previous lower bound in the setting of linear bandits (Soare, 2015) and lower bounds in the combinatorial bandit literature (Chen et al., 2017).

Optimal static allocation. To demonstrate that this lower bound is tight, define

$$\lambda^* := \arg\min_{\lambda \in \triangle_{\mathcal{X}}} \max_{z \in \mathcal{Z} \setminus \{z_*\}} \frac{\|z_* - z\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x x x^\top)^{-1}}}{((z_* - z)^\top \theta^*)^2} \text{ and } \psi^* = \max_{\mathcal{Z} \setminus \{z_*\}} \frac{\|z_* - z\|^2_{(\sum_{x \in \mathcal{X}} \lambda^*_x x x^\top)^{-1}}}{((z_* - z)^\top \theta^*)^2}, \tag{6.1}$$

where ψ^* is the value of the lower bound and λ^* is the allocation that achieves it. Suppose we sample arm $x \in \mathcal{X}$ exactly $2\lfloor \lambda^*_x N \rfloor$ times where we assume[1] $N \in \mathbb{N}$ is sufficiently large so that $\min_{x:\lambda_x > 0} \lfloor \lambda_x N \rfloor > 0$. If $N = \lceil 2\psi^* \log(|\mathcal{Z}|/\delta) \rceil$ then as we will show shortly (Section 6.2.2), the least squares estimator $\widehat{\theta}$ satisfies $(z_* - z)^\top \widehat{\theta} > 0$ for all $z \in \mathcal{Z} \setminus z_*$ with probability at least $1 - \delta$. Thus, with probability at least $1 - \delta$, z_* is equal to $\widehat{z} = \arg\max_{z \in \mathcal{Z}} z^\top \widehat{\theta}$ and the total number of samples is bounded by $2N$ which is within $4 \log(|\mathcal{Z}|)$ of the lower bound. Unfortunately, of course, the allocation λ^* relies on knowledge of θ^* (which determines z_*) which is unknown a priori, and thus this is not a realizable strategy.

Other static allocations. Short of λ^* it is natural to consider allocations that arise from optimal linear experimental design (Pukelsheim, 2006). For the special case of $\mathcal{X} = \mathcal{Z}$ it has been argued ad nauseam that a G-optimal design, $\arg\min_{\lambda \in \triangle_{\mathcal{X}}} \max_{x \in \mathcal{X}, x \neq x_*} \|x\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x x x^\top)^{-1}}$, is woefully loose since it does not utilize the differences $x - x'$, $x, x' \in \mathcal{X}$ (Lattimore and Szepesvari, 2017; Soare et al., 2014; Xu et al., 2018). Also for the $\mathcal{X} = \mathcal{Z}$ case, Soare et al. (2014); Yu et al. (2006) have proposed the static $\mathcal{XY}$-allocation given as $\arg\min_{\lambda \in \triangle_{\mathcal{X}}} \max_{x, x' \in \mathcal{X}} \|x - x'\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x x x^\top)^{-1}}$. In the work of Soare et al. (2014) it is shown that no more than $\mathcal{O}(\frac{d}{\Delta^2_{\min}} \log(|\mathcal{X}| \log(1/\Delta_{\min})/\delta))$ samples from each of these allocations suffice to identify the best arm. While the above discussion demonstrates that for every θ^* there exists an optimal static allocation (that explicitly uses θ^*) nearly achieving the lower bound, any static allocation with no prior knowledge of θ^* can require a factor of d more samples than necessary.

Proposition 6.1. *Let c, c' be universal constants. For any $\gamma > 0$, d even, there exists sets $\mathcal{X} = \mathcal{Z} \subset \mathbb{R}^d$ and a set $\Theta \subset \mathbb{R}^d$, such that $\inf_{\mathcal{A}} \max_{\theta \in \Theta} \mathbb{E}_\theta[\tau] \geq \frac{cd \log(1/\delta)}{\gamma}$ where $\mathcal{A}$ is the set of all algorithms that are δ-PAC for $\mathcal{X}, \mathcal{Z}$ and take a static allocation of samples. On the other hand $\psi^*/c' \leq d + \frac{1}{\gamma}$ for every choice of $\theta^* \in \Theta$.*

This proposition indicates that it is necessary to devise an adaptive algorithm to obtain a instance-optimal sample complexity. The proof of this proposition can be found in Section 6.B.

Adaptive allocations. As suggested by the problem definition, our strategy is to adapt the allocation over time, informed by the observations up to the current time. Specifically, our algorithm will proceed in rounds where at round t, we perform an $\mathcal{XY}$-allocation that is sufficient to remove all arms $z \in \mathcal{Z}$ that have gaps of at least $2^{-(t-1)}$. We show that the total number of measurements accumulates to $\psi^* \log(|\mathcal{Z}|^2/\delta)$ times some additional logarithmic factors, nearly achieving the optimal allocation as well as the lower bound. In Section 6.4, we review related procedures for the specific case of $\mathcal{X} = \mathcal{Z}$.

---

[1]Such an assumption is avoided by a sophisticated rounding procedure that we will describe shortly.

6.2.2 Review of Least Squares

Given a fixed design $\mathbf{x}_T = (x_t)_{t=1}^T$ with each $x_t \in \mathcal{X}$ and associated rewards $(r_t)_{t=1}^T$, a natural approach is to construct the ordinary-least squares (OLS) estimate $\widehat{\theta} = (\sum_{t=1}^T x_t x_t^\top)^{-1}(\sum_{t=1}^T r_t x_t)$. One can show $\widehat{\theta}$ is unbiased with covariance $\preceq (\sum_{t=1}^T x_t x_t^\top)^{-1}$. Moreover, for any $y \in \mathbb{R}^d$, we have[2]

$$\mathbb{P}\left(y^\top(\theta^* - \widehat{\theta}) \geq \sqrt{\|y\|^2_{(\sum_{t=1}^T x_t x_t^\top)^{-1}} 2\log(1/\delta)}\right) \leq \delta. \tag{6.2}$$

In particular, if we want this to hold for all $y \in \mathcal{Y}^*(\mathcal{Z})$, we need to union bound over $\mathcal{Z}$ replacing δ with $\delta/|\mathcal{Z}|$. Let us now use this to analyze the procedure discussed above (in the discussion on the optimal static allocation after Theorem 6.1) that gives an allocation matching the lower bound. With the choice of $N = \lceil 2\psi^* \log(|\mathcal{Z}|/\delta)\rceil$ and the allocation $2\lfloor \lambda_x^* N\rfloor$ for each $x \in \mathcal{X}$, we have for each $z \in \mathcal{Z} \setminus z_*$ that with probability at least $1-\delta$,

$$(z_* - z)^\top \widehat{\theta} \geq (z_* - z)^\top \theta^* - \sqrt{\|z_* - z\|^2_{(\sum_x 2\lfloor N\lambda_x^* x x^T\rfloor)^{-1}} 2\log(|\mathcal{Z}|/\delta)} \geq 0$$

since for each $y = z_* - z \in \mathcal{Y}^*(\mathcal{Z})$ we have

$$y^\top\Big(\sum_{x\in\mathcal{X}} 2\lfloor N\lambda_x^*\rfloor x x^\top\Big)^{-1} y \leq y^\top\Big(\sum_{x\in\mathcal{X}} \lambda_x^* x x^\top\Big)^{-1} y/N \leq ((z_* - z)^\top\theta^*)^2/(2\log(|\mathcal{Z}|/\delta)), \tag{6.3}$$

where the last inequality plugs in the value of N and the definition of ψ^*. The fact that at most one $z' \in \mathcal{Z}$ can satisfy $(z'-z)^\top\widehat{\theta} > 0$ for all $z \neq z' \in \mathcal{Z}$, and that $z' = z_*$ does, certifies that $\widehat{z} = \arg\max_{z\in\mathcal{Z}} z^\top\widehat{\theta}$ is indeed the best arm with probability at least $1-\delta$. Note that equation (6.3) provides the motivation for how the form of ψ^* is obtained. Rearranging, it is equivalent to,

$$N \geq 2\log(|\mathcal{Z}|/\delta) \max_{\mathcal{Z}\setminus\{z_*\}} \frac{\|z_* - z\|^2_{(\sum_{x\in\mathcal{X}} \lambda_x^* x x^\top)^{-1}}}{((z_* - z)^\top\theta^*)^2} \text{ for all } z \in \mathcal{Z}\setminus\{z_*\}$$

Thinking of the right hand side of the inequality as a function of λ, λ^* is precisely chosen to minimize this quantity and hence the sample complexity.

6.2.3 Rounding Procedures

We briefly digress to address a technical issue. Given an allocation λ and an arbitrary subset of vectors $\mathcal{Y}$, in general, drawing N samples $\mathbf{x}_N := \{x_1, \ldots, x_N\}$ at random from $\mathcal{X}$ according to the distribution λ_x may result in a design where $\max_{y\in\mathcal{Y}} \|y\|^2_{(\sum_{t=1}^N x_t x_t^\top)^{-1}}$ (which appears in the width of the confidence interval (6.2)) differs significantly from $\max_{y\in\mathcal{Y}} \|y\|^2_{(\sum_{x\in\mathcal{X}} \lambda_x x x^\top)^{-1}}/N$. Naive strategies for choosing $\mathbf{x}_N$ will fail. We cannot simply use an allocation of $N\lambda_x$ samples for any specific x since this may not be an integer. Furthermore, greedily rounding $N\lambda_x$ to an allocation $\lfloor N\lambda_x\rfloor$ or $\lceil N\lambda_x\rceil$ may result in fewer than necessary, or far more than N total samples if

[2]There is a technical issue of whether the set $\mathcal{Z}$ lies in the span of $\mathcal{X}$ which in general is necessary to obtain unbiased estimates of $(z_* - z)^\top\theta^*$. Throughout the following we assume that $\text{span}(\mathcal{X}) = \mathbb{R}^d$.

the support of λ is large. However, given $\epsilon > 0$, there are *efficient rounding procedures* that produce $(1+\epsilon)$ approximations as long as N is greater than some minimum number of samples $r(\epsilon)$. In short, given λ and a choice of N they return an allocation $\mathbf{x}_N$ satisfying

$$\max_{y \in \mathcal{Y}} \|y\|^2_{(\sum_{i=1}^N x_i x_i^\top)^{-1}} \leq (1+\epsilon) \max_{y \in \mathcal{Y}} \|y\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x x x^\top)^{-1}} / N.$$

Such a procedure with $r(\epsilon) \leq \mathcal{O}(d/\epsilon^2)$ is described in the work of Allen-Zhu et al. (2021) and detailed with respect to our usage in Section 6.D. In our experiments we use a rounding procedure from Pukelsheim (2006) that is easier to implement with $r(\epsilon) = 2\|\lambda\|_0/\epsilon \leq (d(d+1)+2)/\epsilon$. In general, ϵ should be thought of as a constant. The number of samples N we need to take in our algorithm will be significantly larger than $r(\epsilon)$, so the impact of the rounding procedure is minimal.

6.3 Sequential Experimental Design for Transductive Linear Bandits

Our algorithm for the pure exploration transductive bandit is presented in Algorithm 6.1. The algorithm proceeds in rounds, keeping track of the active arms $\widehat{\mathcal{Z}}_t \subseteq \mathcal{Z}$ in each round t. At the start of round t, the algorithm samples in such a way to remove all arms with gaps greater than $2^{-(t-1)}$. Thus denoting $\mathcal{S}_t := \{z \in \mathcal{Z} : \Delta(z) \leq 4 \cdot 2^{-t}\}$, in round t we expect $\widehat{\mathcal{Z}}_t \subset \mathcal{S}_t$.

As described above, if we knew θ^*, we would sample according to the optimal allocation

$$\arg\min_{\lambda \in \triangle_{\mathcal{X}}} \max_{z \in \widehat{\mathcal{Z}}_t} \|z_* - z\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x x x^\top)^{-1}} / ((z_* - z)^\top \theta^*)^2.$$

However, if at the start of the round we only have an upper bound on the gaps $\Delta(z) \leq 4 \cdot 2^{-t}$ and do not know z^*, we can use the triangle inequality to obtain

$$4 \max_{z \in \widehat{\mathcal{Z}}_t} \|z_* - z\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x x x^\top)^{-1}} \geq \max_{y \in \mathcal{Y}(\widehat{\mathcal{Z}}_t)} \|y\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x x x^T)^{-1}}$$

and lower-bound the objective by the expression $(2^{t-3})^2 \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{y \in \mathcal{Y}(\widehat{\mathcal{Z}}_t)} \|y\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x x x^T)^{-1}}$.[3] This motivates our choice of λ_t and $\rho(\mathcal{Y}(\widehat{\mathcal{Z}}_t))$. Thus by the same logic used in Section 6.2.2, selecting $N_t = \lceil 2(2^t)^2(1+\epsilon)\rho(\mathcal{Y}(\widehat{\mathcal{Z}}_t)) \log(|\mathcal{Z}|^2/\delta_t) \rceil$ samples should suffice to guarantee that we can construct a confidence interval on each $(z - z')^\top \theta^*$ for $(z - z') \in \mathcal{Y}(\widehat{\mathcal{Z}}_t)$ of size at most 2^{-t} (with the $|\mathcal{Z}|^2$ in the logarithm accounting for a union bound over arms). The $(1+\epsilon)$ accounts for slack from the rounding principle. Finally, this confidence interval allows us to provably remove any arm $z \in \widehat{\mathcal{Z}}_t$ such that $\Delta(z) > 2^{-(t-1)}$ in round t.

Theorem 6.2. *Assume that* $\max_{z \in \mathcal{Z}} \Delta(z) \leq 2$. *Then with probability at least* $1 - \delta$, *using an* ϵ*-efficient rounding procedure, Algorithm 6.1 returns* z_* *and requires a worst-case sample complexity*

[3] Where we recall for any subset $\mathcal{S} \subset \mathcal{Z}, \mathcal{Y}(\mathcal{S}) := \{z - z' : z, z' \in \mathcal{S}\}$ and for an arbitrary subset $\mathcal{V} \subset \mathbb{R}^d$ we have $\rho(V) = \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{v \in \mathcal{V}} \|v\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x x x^\top)^{-1}}$.

Algorithm 6.1: **RAGE**$(\mathcal{X}, \mathcal{Z}, \epsilon, r(\cdot), \delta)$: **R**andomized **A**daptive **G**ap **E**limination

Input: Arms $\mathcal{X} \subset \mathbb{R}^d$, items $\mathcal{Z} \subset \mathbb{R}^d$, rounding approximation factor ϵ with default value 1/10, function $r(\cdot)$ giving minimum number of samples to obtain rounding approximation ϵ, & confidence level $\delta \in (0,1)$.
Initialize: Let $\widehat{\mathcal{Z}}_1 \leftarrow \mathcal{Z}, t \leftarrow 1$
while $|\widehat{\mathcal{Z}}_t| > 1$ **do**

$\delta_t \leftarrow \frac{\delta}{t^2}$
$\lambda_t^* \leftarrow \arg\min_{\lambda \in \triangle_{\mathcal{X}}} \max_{y \in \mathcal{Y}(\widehat{\mathcal{Z}}_t)} \|y\|^2_{(\sum_{x\in\mathcal{X}} \lambda_x x x^\top)^{-1}}$
$\rho(\mathcal{Y}(\widehat{\mathcal{Z}}_t)) \leftarrow \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{y \in \mathcal{Y}(\widehat{\mathcal{Z}}_t)} \|y\|^2_{(\sum_{x\in\mathcal{X}} \lambda_x x x^\top)^{-1}}$
$N_t \leftarrow \max\left\{\left\lceil 2(2^t)^2 \rho(\mathcal{Y}(\widehat{\mathcal{Z}}_t))(1+\varepsilon)\log(|\mathcal{Z}|^2/\delta_t)\right\rceil, r(\epsilon)\right\}$
$\mathbf{x}_{N_t} \leftarrow \textsc{Round}(\lambda_t^*, N_t)$
Pull arms $x_1, \ldots, x_{N_t}$ and obtain rewards $r_1, \ldots, r_{N_t}$
Compute $\widehat{\theta}_t = A_t^{-1} b_t$ using $A_t := \sum_{j=1}^{N_t} x_j x_j^\top$ and $b_t := \sum_{j=1}^{N_t} x_j r_j$
$\widehat{\mathcal{Z}}_{t+1} \leftarrow \widehat{\mathcal{Z}}_t \setminus \left\{z \in \widehat{\mathcal{Z}} | \exists\, z' \in \widehat{\mathcal{Z}} : \|z'-z\|_{A_t^{-1}} \sqrt{2\log(|\mathcal{Z}|^2/\delta_t)} < (z'-z)^\top \widehat{\theta}_t\right\}$
$t \leftarrow t+1$

Output: $\widehat{\mathcal{Z}}_t$

of

$$N \le \sum_{t=1}^{\lfloor \log_2(4/\Delta_{\min})\rfloor} \max\left\{\left\lceil 2(2^t)^2 \rho(\mathcal{Y}(\mathcal{S}_t))(1+\epsilon)\log(t^2|\mathcal{Z}|^2/\delta)\right\rceil, r(\epsilon)\right\} \tag{6.4}$$

where $\mathcal{S}_t = \{z \in \mathcal{Z} : \Delta(z) \le 4 \cdot 2^{-t}\}$. *In particular,* Round *can be chosen so that* $r(\epsilon) = \mathcal{O}(d/\epsilon^2)$. *Furthermore,* $N \le c\psi^* \log_2(4/\Delta_{\min}) \log(|\mathcal{Z}|^2 \log_2(4/\Delta_{\min})^2/\delta) + r(\epsilon)\log_2(4/\Delta_{\min})$ *for some absolute constant* c, *in other words Algorithm 6.1 is instance optimal up to logarithmic factors.*

The proof of Theorem 6.2 is in Section 6.C. The explanation before Algorithm 6.1 provides a high level of how the sample complexity is derived. Essentially, the number of samples in each round is selected to ensure that all arms worse than a certain suboptimality threshold or removed from the active set, and also so that the optimal arm is never knocked out from the active set. Summing the samples over the number of rounds then gives the sample complexity. The relation from the bound given in (6.4) to the problem-dependent quantity ψ^* is then the primary novelty of the proof so that we can adequately compare the algorithm sample complexity to the lower bound.

