Optimization and Generalization of Minimax Algorithms

by

Sarath Pattathil

B.Tech., Indian Institute of Technology, Bombay (2018)
M.Tech., Indian Institute of Technology, Bombay (2018)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2023

Authored by:	Sarath Pattathil Department of Electrical Engineering and Computer Science July 20, 2023
Certified by:	Asuman Ozdaglar MathWorks Professor of Electrical Engineering and Computer Science Thesis Supervisor
Accepted by:	Leslie A. Kolodziejski Professor of Electrical Engineering and Computer Science Chair, Department Committee on Graduate Students

Optimization and Generalization of Minimax Algorithms

by

Sarath Pattathil

Submitted to the Department of Electrical Engineering and Computer Science
on July 20, 2023, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis explores minimax formulations of machine learning and multi-agent learning problems, focusing on algorithmic optimization and generalization performance. The first part of the thesis delves into the smooth convex-concave minimax problem, providing a unified analysis of widely used algorithms such as Extra-Gradient (EG) and Optimistic Gradient Descent Ascent (OGDA), whose convergence behavior was not systematically understood. We derive convergence rates for these algorithms in the convex-concave setting. We show that these algorithms work effectively due to their approximation of the Proximal Point (PP) method, which converges to the solution at a fast rate, but is impractical to implement. In the next chapter, we expand our study to nonconvex-nonconcave problems. These problems are generally challenging to solve, as a solution may not be well defined, or even if a solution exists, its computation may not be tractable. We identify a class of nonconvex-nonconcave problems that do have well defined and computationally tractable solutions. Leveraging the concepts developed in the first chapter, we design algorithms to efficiently tackle this special class of nonconvex-nonconcave problems. The final part of this thesis addresses the issue of generalization. In many cases, such as GANs and adversarial training, the objective function for finding the saddle point can be written as an expected value over the data distribution. However, since we often do not have direct access to this distribution, we solve the empirical problem instead, which involves averaging over the available dataset. The final chapter aims to evaluate the quality of solutions to the empirical problem compared to the original population problem. Existing metrics like the primal risk, which are used to assess generalization in the minimax setting are found to be inadequate in capturing the generalization of minimax learners. This prompts the proposal of a new metric, the primal gap, which overcomes these limitations. This novel metric is then utilized to investigate the generalization performance of popular algorithms like Gradient Descent Ascent (GDA) and Gradient Descent-Max (GDMax).

Thesis Supervisor: Asuman Ozdaglar
Title: Professor of Electrical Engineering and Computer Science

Acknowledgments

I would like to express my deepest gratitude to my advisor, Asu Ozdaglar, whose energy and enthusiasm was the main driving force behind this thesis. Her invaluable guidance, unwavering support, and continuous encouragement throughout my PhD is what kept me motivated during my journey here. I would like to thank her from the bottom of my heart. I am also deeply thankful to my committee members Costis Daskalakis, Gabriele Farina, and Kaiqing Zhang for providing valuable feedback and comments which have made this thesis better.

My journey at MIT has been so much more fruitful because of the various collaborators that I've had over the years. I would like to thank Daron Acemoglu, Costis Daskalakis, Alireza Fallah, Noah Golowich, Aryan Mokhtari, Francesca Parise, Jiawei Zhang, and Kaiqing Zhang for the several projects that we have worked on, spanning from optimization and reinforcement learning, to epidemic control and network economics.

I am also thankful to LIDS/CSAIL and the MIT community at large for making my time here so enjoyable. I have been fortunate to have had many friends at MIT and I am thankful for all the times we've spent playing soccer, badminton, tennis, taking road trips and just hanging out!

Last but certainly not least, I am deeply indebted to my family. Their unwavering belief in me, their sacrifices, and their constant encouragement have been my pillars of strength. I am forever grateful for their unconditional love, patience, and understanding.

Grants and fundings: My Ph.D. research was generously supported by the MathWorks Engineering fellowship, and the Michael Athans Fellowship. In addition, my research was also supported by the following grants: Connection Science Center funding, Army Research Office grant W911NF-18-1-0407, and Air Force Research Laboratory grant FA8750-19-2-1000.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Socio-technological systems are all around us - in transportation, energy, social media and healthcare. They are evolving at breakneck speeds thanks to abundance of online and sensory data and advances in AI/ML technologies. These decentralized systems also involve multiple agents interacting and learning both about each other's strategies and uncertain changing environments. Motivated by these applications, my research investigates design of algorithms for robust training of ML models and stable learning dynamics in multi-agent systems.

Training such ML models involves the study of robust algorithms. This involves designing algorithms which are robust to input perturbations, domain shifts, and task adaptations. Development of such robust algorithms for multi-agent systems motivates the study of minimax formulations. In particular, given a function $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ (where $\mathcal{X} \subseteq \mathbb{R}^m$ and $\mathcal{Y} \subseteq \mathbb{R}^n$), we consider finding a saddle point of the problem

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \; f(x, y),$$

where a saddle point of this problem is defined as a pair $(x^*, y^*) \in \mathcal{X} \times \mathcal{Y}$ that satisfies

$$f(x^*, y) \leq f(x^*, y^*) \leq f(x, y^*)$$

for all $x \in \mathcal{X}, y \in \mathcal{Y}$. This formulation arises in several areas, including zero-sum games [10], robust optimization [15], robust control [62] and more recently in machine learning in the context of Generative Adversarial Networks (GANs) and Adversarial Training [80].

In Chapter 2, we focus on the unconstrained version of the problem. Furthermore, we assume that the function $f(x, y)$ is *convex-concave*, i.e., for any $y \in \mathbb{R}^n$, the function $f(x, y)$ is a convex function of x and for any $x \in \mathbb{R}^m$, the function $f(x, y)$ is a concave function of y.

Our goal in this chapter is to analyze the convergence rate of some discrete-time gradient based optimization algorithms for finding a saddle point in the convex-concave case. In particular, we focus on Extra-gradient (EG) and Optimistic Gradient Descent Ascent (OGDA) methods because of their widespread use for training GANs (see [34, 75]). We provide the first unified convergence analysis for establishing a sublinear convergence rate of $\mathcal{O}(1/k)$ in terms of the primal-dual gap of the averaged iterates for both OGDA and EG for convex-concave saddle point problems.

In the next chapter, we move our focus onto certain structured nonconvex-nonconcave minimax problems. The specific structure we will be dealing with is the case where the variables x and y, can be written as a softmax parametrization of unconstrained variables θ and ν respectively. The main motivating example for this setting is multi-agent Reinforcement learning (RL).

In this chapter, we study the global convergence of Natural Policy Gradient (NPG) [68], which forms the basis for many popular Policy Gradient (PG) algorithms (e.g., Proximal Policy Optimization (PPO)/Trust Region Policy Optimization (TRPO)), in the parameter space and for multi-agent learning. We are interested in the setting where the agents take *symmetric* roles and operate *independently*, as it does not require a central coordinator and it scales favorably with the number of agents. We develop symmetric variants of the NPG method, both without and with the optimistic updates (similar to the optimistic updates in Chapter 2) and establish the last-iterate global convergence to the Nash equilibrium in the policy parameter space.

