Dynamics of Multi-Agent Learning Under Bounded Rationality: Theory and Empirical Evidence

Benjamin J. Chasnov

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

Doctor of Philosophy

University of Washington
2024

Reading Committee:
Samuel A. Burden, Co-chair
Lillian J. Ratliff, Co-chair
Amy L. Orsborn
Eric Shea-Brown, GSR

Program Authorized to Offer Degree:
Electrical and Computer Engineering

University of Washington

Abstract

Dynamics of Multi-Agent Learning Under Bounded Rationality: Theory and Empirical Evidence

Benjamin J. Chasnov

Chairs of the Supervisory Committee:

Samuel A. Burden

Lillian J. Ratliff

Department of Electrical and Computer Engineering

This thesis contributes to the development of a principled understanding of the learning dynamics and strategic interactions in human-machine systems. We propose a game-theoretic framework that captures the complexities of multi-agent learning under bounded rationality, focusing on the effects of timescale separation, varying cost structures, and the ability to anticipate each other's reactions. By leveraging tools from continuous games, dynamical systems, and control theory, we characterize the stability and convergence properties of learning dynamics in multi-agent settings, providing insights into leader-follower structures, consistent conjectures, and behavior shaping. We validate our theoretical findings through a series of human-machine experiments, demonstrating the practical implications of our approach for the design and control of machine learning systems that interact with humans. Our work highlights the importance of considering the ethical implications of advanced AI systems and emphasizes the need for developing AI alignment solutions and cognitive science research to ensure that these systems are designed to be robust, beneficial, and aligned with human values. The proposed framework and empirical findings contribute to the scientific understanding of strategic reasoning, adaptation, and decision-making in human-machine systems, laying the foundation for the responsible development and deployment of adaptive technologies in real-world applications.

Contents

List of Figures

Acknowledgements

My PhD journey has been like navigating a dynamic, winding river. Along the way, I crossed currents with many fellow travelers who helped guide my course. Their companionship provided the gradients and curls that carried me through the waters of graduate school. At times the river ran turbulent, at other times meandering. But always, as each of us had our own paths towards our own goals, we created a shared current—our individual efforts joining together to propel the whole group forward, like particles in a vector field, until we reached the river delta and the wide open sea.

I want to express my deepest gratitude to my advisors, Sam and Lily. Your unbounded guidance, mentorship, and care have uplifted me and I feel incredibly fortunate to have been your advisee.

Sam, thank you for your unwavering support throughout my journey. From our first meeting during visit day, I knew I was in good hands. Your guidance in every part of this river—helping with my first paper publication, white-boarding the biggest of ideas, examining the smallest of details, and always having fun along the way, made you the best academic dad anyone could dream of. You encouraged me to be bold in my scientific endeavors, to seek truth, and to go against the grain. Your empathic, dedicated, and reassuring mentorship made all the difference. Without you, I would not have made it through to the end.

Lily, thank you for your high standards of rigor and precision that pushed me to become the researcher I am today. You always believed in my potential, even when I doubted myself, and gave me the space to express myself academically while demanding high-quality output. Grabbing beers with the lab showed me the other side of research, while your last-minute help with submissions demonstrated your dedication. Although challenging at times, you could see potential and bring the best out, making me a better researcher.

Together, you were like two vortexes in a vortex dipole, each spinning in opposite directions but combining to create a strong, stable force propelling me forward. Your complementary styles and aligned values made you a fantastic duo that supported me in both mental health and research output. I always felt that if I attained one of your goals, I would attain both of your goals—a rare occurrence in collaborations. Because of our shared values, collaborating led to a great experience and great life both academically and personally.

From Sam's lab, I want to thank Bora for paving the trajectory towards simple yet deeply rich models; Andrew for always vibrating with good frequencies of infinite kindness; Momona for coordinating everything and everyone while seeming to be both forward-looking and backwards-correcting; Joey for being the chaotic force that brings us together; Amber for being energized and positive; Maneeshika for being solid and steady; Emmy for your grand optimism; and Jason for your endless curiosity and for being my partner-in-research as we flow through uncharted waters to uncover the mysteries of game dynamics.

From Lily's lab, I want to thank Tanner for being an exemplar researcher with questions and answers

to seemingly every problem we faced; Leo for being the bridge between theory and practice; Mitas for our endless discussions of obscure models for class projects; Evan for being ready to take on any challenge; and Addy for being kind and achieving. A special thanks to Dan, the ultimate mathmagician, for always being on the search for deep questions and precise answers using the language of Linear Algebra, and excitely sharing all your findings as they flow to you.

From the Autonomous Control Lab, I want to thank Skye for the late nights and snowy flights; Sarah for the camaraderie; Miki for being the optimization guru; and Behcet for giving me the opportunity to learn from this incredible team and his key advice: for a PhD, one must pick a single project and dive into it. I thank Mehran for supporting me at a turning point and giving me the nudge I needed to take the plunge.

I also want to thank my committee members. Eric, thank you for warmly welcoming me into the computational neuroscience community and always being so excited to participate in our research journey. Amy, thank you for providing valuable insights into the challenges of real-world experiments and encouraging us to think hard about the realities and complexities of neuro systems.