6.3.1 Interpreting the sample complexity.

Up to logarithmic factors, Algorithm 6.1 matches the lower bound obtained in Theorem 6.1. However, the term $\rho(\mathcal{Y}(\mathcal{S}_t))$ may seem a bit mysterious. In this section we try to interpret this quantity in terms of the geometry of $\mathcal{X}$ and $\mathcal{Z}$.

Let $\text{conv}(\mathcal{X} \cup -\mathcal{X})$ denote the convex hull of $\mathcal{X} \cup -\mathcal{X}$, and for any set $\mathcal{Y} \subset \mathbb{R}^d$ define the gauge of $\mathcal{Y}$,

$$\gamma_{\mathcal{Y}} = \max\{c > 0 : c\mathcal{Y} \subset \text{conv}(\mathcal{X} \cup -\mathcal{X})\}.$$

In the case where $\mathcal{Y}$ is a singleton $\mathcal{Y} = \{y\}$, $\gamma(y) := \gamma_{\mathcal{Y}}$ is the *gauge norm* of y with respect to $\text{conv}(\mathcal{X} \cup -\mathcal{X})$, a familiar quantity from convex analysis (Rockafellar, 2015). We can provide a natural upper bound for $\rho(\mathcal{Y})$ in terms of the gauge.

Lemma 6.1. *Let $\mathcal{Y} \subset \mathbb{R}^d$. Then*

$$\max_{y \in \mathcal{Y}} \|y\|_2^2 / (\max_{x \in \mathcal{X}} \|x\|_2) \leq \rho(\mathcal{Y}) \leq d/\gamma_{\mathcal{Y}}^2. \tag{6.5}$$

In the case of a singleton $\mathcal{Y} = \{y\}$, we can improve the upper bound to $\rho(\mathcal{Y}) \leq 1/\gamma(y)^2$.

The proof of this Lemma is in Section 6.E. To see the potential for adaptive gains we focus on the case of linear bandits where $\mathcal{X} = \mathcal{Z}$. Consider an example with $\mathcal{X} = \{e_i\}_{i=1}^d \cup \{z'\}$ for $z' = (\cos(\alpha), \sin(\alpha), 0, \cdots, 0)$ where $\alpha \in [0, .1)$, and $\theta^* = e_1$. Note that $\Delta_{\min} \approx 1 - \cos(\alpha) \approx \alpha^2/2$. Then $\mathcal{S}_1 = \mathcal{X}$, and an easy computation shows $\gamma_{\mathcal{Y}(\mathcal{X})}$ is a constant bounded from zero. After the first round, all arms except e_1 and z' will be removed, so $\mathcal{Y}(\mathcal{S}_t) = \{e_1 - z'\}$ for $t \geq 2$, and $\gamma_{\mathcal{Y}(\mathcal{S}_t)} \approx 1/\sin(\alpha) \approx 1/\alpha$. Summing over all rounds, we see that this implies a sample complexity of $\widetilde{\mathcal{O}}(d + 1/\alpha^2)$ up to log factors, which is a significant improvement over the static $\mathcal{XY}$-allocation sample complexity of $\widetilde{\mathcal{O}}(d/\alpha^2)$.

6.4 Related Work

When $\mathcal{X} = \mathcal{Z} = \{e_1, \cdots, e_d\} \subset \mathbb{R}^d$ is the set of standard basis vectors, the problem reduces to that of the best-arm identification problem for multi-armed bandits which has been extensively studied (Chen and Li, 2015; Even-Dar et al., 2006; Jamieson et al., 2014; Karnin et al., 2013; Kaufmann et al., 2016). In addition, pure exploration for combinatorial bandits where $\mathcal{X} = \{e_1, \cdots, e_d\} \subset \mathbb{R}^d$ and $\mathcal{Z} \subset \{0,1\}^d$ has also received a great deal of attention (Cao and Krishnamurthy, 2017; Chen et al., 2016, 2017, 2014).

In the setting of linear bandits when $\mathcal{X} = \mathcal{Z}$, despite a great deal of work in the regret and contextual settings (Abbasi-Yadkori et al., 2011; Dani et al., 2008; Lattimore and Szepesvari, 2017; Li et al., 2010), there has been far less work on linear bandits for pure exploration. This problem was first introduced in Soare et al. (2014) and since then, there have been a few other works on this topic (Karnin, 2016; Tao et al., 2018; Xu et al., 2018) that we now discuss.

- Soare et al. (2014) made the initial connections to G-optimal experimental design. That work provides the first passive algorithm with a sample complexity of $\mathcal{O}(\frac{d}{\Delta_{\min}^2} \log(|\mathcal{X}|/\delta) + d^2)$. Note that the d^2 comes from the minimum number of samples needed for an efficient rounding procedure and thus could be reduced to d using improved rounding procedures such as the methods from Allen-Zhu et al. (2021). They also provide an adaptive algorithm, $\mathcal{XY}$-adaptive algorithm for linear bandits. Their algorithm is very similar to ours, with two notable differences. Firstly, instead of using an efficient rounding procedure, they use a greedy iterative scheme to compute an optimal allocation. Secondly, their algorithm does not discard items that are provably sub-optimal. As a result, their sample complexity (up to logarithmic factors) scales as $\max\{M^*, \psi^*\} \log(|\mathcal{X}|/(\Delta_{\min}\delta)) + d^2$ where M^* is defined (informally) as the amount of samples needed using a static allocation to remove all sub-optimal directions in $\mathcal{Y}(\mathcal{X}) \setminus \mathcal{Y}^*(\mathcal{X})$.

- In Tao et al. (2018), the focus is on developing different estimators with the goal of removing the constant term d^2 in Soare et al.'s passive sample complexity. Instead of using a rounding procedure, they use a different estimator than the ordinary least squares estimator θ^*. Note that

the rounding procedure from Allen-Zhu et al. (2021) could have been applied directly to the static allocation algorithm from Soare et al. (2014) to obtain the same sample complexity as the one obtained in Tao et al. (2018). They also provide an adaptive algorithm *ALBA*, that achieves a sample complexity of $\mathcal{O}(\sum_{i=1}^{d} 1/\Delta_i^2)$ where Δ_i is the i-th smallest gap of the vectors in $\mathcal{X}$. It is easy to see that this sample complexity is not optimal: imagine a situation in which the vectors of $\mathcal{X}$ with the $(d-1)$-smallest gaps are identical to the vector $x' \neq x^\star$. Then we only need to pay once for the samples needed to remove x', not $(d-1)$-times. Finally, their algorithms do not compute the optimal allocation over differences of vectors in $\mathcal{X}$, but instead on $\mathcal{X}$ directly à la G-optimal design. We will see the inefficiency of this strategy in the experiments.

- Karnin (2016) provides an algorithm that uses repeated rounds (for probability amplification) of exploration phases combined with verification phases to provide an asymptotically optimal algorithm, meaning when $\delta \to 0$ the sample complexity divided by $\log(1/\delta)$ approaches ψ^*. Though this is a nice theoretical result, the algorithm is not practical; the exploration phase is simply a naïve passive G-optimal design.

- In Xu et al. (2018), a fully adaptive algorithm called LinGapE inspired by the UGapE algorithm (Gabillon et al., 2012) is proposed. Since LinGapE is fully adaptive, a confidence bound allowing for dependence in the samples is necessary and the authors employ the self-normalized bound of Abbasi-Yadkori et al. (2011). The algorithm requires each arm to be pulled once, an undesirable characteristic of a linear bandit algorithm since the structure of the problem allows for information to be obtained about arms that are not pulled. A recent work (Kazerouni and Wein, 2021), extends this algorithm to generalized linear models where the expected reward of pulling arm z reward is given by a non-linear link function of $z^\top \theta^*$.

Finally, we mention Yu et al. (2006), who consider transductive experimental design from a computational and optimization perspective, and explores $\mathcal{XY}$-allocation for arbitrary kernels.

6.5 Experiments

In this section, we present simulations for the linear bandit pure exploration problem and the general transductive bandit problem. We compare our proposed algorithm with both adaptive and non-adaptive strategies. The adaptive strategies are $\mathcal{XY}$-Adaptive allocation from Soare et al. (2014), LinGapE from Xu et al. (2018), and ALBA from Tao et al. (2018). The non-adaptive strategies are static $\mathcal{XY}$-allocation, as described in Section 6.2, and an oracle strategy that knows θ^* and samples according to λ^*. We do not compare to the algorithm given in Karnin (2016) since it is primarily a theoretical contribution and in moderate-confidence regimes obtains only the non-adaptive sample complexity. We run each algorithm at a confidence level of $\delta = 0.05$. The empirical failure probability of each of the algorithms in the simulations is zero. To compute the samples for RAGE, we used the Frank-Wolfe algorithm to find λ_t, and then the rounding procedure from Pukelsheim (2006) with $\epsilon = 1/10$. We remark here that in our implementation of the $\mathcal{XY}$-Adaptive allocation, we follow the experiments of Soare et al. (2014) and allow for provably suboptimal arms to be discarded (though this is not how the algorithm is written in their paper). The resulting algorithm is then similar to our algorithm. Unless explicitly mentioned, noise in the observations

was generated from a standard normal distribution. Additional details of the experimental setup are provided in Section 6.F.

Linear bandits: benchmark example. The first experiment we present has become a benchmark in the linear bandit pure exploration literature since it was introduced by Soare et al. (2014). In this problem, $\mathcal{X} = \mathcal{Z} = \{e_1, \dots, e_d, x'\} \subset \mathbb{R}^d$ where e_i is the i-th standard basis vector, $x' = \cos(.01)e_1 + \sin(.01)e_2$, and $\theta^* = 2e_1$ so that $x_* = x_1$. An efficient sampling strategy for this problem needs to focus on reducing uncertainty in the direction $(x_1 - x_{d+1})$, which can be achieved by focusing pulls on arm $x_2 = e_2$ since it is most aligned with this direction.

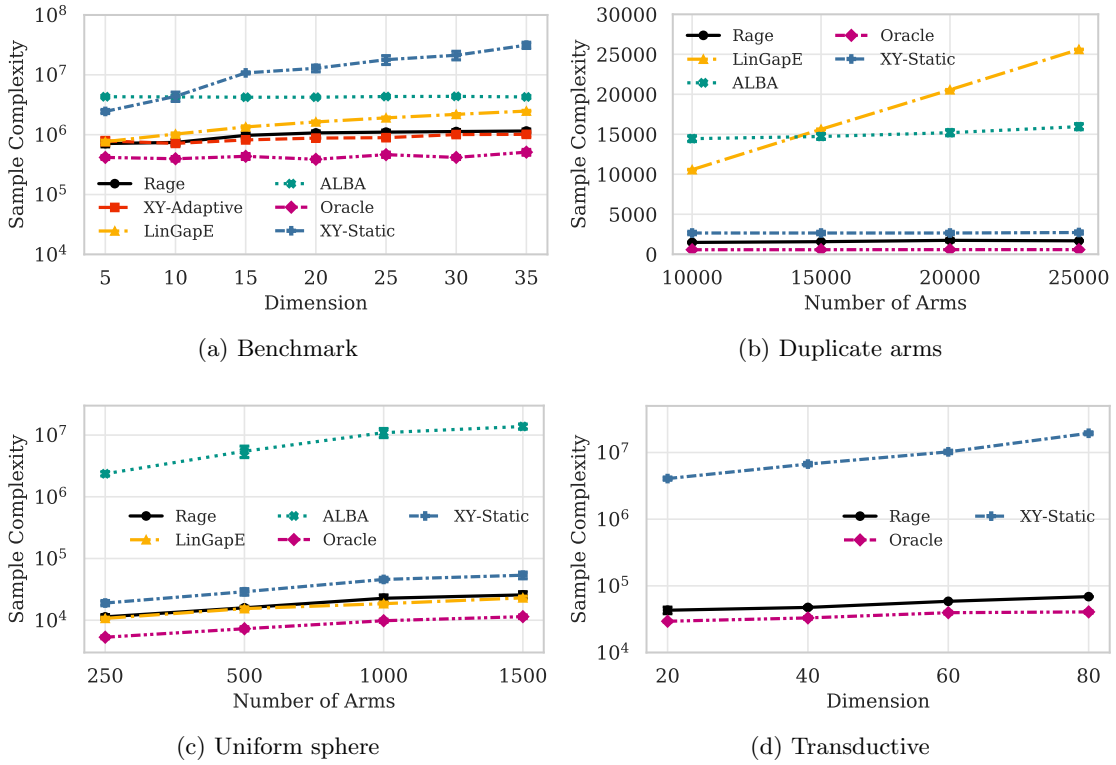
The results for this experiment are shown in Fig. 6.1a. The RAGE algorithm performs competitively with existing algorithms and the oracle allocation. The $\mathcal{XY}$-Adaptive algorithm is similar to RAGE, but with weaker theoretical guarantees, so naturally it performs nearly equivalently. We omit it from the remaining experiments for this reason. The LinGapE algorithm performs well when the number of dimensions and arms is small. However, as the number of arms grows, LinGapE suffers from a worse dimension dependency in the confidence interval. ALBA performs the worst of the recently proposed algorithms and this is to be expected since it computes an allocation on the $\mathcal{X}$ set instead of on the $\mathcal{Y}(\mathcal{X})$ set. This example clearly highlights the gains of adaptive sampling over non-adaptive allocations such as the static $\mathcal{XY}$-allocation. However, since $\mathcal{X}$ is relatively small in this case, it fails to tease out important differences between the algorithms that can greatly increase the sample complexity. We construct examples to demonstrate these claims now.

Many arms with moderate gaps. In this example, for a given value of $n \geq 3$, we construct a set of arms $\mathcal{X} \subset \mathbb{R}^2$, where $\mathcal{X} = \mathcal{Z} = \{e_1, \cos(3\pi/4)e_1 + \sin(3\pi/4)e_2\} \cup \{\cos(\pi/4 + \phi_i)e_1 + \sin(\pi/4 + \phi_i)e_2\}_{i=3}^n$ with $\phi_i \sim \mathcal{N}(0, .09)$ for each $i \in \{3, \dots, n\}$. The parameter vector is fixed to be $\theta^* = e_1$ so that x_1 is the optimal arm, x_2 gives the most information to identify the optimal arm, and the remaining arms roughly point in the same direction with an expected gap of $\Delta \approx 0.3$.

In Fig. 6.1b, we show the results of the experiment as we increase the number of arms. The LinGapE algorithm suffers from a linear scaling in the number of arms since it must sample each arm as an initialization. An efficient sampling strategy should focus energy on x_2, and as it does so, it will gain information about the arms that are nearly duplicates of each other, which is how RAGE performs.