Finally, in Chapter 4, we move our attention from the problem of optimization, to the problem of generalization. Stochastic minimax optimization, a classical and fundamental problem in operations research and game theory, involves solving the following problem:

$$\min_{w \in W} \max_{\theta \in \Theta} E_{z \sim P_z}[f(w, \theta; z)].$$

More recently, such minimax formulations have received increasing attention in machine learning, with significant applications in generative adversarial networks (GANs) [54], adversarial learning [80], and reinforcement learning [27, 30]. Most existing works (including the first two chapters!) have focused on the *optimization* aspect of the problem, However, the optimization aspect is not sufficient to achieve the success of stochastic minimax formulations in machine learning. In particular, as in classical supervised learning, which is usually studied as a *minimization* problem [63], the out-of-sample *generalization* performance is a key metric for evaluating the learned models.

Existing works have studied *primal risk* and/or (variants of) *primal-dual risk* under different convexity and smoothness assumptions of the objective. Primal risk (see formal definition in §4.2) is a natural extension of the definition of risk from minimization problems. Primal-dual risk, on the other hand, is defined similarly but based on the duality gap of the solution.

Although these metrics are natural extensions of generalization metrics from the *minimization* setting, they might not be the most suitable ones for studying generalization in stochastic *minimax* optimization, especially in the *nonconvex* settings that is pervasive in machine/deep learning applications, where the global saddle-point might not exist.

In this final chapter, we first identify the inadequacies of the existing metric, and proposing a new metric, the *primal gap* that overcomes these inadequacies. We then provide generalization error bounds for the newly proposed metric, and discuss how it captures information not included in the other existing metrics.

Structure of the Thesis

The rest of the thesis is organized as follows. In Chapter 2, we study the convergence rates of OGDA and EG for smooth convex-concave minimax problems. Then, in Chapter 3, we move on to study algorithms to solve certain structured nonconvex-nonconcave problems which appear in Reinforcement learning formulations. Finally, in Chapter 4, we study the problem of generalization, and propose a new metric which correctly identifies the generalization capabilities of several popular algorithms which are used to solve minimax problems. Note that, we defer the proofs of all results in the chapter to the appendix which can be found at the end of each chapter.

Chapter 2

Convex-Concave Minimax Problems

2.1 Introduction

In this chapter, we consider finding a saddle point of the problem

$$\min_{x\in\mathbb{R}^m}\max_{y\in\mathbb{R}^n} f(x,y), \tag{2.1.1}$$

We focus on two popular algorithms used to solve this problem: Extra-gradient (EG) and Optimistic Gradient Descent Ascent (OGDA).

EG method is a classical algorithm for solving saddle point problems introduced in [70]. Its linear rate of convergence for smooth and strongly convex-strongly concave functions $f(x,y)$ [1] and bilinear functions, i.e., $f(x,y) = x^\top Ay$ (where A is a square, full rank matrix), was established in [70] as well as the variational inequality literature (see [124] and [41]). Its $\mathcal{O}(1/k)$ convergence rate for the constrained convex-concave setting was first established by [94] under the assumption that the feasible set is convex and compact.[2] [91] established a similar $\mathcal{O}(1/k)$ convergence rate for EG without assuming compactness of the feasible set by using a new termination criterion that relies on enlargement of the operator of the VI reformulation of the saddle point problem defined in [21]. OGDA was introduced by [103],

[1] Note that when we state that $f(x,y)$ is strongly convex-strongly concave, it means that $f(\cdot,y)$ is strongly convex for all $y\in\mathbb{R}^n$ and $f(x,\cdot)$ is strongly concave for all $x\in\mathbb{R}^m$.

[2] The result in [94] shows a $\mathcal{O}(1/k)$ convergence rate for the mirror-prox algorithm which specializes to the EG method for the Euclidean case.

as a variant of the Extragradient method, and has gained popularity recently due to its performance in training GANs (see [34]). To the best of our knowledge, iteration complexity of OGDA for the convex-concave case has not been studied before.

In this chapter, we provide a unified convergence analysis for establishing a sublinear convergence rate of $\mathcal{O}(1/k)$ in terms of the primal-dual gap of the averaged iterates and a saddle point for both OGDA and EG for convex-concave saddle point problems. Our analysis holds for unconstrained problems and does not require boundedness of the feasible set, and it establishes rate results using the primal-dual gap, as used in [94] (suitably redefined for an unconstrained feasible set, see Section 2.5). Therefore, we get convergence of the EG method in unconstrained spaces without using the modified termination (error) criterion proposed in [91]. The key idea of our approach is to view both OGDA and EG iterates as approximations of the iterates of the proximal point method that was first introduced in [83] and later studied in [109]. We would like to add that the idea of interpreting EG as an approximation of the Proximal Point method was first studied in [94]. This paper considers the conceptual mirror prox, which is similar to the proximal point method, and shows that the mirror prox algorithm (of which EG is a special case) provides a good implementable approximation to this method. Further, [91] use a similar interpretation and propose the Hybrid Proximal Extragradient method to establish the convergence of EG in unbounded settings using a different convergence criteria. More recently, [88] study both OGDA and EG as approximations of proximal point method and analyze these algorithms for bilinear and strongly convex-strongly concave problems.

More specifically, we first consider a proximal point method with error and establish some key properties of its iterates. We then focus on OGDA as an approximation of proximal point method and use this connection to show that the iterates of OGDA remain in a compact set. We incorporate this result to prove a sublinear convergence rate of $\mathcal{O}(1/k)$ for the primal-dual gap of the averaged iterates generated by the OGDA update. We next consider EG where two gradient pairs are used in each iteration, one to compute a midpoint and other to find the new iterate using the gradient of the midpoint. Our first step again is to show boundedness of the iterates generated by EG. We then approximate the evolution of the midpoints using a proximal point method and use this approximation to establish $\mathcal{O}(1/k)$ convergence rate for the function value of the averaged iterates generated by EG.

Related Work

Several recent papers have studied the convergence rate of OGDA and EG for the case when the objective function is bilinear or strongly convex-strongly concave. [34] showed the convergence of the OGDA iterates to a neighborhood of the solution when the objective function is bilinear. [75] used a dynamical system approach to prove the linear convergence of the OGDA method for the special case when $f(x, y) = x^\top Ay$ and the matrix A is square and full rank. They also presented a linear convergence rate of the vanilla Gradient Ascent Descent (GDA) method when the objective function $f(x, y)$ is strongly convex-strongly concave. [51] considered a variant of the EG method, relating it to OGDA updates, and showed the linear convergence of the corresponding EG iterates in the case where $f(x, y)$ is strongly convex-strongly concave (though without showing the convergence rate for the OGDA iterates). Optimistic gradient methods have also been studied in the context of convex online learning [29, 107, 108].