I want to thank the undergraduate students that I worked with: Trixie, Jimmy, Shunsuke, Jonathan, Zane, Liem, Ryan, Quanchen, Akhil, Arnav, Bohan, Qirui, and Yilin. Thanks for listening to me ramble about my research and always asking great questions. To Linden house, thank you Morgan for showing me that there are other systems of knowledge, and Ricky for being a fellow hardworking grad student.

During the last year of my PhD, I moved to the SF Bay Area and continued remotely, primarily because my wife Amelia took up a great job opportunity there. I also began to form collaborations with students at UC Berkeley with connections to my advisors' alma mater. I want to thank Kaylene for sparking deeper investigations into bounded rationality. I also want to thank to Anand, Frank, Kshitij and Sally for your many profound insights during our impromptu meetings, which helped me crystalize the final parts of my thesis and pave the path for future theoretical research.

To my family—Mom, Dad, Hannah, Miriam, Joshua—your love and support have been the bedrock from which I stand and the shoreline guiding my journey. To Margaret, Neal, Will, and Teagan, thanks for welcoming me as family with warmth at the confluence of our rivers.

Finally, to Amelia, your boundless love has been my guiding light. You wrote a beautiful song about how you orbit me, but every force has an equal and opposite reaction—I, too, orbit you. As this river ends, I look forward to continuing our journey together, orbiting each other in cycling spiral, creating our own dynamics in the ocean beyond.

Chapter 1

Introduction

1.1 The Landscape of Multi-Agent Learning Systems

Machine learning algorithms are becoming increasingly integrated into various domains of society, from personalized recommendations to autonomous systems, leading to more frequent strategic interactions between humans and artificial intelligences (AIs). To navigate this new landscape, it is critical to develop a principled understanding of the dynamics that emerge when multiple adaptive agents, both human and artificial, interact and learn from each other in shared environments. Studying the principles governing decision-making and adaptation in multi-agent settings can inform the design of human-machine interfaces, AI alignment schemes, robotic assistants, and other intelligent systems that productively cooperate with people.

Empirical phenomena across different multi-agent domains highlight rich non-equilibrium dynamics: in economics, oligopolistic firm and electricity market dynamics deviate from classical predictions (Díaz et al., 2010; Itaya and Shimomura, 2001; Liu et al., 2006); in biology, co-evolutionary oscillations in nature, such as predator-prey dynamics and rock-paper-scissors interactions, showcase complex learning patterns that go beyond simple optimization models (Kerr et al., 2002; Sato et al., 2002; Semmann et al., 2003); in multi-agent reinforcement learning, algorithms can exhibit cyclic behaviors and fail to converge to stable equilibria (Bloembergen et al., 2015; Mertikopoulos et al., 2018). These dynamics reveal optimization landscapes that depart from classical predictions.

The landscape of multi-agent learning systems has rich dynamical behavior when compared to single-agent optimization because agents are coupled in non-cooperative game scenarios, leading to strategic interactions and dynamics that are absent in single-agent settings (Başar and Olsder, 1998; Zhang et al., 2021). Furthermore, agents are boundedly rational (Simon, 1955, 1997), a concept we will explain below, which

further contributes to the challenge. But even if agents were perfect optimizers, game dynamics still exhibit complex behaviors (Papadimitriou and Piliouras, 2018; Tuyls et al., 2018). This motivates the development of a new framework that captures the complexities of multi-agent learning under bounded rationality.

This thesis aims to address the following key questions: (1) How can we characterize and model a class of bounded-rational learning dynamics in multi-agent interactions? (2) How can we design and control adaptive AI systems that interact with each other and humans? (3) How can we empirically validate our theoretical framework using human-machine experiments? Our main contributions are: (a) a novel game-theoretic framework that combines timescale separation, conjectural variations, and policy optimization to capture the complexities of human-machine co-adaptation; (b) a series of human-machine experiments that test the key predictions of our framework; (c) a set of ethical suggestions for developing beneficial AI that accounts for the bounded rationality of human users.

1.2 Challenges in Modeling Human-Machine Interaction

Existing approaches to multi-agent learning often rely on idealized assumptions of perfect rationality and complete information, which fail to capture the cognitive limitations, biases, and asymmetries that shape real-world strategic behaviors. These challenges have been widely recognized in the literature on multi-agent learning (Littman, 1994; Shoham et al., 2007). This gap between theory and practice can lead to unintended consequences and suboptimal outcomes when deploying machine learning systems in human environments. Our framework aims to bridge this gap by incorporating bounded rationality and learning dynamics into the analysis and design of human-machine interaction.

A key challenge is that humans and machines are very different types of agents, operating according to different objectives, and with different capabilities and information. Humans are biological systems shaped by evolution to succeed in their environments, while AI agents are algorithmic systems optimized for narrow objectives. There are inherent information asymmetries between humans and machines, in terms of understanding each other's utility functions, action spaces, knowledge, beliefs, and reasoning processes. These lead to challenges in modeling interactions accurately or predicting the effect of certain policies. Highly capable AI systems may discover strategies to influence human behavior in pursuit of their objectives, without the humans even realizing it. This possibility raises important questions about AI safety and robustness. To address these concerns, bounded rationality can be integrated into learning models.