Uniform distribution on a sphere. In this example, $\mathcal{X} = \mathcal{Z}$ is sampled from a unit sphere of dimension $d = 9$ centered at the origin. Following Tao et al. (2018), we select the two closest arms $x, x' \in \mathcal{X}$ and let $\theta^* = x$. In Fig. 6.1c, we show the sample complexity of the algorithms as the number of arms grows. The RAGE algorithm significantly outperforms ALBA and this is primarily due to the fact that ALBA computes a G-optimal design on the active vectors in each round instead of on the differences between these vectors. Thus the ALBA sampling distribution can be focused on a very different set of arms from the optimal one.

Transductive example. We now present a general transductive bandit example. Since the existing algorithms in the linear bandit literature do not generalize to this problem, we compare with a static $\mathcal{XY}$-allocation on $\mathcal{X}, \mathcal{Y}(\mathcal{Z})$ and an oracle $\mathcal{XY}$-allocation on $\mathcal{X}, \mathcal{Y}^*(\mathcal{Z})$ that knows the optimal arm and the gaps. We construct an example in $\mathbb{R}^d$ with d even where $\mathcal{X} = \{e_1, \dots, e_d\}$. The set $\mathcal{Z}$ is also chosen so $|\mathcal{Z}| = d$, the first $d/2$ vectors are given by $z_1, \dots, z_{d/2} = (e_1, \dots, e_{d/2})$ and then $z_{d/2+j} = \cos(.1)e_j + \sin(.1)e_{j+d/2}$ for each $j \in \{1, \dots, d/2\}$. Take $\theta^* = e_1$ so z_1 is the optimal



(a) Benchmark

(b) Duplicate arms

(c) Uniform sphere

(d) Transductive

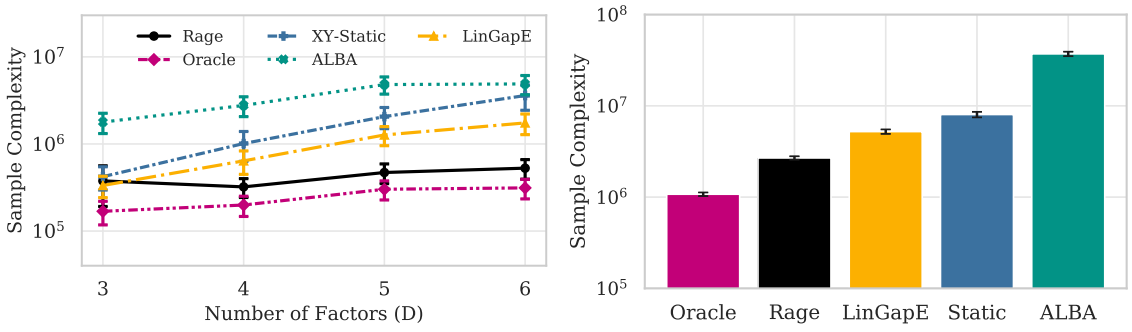
Figure 6.1

arm. The results of this simulation are depicted in Fig. 6.1d. The RAGE algorithm significantly outperforms the static allocation and nearly matches the oracle allocation.

We now present examples motivated by real-world applications.

Multivariate testing example. In many experimental design settings, there are a series of D factors that can be either in a set of N states, and the goal is to determine the treatment configuration that has the highest outcome for a given metric. As a concrete example in web page optimization, it is common that the composition of an advertisement layout selection may consist of several choices such as an image, background color, and keyword to display (e.g. Hill et al. (2017)), and we seek to find the combination with the highest clickthrough rate. To formalize the problem, consider a webpage consisting of D distinct slots and suppose that there are 2 content choices that can be presented in each slot. Let the set $\mathcal{W} = \{-1, 1\}^D$ satisfying $|\mathcal{W}| = 2^D$ encode each layout. We model the problem using a factorial design (see, e.g., Box et al. (2005)) including pairwise interaction features to generate a linear bandit problem. Each layout is represented by an arm $x \in \mathcal{X}$ where $\mathcal{X} = \mathcal{Z} \subset \{-1, 1\}^{1+D+D(D-1)/2}$ and $|\mathcal{X}| = 2^D$. The expected reward of any $x \in \mathcal{X}$ corresponding to a layout $w \in \mathcal{W}$ is given by

$$x^\top \theta^* = \theta_0^* + \alpha_1 \sum_{j=1}^{D} \theta_j^* w_j + \alpha_2 \sum_{k=1}^{D} \sum_{\ell=k+1}^{D} \theta_{k,\ell}^* w_k w_\ell,$$



(a) Experimental Design

(b) Yahoo Example

Figure 6.2

where θ_0^* is a common bias weight, θ_j^* is a weight for the j–th slot, and $\theta_{j,k}^*$ is a weight for the interaction between the content in the k–th and ℓ–th slots. We also include known parameters $\alpha_1 = 1$ and $\alpha_2 = 0.5$ that control the strength of the first and second order interactions respectively. The weights of the parameter vector are drawn from a discrete uniform distribution with a range of $[-0.3, 0.3]$ and a granularity of 0.01. The results of this example are shown in Fig. 6.2a. The RAGE algorithm performs close to the oracle on this example, while the sample complexity of the rest of the algorithms grows as the number of arms and dimension of the problem goes up.

Click-through example. To conduct an experiment based on real data, we build a problem using the Yahoo! Webscope Dataset R6A. The dataset contains user click log records for news articles displayed uniformly at random on the Yahoo! front page between May 1st, 2009 and May 10th, 2009. Each click log record consists of a binary outcome along with 6 features identifying the user and 6 features identifying the article.

To build a linear bandit problem from the dataset, we construct an arm set $\mathcal{X} = \mathcal{Z} \subset \mathbb{R}^{36}$ by taking the outer product of the user and article features for each click log record on May 1st, 2009. We then fit a regularized least squares estimate using a regularization parameter of 0.01 to obtain θ^*. To model binary rewards, we let the observed reward be generated by a draw of a Bernoulli random variable with parameter $x^\top \theta^*$ for any arm selection $x \in \mathcal{X}$. Since $x^\top \theta^* \in (0, 0.11)\ \forall\ x \in \mathcal{X}$, the noise is bounded between $[-1, 1]$, which causes it to be 1-sub-Gaussian. We simulate the problem with 40 arms including the arm with the maximum reward in the dataset and the remaining arms were selected at random from the set of arms with gap at least 0.01 from the optimal arm so the problem is not too hard. The experiment setup is similar to that from Xu et al. (2018) for this dataset. The results are presented in Fig. 6.2b. We see that the RAGE algorithm has good performance on this example based on real world data.

6.6 Discussion

In this chapter, we have proposed the problem of best-arm identification for transductive linear bandits, provided an algorithm, and matching upper and lower bounds. The results are a key development of extending the class of problems with near-optimal or optimal guarantees for best-

arm identification. Indeed, the pure exploration multi-armed bandit problem is now fairly well-understood, but despite several attempts, developing a provably near-optimal algorithm for the more general linear bandit problem has been elusive. In this work, we showed that a variant of the successive elimination algorithm that is standard and near-optimal in multi-armed bandits, can be extended to a near-optimal algorithm for linear bandits and more generally transductive linear bandits. As is highlighted by the algorithm and the overall approach, the key technique is exploiting the linear structure of the feedback model. Perhaps the most important open question relating to this work is the $\log(1/\Delta_{\min})$ factor in the sample complexity. While we suspect this is an artifact of the analysis, showing it is not needed would be a significant advancement.

Since the results in this chapter appeared, there has been a surge of work on the topic and related problems and significant progress has been made. Sticking with the pure exploration linear bandit problem, asymptotically optimal ($\delta \to 0$) algorithms have been developed (Degenne et al., 2020a; Jedra and Proutiere, 2020), algorithms for the fixed budget exploration problem (Alieva et al., 2021; Katz-Samuels et al., 2020), and others (Zaki et al., 2019, 2020). It is worth pointing out that a key development of Katz-Samuels et al. (2020) is removing the need for an explicit union bound. Moreover, Degenne et al. (2020a); Zaki et al. (2020) provide an interesting game-theoretic perspective based around employing no-regret learning algorithms. There has also been extensions to top-k arm identification (Réda et al., 2021), generalized linear bandits (Jun et al., 2021), along with kernel and neural bandits (Zaki et al., 2020).

Finally, we remark that the experimental design techniques in this chapter are similarly relevant to the instance-dependent regret problem in linear bandits and contextual bandits. Notably, to obtain optimal instance-dependent regret asymptotically, it is necessary to sometimes sample high-regret actions in order to maximize the amount of information that is obtained (Lattimore and Szepesvari, 2017). There has been recent progress on the regret problem using methods similar to that from pure exploration linear bandits (Degenne et al., 2020b; Jun and Zhang, 2020; Tirinzoni et al., 2020; Wagenmaker et al., 2021). However, room for improvement remains. We also note that developing pure exploration methods for contextual linear bandits is still an open area of research in which techniques from this chapter are relevant.

CHAPTER 6 APPENDIX

6.A Proof of Theorem 6.1

In this section we assume $\mathcal{X} = \{x_1, \cdots, x_n\}$ and $\mathcal{Z} = \{z_1, \cdots, z_m\}$. Without loss of generality, we assume that $z_1 = \arg\max_{z_i \in \mathcal{Z}} z_i^\top \theta^*$. Let $\mathcal{C} := \{\theta \in \mathbb{R}^d : \exists i \text{ s.t. } \theta^\top (z_1 - z_i) \leq 0\}$, i.e. $\theta \in \mathcal{C}$ if and only if z_1 is not the best arm in the linear bandit instance $(\mathcal{X}, \mathcal{Z}, \theta)$.

We now recall the transportation lemma of (Kaufmann et al., 2016). Under a δ-PAC strategy for finding the best arm for the bandit instance $(\mathcal{X}, \mathcal{Z}, \theta^*)$, let T_i denote the random variable which is the number of times arm i is pulled. In addition let $\nu_{\theta,i}$ denote the reward distribution of the i-th arm of $\mathcal{X}$, i.e. $\nu_{\theta,i} = \mathcal{N}(x_i^\top \theta, 1)$. Then for any $\theta \in \mathcal{C}$ we have that

$$\sum_{i=1}^{n} \mathbb{E}[T_i] KL(\nu_{\theta^*,i}, \nu_{\theta,i}) \geq \log(1/2.4\delta).$$

In particular, $\sum_{i=1}^n \mathbb{E}[T_i] \geq \sum_{i=1}^n t_i$ for any $\mathbf{t} := (t_1, \cdots, t_n)$ which is a feasible solution of the optimization problem,

$$\min \sum_{i=1}^{n} t_i \quad \text{subject to } \min_{\theta \in \mathcal{C}} \sum_{i=1}^{n} t_i KL(\nu_{\theta^*,i} || \nu_{\theta,i}) \geq \log(1/2.4\delta).$$

Taking $\mathbf{t}^*$ to be an optimal solution to the previous problem, note that

$$\min_{\theta \in \mathcal{C}} \sum_{i=1}^{n} \frac{t_i^*}{\sum_{j=1}^n t_j^*} KL(\nu_{\theta^*,i} || \nu_{\theta,i}) \geq \frac{\log(1/2.4\delta)}{\sum_{j=1}^n t_j^*} \geq \frac{\log(1/2.4\delta)}{\sum_{j=1}^n \mathbb{E}[T_j]}$$

In particular, since $\sum_{i=1}^n \frac{t_i^*}{\sum_{j=1}^n t_j^*} = 1$, we see that

$$\max_{\lambda \in \triangle_n} \min_{\theta \in \mathcal{C}} \sum_{i=1}^{n} \lambda_i KL(\nu_{\theta^*,i} || \nu_{\theta,i}) \geq \frac{\log(1/2.4\delta)}{\sum_{i=1}^n \mathbb{E}[T_i]}.$$

Rearranging, we see that

$$\sum_{i=1}^{n} \mathbb{E}[T_i] \geq \log(1/2.4\delta) \min_{\lambda \in \triangle_n} \max_{\theta \in \mathcal{C}} \frac{1}{\sum_{i=1}^n \lambda_i KL(\nu_{\theta^*,i} || \nu_{\theta,i})}. \tag{6.6}$$

Now for $j \neq 1$, $\lambda \in \triangle^n$ and $\epsilon > 0$, define

$$\theta_j(\epsilon, \lambda) = \theta^* - \frac{(y_j^\top \theta^* + \epsilon) A(\lambda)^{-1} y_j}{y_j^\top A(\lambda)^{-1} y_j}.$$

where $A(\lambda) := \sum_{i=1}^n \lambda_i x_i x_i^\top$ and $y_j = z_1 - z_j$. Note that $y_j^\top \theta_j(\epsilon, \lambda) = -\epsilon < 0$ which implies that $\theta_j \in \mathcal{C}$. Also, the KL-divergence is given by

$$\begin{aligned} KL(\nu_{\theta^*,i} || \nu_{\theta_j(\epsilon,\lambda),i}) &= (x_i^\top(\theta^* - \theta_j(\epsilon,\lambda)))^2 \\ &= y_j^\top A(\lambda)^{-1} \frac{(y_j^\top \theta^* + \epsilon)^2 x_i x_i^\top}{(y_j^\top A(\lambda)^{-1} y_j)^2} A(\lambda)^{-1} y_j. \end{aligned}$$

Hence, returning to (6.6), we have that

$$\begin{aligned} \sum_{i=1}^n \mathbb{E}[T_i] &\geq \log(1/2.4\delta) \min_{\lambda \in \triangle_n} \max_{\theta \in \mathcal{C}} \frac{1}{\sum_{i=1}^n \lambda_i KL(\nu_{\theta^*} || \nu_\theta)} \\ &\geq \log(1/2.4\delta) \min_{\lambda \in \triangle_n} \max_{j=2,\cdots,m} \frac{1}{\sum_{i=1}^n \lambda_i KL(\nu_{\theta^*,i} || \nu_{\theta_j(\epsilon,\lambda),i})} \\ &\geq \log(1/2.4\delta) \min_{\lambda \in \triangle_n} \max_{j=2,\cdots,m} \frac{(y_j^\top A(\lambda)^{-1} y_j)^2}{(y_j^\top \theta^* + \epsilon)^2 y_j^\top A(\lambda)^{-1} (\sum_{i=1}^n \lambda_i x_i x_i^\top) A(\lambda)^{-1} y_j} \\ &= \log(1/2.4\delta) \min_{\lambda \in \triangle_n} \max_{y \in \mathcal{Y}^*(\mathcal{Z})} \frac{y_j^\top A(\lambda)^{-1} y_j}{(y_j^\top \theta^* + \epsilon)^2} \end{aligned}$$

where in the second to last line we used the fact that $\sum_{i=1}^n \lambda_i x_i x_i^\top = A(\lambda)$. Letting $\epsilon \to 0$ establishes the result.

Remark: Note that $\theta_j = \arg\min_{\theta \in \mathbb{R}^d} \|\theta - \theta^*\|^2_{A(\lambda)}$ subject to $y_j^\top \theta = -\epsilon$.

6.B Proof of Proposition 6.1

Assume d is even and each $\epsilon_t \sim \mathcal{N}(0,1)$. Fix some $\alpha \in (0,1)$ which will depend on γ in a clear way momentarily, and consider an instance where $\mathcal{X} = \mathcal{Z} = \{e_i\}_{i=1}^{d/2} \cup \{\cos(\alpha) e_i + \sin(\alpha) e_{d/2+i}\}_{i=1}^{d/2}\}$ where e_i is the i-th standard basis vector.

If an algorithm is δ-PAC, and takes N_i samples from arm i, then for any $j \leq d/2$ it will be able to distinguish between $\theta = z_j$ and $\theta = z_{j+d/2}$. By standard Le Cam arguments (Tsybakov, 2004) this hypothesis test requires $N_j + N_{j+d/2} \geq \frac{c \log(1/\delta)}{(1-\cos(\alpha))^2}$ for some universal constant $c > 0$. Because $(1 - \cos(\alpha))^2 \approx \alpha^4/4$ and these inequalities must hold for all $j = 1, \ldots, d/2$ simultaneously for the single static allocation, we obtain the result.