[93] analyzed the (sub)Gradient Descent Ascent (GDA) algorithm for convex-concave saddle point problems when the (sub)gradients are bounded over the constraint set, showing a convergence rate of $\mathcal{O}(1/\sqrt{k})$ in terms of the function value difference of the averaged iterates and a saddle point.

[25] focused on a particular case of the saddle point problem where the coupling term in the objective function is bilinear, i.e., $f(x, y) = G(x) + x^\top Ky - H(y)$ with G and H convex functions. They proposed a proximal point based algorithm which converges at a rate $\mathcal{O}(1/k)$ and further showed linear convergence when the functions G and H are strongly convex. [28] proposed an accelerated variant of this algorithm when G is smooth and showed an optimal rate of $(\frac{L_G}{k^2} + \frac{L_K}{k})$, where L_G and L_K are the smoothness parameters of G and the norm of the linear operator K respectively. When the functions G and H are strongly convex, primal-dual gradient-type methods converge linearly, as shown in [26, 12]. Further, [39] showed that GDA achieves a linear convergence rate in this linearly coupled setting when G is convex and H is strongly convex.

For the case that $f(x, y)$ is strongly concave with respect to y, but possibly nonconvex with respect to x, [111] provided convergence to a first-order stationary point using an algorithm that requires running multiple updates with respect to y at each step.

2.2 Preliminaries

In this section we present properties and notations used in our results.

Definition 2.2.1. A function $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ is L-smooth if it has L-Lipschitz continuous gradients on $\mathbb{R}^n$, i.e., for any $x, \widehat{x} \in \mathbb{R}^n$, we have

$$||\nabla\phi(x) - \nabla\phi(\widehat{x})|| \leq L||x - \widehat{x}||.$$

Definition 2.2.2. A continuously differentiable function $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex on $\mathbb{R}^n$ if for any $x, \widehat{x} \in \mathbb{R}^n$, we have

$$\phi(\widehat{x}) \geq \phi(x) + \nabla\phi(x)^T(\widehat{x} - x).$$

Further, $\phi(x)$ is concave if $-\phi(x)$ is convex.

Definition 2.2.3. The pair (x^*, y^*) is a saddle point of a convex-concave function $f(x, y)$, if for any $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, we have

$$f(x^*, y) \leq f(x^*, y^*) \leq f(x, y^*).$$

Throughout this chapter, we will assume that the following conditions are satisfied.

Assumption 2.2.4. The function $f(x, y)$ is continuously differentiable in x and y. Further, for any $y \in \mathbb{R}^n$, the function $f(x, y)$ is a convex function of x and for any $x \in \mathbb{R}^m$, the function $f(x, y)$ is a concave function of y.

Assumption 2.2.5. The gradient $\nabla_x f(x, y)$, is L_{xx}-Lipschitz with respect to x and L_{xy}-Lipschitz with respect to y and the gradient $\nabla_y f(x, y)$, is L_{yy}-Lipschitz with respect to y and

L_{yx}-Lipschitz with respect to x, i.e.,

$$\begin{aligned}
\|\nabla_x f(x_1, y) - \nabla_x f(x_2, y)\| &\leq L_{xx}\|x_1 - x_2\| \quad \forall\, y,\\
\|\nabla_x f(x, y_1) - \nabla_x f(x, y_2)\| &\leq L_{xy}\|y_1 - y_2\| \quad \forall\, x,\\
\|\nabla_y f(x, y_1) - \nabla_y f(x, y_2)| &\leq L_{yy}\|y_1 - y_2\| \quad \forall\, x,\\
\|\nabla_y f(x_1, y) - \nabla_y f(x_2, y)| &\leq L_{yx}\|x_1 - x_2\| \quad \forall\, y.
\end{aligned}$$

We define $L := 2 \times \max\{L_{xx}, L_{xy}, L_{yx}, L_{yy}\}$. [3]

Assumption 2.2.6. The solution set $\mathcal{Z}^*$ defined as

$$\mathcal{Z}^* := \{[x; y] \in \mathbb{R}^{n+m} : (x, y) \text{ is a saddle point of Problem (2.1.1)}\}, \tag{2.2.1}$$

is nonempty.

In the following sections, we present and analyze three different iterative algorithms for solving the saddle point problem introduced in (2.1.1). The k^{th} iterates of these algorithms are denoted by (x_k, y_k). We denote the averaged (ergodic) iterates by $\widehat{x}_k, \widehat{y}_k$, defined as follows:

$$\widehat{x}_k = \frac{1}{k}\sum_{i=1}^{k} x_i, \qquad \widehat{y}_k = \frac{1}{k}\sum_{i=1}^{k} y_i. \tag{2.2.2}$$

In our convergence analysis, we use a variational inequality approach in which we define the vector $z = [x; y] \in \mathbb{R}^{n+m}$ as our decision variable and define the operator $F : \mathbb{R}^{m+n} \to \mathbb{R}^{m+n}$ as

$$F(z) = [\nabla_x f(x, y); -\nabla_y f(x, y)]. \tag{2.2.3}$$

In the following lemma we characterize the properties of operator F in (2.2.3) when the conditions in Assumptions 2.2.4 and 2.2.5 are satisfied. We would like to emphasize that the following lemma is well-known – see, e.g., [94] – and we state it for completeness.

[3] In this definition we need an additional factor of 2 because in the analysis we use L as the Lipschitz continuity of the operator $F(\cdot) = [\nabla_x f(\cdot); -\nabla_y f(\cdot)]$.

Lemma 2.2.7. Let $F(\cdot)$ be defined as in Equation (2.2.3). Suppose Assumptions 2.2.4 and 2.2.5 hold. Then

(a) F is a monotone operator, i.e., for any $z_1, z_2 \in \mathbb{R}^{m+n}$, we have

$$\langle F(z_1) - F(z_2), z_1 - z_2 \rangle \geq 0.$$

(b) F is an L-Lipschitz continuous operator, i.e., for any $z_1, z_2 \in \mathbb{R}^{m+n}$, we have

$$\|F(z_1) - F(z_2)\| \leq L\|z_1 - z_2\|.$$

(c) For all $z^* \in \mathcal{Z}^*$, we have $F(z^*) = 0$.

According to Lemma 2.2.7, when f is convex-concave and smooth, the operator F defined in (2.2.3) is monotone and Lipschitz. The third result in Lemma 2.2.7 shows that any saddle point of problem (2.1.1) satisfies the first-order optimality condition, i.e $\forall$ $(x^*, y^*) \in \mathcal{Z}^*$, we have:

$$\nabla_x f(x^*, y^*) = 0 \qquad \nabla_y f(x^*, y^*) \tag{2.2.4}$$

Before presenting our main results, we state the following well known result (see for example [94]) which will be used later in the analysis of OGDA and EG. We present the proof here for completeness.