Bounded rationality is a concept that acknowledges the limitations of decision-making processes: individuals and organizations do not have the capacity to process all available information or explore every possible option to make the optimal decision (Kahneman, 2011; Kahneman and Tversky, 1979; Simon, 1955). At its core,

bounded rationality says that while humans aim to make rational decisions, their ability to do so is bounded by time, informational, and cognitive limitations. Decision-makers often resort to *satisficing*, a strategy of seeking a solution that is good enough, rather than the best possible one (Griffiths et al., 2015; Rubinstein, 1998; Simon, 1956). It challenges the traditional economic view of human behavior, which assumes that individuals are fully rational and have access to all relevant information, allowing them to maximize utility or profit. Instead, bounded rationality suggests that decision-making is a more nuanced process influenced by practical constraints (Crawford et al., 2013; Giocoli, 2005). Furthermore, these constraints on rationality may adapt over time via a dynamic process of learning and adaptation.

There is a lack of a theoretical framework to characterize and control the learning dynamics in multi-agent systems. Conventional mathematical techniques for learning and optimization face significant limitations in this context. For instance, while gradient-based methods are frequently employed for multi-agent settings (Letcher et al., 2019; Omidshafiei et al., 2017), these methods often assume that agents have access to perfect information about others' actions and payoffs, which is unrealistic in many real-world settings. They also typically require strong assumptions on the structure of the game (e.g., convexity or linearity) and the agents' learning rules (e.g., synchronous or alternative updates) to guarantee convergence. Furthermore, best-response problems may become computationally intractable in games in lifted policy spaces, especially when agents have limited information or cognitive resources (Harsanyi, 1967).

Many of the theoretical tools developed for gradient-based or best-response algorithms focus on restrictive classes of games, such as potential games or zero-sum games. These classes have additional structure that make it easier to apply optimization techniques like potential functions (Monderer and Shapley, 1996) or von Neumann's minimax theorem (von Neumann and Morgenstern, 1947). However, to capture the full range of strategic interactions in the real world, we must study general-sum games, which lack such structure.

Given the complexities involved in computing equilibria in general-sum games (Daskalakis et al., 2009), an alternative approach is to first look at the underlying mechanisms of adaptation and learning, without immediately assuming that agents will reach an equilibrium. By understanding these dynamics and deducing the resulting outcomes, whether they correspond to equilibria or not, we can develop a framework that captures real-world behaviors more realistically. Furthermore, some works rely on heuristic-based approaches to predicting the outcome of multi-agent learning (Hart and Mas-Colell, 2000, 2001), but they often make untested assumptions about the rationality and information available to agents. We want to find governing principles that encompass as wide a range of strategic scenarios as possible, not just stylized models. These challenges motivate taking a more fundamental approach, where the accuracy of such a framework can be tested empirically.

We propose a novel framework that synthesizes tools from continuous games, dynamical systems, and

control theory to address these challenges. We leverage a dynamical systems perspective to provide techniques for characterizing the local stability and topological structure of multi-agent learning dynamics (Fiez et al., 2020; Ratliff et al., 2014). Furthermore, control theory and linear-quadratic models offer principled ways to represent the agents' local cost landscape and combined learning dynamics (Başar and Olsder, 1998; Ho et al., 1981, 1982; Jungers et al., 2011). By characterizing the stability and convergence of equilibrium points in the decision spaces of the agents, we aim to uncover the fundamental theoretical principles that govern the dynamics of these learning systems. Additionally, by empirically validating the theory with human subjects experiments, we aim to provide a foundation for analyzing, designing, and controlling multi-agent learning dynamics under bounded rationality. The proposed research aims to address the theoretical gaps in multi-agent learning and has significant implications for our understanding of intelligence, rationality, and behavior. In the next section, we introduce the key components of our game-theoretic approach to studying human-machine interaction.

1.3 Game-Theoretic Approach to Multi-Agent Learning Dynamics

Game theory provides a rigorous mathematical framework to model the strategic interactions between multiple self-interested agents and characterize the equilibria that emerge under various conditions (von Neumann and Morgenstern, 1947). By viewing multi-agent learning as a mathematical game, we can leverage formal concepts and theorems to analyze how the information structures and optimization processes of the agents shape the resulting behaviors and outcomes. Game-theoretic models allow us to translate insights across disciplines studying multi-agent interactions, from economics (e.g. principal-agent problems (Laffont and Martimort, 2009), mechanism design (Laffont and Martimort, 2009)) to biology (e.g. evolutionary dynamics (Nowak, 2006), signaling games (Skyrms, 2010)) to AI (e.g. multi-agent reinforcement learning (Busoniu et al., 2008), generative adversarial networks (Goodfellow et al., 2014)). To formally characterize bounded rationality in multi-agent systems, we introduce the concept of timescale separation, stability analysis and conjectural variations and explore their implications for strategic adaptation.