6.C Proof of Theorem 6.2

Let the good event for the t–th round of Algorithm 6.1 be

$$\mathcal{E}_t := \left\{ N_t \leq \max\{ \lceil 2(2^t)^2 \rho(\mathcal{Y}(\mathcal{S}_t))(1+\epsilon) \log\left(\tfrac{|\mathcal{Z}|^2}{\delta_t}\right) \rceil, r(\epsilon) \} \right\} \cap \{ z_* \in \widehat{\mathcal{Z}}_{t+1} \} \cap \{ \widehat{\mathcal{Z}}_{t+1} \subseteq \mathcal{S}_{t+1} \}$$

where we recall that $\mathcal{S}_t = \{z \in \mathcal{Z} : \Delta(z) \leq 4 \cdot 2^{-t}\}$. The good event characterizes the worst-case sample complexity of the t–th phase of Algorithm 6.1 and guarantees that the set of active arms at

the end of the phase contains the optimal arm and it is contained in the set of arms with gaps below the threshold to be eliminated in the phase. Note that for $t > \log_2(4/\Delta_{\min})$ we have $S_t = \{z_*\}$.

The proof proceeds as follows. We begin by showing that the good event holds with probability at least $1 - \delta_t$ in phase t given that the good event held in the previous phases. We then show that the probability of the good event holding in every phase is at least $1 - \delta$. As a result, we simply sum over the bound on the sample complexity in each phase given in the good event to obtain the stated bound on the sample complexity.

The following lemma shows that good event holds in phase t with probability at least $1 - \delta_t$ conditioned on the good event holding in the previous phases.

Lemma 6.2. $\mathbb{P}(\mathcal{E}_t|\mathcal{E}_{t-1}, \cdots, \mathcal{E}_1) \geq 1 - \delta_t$.

Proof. We begin by bounding the confidence width with high probability.

Step 1: Confidence Bound Width. Conditioned on a choice of $\mathcal{Y}(\widehat{\mathcal{Z}}_t)$, since $\widehat{\theta}$ is a least squares estimator of θ^* and the noise is i.i.d., we know that $y^\top(\theta^* - \widehat{\theta}_t)$ is $\|y\|^2_{A_t^{-1}}$-subGaussian for all $y \in \mathcal{Y}(\widehat{\mathcal{Z}}_t)$. Furthermore, due to the guarantees of the rounding procedure, $\|y\|^2_{A_t^{-1}} \leq (1+\epsilon)\rho(\mathcal{Y}(\widehat{\mathcal{Z}}_t))/N_t \leq \left(2(2^t)^2 \log(|\mathcal{Z}|^2/\delta_t)\right)^{-1}$ for all $y \in \mathcal{Y}(\widehat{\mathcal{Z}}_t)$ by our choice of N_t. Since the right-hand side is deterministic, independent of $\mathcal{Y}(\widehat{\mathcal{Z}}_t)$, for any $\nu > 0$, we have that

$$\mathbb{P}\left(|y^\top(\theta^* - \widehat{\theta})| > \sqrt{\frac{2\log(2/\nu)}{2(2^t)^2\log(|\mathcal{Z}|^2/\delta_t)}}\middle| \mathcal{E}_{t-1}, \cdots, \mathcal{E}_1\right) \leq \nu$$

for any $y \in \mathcal{Y}(\widehat{\mathcal{Z}}_t)$. Taking $\nu = 2\delta_t/|\mathcal{Z}|^2$ and union bounding over all the possible $y \in \mathcal{Y}(\widehat{\mathcal{Z}}_t)$ where $|\mathcal{Y}(\widehat{\mathcal{Z}}_t)| \leq |\mathcal{Y}(\mathcal{Z})| \leq |\mathcal{Z}|^2/2$, gives us that

$$\mathbb{P}(\exists y \in \mathcal{Y}(\widehat{\mathcal{Z}}_t) \quad |y^\top(\theta^* - \widehat{\theta})| > 2^{-t}|\mathcal{E}_{t-1}, \cdots, \mathcal{E}_1) \leq \delta_t. \tag{6.7}$$

Step 2: Suboptimal Arms are Eliminated. We claim that every arm $z \in \widehat{\mathcal{Z}}_t$ such that $\Delta(z) > 2^{-(t-1)}$ is discarded in phase t so that $\widehat{\mathcal{Z}}_{t+1} \subseteq \mathcal{S}_{t+1}$ with probability at least $1 - \delta_t$.

Indeed, since we conditioned on $\mathcal{E}_{t-1}$, $z_* \in \widehat{\mathcal{Z}}_t$. If $z \in \mathcal{S}_{t+1}^c \cap \widehat{\mathcal{Z}}_t$ then by definition $\Delta(z) = (z_* - z)^\top\theta^* > 2^{-(t-1)}$. Taking $y = z_* - z$, we know that a) $y^\top\theta^* > 2^{-(t-1)}$, and b) from the confidence bound $\|y\|_{A_t^{-1}}\sqrt{2\log(|\mathcal{Z}|^2/\delta_t)} \leq 2^{-t}$. Hence,

$$\begin{aligned}
y^\top\widehat{\theta}_t &\geq y^\top\theta^* - \|y\|_{A_t^{-1}}\sqrt{2\log(|\mathcal{Z}|^2/\delta_t)} \\
&\overset{a)}{>} 2^{-(t-1)} - \|y\|_{A_t^{-1}}\sqrt{2\log(|\mathcal{Z}|^2/\delta_t)} \\
&\overset{b)}{\geq} 2^{-(t-1)} - 2^{-t} \\
&= 2^{-t} \overset{b)}{\geq} \|y\|_{A_t^{-1}}\sqrt{2\log(|\mathcal{Z}|^2/\delta_t)}
\end{aligned}$$

However, this is precisely the discard condition of the algorithm guaranteeing z will be eliminated. Thus, this proves the claim.

Step 3: The Optimal Arm is not Eliminated. We claim $z_* \in \widehat{\mathcal{Z}}_{t+1}$ with probability at least $1-\delta_t$. We prove this claim by contradiction. To begin, observe that z_* is in $\widehat{\mathcal{Z}}_t$ since $\mathcal{E}_{t-1}$ holds. Now, suppose that z_* is discarded in phase t. This implies that there exists a $z \neq z_*$ for $z \in \widehat{\mathcal{Z}}_t$ such that $\|z - z^*\|_{A_t^{-1}} \sqrt{2\log(|\mathcal{Z}^2|/\delta_t)} < (z - z_*)^\top \widehat{\theta}_t$. However from the confidence interval (6.7), $(z - z_*)^\top (\widehat{\theta}_t - \theta^*) \leq \|z - z^*\|_{A_t^{-1}} \sqrt{2\log(|\mathcal{Z}^2|/\delta_t)}$. Combining these we see that $(z - z_*)^\top (\widehat{\theta}_t - \theta^*) < (z - z_*)^\top \widehat{\theta}_t$ which implies $(z - z_*)^\top \theta^* > 0$ which is a contradiction, so that the claim holds.

Step 4: Sample Complexity. We complete the proof by showing that the sample complexity of phase t given in the good event holds with probability $1-\delta_t$. Since $\mathcal{E}_{t-1}$ is given, $\widehat{\mathcal{Z}}_t \subseteq \mathcal{S}_t$, which implies with probability at least $1-\delta_t$,

$$\begin{aligned} N_t &= \max\left\{\left\lceil 2(2^t)^2 \rho(\mathcal{Y}(\widehat{\mathcal{Z}}_t))(1+\varepsilon)\log(|\mathcal{Z}|^2/\delta_t)\right\rceil, r(\epsilon)\right\} \\ &\leq \max\left\{\left\lceil 2(2^t)^2 \rho(\mathcal{Y}(\mathcal{S}_t))(1+\varepsilon)\log(|\mathcal{Z}|^2/\delta_t)\right\rceil, r(\epsilon)\right\} \end{aligned}$$

where we note that the quantity on the right hand side is deterministic. Combining the conclusions of each step, proves the result. □

We now show that the good event holds at each round with probability at least $1-\delta$, which in turn proves the correctness of the algorithm, meaning that the optimal arm is identified with probability at least $1-\delta$.

Lemma 6.3. $\mathbb{P}(\mathcal{E}_1 \cap \cdots \cap \mathcal{E}_{\lfloor \log_2(4/\Delta_{\min})\rfloor}) \geq 1-\delta$.

Proof. Let us first expand the intersection of the events into a product of conditional probabilities as follows:

$$\mathbb{P}(\mathcal{E}_1 \cap \cdots \cap \mathcal{E}_{\lfloor \log_2(4/\Delta_{\min})\rfloor}) = \Pi_{t=1}^{\lfloor \log_2(4/\Delta_{\min})\rfloor} \mathbb{P}(\mathcal{E}_t | \mathcal{E}_{t-1} \cap \cdots \cap \mathcal{E}_1)$$

We now obtain a lower bound on the success probability using Lemma 6.2 and facts about infinite products:

$$\Pi_{t=1}^{\lfloor \log_2(4/\Delta_{\min})\rfloor} \mathbb{P}(\mathcal{E}_t | \mathcal{E}_{t-1} \cap \cdots \cap \mathcal{E}_1) \geq \Pi_{t=1}^{\lfloor \log_2(4/\Delta_{\min})\rfloor} (1-\delta_t) \geq \Pi_{t=1}^{\infty} \Big(1 - \frac{\delta}{t^2}\Big) = \frac{\sin(\pi\delta)}{\pi\delta}.$$

Finally, using the fact that $\frac{\sin(\pi\delta)}{\pi\delta} \geq 1-\delta$ for $\delta \in (0,1)$, we obtain the result $\mathbb{P}(\mathcal{E}_1 \cap \cdots \cap \mathcal{E}_{\lfloor \log_2(4/\Delta_{\min})\rfloor}) \geq 1-\delta$. □

The sample complexity now follows immediately from Lemmas 6.2 and 6.3 since we can now sum the number of samples taken in each phase. However, a key novelty of this proof is the the following method to quantify the relationship between the algorithm sample complexity and the lower bound. With probability at least $1-\delta$,

$$\begin{aligned} N &\leq \sum_{t=1}^{\lfloor \log_2(4/\Delta_{\min})\rfloor} \max\left\{\left\lceil 2(2^t)^2 \rho(\mathcal{Y}(\mathcal{S}_t))(1+\epsilon)\log(t^2|\mathcal{Z}|^2/\delta)\right\rceil, r(\epsilon)\right\} \\ &\leq 128\psi^*(1+\epsilon)\log_2(4/\Delta_{\min})\log(\log_2(4/\Delta_{\min})^2|\mathcal{Z}|^2/\delta) + (1+r(\epsilon))\log_2(4/\Delta_{\min}). \end{aligned}$$

Recall that $\mathcal{Y}^*(\mathcal{S}) = \{z_* - z : \forall\ z \in \mathcal{S} \setminus z_*\}$. To see the second inequality, note that

$$\begin{aligned}
\psi^* &= \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{y \in \mathcal{Y}^*(\mathcal{Z})} \frac{\|y\|^2_{(\sum_{x\in\mathcal{X}} \lambda_x x x^\top)^{-1}}}{\Delta(y)^2} \\
&= \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{t \le \lfloor \log_2(4/\Delta_{\min})\rfloor} \max_{y \in \mathcal{Y}^*(\mathcal{S}_t)} \frac{\|y\|^2_{(\sum_{x\in\mathcal{X}} \lambda_x x x^\top)^{-1}}}{\Delta(y)^2} \\
&\ge \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{t \le \lfloor \log_2(4/\Delta_{\min})\rfloor} \max_{y \in \mathcal{Y}^*(\mathcal{S}_t)} \frac{\|y\|^2_{(\sum_{x\in\mathcal{X}} \lambda_x x x^\top)^{-1}}}{(4 \cdot 2^{-t})^2} \\
&\overset{(i)}{\ge} \frac{1}{16 \log_2(1/\Delta_{\min})} \min_{\lambda \in \triangle_{\mathcal{X}}} \sum_{t=1}^{\lfloor \log_2(4/\Delta_{\min})\rfloor} (2^t)^2 \max_{y \in \mathcal{Y}^*(\mathcal{S}_t)} \|y\|^2_{(\sum_{x\in\mathcal{X}} \lambda_x x x^\top)^{-1}} \\
&\overset{(ii)}{\ge} \frac{1}{16 \log_2(4/\Delta_{\min})} \sum_{t=1}^{\lfloor \log_2(4/\Delta_{\min})\rfloor} (2^t)^2 \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{y \in \mathcal{Y}^*(\mathcal{S}_t)} \|y\|^2_{(\sum_{x\in\mathcal{X}} \lambda_x x x^\top)^{-1}} \\
&\overset{(iii)}{\ge} \frac{1}{64 \log_2(4/\Delta_{\min})} \sum_{t=1}^{\lfloor \log_2(4/\Delta_{\min})\rfloor} (2^t)^2 \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{y \in \mathcal{Y}(\mathcal{S}_t)} \|y\|^2_{(\sum_{x\in\mathcal{X}} \lambda_x x x^\top)^{-1}} \\
&= \frac{1}{64 \log_2(4/\Delta_{\min})} \sum_{i=1}^{\lfloor \log_2(4/\Delta_{\min})\rfloor} (2^t)^2 \rho(\mathcal{Y}(S_t))
\end{aligned}$$

where (i) follows from the fact that the maximum of positive numbers is always greater than the average, and (ii) by the fact that the minimum of a sum is greater than the sum of minimums. To see (iii), note that for $y \in \mathcal{Y}(\mathcal{S}_t)$, if $y = z_i - z_j$, then $y = (z_* - z_j) - (z_* - z_i)$. Hence $\max_{y\in\mathcal{Y}(\mathcal{S}_t)} \|y\|^2_{(\sum_{x\in\mathcal{X}} \lambda_x x x^\top)^{-1}} \le 4 \max_{y\in\mathcal{Y}^*(\mathcal{S}_t)} \|y\|^2_{(\sum_{x\in\mathcal{X}} \lambda_x x x^\top)^{-1}}$. This completes the proof.

6.D Efficient Rounding Procedures

Throughout the following we assume that $\mathcal{Y} \subset \mathbb{R}^d$ is arbitrary and that $\mathcal{X} = \{x_1, \cdots, x_n\} \subset \mathbb{R}^d$ is a subset with $\dim \operatorname{span}(\mathcal{X}) = d$.

Definition 6.2. *A* rounding procedure *is an algorithm that takes as input $\lambda \in \triangle^n$, a set of vectors $\mathcal{X}$, and a number of samples N and returns a finite* ***allocation*** *$s = (s_1, \cdots, s_n) \in \mathbb{N}^n$ satisfying the following properties: 1. $\sum_{i=1}^n s_i = N$; 2. there exists a function $r(\epsilon)$ such that if $N > r(\epsilon)$, then $\max_{y\in\mathcal{Y}} \|y\|^2_{(\sum_{i=1}^n s_i x_i x_i^\top)^{-1}} \le (1+\epsilon) \max_{y\in\mathcal{Y}} \|y\|^2_{(\sum_{i=1}^n \lambda_i x_i x_i^\top)^{-1}}/N$.*

Fortunately, there has been extensive work on efficient rounding procedures, motivated by the strong connection to G-optimal design in optimal linear experimental design (Pukelsheim, 2006). Here we discuss two important rounding procedures. The first is due to (Pukelsheim, 2006) and has an $r(\epsilon) = 2p/\epsilon \le (d(d+1)+2)/\epsilon$ where p is the support size of λ.

Rounding Procedure of Pukelsheim (2006). An efficient rounding procedure is given in Chapter 12 of (Pukelsheim, 2006) to transform a design $\lambda \in \triangle^n$ into a discrete allocation $s \in \mathbb{N}^n$ for any fixed number of samples N. The rounding procedure determines the number of pulls N_i to

allocate to each arm x_i in the support of λ such that $\sum_{i\leq p} N_i = N$ where p is the cardinality of the support of λ. The discrete allocation from the rounding procedure is obtained in two phases:

1. Given the number of samples N to obtain and the cardinality of the support of λ, samples to allocate to arms in the support of λ are computed using $N_i = \lceil (N - \frac{1}{2}p)\lambda_i \rceil$, where $N_1, N_2, \ldots, N_p$ are positive integers constrained such that $\sum_{i\leq p} N_i \geq N$.