Proposition 2.2.8. Recall the definition of the operator $F(\cdot)$ in (2.2.3) and the points $\widehat{x}_k, \widehat{y}_k$ in (2.2.2). Suppose Assumptions 2.2.4 and 2.2.6 hold. Then for any $z = [x; y] \in \mathbb{R}^{m+n}$, we have

$$f(\widehat{x}_N, y) - f(x, \widehat{y}_N) \leq \frac{1}{N} \sum_{k=1}^{N} F(z_k)^\top (z_k - z) \tag{2.2.5}$$

Proof. Using the definition of the operator F, we can write

$$\frac{1}{N} \sum_{k=1}^{N} F(z_k)^\top (z_k - z) = \frac{1}{N} \sum_{k=1}^{N} [\nabla_x f(x_k, y_k)^\top (x_k - x) + \nabla_y f(x_k, y_k)^\top (y - y_k)]$$

$$\begin{aligned} &\geq \frac{1}{N}\sum_{k=1}^{N}[f(x_k, y_k) - f(x, y_k) + f(x_k, y) - f(x_k, y_k)] \\ &= \frac{1}{N}\sum_{k=1}^{N}[f(x_k, y) - f(x, y_k)], \qquad (2.2.6) \end{aligned}$$

where the inequality holds due to the fact that f is convex-concave. Using convexity of f with respect to x and concavity of f with respect to y, we have

$$\frac{1}{N}\sum_{k=1}^{N} f(x_k, y) \geq f(\widehat{x}_N, y), \qquad \frac{1}{N}\sum_{k=1}^{N} f(x, y_k) \leq f(x, \widehat{y}_N). \qquad (2.2.7)$$

Combining inequalities (2.2.6) and (2.2.7) yields

$$\frac{1}{N}\sum_{k=1}^{N} F(z_k)^\top (z_k - z) \geq f(\widehat{x}_N, y) - f(x, \widehat{y}_N),$$

completing the proof. □

2.3 Proximal point method with error

One of the classical algorithms studied for solving the saddle point problem in (2.1.1) is the Proximal Point (PP) method, introduced in [83] and studied in [109]. The PP method generates the iterate $\{x_{k+1}, y_{k+1}\}$ which is defined as the unique solution to the saddle point problem[4]

$$\min_{x\in\mathbb{R}^m} \max_{y\in\mathbb{R}^n} \left\{ f(x, y) + \frac{1}{2\eta}\|x - x_k\|^2 - \frac{1}{2\eta}\|y - y_k\|^2 \right\}. \qquad (2.3.1)$$

It can be verified that if the pair $\{x_{k+1}, y_{k+1}\}$ is the solution of problem (2.3.1), then x_{k+1} and y_{k+1} satisfy

$$x_{k+1} = \operatorname*{argmin}_{x\in\mathbb{R}^m} \left\{ f(x, y_{k+1}) + \frac{1}{2\eta}\|x - x_k\|^2 \right\}, \qquad (2.3.2)$$

[4] Again $\{x_{k+1}, y_{k+1}\}$ is unique since the objective function of problem (2.3.1) is strongly convex in x and strongly concave in y

$$y_{k+1} = \underset{y \in \mathbb{R}^n}{\operatorname{argmax}} \left\{ f(x_{k+1}, y) - \frac{1}{2\eta} \|y - y_k\|^2 \right\}. \tag{2.3.3}$$

Using the optimality conditions of the updates in (2.3.2) and (2.3.3) (which are necessary and sufficient since the problems in (2.3.2) and (2.3.3) are strongly convex and strongly concave, respectively), the update of the PP method for the saddle point problem in (2.1.1) can be written as

$$\begin{aligned} x_{k+1} &= x_k - \eta \nabla_x f(x_{k+1}, y_{k+1}), \\ y_{k+1} &= y_k + \eta \nabla_y f(x_{k+1}, y_{k+1}). \end{aligned} \tag{2.3.4}$$

It is well-known that the proximal point method achieves a sublinear rate of $\mathcal{O}(1/k)$ when k is the number of iterations for convex minimization and for solving monotone variational inequalities (see [56, 57, 19, 122, 94]). Note that [94] in fact analyzed the conceptual mirror prox (the proximal point method) as a building block to analyze the mirror-prox algorithm. For completeness, we present the convergence rate of the proximal point method for convex-concave saddle point problems in the following theorem (see Appendix 2.8.1 for the proof).

Theorem 2.3.1. Suppose Assumption 2.2.4 holds. Let $\{x_k, y_k\}$ be the iterates generated by the updates in (2.3.4). Consider the definition of the averaged iterates $\widehat{x}_k, \widehat{y}_k$ in (2.2.2). Then for all $k \geq 1$, we have

$$|f(\widehat{x}_k, \widehat{y}_k) - f(x^*, y^*)| \leq \frac{\|x_0 - x^*\|^2 + \|y_0 - y^*\|^2}{\eta k}. \tag{2.3.5}$$

The result in Theorem 2.3.1 shows that by following the update of proximal point method the gap between the function value for the averaged iterates $(\widehat{x}_k, \widehat{y}_k)$ and the function value for a saddle point (x^*, y^*) of the problem (2.1.1) approaches zero at a sublinear rate of $\mathcal{O}(1/k)$.

Our goal is to provide similar convergence rate estimates for OGDA and EG using the fact that these two methods can be interpreted as approximate versions of the proximal point method. To do so, let us first rewrite the update of the proximal point method given in

(2.3.4) as

$$z_{k+1} = z_k - \eta F(z_{k+1}), \tag{2.3.6}$$

where $z = [x; y] \in \mathbb{R}^{m+n}$ and the operator F is defined in (2.2.3). In the following proposition, we establish a relation for the iterates of a proximal point method with error. This relation will be used later for our analysis of OGDA and EG methods.

Proposition 2.3.2. Consider the sequence of iterates $\{z_k\} \in \mathbb{R}^{n+m}$ generated by the following update

$$z_{k+1} = z_k - \eta F(z_{k+1}) + \varepsilon_k, \tag{2.3.7}$$

where $F : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ is a monotone and Lipschitz continuous operator, $\varepsilon_k \in \mathbb{R}^{n+m}$ is an arbitrary vector, and η is a positive constant. Then for any $z \in \mathbb{R}^{n+m}$ and for each $k \geq 1$ we have

$$\begin{aligned} &F(z_{k+1})^\top (z_{k+1} - z) \\ &\quad = \frac{1}{2\eta}\|z_k - z\|^2 - \frac{1}{2\eta}\|z_{k+1} - z\|^2 - \frac{1}{2\eta}\|z_{k+1} - z_k\|^2 + \frac{1}{\eta}{\varepsilon_k}^\top (z_{k+1} - z). \end{aligned} \tag{2.3.8}$$

Proof. According to the update in (2.3.7), we can show that for any $z \in \mathbb{R}^{m+n}$ we have

$$\begin{aligned} \|z_{k+1} - z\|^2 = \|z_k - z\|^2 - 2\eta(z_k - z)^\top F(z_{k+1}) + \eta^2\|F(z_{k+1})\|^2 + \|\varepsilon_k\|^2 \\ + 2{\varepsilon_k}^\top (z_k - z - \eta F(z_{k+1})). \end{aligned} \tag{2.3.9}$$