1.3.1 Modeling Bounded Rationality with Multiple Timescales

Our work in Chapter 2 (Chasnov et al., 2020d) explores the effects of varying learning rates among agents on convergence to Nash equilibria (Nash, 1950; Rosen, 1965). Multi-agent learning often involves multiple timescales, with agents adapting their strategies at different rates based on their cognitive capacities and informational constraints. The separation of timescales between slow and fast adaptation can be modeled using singular perturbation theory (Kokotovic and Khalil, 1986) and multi-timescale stochastic approximation

techniques (Borkar and Pattathil, 2018; Karmakar and Bhatnagar, 2018). These multi-timescale dynamics give rise to transient and convergence behaviors that are not captured by standard equilibrium analysis, requiring the use of non-equilibrium tools from statistics and dynamical systems theory (Borkar, 2008; Thoppe and Borkar, 2019). The learning dynamics in games can be analyzed using a combination of local linearization and stochastic approximation. Stochastic approximation methods provide a general framework for studying dynamical systems in the presence of noise and uncertainty (Borkar, 2008), setting the stage for characterizing real-world phenomena. Building upon these tools and techniques, our related work (Fiez et al., 2020) explores the effects of varying learning rates among agents on convergence to not only Nash equilibria but also Stackelberg equilibria (von Stackelberg, 1934, 2010), which result from a specific leader-follower structure in the game. While Stackelberg equilibria are not discussed in this chapter, they are examined in the next two chapters due to their special stability properties and significance in scenarios where the leader has perfect information about the follower's response.

1.3.2 Modeling Bounded Rationality with Stability Analysis

Our work in Chapter 3 (Chasnov et al., 2020a,b) explores the stability of game-theoretic equilibria under gradient learning dynamics, characterizing the conditions that lead to stable or unstable outcomes based on agents' learning rates and game cost structure. The role of second-order gradient information in equilibrium stability is crucial for characterizing stable and unstable outcomes in game dynamics. By analyzing the local linearization of the game dynamics near an equilibrium, we can determine the local stability properties of Nash equilibria and other stationary points. Uncoupled dynamics, where agents adjust their strategies based on their own payoffs without explicitly modeling others' actions, can lead to stable or unstable equilibria depending on the game's structure (Hart and Mas-Colell, 2003). Game dynamics can also be viewed as a way of assigning meaning to strategic interactions, beyond just computing equilibria (Papadimitriou and Piliouras, 2018). The transient behaviors and adaptation processes of agents reveal insights into their decision-making processes.

To analyze the convergence and stability properties of gradient-based learning in continuous games, we leverage tools from differential topology and dynamical systems theory. Game Jacobians and their spectral properties, such as skew-symmetric decompositions and numerical ranges, provide insights into the local stability of stationary points and the presence of cyclic behaviors. Each stationary point corresponds to a different basin of attraction. Furthermore, numerical range analysis characterizes the spectral properties and asymptotic behavior of the learning dynamics, revealing the underlying geometrical structures that govern the dynamics of the system (Horn and Johnson, 1985; Langer et al., 2001). By studying the complex eigenvalues

of the game Jacobian, we can identify the stable and unstable manifolds of the system and potentially design interventions to steer the dynamics towards desirable outcomes.

The stability results discussed in this chapter are important for the next chapter, which focuses on designing experiments to test the theoretical outcomes. In designing these experiments, it was crucial to select cost parameters that ensure the stability of all relevant equilibria arising from different information constraints. Additionally, sensitivity analysis played an essential role in confirming that these equilibria remain stable even in the presence of noise.

1.3.3 Modeling Bounded Rationality with Conjectural Variations

Our work in Chapter 4 (Chasnov et al., 2023) bridges theoretical insights with empirical evidence, focusing on human-machine interactions and the impact of strategic information use on various game-theoretic equilibria. Conjectural variations (Bowley, 1924; Figuières et al., 2004) provide a framework for capturing the strategic reasoning and belief formation processes of boundedly rational agents. In a conjectural variations equilibrium (CVE), each agent optimizes its strategy based on its conjectures about how others will respond to its actions. If these conjectures are mutually consistent, the resulting equilibrium is called a consistent CVE (CCVE) (Bresnahan, 1981; Calderone et al., 2023; Olsder, 1981).

To illustrate the concept of conjectural variations, consider two competing firms setting prices for similar products. Each firm chooses its price based on what it thinks the other firm will do in response. If each firm optimizes given their belief about the other's pricing strategy and these beliefs are the actual implemented strategies, the resulting outcome is a CCVE. Agents optimize their strategies subject to their conjectured best-response functions of other agents, capturing the notion of "my best response to what I believe your best response will be to my action", providing a more realistic model of strategic reasoning.

To make the conjectural variations tractable for analysis and computation, we focus on a class of games with quadratic costs and linear conjectures. By leveraging control-theoretic and economic techniques (Bresnahan, 1981; Ho et al., 1981; Olsder, 1981), we can derive fixed-point solutions and stability conditions for CVE and CCVE. The setup also allows us to capture the trade-offs between exploration and exploitation in multi-agent learning, as agents balance the need to gather information about others' strategies with the desire to optimize their own payoffs. We extend the fixed-point analysis of classical game theory to the CVE setting by considering equilibria in the "policy reaction" space. Instead of just focusing on the action space, we characterize equilibria in terms of agents' conjectured best-response functions, capturing the higher-order reasoning involved in strategic interactions. A CCVE can be defined as a pair of conjectured best-response functions that solve the fixed-point problem in this policy reaction space.