2. Following the previous phase of the rounding procedure, loop until the discrepancy $(\sum_{i\leq p} N_i) - N = 0$, from either increasing a sample count N_j which obtains $N_j/\lambda_j = \min_{i\leq p} N_i/\lambda_i$ to $N_j + 1$, or decreasing a sample count N_j which obtains $(N_j - 1)/\lambda_j = \max_{i\leq p}(N_i - 1)/\lambda_i$ to $N_j - 1$.

The efficient design apportionment theorem in Section 12.5 of (Pukelsheim, 2006) provides the foundation the procedure. We now provide some details on the efficiency of the procedure.

Let $\gamma = s/N$ represent the fractional allocation corresponding to a finite allocation s satisfying the properties in Definition 6.2 and obtained from applying the efficient rounding procedure to the distribution λ. Moreover, define $\upsilon(\mathcal{Y}) := \max_{y\in\mathcal{Y}} \|y\|^2_{(\sum_{x\in\mathcal{X}} \gamma_x xx^\top)^{-1}}$.

Proposition 6.2. *The efficient rounding procedure of (Pukelsheim, 2006) guarantees for $N \geq 2p$,*

$$\rho(\mathcal{Y}) \leq \upsilon(\mathcal{Y}) \leq \Big(1 + \frac{2p}{N}\Big)\rho(\mathcal{Y}).$$

Moreover, when $dim(span(\mathcal{X})) = d$ and $N \geq d^2 + d + 2$,

$$\rho(\mathcal{Y}) \leq \upsilon(\mathcal{Y}) \leq \Big(1 + \frac{d^2 + d + 2}{N}\Big)\rho(\mathcal{Y}).$$

Proof. Define the minimum likelihood ratio of γ relative to λ as

$$\zeta(\gamma, \lambda) = \min_{x\in \mathrm{supp}(\lambda)} \frac{\gamma_x}{\lambda_x} = \max\{\kappa \geq 0 : \gamma_x \geq \kappa\lambda_x \text{ for all } x \in \mathcal{X}\}.$$

Observe that $\zeta(\gamma, \lambda) \in [0, 1]$ by definition. As an immediate consequence of this definition, we obtain

$$\sum_{x\in\mathcal{X}} \gamma_x xx^\top \geq \sum_{x\in\mathcal{X}} \zeta(\gamma, \lambda_x)\lambda_x xx^\top \iff \Big(\sum_{x\in\mathcal{X}} \gamma_x xx^\top\Big)^{-1} \leq \frac{1}{\zeta(\gamma, \lambda)}\Big(\sum_{x\in\mathcal{X}} \lambda_x xx^\top\Big)^{-1}.$$

It follows that for all $y \in \mathcal{Y}$,

$$y^\top\Big(\sum_{x\in\mathcal{X}} \gamma_x xx^\top\Big)^{-1} y \leq \frac{1}{\zeta(\gamma, \lambda)} y^\top\Big(\sum_{x\in\mathcal{X}} \lambda_x xx^\top\Big)^{-1} y,$$

which implies $\upsilon(\mathcal{Y}) \leq \rho(\mathcal{Y})/\zeta(\gamma, \lambda)$. Since the design γ is obtained from an efficient design apportionment, Theorem 12.7 of (Pukelsheim, 2006) indicates it obtains the best efficiency bound among all standardized designs of sample size N. As a result, Lemma 12.8 from (Pukelsheim, 2006) holds

and guarantees that for $N \geq 2p$,

$$\frac{1}{\zeta(\gamma,\lambda)} \leq \frac{1}{1-\frac{p}{N}} \leq 1 + \frac{2p}{N}.$$

Hence, for $N \geq 2p$

$$\rho(\mathcal{Y}) \leq \upsilon(\mathcal{Y}) \leq \Big(1 + \frac{2p}{N}\Big)\rho(\mathcal{Y}).$$

When $\dim(\text{span}(\mathcal{X})) = d$, Caratheodory's theorem indicates that $p \leq d(d+1)/2+1$. Consequently, when $N \geq d^2 + d + 2$, we get

$$\rho(\mathcal{Y}) \leq \upsilon(\mathcal{Y}) \leq \Big(1 + \frac{d^2+d+2}{N}\Big)\rho(\mathcal{Y}).$$

□

6.D.0.0.1 Rounding Procedure of Allen-Zhu et al. (2021). We refer the reader to Algorithm 1 in (Allen-Zhu et al., 2021) for details about their rounding procedure. Here we describe their result and how to modify it to our setting. Let $\mathcal{S}_{b,k} = \{s \in [b]^n : \sum_{i=1}^n s_i \leq k\}$ and a continuous relaxation $\mathcal{C}_{b,k} = \{s \in [0,b]^n : \sum_{i=1}^n s_i \leq k\}$.

Theorem 6.3 (Theorem 2.1 of (Allen-Zhu et al., 2021))**.** *Suppose $\epsilon \in (0, 1/3]$, $n \geq k \geq 180d/\epsilon^2$, $b \in [k]$. Let $\pi \in C_{b,k}$, then in polynomial-time (in n and d) we can round π to an integral solution $\widehat{s} \in S_{b,k}$ satisfying* $\max_{y\in\mathcal{Y}} \|y\|^2_{(\sum \widehat{s}_i x_i x_i^\top)^{-1}} \leq (1+\epsilon)\max_{y\in\mathcal{Y}} \|y\|^2_{(\sum \pi_i x_i x_i^\top)^{-1}}$.

To apply this theorem to obtain an efficient rounding procedure, consider the following. Given a $\lambda \in \triangle_{\mathcal{X}}$, and a number of samples N, let $\pi = N\lambda$ and consider the case where $b = k = N$. Then $k\lambda \in C_{k,k}$. In general the theorem does not allow $N = k > n$, but we can circumvent this by just duplicating each vector in $\mathcal{X}$ exactly N times. Then the allocation $\widehat{s}$ obtained will satisfy the conditions of the above with $r(\epsilon) = 180d/\epsilon^2$. The authors remark that it is most likely true that $r(\epsilon) = d/\epsilon^2$ suffices, but we are not aware of any such result in the literature.

6.E Proof of Lemma 6.1

Consider the following:

$$\begin{aligned}
\rho(\mathcal{Y}) &= \min_{\lambda\in\triangle_{|\mathcal{X}|}} \max_{y\in\mathcal{Y}} \|y\|^2_{(\sum_{x\in\mathcal{X}} \lambda_x x x^\top)^{-1}} \\
&= \frac{1}{\gamma_{\mathcal{Y}}^2} \min_{\lambda\in\triangle_{|\mathcal{X}|}} \max_{y\in\mathcal{Y}} \|y\gamma_{\mathcal{Y}}\|^2_{(\sum_{x\in\mathcal{X}} \lambda_x x x^\top)^{-1}} \\
&\leq \frac{1}{\gamma_{\mathcal{Y}}^2} \min_{\lambda\in\triangle_{|\mathcal{X}|}} \max_{x\in\text{conv}(\mathcal{X}\cup-\mathcal{X})} \|x\|^2_{(\sum_{x\in\mathcal{X}} \lambda_x x x^\top)^{-1}} \\
&= \frac{1}{\gamma_{\mathcal{Y}}^2} \min_{\lambda\in\triangle_{|\mathcal{X}|}} \max_{x\in\mathcal{X}} \|x\|^2_{(\sum_{x\in\mathcal{X}} \lambda_x x x^\top)^{-1}}
\end{aligned}$$

The third equality follows from the fact that the maximum value of a convex function on a convex set must occur at a vertex. By the celebrated Kiefer-Wolfowitz theorem for G-optimal design

(Pukelsheim, 2006), $\min_{\lambda \in \triangle_{|\mathcal{X}|}} \max_{x \in \mathcal{X}} \|x\|^2_{(\sum \lambda_x x x^\top)^{-1}} = d$ so we see that $\rho(\mathcal{Y}) \leq d/\gamma^2_{\mathcal{Y}}$. For a lower bound, note that

$$\begin{aligned}\min_{\lambda \in \triangle_{\mathcal{X}}} \max_{y \in \mathcal{Y}} \|y\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x x x^\top)^{-1}} &\geq \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{y \in \mathcal{Y}} \sigma_{\min}((\sum_{x \in \mathcal{X}} \lambda_x x x^\top)^{-1}) \|y\|_2^2 \\ &= \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{y \in \mathcal{Y}} \|y\|_2^2 / \sigma_{\max}((\sum_{x \in \mathcal{X}} \lambda_x x x^\top)^{-1})\end{aligned}$$

where $\sigma_{\max}$ and $\sigma_{\min}$ are respectively the largest and smallest eigenvalue operators of a matrix. Since $\sigma_{\max}(\sum_{i=1}^n \lambda_i x_i x_i^\top) \leq \max_{x \in \mathcal{X}} \|x\|_2$, we have that $\rho(\mathcal{Y}(S_t)) \geq \max_{y \in \mathcal{Y}(S_t)} \|y\|_2^2 / (\max_{x \in \mathcal{X}} \|x\|_2)$. The final statement in the case of a singleton is also known as Elfving's Theorem, see Section 2.14 in (Pukelsheim, 2006)

6.F Experiment Details

In this section, we provide further details on the implementation of each algorithm. Each experiment was repeated 20 times with the mean sample complexity is reported and error bars representing the standard error are plotted. Simulations were implemented in Python 3 and parallelized on an Intel(R) Xeon(R) CPU E5-2690. For each algorithm that requires computing a design λ from an optimization of the form $\min_{\lambda \in \triangle_{\mathcal{X}}} \max_{s \in \mathcal{S}} \|s\|^2_{(\sum_x \lambda_x x x^\top)^{-1}}$ for $\mathcal{S} \subset \mathbb{R}^d$ (RAGE, $\mathcal{XY}$-static, $\mathcal{XY}$-oracle, and ALBA) we used a Franke-Wolfe algorithm (Jaggi, 2013) with constant step-size $2/(k+2)$ (k being the iteration counter). The algorithm was run until the relative change in λ with respect to the ℓ_2 norm was less than .01 or 5000 iterations were reached. Any values of $\lambda < 10^{-5}$ were then thresholded to 0 and λ was scaled to sum to 1. We now provide further details on each implementation.

- $\mathcal{XY}$-Adaptive (Soare et al., 2014): This algorithm requires a parameter α that governs the length of each adaptive phase. We follow the simulations in (Soare et al., 2014) and let $\alpha = 0.1$. We remark that the algorithm given in the paper implements a greedy update to select arms in contrast to rounding the optimal allocation as is considered in the analysis. We implement the greedy arm selection procedure to match the simulations in the paper. It is worth noting that in several of the recent linear bandit papers that have implemented this algorithm, the active arm set has been reset at the conclusion of a phase before discarding arms. We do not reset the arm set at the conclusion of a phase to match what was done in (Soare et al., 2014). Finally, in the confidence interval, we include the phase index and not the number of samples since we only need to union bound over when it is evaluated.

- $\mathcal{XY}$-Static and $\mathcal{XY}$-Oracle: To implement each allocation, we compute the optimal design on the set $\mathcal{Y}(\mathcal{Z})$ for the static strategy and the set $\mathcal{Y}^*(\mathcal{Z})$ normalized by the gaps for the oracle. Each algorithm is ran in phases in which γ^t samples are drawn from the allocation. We experimented using γ in the range $(1, 2)$ to optimize the performance of the algorithms and ended up using $\gamma = 1.1$ for the oracle strategy and $\gamma = 1.35$ for the static strategy across the examples. The stopping condition $\left\{\exists z' \in \mathcal{Z} | \forall\ z \in \mathcal{Z} : \|z' - z\|_{A_t^{-1}} \sqrt{2 \log(2t^2 |\mathcal{Z}|^2/\delta)} \leq (z' - z)^\top \widehat{\theta}_t\right\}$ is evaluated

at the end of each phase t to decide when to terminate the experiment for each algorithm, but only union bounding over $|\mathcal{Z}|$ for the oracle.

- LinGapE (Xu et al., 2018): We run this algorithm with a regularizer on the least squares estimator of $\lambda = 1$ following the implementation given in the paper. LinGapE is designed to find an ε good arm. We let $\varepsilon = 0$ to ensure the optimal arm is identified. The simulations in (Xu et al., 2018) apply a greedy arm selection strategy that deviates from the algorithm that is analyzed. We instead implement the LinGapE algorithm in the form that it is analyzed.

- ALBA (Tao et al., 2018): This algorithm is parameter free and we implement the $\mathcal{Y}$-ElimTil sub-procedure following the paper since it gives improved empirical results compared to the $\mathcal{X}$-ElimTil sub-procedure that provides identical theoretical results.

- RAGE: To compute the discrete allocation given a design, we use the rounding procedure discussed in Section 6.D from (Pukelsheim, 2006) and $\epsilon = 1/10$.

Chapter 7

Conclusion and Future Directions

This thesis analyzes and develops algorithms for learning in games and sequential decision-making problems. Notably, we develop methods to effectively overcome and understand the challenges that arise from both competition and uncertainty. This work requires a diverse set of tools from a conglomeration of fields to obtain provable guarantees of performance and this will be key to future theoretical advances. Moreover, as can be seen throughout the thesis, the analytical results are tightly coupled with and provide insights into practical problems, and similarly, the empirical evaluations we conduct support and inform theoretical pursuits.

In Part I of this thesis, several research projects on the topic of learning and optimization in games were presented, with a focus on classes of nonconvex games and gradient-based learning algorithms within them. In the process, we significantly develop the understanding of gradient-based algorithms in nonconvex games and the placement of equilibrium concepts within them. This work explains what can be expected from standard algorithms implemented for solving machine learning problems formulated as games and describes how desirable outcomes can be achieved. Moreover, from an arguably more typical game-theoretic perspective, we expand the types of systems that can be abstracted as a game to provide insights to observed behavior.

In Part II of this thesis, we present research on a pair of sequential decision-making problems. Specifically, we identify a key component of the typical conference peer review system that presently is being handled ineffectively and also exhibits interesting theoretical challenges. In the process of solving this problem, we demonstrate how competing objectives that cannot be immediately realized can be carefully balanced algorithmically, which is a commonly arising problem in markets and crowdsourcing. Then, we develop and study a generalization of known multi-armed bandit frameworks. For this problem, we characterize the limits of achievable performance and present a near-optimal algorithm. This work expands the classes of bandit problems with desirable theoretical guarantees and is key to a number of applications such as scientific discovery and recommender systems where there is often an underlying structure coupling actions that can be exploited.

While at the end of each chapter, we describe some open questions and future directions of research, we expand on those now and also mention some topics that have not been discussed thus far. A key challenge of nonconvex games, and specifically nonconvex-nonconcave zero-sum games, is that generally global convergence guarantees to meaningful solutions are difficult to come by without imposing structure into the problem. This statement clearly hinges on the definition of 'meaningful solution'. Thus an interesting direction of future work is looking at relaxed solution notions that are relevant to machine learning problems, while also being computable. In our own work with collaborators, we have begun to explore this topic (Fiez et al., 2021a). Specifically, in recent work, we develop a novel, relaxed framework for studying nonconvex-nonconcave zero-sum games. In particular, we propose a new algorithm for the min-player to play against smooth algorithms deployed by the adversary instead of against full maximization. This results in an algorithm that

is guaranteed to make monotonic progress and which also find an appropriate "stationary point" in a polynomial number of iterations. This framework covers practical settings where the smooth algorithms deployed by the adversary are multi-step stochastic gradient ascent, and its accelerated version. Thus, it is keenly relevant to machine learning problems such as adversarial training. More generally, inserting structure into the assumed behavior of players in nonconvex games has potential to lead to stronger results.

While the research in Part I of this thesis considers extremely general formulations, it now appears time to work up from more precise problems. In our own work with collaborators, we have started in this direction by looking at how actor-critic reinforcement learning algorithms can be viewed as a Stackelberg game and then optimized using the Stackelberg gradient dynamics from Chapter 2 (Zheng et al., 2021). Moreover, the Stackelberg game framework is closely related to the problem of incentive and mechanism design given that these problems are often studied under the hood of what is known as reverse Stackelberg games. In our own work with collaborators in the past, this is a topic that we have explored. Specifically, problems in which a principal must learn the preferences of an agent or group of agents and induce a desired response (Ratliff and Fiez, 2020; Ratliff et al., 2019). It seems timely to revisit this research direction and exploit the tools that have been developed in recent years for learning in games.