We add and subtract the inner product $2\eta z_{k+1}^\top F(z_{k+1})$ to the right hand side and regroup the terms to obtain

$$\begin{aligned} \|z_{k+1} - z\|^2 = \|z_k - z\|^2 - 2\eta(z_{k+1} - z)^\top F(z_{k+1}) - 2\eta(x_k - x_{k+1})^\top F(z_{k+1}) \\ + \eta^2\|F(z_{k+1})\|^2 + \|\varepsilon_k\|^2 + 2{\varepsilon_k}^T (z_k - z - \eta F(z_{k+1})). \end{aligned} \tag{2.3.10}$$

Replacing $F(z_{k+1})$ with $(1/\eta)(-z_{k+1}+z_k+\varepsilon_k)$, we obtain

$$\begin{aligned}
&\|z_{k+1}-z\|^2 \\
&= \|z_k-z\|^2 - 2\eta(z_{k+1}-z)^\top F(z_{k+1}) + 2(z_k-z_{k+1})^\top(z_{k+1}-z_k-\varepsilon_k) \\
&\qquad + \|z_{k+1}-z_k-\varepsilon_k\|^2 + \|\varepsilon_k\|^2 + 2{\varepsilon_k}^T(z_{k+1}-z-\varepsilon_k) \\
&= \|z_k-z\|^2 - 2\eta(z_{k+1}-z)^\top F(z_{k+1}) - \|z_{k+1}-z_k\|^2 + 2{\varepsilon_k}^T(z_{k+1}-z). \qquad (2.3.11)
\end{aligned}$$

On rearranging the terms, we obtain the following inequality:

$$\begin{aligned}
&F(z_{k+1})^\top(z_{k+1}-z) \\
&\quad = \frac{1}{2\eta}\|z_k-z\|^2 - \frac{1}{2\eta}\|z_{k+1}-z\|^2 - \frac{1}{2\eta}\|z_{k+1}-z_k\|^2 + \frac{1}{\eta}{\varepsilon_k}^T(z_{k+1}-z), \qquad (2.3.12)
\end{aligned}$$

and the proof is complete. □

2.4 Optimistic Gradient Descent Ascent

In this section, we focus on analyzing the performance of optimistic gradient descent ascent (OGDA) for finding a saddle point of a general smooth convex-concave function. It has been shown that the OGDA method achieves the same iteration complexity as the proximal point method for both strongly convex-strongly concave and bilinear problems; see [75], [51], [88]. However, its iteration complexity for a general smooth convex-concave case has not been established to the best of our knowledge. In this section, we show that the function value of the averaged iterate generated by the OGDA method converges to the function value at a saddle point at a rate of $\mathcal{O}(1/k)$, which matches the convergence rate of the proximal point method shown in Theorem 2.3.1.

Given a stepsize $\eta > 0$, the OGDA method updates the iterates x_k and y_k for each $k \geq 0$ as

$$\begin{aligned}
x_{k+1} &= x_k - 2\eta\nabla_x f\left(x_k, y_k\right) + \eta\nabla_x f\left(x_{k-1}, y_{k-1}\right), \\
y_{k+1} &= y_k + 2\eta\nabla_y f\left(x_k, y_k\right) - \eta\nabla_y f\left(x_{k-1}, y_{k-1}\right) \qquad (2.4.1)
\end{aligned}$$

with the initial conditions $x_0 = x_{-1}$ and $y_0 = y_{-1}$. The main difference between the updates of OGDA in (2.4.1) and the gradient descent ascent (GDA) method is in the additional "momentum" terms $-\eta(\nabla_x f(x_k, y_k) - \nabla_x f(x_{k-1}, y_{k-1}))$ and $\eta(\nabla_y f(x_k, y_k) - \nabla_y f(x_{k-1}, y_{k-1}))$. This additional term makes the update of OGDA a better approximation to the update of the proximal point method compared to the update of the GDA; for more details we refer readers to Proposition 1 in [88].

To establish the convergence rate of OGDA for convex-concave problems, we first illustrate the connection between the updates of proximal point method and OGDA. Note that using the definitions of the vector $z = [x; y] \in \mathbb{R}^{n+m}$ and the operator $F(z) = [\nabla_x f(x, y); -\nabla_y f(x, y)] \in \mathbb{R}^{n+m}$, we can rewrite the update of the OGDA algorithm at iteration k as

$$z_{k+1} = z_k - 2\eta F(z_k) + \eta F(z_{k-1}). \tag{2.4.2}$$

Considering this expression, we can also write the update of OGDA as an approximation of the proximal point update, i.e.,

$$z_{k+1} = z_k - \eta F(z_{k+1}) + \varepsilon_k, \tag{2.4.3}$$

where the error vector ε_k is given by

$$\varepsilon_k = \eta[(F(z_{k+1}) - F(z_k)) - (F(z_k) - F(z_{k-1}))]. \tag{2.4.4}$$

To derive the convergence rate of OGDA for the unconstrained problem in (2.1.1), we first use the result in Proposition 2.3.2 to derive a result for the specific case of OGDA updates. We then show that the iterates generated by the OGDA method remain in a bounded set. This is done in the following lemma (Note that boundedness of OGDA iterates can be deduced from [103], whereas a result similar to Lemma 2.4.1(b) was shown in a recent independent paper by [82]).

Lemma 2.4.1. Let $\{z_k\}$ be the iterates generated by the optimistic gradient descent ascent (OGDA) method introduced in (2.4.2) with the initial conditions $x_0 = x_{-1}$ and $y_0 = y_{-1}$ (i.e. $z_0 = z_{-1}$). If Assumptions 2.2.4, 2.2.5, and 2.2.6 hold and the stepsize η satisfies the condition $0 < \eta \leq \frac{1}{2L}$, then:

(a) The iterates $\{z_k\}$ satisfy the following relation:

$$\begin{aligned} & F(z_{k+1})^\top (z_{k+1} - z) \\ & \leq \frac{1}{2\eta}\|z_k - z\|^2 - \frac{1}{2\eta}\|z_{k+1} - z\|^2 - \frac{L}{2}\|z_{k+1} - z_k\|^2 + \frac{L}{2}\|z_k - z_{k-1}\|^2 \\ & \qquad + (F(z_{k+1}) - F(z_k))^\top (z_{k+1} - z) - (F(z_k) - F(z_{k-1}))^\top (z_k - z). \end{aligned} \tag{2.4.5}$$

(b) The iterates $\{z_k\}$ stay within the compact set $\mathcal{D}$ defined as

$$\mathcal{D} := \{(x, y) \mid \|x - x^*\|^2 + \|y - y^*\|^2 \leq 2\left(\|x_0 - x^*\|^2 + \|y_0 - y^*\|^2\right)\}, \tag{2.4.6}$$

where $(x^*, y^*) = z^* \in \mathcal{Z}^*$ is a saddle point of the problem defined in (2.1.1).