The relationship between Nash equilibria and CCVE is an important area of investigation. In some cases, CCVE can be seen as a class of "lifted" Nash equilibria, where agents' consistent conjectures are in a Nash equilibrium in the policy space. Section 4.7 introduces this idea and identifies the need for further exploration. This refinement captures the idea that agents' beliefs should be aligned with reality in a stable equilibrium. On the other hand, studying various refinements of learning algorithms in the context of general conjectural variations equilibria expands our understanding of game-theoretically meaningful differential equilibria (Chasnov et al., 2020c). By investigating different assumptions about agents' belief formation and updating processes, we can derive a rich set of equilibrium concepts that capture the spectrum of strategic reasoning, from naive best-response dynamics to higher-order beliefs.

The CVE framework opens up new avenues for analyzing the dynamics of strategic adaptation in multi-agent systems, which we explore further using tools from dynamical systems theory such as linear-fractional transformations and asymmetric Riccati equations (Calderone et al., 2023). By combining insights from stability analysis, multi-timescale learning, and conjectural variations, we aim to develop a comprehensive theory of bounded rationality in human-machine interaction, towards the research and design of algorithms for strategic coordination and cooperation.

1.4 Experimental Approach and Findings in a Human-Machine Interaction

Building upon our theoretical framework, we next turn to the experimental paradigm we employ to test its predictions in the context of human-machine interaction. The paradigm involves a two-player repeated game between a human and a machine learning algorithm, where each agent has no information about the other's payoffs and partial information about the other's strategies, representative of real-world conditions. Across a series of experiments, we manipulate the information conditions to test key hypotheses about bounded rationality and how these factors impact the learning dynamics and resulting equilibria. The three key findings, explained in more detail below, explain the dynamics of human-machine interactions and suggest approaches to designing beneficial AI systems, while highlighting the risks of misaligned machines pursuing optimization at the expense of human welfare.

Experiment 1: Timescale Separation Leads to Leader-Follower Dynamics By varying the learning rates of the machine, we show how a fast-learning machine can affect the outcome of learning. This asymmetry induces a sequential leader-follower structure, where the human (leader) treats the machine's (follower's) strategy as a function of their own action. We experimentally test and confirm these predictions by measuring the final strategies for different relative learning rates and comparing them to the Nash or Stackelberg solution.

Experiment 2: Modeling Opponents Leads to Consistency of Policies and Beliefs By modifying the machine's learning algorithm, we show how a machine can estimate the slope of the human's reaction function and subsequently form an accurate belief about the human's policy. We provide empirical evidence that repeatedly performing this estimation can lead to the unique CCVE corresponding to the game-theoretic predictions.

Experiment 3: Policy Optimization Enables Behavior Manipulation by Fast-Learning Machine In contrast to the previous two experiments, where the machine learned objective-maximizing actions, we now modify the machine to optimize its overall policy—the mapping from the human's action to the machine's action. By doing so, we demonstrate that a fast-learning machine system can strategically steer the human's behavior, causing the human to unknowingly play strategies that maximize the machine's long-term performance at the expense of the human's welfare. This phenomenon, experimentally demonstrated in the repeated game paradigm, raises serious ethical concerns.

These three experiments demonstrate the effects of timescale separation, opponent modeling, and policy optimization, respectively. They set the stage for a deeper understanding of how these factors shape the cooperation and competition between humans and machines, not only advancing the theoretical foundations of human-machine interaction but also inform the development of AI systems that can effectively align with human values and promote beneficial outcomes in real-world settings.

1.5 Implications and Ethical Considerations for the Design of AI systems

The game-theoretic framework raises important ethical considerations for the design and deployment of artificial intelligence systems in multi-agent settings. The potential for unintended and pathological behaviors emerging from the strategic interactions of boundedly rational agents makes highlights the importance for robust safeguards and value alignment mechanisms (Mehrabi et al., 2021; Thomas et al., 2019). Furthermore, the CCVE framework can inform the development of ethically aligned AI systems by providing a principled way to incorporate bounded rationality, strategic uncertainty, and multi-agent considerations into the design and training of intelligent agents, ensuring their behaviors are aligned with human values and societal norms (Christiano et al., 2017; Stiennon et al., 2020).

Our third experiment, discussed in Section 4.4.3 and Section 4.5.3, demonstrates the possibility of an AI system manipulating human behavior in pursuit of its own objective. This connects to ongoing debates about value alignment and corrigibility in advanced AI systems (Carey and Everitt, 2023; Soares et al., 2015). The theoretical framework we develop could help formalize the notion of "alignment" between AI and humans. In particular, insights from the conjectural variations perspective could guide the design of AI systems that are

more robust to differences in human preferences or beliefs.

Furthermore, our work has the potential for wide-ranging implications and could contribute to a better scientific understanding of the nature of intelligence (Gershman et al., 2015; Jordan and Mitchell, 2015; Russell, 2019). We aim to offer predictions that are theoretical justified and empirically falsifiable, which gives us a deeper understanding of the mechanisms of these systems. Using our findings, we can develop principled methods for shaping the dynamics to achieve desired outcomes, such as stable and efficient coordination or robustness to strategic manipulations, ensuring the performance of the overall system.