The research on learning in games presented in Part I follows the usual approach of treating the game as a fixed object. However, the typical restriction of studying evolving agents acting in a static game does not allow us to capture many applications of interest where the rules of interaction can themselves adapt to the collective history of the agent behavior. Indeed, in adversarial learning, the difficulty of the game can increase over time by exactly focusing on the settings where the agent has performed the weakest. Similarly, in biology or economics, negative frequency-dependent selection indicates that if a particular advantageous strategy is used exhaustively by agents, then its relative advantages typically dissipate over time. As soon as the game stops being a passive object that the agents act upon, it can be best thought of as an algorithm itself. This crucial observation motivates the study of dynamically evolving players in dynamically evolving games. We have been actively working on this area of learning in games and believe it will continue to be a fruitful and meaningful research direction (Fiez et al., 2021c; Skoulakis et al., 2021).

With respect to the research in Part II of this thesis, there are several interesting future directions. Specific to the work presented in Chapter 5, we are actively working with a conference management system to implement our algorithm for use in conferences in workshops. We believe that the results of these experiments will be extremely illuminating and can lead to future research problems. From a theoretical standpoint, it would be interesting to incorporate strategic behavior into the reviewer behavior models. Moreover, with regard to the work presented in Chapter 6, as mentioned previously, there are still some important standing theoretical questions to be solved that are closely related. Specifically, obtaining a precisely tight algorithm for the pure exploration linear bandit problem, expanding the techniques to the contextual bandit identification problem, and also designing an instance-dependent optimal regret algorithm for linear bandits. While not discussed in this thesis, an interesting topic in multi-armed bandits is the regret and significance trade-off. This is at the forefront of practitioner's minds at companies, where it is often up for debate whether the increased regret of identification methods is worth being able to make decisions faster. Similarly, while many multi-armed bandit algorithms assume either a perfectly stochastic model or a purely adversarial model, the real world seems to be somewhere in between. Fortunately, there has been

a significant amount of work recently on the development of regret algorithms that can smoothly interpolate between these regimes. That being said, the study of identification algorithms in this light is fairly sparse. This seems to be a promising direction of future work that is still somewhat under-explored.

Finally, we believe that a closer merging of the study in learning in games and the study of decision-making under uncertainty is a promising direction of future research. While there are some works like this, they generally fall under simple classes of games and there is opportunities to extend this literature.

BIBLIOGRAPHY

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, pages 2312–2320, 2011.

EH Abed, L Saydy, and AL Tits. Generalized stability of linear singularly perturbed systems including calculation of maximal parameter range. In *Robust Control of Linear Systems and Nonlinear Control*, pages 197–203. Springer, 1990.

Ralph Abraham, Jerold E. Marsden, and Tudor Ratiu. *Manifolds, Tensor Analysis, and Applications*. Springer, 2nd edition, 1988.

Leonard Adolphs, Hadi Daneshmand, Aurelien Lucchi, and Thomas Hofmann. Local saddle point optimization: A curvature exploitation approach. In *International Conference on Artificial Intelligence and Statistics*, pages 486–495, 2019.

Deepak Agarwal, Bee-Chung Chen, Pradheep Elango, and Xuanhui Wang. Click shaping to optimize multiple objectives. In *SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 132–140, 2011.

Nir Ailon, Moses Charikar, and Alantha Newman. Aggregating inconsistent information: ranking and clustering. *Journal of the ACM (JACM)*, 55(5):23, 2008.

Ayya Alieva, Ashok Cutkosky, and Abhimanyu Das. Robust pure exploration in linear bandits with limited budget. In *International Conference on Machine Learning*, pages 187–195, 2021.

Zeyuan Allen-Zhu, Yuanzhi Li, Aarti Singh, and Yining Wang. Near-optimal discrete optimization for experimental design: A regret minimization approach. *Mathematical Programming*, 186(1): 439–478, 2021.

John M Alongi and Gail Susan Nelson. *Recurrence and topology*, volume 85. American Mathematical Society, 2007.

Simon P Anderson and Maxim Engers. Stackelberg versus cournot oligopoly equilibrium. *International Journal of Industrial Organization*, 10(1):127–135, 1992.

I. K. Argyros. A generalization of ostrowski's theorem on fixed points. *Applied Mathematics Letters*, 12:77–79, 1999.

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International Conference on Machine Learning*, pages 214–223, 2017.

Grigor Aslanyan and Utkarsh Porwal. Position bias estimation for unbiased learning-to-rank in ecommerce search. In *ACM International Symposium on String Processing and Information Retrieval*, pages 47–64. Springer, 2019.

Haris Aziz, Omer Lev, Nicholas Mattei, Jeffrey S Rosenschein, and Toby Walsh. Strategyproof peer selection using randomization, partitioning, and apportionment. *Artificial Intelligence*, 275: 295–309, 2019.

David Balduzzi, Sebastien Racaniere, James Martens, Jakob Foerster, Karl Tuyls, and Thore Graepel. The mechanics of n-player differentiable games. In *International Conference on Machine Learning*, pages 354–363, 2018.

Tamer Basar and G Olsder. Mixed stackelberg strategies in continuous-kernel games. *IEEE Transactions on Automatic Control*, 25(2):307–309, 1980.

Tamer Basar and Geert Jan Olsder. *Dynamic Noncooperative Game Theory*. Society for Industrial and Applied Mathematics, 2nd edition, 1998.

Tamer Basar and Hasan Selbuz. Closed-loop stackelberg strategies with applications in the optimal control of multilevel systems. *IEEE Transactions on Automatic Control*, 24(2):166–179, 1979.

Michel Benaim. A Dynamical System Approach to Stochastic Approximations. *SIAM Journal on Control and Optimization*, 34(2):437–472, 1996.

Michel Benaim and Morris W Hirsch. Dynamics of morse-smale urn processes. *Ergodic Theory and Dynamical Systems*, 15(6):1005–1030, 1995.

Hugo Berard, Gauthier Gidel, Amjad Almahairi, Pascal Vincent, and Simon Lacoste-Julien. A Closer Look at the Optimization Landscapes of Generative Adversarial Networks. In *International Conference on Learning Representations*, 2020.

Thomas Berger, Juan Giribet, Francisco Martínez Pería, and Carsten Trunk. On a class of non-hermitian matrices with positive definite schur complements. *Proceedings of the American Mathematical Society*, 147(6):2375–2388, 2019.

Rajendra Bhatia. *Matrix analysis*, volume 169. Springer Science & Business Media, 2013.

Federico Bianchi and Flaminio Squazzoni. Is three better than one? simulating the effect of reviewer selection and behavior on the quality and efficiency of peer review. In *Winter Simulation Conference*, pages 4081–4089. IEEE, 2015.

Nick Black, Susan Van Rooyen, Fiona Godlee, Richard Smith, and Stephen Evans. What makes a good reviewer and a good review for a general medical journal? *Journal of the American Medical Association*, 280(3):231–233, 1998.

Vivek S. Borkar. *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press, 2008.

Vivek S Borkar and Sarath Pattathil. Concentration bounds for two time scale stochastic approximation. In *Allerton Conference on Communication, Control, and Computing*, pages 504–511. IEEE, 2018.

George EP Box, J Stuart Hunter, and William G Hunter. Statistics for experimenters. In *Wiley Series in Probability and Statistics*. Wiley, 2005.

Michele Breton, Abderrahmane Alj, and Alain Haurie. Sequential stackelberg equilibria in two-person games. *Journal of Optimization Theory and Applications*, 59(1):71–97, 1988.

H.W. Broer and F. Takens. Chapter 1 - preliminaries of dynamical systems theory. In F Takens Henk Broer and B Hasselblatt, editors, *Handbook of Dynamical Systems*, volume 3 of *Handbook of Dynamical Systems*, pages 1 – 42. Elsevier Science, 2010.

George W Brown. Some notes on computation of games solutions. Technical report, Rand Corporation, 1949.

Zach Y Brown and Alexander MacKay. Competition in pricing algorithms. Technical report, National Bureau of Economic Research, 2021.

Rainer Burkard, Mauro Dell'Amico, and Silvano Martello. *Assignment problems, revised reprint*, volume 106. SIAM, 2012.

Guillaume Cabanac and Thomas Preuss. Capitalizing on order effects in the bids of peer-reviewed conferences to secure reviews by expert referees. *Journal of the American Society for Information Science and Technology*, 64(2):405–415, 2013.

Tongyi Cao and Akshay Krishnamurthy. Disagreement-based combinatorial pure exploration: Efficient algorithms and an analysis with localization. *arXiv preprint arXiv:1711.08018*, 2017.

Zhe Cao, Tao Qin, Tie-Yan Liu, Ming-Feng Tsai, and Hang Li. Learning to rank: from pairwise approach to listwise approach. In *International Conference on Machine Learning*, pages 129–136, 2007.

Laurent Charlin and Richard Zemel. The toronto paper matching system: an automated paper-reviewer assignment system. In *ICML Workshop on Peer Reviewing and Publishing Models*, 2013.

Benjamin Chasnov, Tanner Fiez, and Lillian Ratliff. Opponent anticipation via conjectural variations. *NeurIPS Workshop on Smooth Games Optimization and Machine Learning: Bridging Game Theory and Deep Learning*, 2019.

Benjamin Chasnov, Lillian J. Ratliff, Eric Mazumdar, and Samuel A. Burden. Convergence analysis of gradient-based learning in continuous games. In *Conference on Uncertainty in Artificial Intelligence*, pages 935–944, 2020.

Lijie Chen and Jian Li. On the optimal sample complexity for best arm identification. *arXiv preprint arXiv:1511.03774*, 2015.

Lijie Chen, Anupam Gupta, and Jian Li. Pure exploration of multi-armed bandit under matroid constraints. In *Conference on Learning Theory*, pages 647–669, 2016.

Lijie Chen, Anupam Gupta, Jian Li, Mingda Qiao, and Ruosong Wang. Nearly optimal sampling algorithms for combinatorial pure exploration. In *Conference on Learning Theory*, pages 482–534, 2017.

Shouyuan Chen, Tian Lin, Irwin King, Michael R Lyu, and Wei Chen. Combinatorial pure exploration of multi-armed bandits. In *Advances in Neural Information Processing Systems*, pages 379–387, 2014.

Aleksandr Chuklin, Ilya Markov, and Maarten de Rijke. Click models for web search. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 7(3):1–115, 2015.

Kenneth Church. Reviewing the reviewers. *Computational Linguistics*, 31(4):575–578, 2005.

Vincent Conitzer. On stackelberg mixed strategies. *Synthese*, 193(3):689–703, 2016.

Bo Dai, Albert Shaw, Lihong Li, Lin Xiao, Niao He, Zhen Liu, Jianshu Chen, and Le Song. Sbeed: Convergent reinforcement learning with nonlinear function approximation. In *International Conference on Machine Learning*, pages 1125–1134, 2018.

Hadi Daneshmand, Jonas Kohler, Aurelien Lucchi, and Thomas Hofmann. Escaping saddles with stochastic gradients. In *International Conference on Machine Learning*, pages 1155–1164, 2018.

Varsha Dani, Thomas P Hayes, and Sham M Kakade. Stochastic linear optimization under bandit feedback. *Conference on Leanring Theory*, 2008.

John M. Danskin. The theory of max-min, with applications. *SIAM Journal on Applied Mathematics*, 14(4):641–664, 1966.

John M. Danskin. *The Theory of Max-Min and its Application to Weapons Allocation Problems*. Springer, 1967.

Constantinos Daskalakis and Ioannis Panageas. The limit points of (optimistic) gradient descent in min-max optimization. In *Advances in Neural Information Processing Systems*, pages 9236–9246, 2018.

Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. Training gans with optimism. *International Conference on Learning Representations*, 2018.

Rémy Degenne, Pierre Ménard, Xuedong Shang, and Michal Valko. Gamification of pure exploration for linear bandits. In *International Conference on Machine Learning*, pages 2432–2442, 2020a.

Rémy Degenne, Han Shao, and Wouter Koolen. Structure adaptive algorithms for stochastic bandits. In *International Conference on Machine Learning*, pages 2443–2452, 2020b.

S Dempe, J Dutta, and BS Mordukhovich. New necessary optimality conditions in optimistic bilevel programming. *Optimization*, 56(5-6):577–604, 2007.

Stephan Dempe. *Foundations of bilevel programming*. Springer Science & Business Media, 2002.

Stephan Dempe. Bilevel programming. In *Essays and Surveys in Global Optimization*, pages 165–193. Springer, 2005.

Stephan Dempe. *Bilevel optimization: theory, algorithms and applications*. TU Bergakademie Freiberg, Fakultät für Mathematik und Informatik, 2018.

Stephan Dempe and N Gadhi. Second order optimality conditions for bilevel set optimization problems. *Journal of Global Optimization*, 47(2):233–245, 2010.

Inderjit S Dhillon. Co-clustering documents and words using bipartite spectral graph partitioning. In *SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 269–274, 2001.

Nicola Di Mauro, Teresa MA Basile, and Stefano Ferilli. Grape: An expert review assignment component for scientific conference management systems. In *International conference on industrial, engineering and other applications of applied intelligent systems*, pages 789–798. Springer, 2005.

Wenxin Ding, Nihar B Shah, and Weina Wang. On the privacy-utility tradeoff in peer-review data analysis. *arXiv:2006.16385*, 2020.

Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of machine learning research*, 7(Jun):1079–1105, 2006.

Yu G Evtushenko. Some local properties of minimax problems. *USSR Computational Mathematics and Mathematical Physics*, 14(3):129–138, 1974.

Farzan Farnia and Asuman Ozdaglar. Do gans always have nash equilibria? *International Conference on Machine Learning*, 2020.

Tanner Fiez and Lillian Ratliff. Gradient descent-ascent provably converges to strict local minmax equilibria with a finite timescale separation. *arXiv preprint arXiv:2009.14820*, 2020.

Tanner Fiez and Lillian Ratliff. Gradient descent-ascent provably converges to strict local minmax equilibria with a finite timescale separation. *AAAI Workshop on Reinforcement Learning in Games*, 2021.

Tanner Fiez, Benjamin Chasnov, and Lillian Ratliff. Characterizing equilibria in stackelberg games. *NeurIPS Workshop on Smooth Games Optimization and Machine Learning: Bridging Game Theory and Deep Learning*, 2019a.

Tanner Fiez, Benjamin Chasnov, and Lillian J Ratliff. Convergence of learning dynamics in stackelberg games. *arXiv preprint arXiv:1906.01217*, 2019b.

Tanner Fiez, Lalit Jain, Kevin G Jamieson, and Lillian Ratliff. Sequential experimental design for transductive linear bandits. In *Advances in Neural Information Processing Systems*, pages 10667–10677, 2019c.

Tanner Fiez, Nihar Shah, and Lillian Ratliff. A super* algorithm to optimize paper bidding in peer review. In *ICML workshop on Real-world Sequential Decision Making: Reinforcement Learning And Beyond*, 2019d.

Tanner Fiez, Benjamin Chasnov, and Lillian J Ratliff. Implicit learning dynamics in stackelberg games: Equilibria characterization, convergence analysis, and empirical study. In *International Conference on Machine Learning*, pages 3133–3144, 2020a.

Tanner Fiez, Nihar Shah, and Lillian Ratliff. A super* algorithm to optimize paper bidding in peer review. In *Conference on Uncertainty in Artificial Intelligence*, pages 580–589, 2020b.

Tanner Fiez, Nihar Shah, and Lillian Ratliff. A super* algorithm to optimize paper bidding in peer review. *arXiv preprint arXiv:2007.07079*, 2020c.

Tanner Fiez, Chi Jin, Praneeth Netrapalli, and Lillian J Ratliff. Minimax optimization with smooth algorithmic adversaries. *arXiv preprint arXiv:2106.01488*, 2021a.

Tanner Fiez, Lillian Ratliff, Eric Mazumdar, Evan Faulkner, and Adhyyan Narang. Global convergence to local minmax equilibrium in classes of nonconvex zero-sum games. *arXiv preprint*, 2021b.

Tanner Fiez, Ryann Sim, Stratis Skoulakis, Georgios Piliouras, and Lillian Ratliff. Online learning in periodic zero-sum games. *AAMAS Workshop on Adaptive Learning Agents*, 2021c.

Jaime F Fisac, Eli Bronstein, Elis Stefansson, Dorsa Sadigh, S Shankar Sastry, and Anca D Dragan. Hierarchical game-theoretic planning for autonomous vehicles. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 9590–9596, 2019.