Proof. Since OGDA iterates satisfy Equation (2.4.3) with the error vector ε_k given in Equation (2.4.4), using Proposition 2.3.2 with this error vector ε_k leads to

$$\begin{aligned} & F(z_{k+1})^\top (z_{k+1} - z) \\ & = \frac{1}{2\eta}\|z_k - z\|^2 - \frac{1}{2\eta}\|z_{k+1} - z\|^2 - \frac{1}{2\eta}\|z_{k+1} - z_k\|^2 \\ & \quad + (F(z_{k+1}) - F(z_k))^\top (z_{k+1} - z) - (F(z_k) - F(z_{k-1}))^\top (z_{k+1} - z). \end{aligned} \tag{2.4.7}$$

We add and subtract the inner product $(F(z_k) - F(z_{k-1}))^\top (z_k - z)$ to the right hand side of the preceding relation to obtain

$$\begin{aligned} & F(z_{k+1})^\top (z_{k+1} - z) \\ & = \frac{1}{2\eta}\|z_k - z\|^2 - \frac{1}{2\eta}\|z_{k+1} - z\|^2 - \frac{1}{2\eta}\|z_{k+1} - z_k\|^2 \\ & \quad + (F(z_{k+1}) - F(z_k))^\top (z_{k+1} - z) - (F(z_k) - F(z_{k-1}))^\top (z_k - z) \\ & \quad + (F(z_k) - F(z_{k-1}))^\top (z_k - z_{k+1}). \end{aligned} \tag{2.4.8}$$

Note that $(F(z_k) - F(z_{k-1}))^\top (z_k - z_{k+1})$ can be upper bounded by

$$\begin{aligned}
(F(z_k) - F(z_{k-1}))^\top (z_k - z_{k+1}) &\leq \|F(z_k) - F(z_{k-1})\| \|z_k - z_{k+1}\| \\
&\leq L\|z_k - z_{k-1}\| \|z_k - z_{k+1}\| \\
&\leq \frac{L}{2}\|z_k - z_{k-1}\|^2 + \frac{L}{2}\|z_k - z_{k+1}\|^2, \qquad (2.4.9)
\end{aligned}$$

where the second inequality holds due to Lipschitz continuity of the operator F (Lemma 2.2.7(b)) and the last inequality holds due to Young's inequality.[5] Replacing $(F(z_k) - F(z_{k-1}))^\top (z_k - z_{k+1})$ in (2.4.8) by its upper bound in (2.4.9) yields

$$\begin{aligned}
&F(z_{k+1})^\top (z_{k+1} - z) \\
&\leq \frac{1}{2\eta}\|z_k - z\|^2 - \frac{1}{2\eta}\|z_{k+1} - z\|^2 - \frac{1}{2\eta}\|z_{k+1} - z_k\|^2 \\
&\quad + (F(z_{k+1}) - F(z_k))^\top (z_{k+1} - z) - (F(z_k) - F(z_{k-1}))^\top (z_k - z) \\
&\quad + \frac{L}{2}\|z_k - z_{k-1}\|^2 + \frac{L}{2}\|z_{k+1} - z_k\|^2 \\
&\leq \frac{1}{2\eta}\|z_k - z\|^2 - \frac{1}{2\eta}\|z_{k+1} - z\|^2 - \frac{L}{2}\|z_{k+1} - z_k\|^2 + \frac{L}{2}\|z_k - z_{k-1}\|^2 \\
&\quad + (F(z_{k+1}) - F(z_k))^\top (z_{k+1} - z) - (F(z_k) - F(z_{k-1}))^\top (z_k - z), \qquad (2.4.10)
\end{aligned}$$

where the second inequality follows as $\eta \leq 1/2L$ and therefore $-\frac{1}{2\eta}\|z_{k+1} - z_k\|^2 \leq -L\|z_{k+1} - z_k\|^2$. This completes the proof of Part (a) of the lemma. Now, taking the sum of the preceding relation from $k = 0, \cdots, N-1$, we obtain

$$\begin{aligned}
&\sum_{k=0}^{N-1} F(z_{k+1})^\top (z_{k+1} - z) \\
&\leq \frac{1}{2\eta}\|z_0 - z\|^2 - \frac{1}{2\eta}\|z_N - z\|^2 - \frac{L}{2}\|z_N - z_{N-1}\|^2 + \frac{L}{2}\|z_0 - z_{-1}\|^2 \\
&\qquad + (F(z_N) - F(z_{N-1}))^\top (z_N - z) - (F(z_0) - F(z_{-1}))^\top (z_0 - z). \qquad (2.4.11)
\end{aligned}$$

[5] We use the following form of Young's inequality throughout this chapter:

$$a^\top b \leq \frac{\|a\|^2}{2} + \frac{\|b\|^2}{2}$$

Now set $z = z^*$, where $z^* \in \mathcal{Z}^*$, to obtain

$$\begin{aligned}
&\sum_{k=0}^{N-1} F(z_{k+1})^\top (z_{k+1} - z^*) \\
&\leq \frac{1}{2\eta}\|z_0 - z^*\|^2 - \frac{1}{2\eta}\|z_N - z^*\|^2 - \frac{L}{2}\|z_N - z_{N-1}\|^2 + \frac{L}{2}\|z_0 - z_{-1}\|^2 \\
&\qquad + (F(z_N) - F(z_{N-1}))^\top (z_N - z^*) - (F(z_0) - F(z_{-1}))^\top (z_0 - z^*). \qquad (2.4.12)
\end{aligned}$$

Note that each term of the summand in the sum in the left is nonnegative due to monotonicity of F and therefore the sum is also nonnegative. Further, we know that $z_0 = z_{-1}$. Using these observations we can write

$$\begin{aligned}
0 &\leq \frac{1}{2\eta}\|z_0 - z^*\|^2 - \frac{1}{2\eta}\|z_N - z^*\|^2 - \frac{L}{2}\|z_N - z_{N-1}\|^2 \\
&\quad + (F(z_N) - F(z_{N-1}))^\top (z_N - z^*). \qquad (2.4.13)
\end{aligned}$$

Using Lipschitz continuity of the operator $F(\cdot)$ (Lemma 2.2.7(b)) and Young's inequality in the preceding relation, we have

$$\begin{aligned}
0 &\leq \frac{1}{2\eta}\|z_0 - z^*\|^2 - \frac{1}{2\eta}\|z_N - z^*\|^2 - \frac{L}{2}\|z_N - z_{N-1}\|^2 \\
&\quad + L\|z_N - z_{N-1}\|\|z_N - z^*\| \\
&\leq \frac{1}{2\eta}\|z_0 - z^*\|^2 - \frac{1}{2\eta}\|z_N - z^*\|^2 - \frac{L}{2}\|z_N - z_{N-1}\|^2 \\
&\quad + \frac{L}{2}\|z_N - z_{N-1}\|^2 + \frac{L}{2}\|z_N - z^*\|^2 \\
&\leq \frac{1}{2\eta}\|z_0 - z^*\|^2 - \frac{1}{2\eta}\|z_N - z^*\|^2 + \frac{L}{2}\|z_N - z^*\|^2 \qquad (2.4.14)
\end{aligned}$$