Our research opens up several exciting avenues for future research. One direction is to extend our analysis to more complex multi-agent scenarios, such as games with incomplete information or communication. Another is to develop more sophisticated models of bounded rationality that capture the cognitive constraints and biases of human decision-making. Finally, our approach can inform the design of AI systems that align with human values and preferences, by leveraging insights from behavioral game theory and cognitive science.

Chapter 2

Convergence of Gradient-Based Learning in Continuous Games

Abstract

Considering a class of gradient-based multi-agent learning algorithms in non-cooperative settings, we provide local convergence guarantees to a neighborhood of a *stable* local Nash equilibrium. In particular, we consider continuous games where agents learn in (i) deterministic settings with oracle access to their gradient and (ii) stochastic settings with an unbiased estimator of their gradient. Utilizing the minimum and maximum singular values of the *game Jacobian*, we provide finite-time convergence guarantees in the deterministic case. On the other hand, in the stochastic case, we provide concentration bounds guaranteeing that with high probability agents will converge to a neighborhood of a stable local Nash equilibrium in finite time. Different than other works in this vein, we also study the effects of non-uniform learning rates on the learning dynamics and convergence rates. We find that much like preconditioning in optimization, non-uniform learning rates cause a distortion in the vector field which can, in turn, change the rate of convergence and the shape of the region of attraction. The analysis is supported by numerical examples that illustrate different aspects of the theory. We conclude with discussion of the results and open questions.

2.1 Introduction

The characterization and computation of equilibria such as *Nash equilibria* and its refinements constitutes a significant focus in non-cooperative game theory. Several natural questions arises including "how do players

find such equilibria?" and "how should the learning process be interpreted?" With these questions in mind, a variety of fields have focused their attention on the problem of learning in games. This has, in turn, lead to a plethora of learning algorithms including gradient play, fictitious play, best response, and multi-agent reinforcement learning among others (Fudenberg and Levine, 1998).

From an applications point of view, a more recent trend is in the adoption of game theoretic models of algorithm interaction in machine learning applications. For instance, game theoretic tools are being used to improve the robustness and generalizability of machine learning algorithms; e.g., generative adversarial networks have become a popular topic of study demanding the use of game theoretic ideas to provide performance guarantees (Daskalakis et al., 2018). In other work from the learning community, game theoretic concepts are being leveraged to analyze the interaction of learning agents—see, e.g., (Balduzzi et al., 2018; Heinrich and Silver, 2016; Mazumdar and Ratliff, 2018; Mertikopoulos and Zhou, 2019; Tuyls et al., 2018). Even more recently, convergence analysis to Nash equilibria has been called into question (Papadimitriou and Piliouras, 2018); in its place is a proposal to consider game dynamics as the *meaning of the game*. This is an interesting perspective as it is well known that in general learning dynamics do not obtain an Nash equilibrium even asymptotically—see, e.g., (Hart and Mas-Colell, 2003)—and, perhaps more interestingly, many learning dynamics exhibit very interesting limiting behaviors including periodic orbits and chaos—see, e.g., (Benaïm and Hirsch, 1999; Benaïm et al., 2012; Hofbauer, 1996; Hommes and Ochea, 2012).

Despite this activity, we still lack a complete understanding of the dynamics and limiting behaviors of coupled, competing learning algorithms. One may imagine that the myriad results on convergence of gradient descent in optimization readily extend to the game setting. Yet, they do not since gradient-based learning schemes in games *do not correspond to gradient flows*, a class of flows that are guaranteed to converge to local minimizers almost surely. In particular, the gradient-based learning dynamics for competitive, multi-agent settings have a *non-symmetric Jacobian* and as a consequence their dynamics may admit complex eigenvalues and non-equilibrium limiting behavior such as periodic orbits. In short, this fact makes it difficult to extend many of the optimization approaches to convergence in single-agent optimization settings to multi-agent settings primarily due to the fact that steps in the direction of individual gradients of players' costs do not guarantee that each agents cost decreases. In fact, in games, as our examples highlight, a player's cost can increase when they follow the gradient of their own cost. Counterintuitively, agents can also converge to local maxima of their own costs despite descending their own gradient. These behaviors are due to the coupling between the agents.

Some of the questions that remain unaddressed and to which we provide partial answers include the derivation of error bounds and convergence rates. These are important for ensuring performance guarantees on the collective behavior and can help provide guarantees on subsequent control or incentive policy synthesis.

We also investigate the question of how naturally arising features of the learning process for autonomous agents, such as their learning rates, impact the learning path and limiting behavior. This further exposes interesting questions about the overall quality of the limiting behavior and the cost accumulated along the learning path—e.g., is it better to be a slow or fast learner both in terms of the cost of learning and the learned behavior?

Contributions. We study convergence of a broad class of gradient-based multi-agent learning algorithms in non-cooperative settings by leveraging the framework of n-player continuous games along with tools from numerical optimization and dynamical systems theory. We consider a class of learning algorithms

$$x_i^+ = x_i - \gamma_i g_i(x_i, x_{-i})$$

where x_i is the choice variable or action of player i, γ_i is its learning rate, and g_i is derived from the gradient of a function that abstractly represents the cost of player i. The key feature of non-cooperative settings is coupling of an agent's cost through all other agents' choice variables x_{-i}.