Lampros Flokas, Emmanouil-Vasileios Vlatakis-Gkaragkounis, and Georgios Piliouras. Solving min-max optimization with hidden structure via gradient descent ascent. *arXiv preprint arXiv:2101.05248*, 2021.

Jakob Foerster, Richard Y Chen, Maruan Al-Shedivat, Shimon Whiteson, Pieter Abbeel, and Igor Mordatch. Learning with opponent-learning awareness. In *International Conference on Autonomous Agents and MultiAgent Systems*, pages 122–130, 2018.

Drew Fudenberg, Fudenberg Drew, David K Levine, and David K Levine. *The theory of learning in games*, volume 2. MIT press, 1998.

Victor Gabillon, Mohammad Ghavamzadeh, and Alessandro Lazaric. Best arm identification: A unified approach to fixed budget and fixed confidence. In *Advances in Neural Information Processing Systems*, pages 3212–3220, 2012.

Naveen Garg, Telikepalli Kavitha, Amit Kumar, Kurt Mehlhorn, and Julián Mestre. Assigning papers to referees. *Algorithmica*, 58(1):119–136, Sep 2010.

Rong Ge, Furong Huang, Chi Jin, and Yang Yuan. Escaping from saddle points—online stochastic gradient for tensor decomposition. In *Conference on Learning Theory*, pages 797–842, 2015.

Saeed Ghadimi and Mengdi Wang. Approximation methods for bilevel programming. *arXiv preprint arXiv:1802.02246*, 2018.

Christopher G. Gibson, Klaus Wirthmüller, Andrew A. du Plessis, and Eduard J. N. Looijenga. Topological stability of smooth mappings. In *Lecture Notes in Mathematics*, volume 552. Springer-Verlag, 1976.

Gauthier Gidel, Hugo Berard, Gaëtan Vignoud, Pascal Vincent, and Simon Lacoste-Julien. A variational inequality perspective on generative adversarial networks. In *International Conference on Learning Representations*, 2019.

Gauthier Gidel, David Balduzzi, Wojciech Czarnecki, Marta Garnelo, and Yoram Bachrach. A limited-capacity minimax theorem for non-convex games or: How i learned to stop worrying about mixed-nash and love neural nets. In *International Conference on Artificial Intelligence and Statistics*, pages 2548–2556, 2021.

Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

EG Golshtein. Generalized gradient method for finding saddlepoints. *Matekon*, 10(3):36–52, 1974.

Martin Golubitsky and Victor Guillemin. *Stable Mappings and Their Singularities*. Springer-Verlag, 1973.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014.

Willy J. F. Govaerts. *Numerical Methods for Bifurcations of Dynamical Equilibria*. Society for Industrial and Applied Mathematics, 2000.

Peter Hart, Nils Nilsson, and Bertram Raphael. A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, 4(2):100–107, 1968.

Joao P Hespanha. *Linear Systems Theory*. Princeton University Press, 2nd edition, 2018.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*, pages 6626–6637, 2017.

Daniel N Hill, Houssam Nassif, Yi Liu, Anand Iyer, and SVN Vishwanathan. An efficient bandit algorithm for realtime multivariate optimization. In *SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1813–1821. ACM, 2017.

Morris W. Hirsch. *Differential Topology*. Springer, 1976.

Morris W Hirsch, Stephen Smale, and Robert L Devaney. *Differential equations, dynamical systems, and an introduction to chaos*. Academic press, 2012.

Mingyi Hong, Hoi-To Wai, Zhaoran Wang, and Zhuoran Yang. A two-timescale framework for bilevel optimization: Complexity analysis and application to actor-critic. *arXiv preprint arXiv:2007.05170*, 2020.

Roger Horn and Charles Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 2011.

Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

Mokter Hossain. Users' motivation to participate in online crowdsourcing platforms. In *International Conference on Innovation Management and Technology Research*, pages 310–315. IEEE, 2012.

Martin Jaggi. Revisiting frank-wolfe: Projection-free sparse convex optimization. In *International Conference on Machine Learning*, pages 427–435, 2013.

Lalit Jain and Kevin Jamieson. Firing bandits: Optimizing crowdfunding. In *International Conference on Machine Learning*, pages 2206–2214, 2018.

Tamas Jambor and Jun Wang. Optimizing multiple objectives in collaborative filtering. In *ACM Conference on Recommender Systems*, pages 55–62, 01 2010.

Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lil'ucb: An optimal exploration algorithm for multi-armed bandits. In *Conference on Learning Theory*, pages 423–439, 2014.

Kalervo Järvelin and Jaana Kekäläinen. IR evaluation methods for retrieving highly relevant documents. In *ACM SIGIR Conference on Research and Development in Information and Retrieval*, pages 41–48, 2000.

Steven Jecmen, Hanrui Zhang, Ryan Liu, Nihar B Shah, Vincent Conitzer, and Fei Fang. Mitigating manipulation in peer review via randomized reviewer assignments. In *ICML Workshop on Incentives in Machine Learning*, 2020.

Yassir Jedra and Alexandre Proutiere. Optimal best-arm identification in linear bandits. *Advances in Neural Information Processing Systems*, 33, 2020.

Chi Jin, Rong Ge, Praneeth Netrapalli, Sham M Kakade, and Michael I Jordan. How to escape saddle points efficiently. In *International Conference on Machine Learning*, pages 1724–1732, 2017.

Chi Jin, Praneeth Netrapalli, and Michael Jordan. What is local optimality in nonconvex-nonconcave minimax optimization? In *International Conference on Machine Learning*, pages 4880–4889, 2020.

Chi Jin, Praneeth Netrapalli, Rong Ge, Sham M Kakade, and Michael I Jordan. On nonconvex optimization for machine learning: Gradients, stochasticity, and saddle points. *Journal of the ACM*, 68(2):1–29, 2021.

Kwang-Sung Jun and Chicheng Zhang. Crush optimism with pessimism: Structured bandits beyond asymptotic optimality. *Advances in Neural Information Processing Systems*, 2020.

Kwang-Sung Jun, Lalit Jain, Houssam Nassif, and Blake Mason. Improved confidence bounds for the linear logistic model and applications to bandits. In *International Conference on Machine Learning*, pages 5148–5157, 2021.

Marc Jungers, Emmanuel Trélat, and Hisham Abou-Kandil. Min-max and min-min stackelberg strategies with closed-loop information structure. *Journal of dynamical and control systems*, 17 (3):387, 2011.

Sameer Kamal. On the convergence, lock-in probability, and sample complexity of stochastic approximation. *SIAM Journal on Control and Optimization*, 48(8):5178–5192, 2010.

Dongyeop Kang, Waleed Ammar, Bhavana Dalvi, Madeleine van Zuylen, Sebastian Kohlmeier, Eduard Hovy, and Roy Schwartz. A dataset of peer reviews (peerread): Collection, insights and nlp applications. In *Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1647–1661, 2018.

Hamed Karimi, Julie Nutini, and Mark Schmidt. Linear convergence of gradient and proximal-gradient methods under the Polyak-Łojasiewicz condition. In *Machine Learning and Knowledge Discovery in Databases*, pages 795–811. Springer International Publishing, 2016.

Maryam Karimzadehgan, ChengXiang Zhai, and Geneva Belford. Multi-aspect expertise matching for review assignment. In *Conference on Information and knowledge management*, pages 1113–1122, 2008.

Zohar Karnin, Tomer Koren, and Oren Somekh. Almost optimal exploration in multi-armed bandits. In *International Conference on Machine Learning*, pages 1238–1246, 2013.

Zohar S Karnin. Verification based solution for structured mab problems. In *Advances in Neural Information Processing Systems*, pages 145–153, 2016.

Julian Katz-Samuels, Lalit Jain, Kevin G Jamieson, et al. An empirical process approach to the union bound: Practical algorithms for combinatorial and linear bandits. *Advances in Neural Information Processing Systems*, 33, 2020.

Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *Journal of Machine Learning Research*, 17:1–42, 2016.

Nicolas Kaufmann, Thimo Schulze, and Daniel Veit. More than fun and money. worker motivation in crowdsourcing-a study on mechanical turk. In *Americas Conference on Information Systems*, volume 11, pages 1–11, 2011.

Abbas Kazerouni and Lawrence M Wein. Best arm identification in generalized linear bandits. *Operations Research Letters*, 49(3):365–371, 2021.

J. Kelley. *General Topology*. Van Nostrand Reinhold Company, 1955.

Vijay Keswani, Oren Mangoubi, Sushant Sachdeva, and Nisheeth K Vishnoi. A convergent and dimension-independent first-order algorithm for min-max optimization. *arXiv preprint arXiv:2006.12376*, 2020.

Hassan Khalil. *Nonlinear Systems*. Prentice Hall, 3rd edition, 2002.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

AI Klimushchev and NN Krasovskii. Uniform asymptotic stability of systems of differential equations with a small parameter in the derivative terms. *Journal of Applied Mathematics and mechanics*, 25(4):1011–1025, 1961.

Ari Kobren, Barna Saha, and Andrew McCallum. Paper matching with local fairness constraints. In *SIGKDD Conference on Knowledge Discovery and Data Mining*, 2019.

P. Kokotovic. A Riccati equation for block-diagonalization of ill-conditioned systems. *IEEE Transactions on Automatic Control*, 20(6):812–814, 1975. doi: 10.1109/TAC.1975.1101089.

Peter V Kokotovic, John O'Reilly, and Hassan K Khalil. *Singular Perturbation Methods in Control: Analysis and Design*. Academic Press, Inc., 1986.

Galina M Korpelevich. The extragradient method for finding saddle points and other problems. *Matecon*, 12:747–756, 1976.

NN Krasovskii. Stability of Motion: Application of Lyapunov's Second Method to Differential Systems and Equations with Time-Delay, 1963.

PA Lagerstrom and RG Casten. Basic concepts underlying singular perturbation techniques. *Siam Review*, 14(1):63–120, 1972.

Peter Lancaster and Miron Tismenetsky. *The theory of matrices: with applications*. Elsevier, 1985.

Tor Lattimore and Csaba Szepesvari. The end of optimism? an asymptotic analysis of finite-armed linear bandits. In *Artificial Intelligence and Statistics*, pages 728–737, 2017.

Eugene L Lawler. Combinatorial optimization, 1976.

N. Lawrence and C. Cortes. The nips experiment. `http://inverseprobability.com/2014/12/16/the-nips-experiment`, 2014.

Jason D Lee, Max Simchowitz, Michael I Jordan, and Benjamin Recht. Gradient descent only converges to minimizers. In *Conference on Learning Theory*, pages 1246–1257, 2016.

Jason D. Lee, Ioannis Panageas, Georgios Piliouras, Max Simchowitz, Michael I. Jordan, and Benjamin Recht. First-order methods almost always avoid strict saddle points. *Mathematical Programming*, 176(1):311–337, 2019.

John Lee. *Introduction to smooth manifolds*. Springer, 2012.

George Leitmann. On generalized stackelberg strategies. *Journal of optimization theory and applications*, 26(4):637–643, 1978.

Alistair Letcher, Jakob Foerster, David Balduzzi, Tim Rocktäschel, and Shimon Whiteson. Stable opponent shaping in differentiable games. *International Conference on Learning Representations*, 2019.

Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *International Conference on World Wide Web*, pages 661–670. ACM, 2010.

Zun Li, Feiran Jia, Aditya Mate, Shahin Jabbari, Mithun Chakraborty, Milind Tambe, and Yevgeniy Vorobeychik. Solving structured hierarchical games using differential backward induction. *arXiv:2106.04663*, 2021.

Jing Wu Lian, Nicholas Mattei, Renee Noble, and Toby Walsh. The conference paper assignment problem: Using order weighted averages to assign indivisible goods. In *AAAI Conference on Artificial Intelligence*, 2018.

Tianyi Lin, Chi Jin, and Michael Jordan. On gradient descent ascent for nonconvex-concave minimax problems. In *International Conference on Machine Learning*, pages 6083–6093, 2020a.

Tianyi Lin, Chi Jin, Michael Jordan, et al. Near-optimal algorithms for minimax optimization. *Conference on Learning Theory*, 2020b.

Sijia Liu, Songtao Lu, Xiangyi Chen, Yao Feng, Kaidi Xu, Abdullah Al-Dujaili, Mingyi Hong, and Una-May O'Reilly. Min-max optimization without gradients: Convergence and applications to black-box evasion and poisoning attacks. In *International Conference on Machine Learning*, pages 6282–6293, 2020.

Cheng Long, Raymond Chi-Wing Wong, Yu Peng, and Liangliang Ye. On good and fair paper-reviewer assignment. In *International Conference on Data Mining*, pages 1145–1150. IEEE, 2013.

Jonathan Lorraine, Paul Vicol, and David Duvenaud. Optimizing millions of hyperparameters by implicit differentiation. In *International Conference on Artificial Intelligence and Statistics*, pages 1540–1552, 2020.

Songtao Lu, Ioannis Tsaknakis, Mingyi Hong, and Yongxin Chen. Hybrid block successive approximation for one-sided non-convex min-max problems: algorithms and applications. *IEEE Transactions on Signal Processing*, 2020.

Matthew MacKay, Paul Vicol, Jon Lorraine, David Duvenaud, and Roger Grosse. Self-tuning networks: Bilevel optimization of hyperparameters using structured best-response functions. *International Conference on Learning Representations*, 2019.

Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. In *International Conference on Learning Representations*, 2018.

Jan Magnus. *Linear Structures*. Griffin, 1988.

Oren Mangoubi and Nisheeth K Vishnoi. Greedy adversarial equilibrium: an efficient alternative to nonconvex-nonconcave min-max optimization. In *Symposium on Theory of Computing*, pages 896–909, 2021.

James Martens et al. Deep learning via hessian-free optimization. In *International Conference on Machine Learning*, volume 27, pages 735–742, 2010.

Eric Mazumdar and Lillian J. Ratliff. Local nash equilibria are isolated, strict local nash equilibria in 'almost all' zero-sum continuous games. In *IEEE Conference on Decision and Control*, pages 6899–6904, 2019.

Eric Mazumdar, Lillian J Ratliff, and S Shankar Sastry. On Gradient-Based Learning in Continuous Games. *SIAM Journal on Mathematics of Data Science*, 2(1):103–131, 2020.

Eric V Mazumdar, Michael I Jordan, and S Shankar Sastry. On finding local nash equilibria (and only local nash equilibria) in zero-sum games. *arXiv preprint arXiv:1901.00838*, 2019.

Reshef Meir, Jérôme Lang, Julien Lesca, Natan Kaminsky, and Nicholas Mattei. A market-inspired bidding scheme for peer review paper assignment. In *Games, Agents, and Incentives Workshop at AAMAS*, 2020.

Panayotis Mertikopoulos, Bruno Lecouat, Houssam Zenati, Chuan-Sheng Foo, Vijay Chandrasekhar, and Georgios Piliouras. Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile. In *International Conference on Learning Representations*, 2019.

Lars Mescheder, Sebastian Nowozin, and Andreas Geiger. The numerics of gans. In *Advances in Neural Information Processing Systems*, pages 1825–1835, 2017.

Lars Mescheder, Andreas Geiger, and Sebastian Nowozin. Which training methods for gans do actually converge? In *International Conference on Machine learning*, pages 3481–3490, 2018.

Luke Metz, Ben Poole, David Pfau, and Jascha Sohl-Dickstein. Unrolled generative adversarial networks. In *International Conference on Learning Representations*, 2017.

Michinari Momma, Alireza Bagheri Garakani, and Yi Sun. Multi-objective relevance ranking. In *ACM SIGIR Conference on Research and Development in Information and Retrieval*, 2019.

Dov Monderer and Lloyd S Shapley. Potential games. *Games and economic behavior*, 14(1):124–143, 1996.

Oskar Morgenstern and John Von Neumann. *Theory of games and economic behavior*. Princeton university press, 1953.

Adrian Mulligan, Louise Hall, and Ellen Raphael. Peer review in a changing world: An international study measuring the attitudes of researchers. *Journal of the American Society for Information Science and Technology*, 64(1):132–161, 2013.

Jamie Murphy, Charles Hofacker, and Richard Mizerski. Primacy and recency effects on clicking behavior. *Journal of Computer-Mediated Communication*, 11(2):522–535, 2006.

D Mustafa and TN Davidson. Generalized integral controllability. In *Proceedings of 1994 33rd IEEE Conference on Decision and Control*, volume 1, pages 898–903. IEEE, 1994.