Regrouping the terms gives us

$$\|z_N - z^*\|^2 \leq \frac{1}{(1 - \eta L)}\|z_0 - z^*\|^2. \qquad (2.4.15)$$

Using the condition $\eta \leq 1/2L$, it follows that for any iterate N we have

$$\|z_N - z^*\|^2 \leq 2\|z_0 - z^*\|^2, \qquad (2.4.16)$$

and the claim in Part (b) follows. □

According to Lemma 2.4.1, the sequence of iterates $\{x_k, y_k\}$ generated by OGDA method stays within a closed and bounded convex set. We use this result to prove a sublinear convergence rate of $\mathcal{O}(1/k)$ for the function value of the averaged iterates generated by OGDA to the function value at a saddle point, for smooth and convex-concave functions in the following theorem.

Theorem 2.4.2. Suppose Assumptions 2.2.4, 2.2.5 and 2.2.6 hold. Let $\{x_k, y_k\}$ be the iterates generated by the OGDA updates in (2.4.1). Let the initial conditions satisfy $x_0 = x_{-1}$ and $y_0 = y_{-1}$. Consider the definition of the averaged iterates $\widehat{x}_N, \widehat{y}_N$ in (2.2.2) and the compact convex set $\mathcal{D}$ in (2.4.6). If the stepsize η satisfies the condition $0 < \eta \le 1/2L$, then for all $N \ge 1$, we have

$$\left[\max_{y:(\widehat{x}_N, y)\in\mathcal{D}} f(\widehat{x}_N, y) - f^\star\right] + \left[f^\star - \min_{x:(x,\widehat{y}_N)\in\mathcal{D}} f(x, \widehat{y}_N)\right] \le \frac{D(8L + \frac{1}{2\eta})}{N}, \tag{2.4.17}$$

where $f^\star = f(x^*, y^*)$ and $D = \|x_0 - x^*\|^2 + \|y_0 - y^*\|^2$.

Proof. From Lemma 2.4.1(a), we have that the iterates generated by the OGDA method satisfy Equation (2.4.5). On taking the sum of this relation from $k = 0, \cdots, N-1$, we obtain for any z

$$\begin{aligned}
&\sum_{k=0}^{N-1} F(z_{k+1})^\top (z_{k+1} - z) \\
&\le \frac{1}{2\eta}\|z_0 - z\|^2 - \frac{1}{2\eta}\|z_N - z\|^2 - \frac{L}{2}\|z_N - z_{N-1}\|^2 + \frac{L}{2}\|z_0 - z_{-1}\|^2 \\
&\qquad + (F(z_N) - F(z_{N-1}))^\top (z_N - z) - (F(z_0) - F(z_{-1}))^\top (z_0 - z).
\end{aligned} \tag{2.4.18}$$

Note that for any $z_1, z_2 \in \mathcal{D}$, we have:

$$\begin{aligned}
\|z_1 - z_2\|^2 &\le 2\|z_1 - z^*\|^2 + 2\|z_2 - z^*\|^2 \\
&\le 4\|z_0 - z^*\|^2 + 4\|z_0 - z^*\|^2 \\
&\le 8D
\end{aligned} \tag{2.4.19}$$

where we have used the fact that $\|z - z^*\|^2 \leq 2\|z_0 - z^*\|^2 \ \forall \ z \in \mathcal{D}$ along with the fact that $\forall \ a, b \in \mathbb{R}^d$, $\|a + b\|^2 \leq 2\|a\|^2 + 2\|b\|^2$. As $z_{-1} = z_0$ and $\eta \leq 1/2L$, for any $z \in \mathcal{D}$ we have

$$\begin{aligned} \frac{1}{N}\sum_{k=0}^{N-1} F(z_{k+1})^\top(z_{k+1} - z) &\leq \frac{\frac{1}{2\eta}\|z_0 - z\|^2 + (F(z_N) - F(z_{N-1}))^\top(z_N - z)}{N} \\ &\leq \frac{D(8L + \frac{1}{2\eta})}{N}. \end{aligned} \tag{2.4.20}$$

This inequality follows since:

$$\begin{aligned} (F(z_N) - F(z_{N-1}))^\top(z_N - z) &\leq \|F(z_N) - F(z_{N-1})\|\|z_N - z\| \\ &\leq L\|z_N - z_{N-1}\|\|z_N - z\| \end{aligned} \tag{2.4.21}$$

and for any $x, y \in \mathcal{D}$, we have:

$$\begin{aligned} \|x - y\| &\leq \|x - z^*\| + \|y - z^*\| \\ &\leq 2\sqrt{2D} \end{aligned} \tag{2.4.22}$$

Therefore, we have:

$$(F(z_N) - F(z_{N-1}))^\top(z_N - z) \leq 8LD \tag{2.4.23}$$

which immediately gives us Inequality (2.4.20). Combining relation (2.4.20) with Proposition 2.2.8 we have that for all $x, y \in \mathcal{D}$

$$f(\widehat{x}_N, y) - f(x, \widehat{y}_N) \leq \frac{D(8L + \frac{1}{2\eta})}{N}. \tag{2.4.24}$$

which gives us the following convergence rate estimate:

$$\left[\max_{y:(\widehat{x}_N, y)\in\mathcal{D}} f(\widehat{x}_N, y) - f^\star\right] + \left[f^\star - \min_{x:(x,\widehat{y}_N)\in\mathcal{D}} f(x, \widehat{y}_N)\right] \leq \frac{D(8L + \frac{1}{2\eta})}{N},$$

where $f^\star = f(x^*, y^*)$. □

Note that convergence in Theorem 2.4.2 is shown in terms of the Primal-Dual gap $\max_{y:(\widehat{x}_N,y)\in\mathcal{D}} f(\widehat{x}_N, y) - \min_{x:(x,\widehat{y}_N)\in\mathcal{D}} f(x, \widehat{y}_N)$ which is a common measure to capture closeness to the solution in convex-concave setting (see [94]). Indeed, the duality gap is zero if and only if $(\widehat{x}_N, \widehat{y}_N)$ is a saddle point of the problem. The primal-dual gap also has the following game theoretic interpretation. If Player x is playing $\widehat{x}_N$, then $\max_{y:(\widehat{x}_N,y)\in\mathcal{D}} f(\widehat{x}_N, y)$ quantifies how much Player y can gain by playing an action in the set $\mathcal{D}$. Similarly, if Player y is playing $\widehat{y}_N$, then $-\min_{x:(x,\widehat{y}_N)\in\mathcal{D}} f(x, \widehat{y}_N)$ quantifies how much Player x can gain by playing an action in $\mathcal{D}$. Therefore, the quantity $\max_{y:(\widehat{x}_N,y)\in\mathcal{D}} f(\widehat{x}_N, y) - \min_{x:(x,\widehat{y}_N)\in\mathcal{D}} f(x, \widehat{y}_N)$ is a measure of the sum of how much each player can gain if they unilaterally deviate from the strategy $(\widehat{x}_N, \widehat{y}_N)$. This goes to zero at the Nash Equilibrium (saddle point), where no player can gain by unilaterally deviating from the equilibrium strategy.