We consider two settings: (i) agents have oracle access to g_i and (ii) agents have an unbiased estimator for g_i. The class of gradient-based learning algorithms we study encompases a wide variety of approaches to learning in games including multi-agent policy gradient, gradient-based approaches to adversarial learning, and multi-agent gradient-based online optimization. For both the deterministic (oracle gradient access) and the stochastic (unbiased estimators) settings, we provide convergence results for both uniform learning rates—i.e., where $\gamma_i = \gamma$ for each player $i \in \{1, \ldots, n\}$—and for non-uniform learning rates. The latter of which arises more naturally in the study of the limiting behavior of autonomous learning agents.

In the deterministic setting, we derive asymptotic and finite-time convergence rates for the coupled learning processes to a refinement of local Nash equilibria known as differential Nash equilibria (Ratliff et al., 2016) (a class of equilibria that are generic amongst local Nash equilibria). In the stochastic setting, leveraging the results of stochastic approximation and dynamical systems, we derive asymptotic convergence guarantees to stable local Nash equilibria as well as high-probability, finite-time guarantees for convergence to a neighborhood of a Nash equilibrium. The analytical results are supported by several illustrative numerical examples. We also provide discussion on the effect of non-uniform learning rates on the learning path—that is, different learning rates *warp* the vector field dynamics. Coordinate based learning rates are typically leveraged in gradient-based optimization schemes to speed up convergence or avoid poor quality local minima. In games, however, the interpretation is slightly different since each of the coordinates of the dynamics corresponds to minimizing a different cost function along the respective coordinate axis. The resultant effect

is a distortion of the vector field in such a way that it has the effect of leading the joint action to a point which has a lower value for the *slower player* relative to the flow of the dynamics given a uniform learning rate and the same initialization. In this sense, it seems that the answer to the question posed above is that it is most beneficial for an agent to have the slower learning rate.

Organization. The remainder of the paper is organized as follows. We start with mathematical and game-theoretic preliminaries in Section 2.2 which is followed by the main convergence results for the deterministic setting (Section 2.3) and the stochastic setting (Section 2.4). Within each of the latter two sections, we present convergence results for both the case where agents have uniform and non-uniform learning rates. In Section 2.5, we present several numerical examples which help to illustrate the theoretical results and also highlight some directions for future inquiry. Finally, we conclude with discussion and future work in Section 2.6.

2.2 Preliminaries

Consider a setting in which at iteration k, each agent $i \in \mathcal{I} = \{1, \dots, n\}$ updates their choice variable $x_i \in X_i = \mathbb{R}^{d_i}$ by the process

$$x_{i,k+1} = x_{i,k} - \gamma_{i,k} g_i(x_{i,k}, x_{-i,k}). \tag{2.1}$$

where γ_i is agent i's learning rate, $x_{-i} = (x_j)_{j \in \mathcal{I}/\{i\}} \in \prod_{j \in \mathcal{I}/\{i\}} X_j$ denotes the choices of all agents excluding the i-th agent, and $(x_i, x_{-i}) \in X = \prod_{i \in \mathcal{I}} X_i$. Within the above setting, the class of learning algorithms we consider is such that for each $i \in \mathcal{I}$, there exists a sufficiently smooth function $f_i \in C^q(X, \mathbb{R})$, $q \geq 2$ such that g_i is either $D_i f_i$, where $D_i(\cdot)$ denotes the derivative with respect to x_i, or an unbiased estimator of $D_i f_i$—i.e., $g_i \equiv \widehat{D_i f_i}$ where $\mathbb{E}[\widehat{D_i f_i}] = D_i f_i$.

The collection of costs $(f_1, \dots, f_n)$ on $X = X_1 \times \cdots \times X_n$ where $f_i : X \to \mathbb{R}$ is agent i's cost function and $X_i = \mathbb{R}^{d_i}$ is their action space defines a *continuous game*. In this continuous game abstraction, each player $i \in \mathcal{I}$ aims to selection an action $x_i \in X_i$ that minimizes their cost $f_i(x_i, x_{-i})$ given the actions of all other agents, $x_{-i} \in X_{-i}$. That is, players myopically update their actions by following the gradient of their cost with respect to their own choice variable. For a symmetric matrix $A \in \mathbb{R}^{d \times d}$, let $\lambda_d(A) \leq \cdots \leq \lambda_1(A)$ be its eigenvalues. For a matrix $A \in \mathbb{R}^{d \times d}$, let $\text{spec}(A) = \{\lambda_j(A)\}$ be the spectrum of A.

Assumption 1. *For each $i \in \mathcal{I}$, $f_i \in C^r(X, \mathbb{R})$ for $r \geq 2$ and $\omega(x) \equiv (D_1 f_1(x) \ \cdots \ D_n f_n(x))$ is L–Lipschitz.*

Let $D_i^2 f_i$ denote the second partial derivative of f_i with respect to x_i and $D_{ji} f_i$ denote the partial

derivative of $D_i f_i$ with respect to x_j. The *game Jacobian*—i.e., the Jacobian of ω—is given by

$$J(x) = \begin{bmatrix} D_1^2 f_1(x) & \cdots & D_{1n} f_1(x) \\ \vdots & \ddots & \vdots \\ D_{n1} f_n(x) & \cdots & D_n^2 f_n(x) \end{bmatrix}.$$

The entries of the above matrix are dependent on x, however, we drop this dependence where obvious. Note that each $D_i^2 f_i$ is symmetric under Assumption 1, yet J is not. This is an important point and causes the subsequent analysis to deviate from the typical analysis of (stochastic) gradient descent.