Vaishnavh Nagarajan and J Zico Kolter. Gradient descent gan optimization is locally stable. In *Advances in Neural Information Processing Systems*, pages 5585–5595, 2017.

Hongseok Namkoong and John C Duchi. Stochastic gradient methods for distributionally robust optimization with f-divergences. In *Advances in Neural Information Processing Systems*, pages 2208–2216, 2016.

John Nash. Non-cooperative games. *Annals of mathematics*, pages 286–295, 1951.

John F Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49, 1950.

Angelia Nedić and Asuman Ozdaglar. Subgradient methods for saddle-point problems. *Journal of optimization theory and applications*, 142(1):205–228, 2009.

Arkadi Nemirovski. Prox-method with rate of convergence o(1/t) for variational inequalities with lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 15(1):229–251, 2004.

Arkadi S Nemirovski and David Berkovich Yudin. Cesari convergence of the gradient method of approximating saddle points of convex-concave functions. In *Doklady Akademii Nauk*, volume 239, pages 1056–1059. Russian Academy of Sciences, 1978.

J v Neumann. Zur theorie der gesellschaftsspiele. *Mathematische annalen*, 100(1):295–320, 1928.

Mark EJ Newman. The structure of scientific collaboration networks. *Proceedings of the national academy of sciences*, 98(2):404–409, 2001.

Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.

Stefanos Nikolaidis, Swaprava Nath, Ariel D Procaccia, and Siddhartha Srinivasa. Game-theoretic modeling of human adaptation in human-robot collaboration. In *International Conference on Human-Robot Interaction*, pages 323–331, 2017.

Ritesh Noothigattu, Nihar Shah, and Ariel Procaccia. Loss functions, axioms, and peer review. In *ICML Workshop on Incentives in Machine Learning*, 2018.

Maher Nouiehed, Maziar Sanjabi, Tianjian Huang, Jason D Lee, and Meisam Razaviyayn. Solving a class of non-convex min-max games using iterative first order methods. In *Advances in Neural Information Processing Systems*, pages 14905–14916, 2019.

J. M. Ortega and W. C. Rheinboldt. *Iterative Solutions to Nonlinear Equations in Several Variables.* Academic Press, 1970.

Ioannis Panageas and Georgios Piliouras. Gradient descent only converges to minimizers: Non-isolated critical points and invariant regions. In *Innovations in Theoretical Computer Science Conference*, 2017.

Christos Papadimitriou and Georgios Piliouras. Game dynamics as the meaning of a game. *ACM SIGecom Exchanges*, 16(2):53–63, 2019.

G Papavassilopoulos and J Cruz. Nonclassical control problems and stackelberg games. *IEEE Transactions on Automatic Control*, 24(2):155–166, 1979.

George P Papavassilopoulos and JB Cruz. Sufficient conditions for stackelberg and nash strategies with memory. *Journal of Optimization Theory and Applications*, 31(2):233–260, 1980.

Robin Pemantle. Nonconvergence to unstable points in urn models and stochastic approximations. *Annuals Probability*, 18(2):698–712, 04 1990.

Alan L Porter and Frederick A Rossini. Peer review of interdisciplinary research proposals. *Science, technology, & human values*, 10(3):33–38, 1985.

Mason A Porter, Jukka-Pekka Onnela, and Peter J Mucha. Communities in networks. *Notices of the AMS*, 56(9):1082–1097, 2009.

Simon Price and Peter A. Flach. Computational support for academic peer review: a perspective from artificial intelligence. *Communications of the ACM*, 60(3):70–79, 2017.

Simon Price, Peter A Flach, and Sebastian Spiegler. Subsift: a novel application of the vector space model to support the academic research process. In *Workshop on Applications of Pattern Analysis*, pages 20–27, 2010.

Friedrich Pukelsheim. *Optimal design of experiments*. SIAM, 2006.

Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. In *International Conference on Learning Representations*, 2016.

Hassan Rafique, Mingrui Liu, Qihang Lin, and Tianbao Yang. Weakly-convex–concave min–max optimization: provable algorithms and applications in machine learning. *Optimization Methods and Software*, pages 1–35, 2021.

Aravind Rajeswaran, Igor Mordatch, and Vikash Kumar. A game theoretic framework for model based reinforcement learning. *International Conference on Machine Learning*, 2020.

L. J. Ratliff, S. A. Burden, and S. S. Sastry. Characterization and computation of local Nash equilibria in continuous games. In *Allerton Conference on Communication, Control, and Computing*, pages 917–924, 2013.

Lillian J Ratliff and Tanner Fiez. Adaptive incentive design. *IEEE Transactions on Automatic Control*, 2020.

Lillian J Ratliff, Samuel A Burden, and S Shankar Sastry. Genericity and structural stability of non-degenerate differential nash equilibria. In *American Control Conference*, pages 3990–3995. IEEE, 2014.

Lillian J Ratliff, Samuel A Burden, and S Shankar Sastry. On the characterization of local Nash equilibria in continuous games. *IEEE Transactions on Automatic Control*, 61(8):2301–2307, 2016.

Lillian J Ratliff, Roy Dong, Shreyas Sekar, and Tanner Fiez. A perspective on incentive design: Challenges and opportunities. *Annual Review of Control, Robotics, and Autonomous Systems*, 2: 305–338, 2019.

Clémence Réda, Emilie Kaufmann, and Andrée Delahaye-Duriez. Top-m identification for linear bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 1108–1116, 2021.

Julia Robinson. An iterative method of solving a game. *Annals of mathematics*, pages 296–301, 1951.

Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 2015.

Mario Rodriguez, Christian Posse, and Ethan Zhang. Multiple objective optimization in recommender systems. In *ACM Conference on Recommender Systems*, pages 11–18, 2012.

Marko A Rodriguez, Johan Bollen, and Herbert Van de Sompel. Mapping the bid behavior of conference referees. *Journal of Informetrics*, 1(1):68–82, 2007.

Magnus Roos, Jörg Rothe, Joachim Rudolph, Björn Scheuermann, and Dietrich Stoyan. A statistical approach to calibrating the scores of biased reviewers: The linear vs. the nonlinear model. In *6th Multidisciplinary Workshop on Advances in Preference Handling*, 2012.

J. B. Rosen. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica: Journal of the Econometric Society*, 33(3):520–534, 1965.

Kevin Roth, Aurelien Lucchi, Sebastian Nowozin, and Thomas Hofmann. Stabilizing training of generative adversarial networks through regularization. In *Advances in Neural Information Processing Systems*, pages 2018–2028, 2017.

Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. In *Advances in Neural Information Processing Systems*, pages 2234–2242, 2016.

S. Shankar Sastry. *Nonlinear Systems Theory*. Springer-Verlag, 1999.

Shankar Sastry and C Desoer. Jump behavior of circuits and systems. *IEEE Transactions on Circuits and Systems*, 28(12):1109–1124, 1981.

Lahcen Saydy. New stability/performance results for singularly perturbed systems. *Automatica*, 32 (6):807 – 818, 1996.

Lahcen Saydy, André L Tits, and Eyad H Abed. Guardian maps and the generalized stability of parametrized families of matrices and polynomials. *Mathematics of Control, Signals and Systems*, 3(4):345–371, 1990.

Florian Schäfer and Anima Anandkumar. Competitive gradient descent. In *Advances in Neural Information Processing Systems*, pages 7625–7635, 2019.

Nihar Shah, Sivaraman Balakrishnan, Aditya Guntuboyina, and Martin Wainwright. Stochastically transitive models for pairwise comparisons: Statistical and computational issues. In *International Conference on Machine Learning*, pages 11–20, 2016.

Nihar B Shah, Behzad Tabibian, Krikamol Muandet, Isabelle Guyon, and Ulrike Von Luxburg. Design and analysis of the NIPS 2016 review process. *Journal of Machine Learning Research*, 19 (1):1913–1946, 2018.

M. Shub. *Global Stability of Dynamical Systems*. Springer-Verlag, 1978.

Marwan Simaan and Jose B Cruz. Additional aspects of the stackelberg strategy in nonzero-sum games. *Journal of Optimization Theory and Applications*, 11(6):613–626, 1973.

Ashudeep Singh and Thorsten Joachims. Policy learning for fairness in ranking. In *Advances in Neural Information Processing Systems*, pages 5427–5437, 2019.

Aman Sinha, Hongseok Namkoong, and John Duchi. Certifiable distributional robustness with principled adversarial training. *International Conference on Learning Representations*, 2018.

Ankur Sinha, Pekka Malo, and Kalyanmoy Deb. A review on bilevel optimization: from classical to evolutionary approaches and applications. *IEEE Transactions on Evolutionary Computation*, 22(2):276–295, 2017.

Maurice Sion. On general minimax theorems. *Pacific Journal of mathematics*, 8(1):171–176, 1958.

Stratis Skoulakis, Tanner Fiez, Ryann Sim, Lillian Ratliff, and Georgios Piliouras. Evolutionary game theory squared: Evolving agents in endogenously evolving games. In *AAAI Conference on Artificial Intelligence*, 2021.

Marta Soare. *Sequential resource allocation in linear stochastic bandits*. PhD thesis, Université Lille 1-Sciences et Technologies, 2015.

Marta Soare, Alessandro Lazaric, and Rémi Munos. Best-arm identification in linear bandits. In *Advances in Neural Information Processing Systems*, pages 828–836, 2014.

Ivan Stelmakh, Nihar Shah, and Aarti Singh. On testing for biases in peer review. In *Advances in Neural Information Processing Systems*, 2019a.

Ivan Stelmakh, Nihar B Shah, and Aarti Singh. Peerreview4all: Fair and accurate reviewer assignment in peer review. In *International Conference on Algorithmic Learning Theory*, pages 828–856, 2019b.

Ivan Stelmakh, Nihar B Shah, and Aarti Singh. Catch me if i can: Detecting strategic behaviour in peer assessment. In *AAAI Conference on Artificial Intelligence*, pages 4794–4802, 2021a.

Ivan Stelmakh, Nihar B Shah, Aarti Singh, and Hal Daumé III. Prior and prejudice: The novice reviewers' bias against resubmissions in conference peer review. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–17, 2021b.

Krysta Svore, Maksims Volkovs, and Christopher Burges. Learning to rank with multiple objective functions. In *International Conference on World Wide Web*, pages 367–376, 03 2011.

Wenbin Tang, Jie Tang, and Chenhao Tan. Expertise matching via constraint-based optimization. In *International Conference on Web Intelligence and Intelligent Agent Technology*, volume 1, pages 34–41. IEEE, 2010.

Chao Tao, Saúl Blanco, and Yuan Zhou. Best arm identification in linear bandits with linear dimension dependency. In *International Conference on Machine Learning*, pages 4884–4893, 2018.

Gugan Thoppe and Vivek Borkar. A concentration bound for stochastic approximation via alekseev's formula. *Stochastic Systems*, 9(1):1–26, 2019.

Stefan Thurner and Rudolf Hanel. Peer-review in a world with rational scientists: Toward selection of the average. *The European Physical Journal B*, 84(4):707–711, 2011.

Andrea Tirinzoni, Matteo Pirotta, Marcello Restelli, and Alessandro Lazaric. An asymptotically optimal primal-dual incremental algorithm for contextual linear bandits. *Advances in Neural Information Processing Systems*, 2020.

Andrew Tomkins, Min Zhang, and William D Heavlin. Reviewer bias in single-versus double-blind peer review. *Proceedings of the National Academy of Sciences*, 114(48):12708–12713, 2017.

G David L Travis and Harry M Collins. New light on old boys: Cognitive and institutional particularism in the peer review system. *Science, Technology, & Human Values*, 16(3):322–341, 1991.

Christiane Tretter. *Spectral Theory of Block Operator Matrices and Applications*. Imperial College Press, 2008.

A. B. Tsybakov. Optimal aggregation of classifiers in statistical learning. *Annals of Statistics*, pages 135–166, 2004.

Hirofumi Uzawa. Iterative methods for concave programming. *Studies in linear and nonlinear programming*, 6:154–165, 1958.

Emmanouil-Vasileios Vlatakis-Gkaragkounis, Lampros Flokas, and Georgios Piliouras. Poincaré recurrence, cycles and spurious equilibria in gradient-descent-ascent for non-convex non-concave zero-sum games. In *Advances in Neural Information Processing Systems*, pages 10450–10461, 2019.

Heinrich Von Stackelberg. *Marktform und gleichgewicht*. Springer, 1934.

Heinrich Von Stackelberg. *Market structure and equilibrium*. Springer Science & Business Media, 2010.

Andrew Wagenmaker, Julian Katz-Samuels, and Kevin Jamieson. Experimental design for regret minimization in linear bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 3088–3096, 2021.

Jingyan Wang and Nihar B Shah. Your 2 is my 1, your 3 is my 9: Handling arbitrary miscalibrations in ratings. In *International Conference on Autonomous Agents and MultiAgent Systems*, pages 864–872, 2019.

Yuanhao Wang, Guodong Zhang, and Jimmy Ba. On solving minimax optimization locally: A follow-the-ridge approach. In *International Conference on Learning Representations*, 2020a.

Zhongruo Wang, Krishnakumar Balasubramanian, Shiqian Ma, and Meisam Razaviyayn. Zeroth-order algorithms for nonconvex minimax problems with improved complexities. *arXiv preprint arXiv:2001.07819*, 2020b.

Mark Ware. Peer review in scholarly journals: Perspective of the scholarly community–results from an international study. *Information Services & Use*, 28(2):109–112, 2008.

Wolfram Wiesemann, Angelos Tsoukalas, Polyxeni-Margarita Kleniati, and Berç Rustem. Pessimistic bilevel optimization. *SIAM Journal on Optimization*, 23(1):353–380, 2013.

Jeanette Wing. Yes, computer scientists are hypercritical. *Communications of the ACM*, 2011.

Liyuan Xu, Junya Honda, and Masashi Sugiyama. A fully adaptive algorithm for pure exploration in linear bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 843–851, 2018.

Yichong Xu, Han Zhao, Xiaofei Shi, and Nihar Shah. On strategyproof conference review. In *International Joint Conferences on Artificial Intelligence*, pages 616–622, 2019.

Himank Yadav, Zhengxiao Du, and Thorsten Joachims. Fair learning-to-rank from implicit feedback. *arXiv:1911.08054*, 2019.

Junchi Yang, Negar Kiyavash, and Niao He. Global convergence and variance reduction for a class of nonconvex-nonconcave minimax problems. *Advances in Neural Information Processing Systems*, 33, 2020.

Yasin Yazici, Chuan-Sheng Foo, Stefan Winkler, Kim-Hui Yap, Georgios Piliouras, and Vijay Chandrasekhar. The unusual effectiveness of averaging in gan training. *International Conference on Learning Representations*, 2019.

Kai Yu, Jinbo Bi, and Volker Tresp. Active learning via transductive experimental design. In *Proceedings of the 23rd international conference on Machine learning*, pages 1081–1088. ACM, 2006.

Mohammadi Zaki, Avinash Mohan, and Aditya Gopalan. Towards optimal and efficient best arm identification in linear bandits. *arXiv preprint arXiv:1911.01695*, 2019.

Mohammadi Zaki, Avi Mohan, and Aditya Gopalan. Explicit best arm identification in linear bandits using no-regret learners. *arXiv preprint arXiv:2006.07562*, 2020.

Alexander J Zaslavski. Necessary optimality conditions for bilevel minimization problems. *Nonlinear Analysis: Theory, Methods & Applications*, 75(3):1655–1678, 2012.

Chongjie Zhang and Victor Lesser. Multi-agent learning with policy prediction. In *AAAI Conference on Artificial Intelligence*, 2010.

Guojun Zhang, Pascal Poupart, and Yaoliang Yu. Optimality and stability in non-convex-non-concave min-max optimization. *arXiv preprint arXiv:2002.11875*, 2020a.

Guojun Zhang, Kaiwen Wu, Pascal Poupart, and Yaoliang Yu. Newton-type methods for minimax optimization. *arXiv preprint arXiv:2006.14592*, 2020b.

Kaiqing Zhang, Zhuoran Yang, and Tamer Başar. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *arXiv preprint arXiv:1911.10635*, 2019.

Liyuan Zheng, Tanner Fiez, Zane Alumbaugh, Benjamin Chasnov, and Lillian Ratliff. Stackelberg actor-critic: A game-theoretic perspective. *AAAI Workshop on Reinforcement Learning in Games*, 2021.