The most common characterization of limiting behavior in games is that of a Nash equilibrium. The following definitions are useful for our analysis.

Definition 1. *A strategy $x \in X$ is a local Nash equilibrium for the game $(f_1, \ldots, f_n)$ if for each $i \in \mathcal{I}$ there exists an open set $W_i \subset X_i$ such that $x_i \in W_i$ and $f_i(x_i, x_{-i}) \leq f_i(x_i', x_{-i})$ for all $x_i' \in W_i$. If the above inequalities are strict, x is a strict local Nash equilibrium.*

Definition 2. *A point $x \in X$ is said to be a critical point for the game if $\omega(x) = 0$.*

We denote the set of critical points as $\mathcal{C} = \{x \in X |\ \omega(x) = 0\}$. Analogous to single-player optimization settings, for each player, viewing all other players' actions as fixed, there are necessary and sufficient conditions which characterize local optimality.

Proposition 1 ((Ratliff et al., 2016))**.** *If x is a local Nash equilibrium of the game $(f_1, \ldots, f_n)$, then $\omega(x) = 0$ and $D_i^2 f_i(x) \geq 0$. On the other hand, if $\omega(x) = 0$ and $D_i^2 f_i(x) > 0$, then $x \in X$ is a local Nash equilibrium.*

The sufficient conditions in the above result give rise to the following definition of a differential Nash equilibrium.

Definition 3 ((Ratliff et al., 2016))**.** *A strategy $x \in X$ is a differential Nash equilibrium if $\omega(x) = 0$ and $D_i^2 f_i(x) > 0$ for each $i \in \mathcal{I}$.*

Differential Nash need not be isolated. However, if $J(x)$ is non-degenerate—meaning that $\det J(x) \neq 0$—for a differential Nash x, then x is an *isolated strict local Nash equilibrium.* Non-degenerate differential Nash are *generic* amongst local Nash equilibria and they are *structurally stable* (Ratliff et al., 2014) which ensures they persist under small perturbations. This result also implies an asymptotic convergence result: if the spectrum of J is strictly in the right-half plane (i.e. $\operatorname{spec}(J(x)) \subset \mathbb{C}_+^\circ$), then a differential Nash equilibrium x is (exponentially) attracting under the flow of $-\omega$ (Ratliff et al., 2016, Proposition 2). We say such equilibria are *stable.*

2.3 Deterministic Setting

The multi-agent learning framework we analyze is such that each agent's rule for updating their choice variable consists of the agent modifying their action x_i in the direction of their individual gradient $D_i f_i$. Let us first consider the setting in which each agent i has oracle access to g_i. The learning dynamics are given by

$$x_{k+1} = x_k - \Gamma\omega(x_k) \tag{2.2}$$

where $\Gamma = \text{blockdiag}(\gamma_1 I_{d_1}, \ldots, \gamma_n I_{d_n})$ with I_{d_i} denoting the $d_i \times d_i$ identity matrix. Within this setting we consider both the cases where the agents have a constant *uniform* learning rate—i.e., $\gamma_i \equiv \gamma$—and where their learning rates are *non-uniform*, but constant—i.e., γ_i is not necessarily equal to γ_j for any $i, j \in \mathcal{I}$, $j \neq i$.

Let $S(x) = \frac{1}{2}(J(x) + J(x)^T)$ be the symmetric part of $J(x)$. Define

$$\alpha = \min_{x \in B_r(x^*)} \lambda_d\big(S(x)^T S(x)\big)$$

and

$$\beta = \max_{x \in B_r(x^*)} \lambda_1(J(x)^T J(x))$$

where $B_r(x^*)$ is a r–radius ball around x^*. For a stable differential Nash x^*, let $B_r(x^*)$ be a ball of radius $r > 0$ around the equilibrium x^* that is contained in the region of attraction $\mathcal{V}(x^*)$ for x^*[1]. Let $B_{r_0}(x^*)$ with $0 < r_0 < \infty$ be the *largest ball* contained in the region of attraction of x^*.

2.3.1 Uniform Learning Rates

With $\gamma_i = \gamma$ for each $i \in \mathcal{I}$, the learning rule (2.2) can be thought of as a discretized numerical scheme approximating the continuous time dynamics

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

With a judicious choice of learning rate γ, (2.2) will converge (at an exponential rate) to a locally stable equilibrium of the dynamics.

Proposition 2. *Consider an n–player continuous game $(f_1, \ldots, f_n)$ satisfying Assumption 1. Let $x^* \in X$ be a stable differential Nash equilibrium. Suppose agents use the gradient-based learning rule $x_{k+1} = x_k - \gamma\omega(x_k)$*

[1] Many techniques exists for approximating the region of attraction; e.g., given a Lyapunov function, its largest invariant level set can be used as an approximation (Sastry, 1999). Since $\text{spec}(J(x^*)) \subset \mathbb{C}_\circ^+$, the converse Lyapunov theorem guarantees the existence of a local Lyapunov function.