Also, note that the result in Theorem 2.4.2 also implies that $|f(\widehat{x}_N, \widehat{y}_N) - f^*| \leq 9LD/N$ as we show in the following corollary.

Corollary 2.4.3. Suppose Assumptions 2.2.4, 2.2.5 and 2.2.6 hold. Let $\{x_k, y_k\}$ be the iterates generated by the OGDA updates in (2.4.1). Consider the definition of the averaged iterates $\widehat{x}_N, \widehat{y}_N$ in (2.2.2). If the stepsize η satisfies the condition $0 < \eta \leq 1/2L$, then for all $N \geq 1$, we have

$$|f(\widehat{x}_N, \widehat{y}_N) - f^\star| \leq \frac{D(8L + \frac{1}{2\eta})}{N},$$

where $f^\star = f(x^*, y^*)$.

Proof. Note that $[\max_{y:(\widehat{x}_N,y)\in\mathcal{D}} f(\widehat{x}_N, y) - f^\star]$ and $[f^\star - \min_{x:(x,\widehat{y}_N)\in\mathcal{D}} f(x, \widehat{y}_N)]$ are both nonnegative. To verify note that

$$\max_{y:(\widehat{x}_N,y)\in\mathcal{D}} f(\widehat{x}_N, y) \geq f(\widehat{x}_N, y^*) \geq f(x^*, y^*)$$

and

$$\min_{x:(x,\widehat{y}_N)\in\mathcal{D}} f(x, \widehat{y}_N) \leq f(x^*, \widehat{y}_N) \leq f(x^*, y^*)$$

(since $(x^*, y^*) \in \mathcal{D}$). Further, note that $(\widehat{x}_N, \widehat{y}_N)$ belongs to the set $\mathcal{D}$. Hence, it yields

$$f(\widehat{x}_N, \widehat{y}_N) - f^\star \leq \max_{y:(\widehat{x}_N,y)\in\mathcal{D}} f(\widehat{x}_N, y) - f^\star \leq \frac{D(8L + \frac{1}{2\eta})}{N}.$$

Also, we can show that

$$f^\star - f(\widehat{x}_N, \widehat{y}_N) \leq f^\star - \min_{x:(x,\widehat{y}_N)\in\mathcal{D}} f(x, \widehat{y}_N) \leq \frac{D(8L + \frac{1}{2\eta})}{N}.$$

Therefore, $|f(\widehat{x}_N, \widehat{y}_N) - f^\star| \leq \frac{D(8L+\frac{1}{2\eta})}{N}$. □

The result in Corollary 2.4.3 shows that the function value of the averaged iterates generated by OGDA converges to the function value at a saddle point of problem (2.1.1) at a sublinear rate of $\mathcal{O}(1/k)$ when the function is smooth and convex-concave. To the best of our knowledge, this is the first non-asymptotic complexity bound for OGDA for the convex-concave setting. Moreover, note that without computing any extra gradient evaluation, i.e., computing only one gradient per iteration with respect to x and y, OGDA recovers the convergence rate of proximal point method.

2.5 Extragradient Method

In this section, we consider finding a saddle point of a general smooth convex-concave function using the Extra-gradient (EG) method. Similar to our analysis of the OGDA method, we show that by interpreting the EG method as an approximation of the proximal point method it is possible to establish a convergence rate of $\mathcal{O}(1/k)$ through a simple analysis.

Consider the update of EG in which we first compute a set of mid-point iterates $\{x_{k+\frac{1}{2}}, y_{k+\frac{1}{2}}\}$ using the gradients with respect to x and y at the current iterate

$$\begin{aligned} x_{k+\frac{1}{2}} &= x_k - \eta\nabla_x f(x_k, y_k), \\ y_{k+\frac{1}{2}} &= y_k + \eta\nabla_y f(x_k, y_k). \end{aligned} \tag{2.5.1}$$

Then, we compute the next iterates of the EG method $\{x_{k+1}, y_{k+1}\}$ using the gradients at the mid-points $\{x_{k+\frac{1}{2}}, y_{k+\frac{1}{2}}\}$, i.e.,

$$\begin{aligned} x_{k+1} &= x_k - \eta\nabla_x f(x_{k+\frac{1}{2}}, y_{k+\frac{1}{2}}), \\ y_{k+1} &= y_k + \eta\nabla_y f(x_{k+\frac{1}{2}}, y_{k+\frac{1}{2}}). \end{aligned} \tag{2.5.2}$$

We aim to show that EG, similar to OGDA, can be analyzed for convex-concave problems by considering it as an approximation of the proximal point. To do so, let us use the notation $z = [x; y] \in \mathbb{R}^{n+m}$ and $F(z) = [\nabla_x f(x, y); -\nabla_y f(x, y)] \in \mathbb{R}^{n+m}$ to write the update of EG as

$$\begin{aligned} z_{k+\frac{1}{2}} &= z_k - \eta F(z_k), \\ z_{k+1} &= z_k - \eta F(z_{k+\frac{1}{2}}). \end{aligned} \tag{2.5.3}$$

To better highlight the connection between proximal point and EG, let us focus on the expression for the update of the mid-point iterates in EG. Considering the updates in (2.5.3), we have

$$\begin{aligned} z_{k+\frac{1}{2}} &= z_k - \eta F(z_k), \\ &= z_{k-1} - \eta F(z_{k-\frac{1}{2}}) - \eta F(z_k) \\ &= z_{k-\frac{1}{2}} + \eta F(z_{k-1}) - \eta F(z_{k-\frac{1}{2}}) - \eta F(z_k) \end{aligned}$$

where the second equality follows by replacing z_k by its update $z_{k-1} - \eta F(z_{k-\frac{1}{2}})$, and the second equality follows by considering the update $z_{k-\frac{1}{2}} = z_{k-1} - \eta F(z_{k-1})$. Therefore, rearranging this equation, we can rewrite the updates as

$$z_{k+\frac{1}{2}} = z_{k-\frac{1}{2}} - \eta F(z_{k-\frac{1}{2}}) - \eta(F(z_k) - F(z_{k-1})). \tag{2.5.4}$$

One can consider the expression $F(z_k) - F(z_{k-1})$ as an approximation of the variation $F(z_{k+\frac{1}{2}}) - F(z_{k-\frac{1}{2}})$. To be more precise, if we assume that the variations $F(z_k) - F(z_{k-1})$ and $F(z_{k+\frac{1}{2}}) - F(z_{k-\frac{1}{2}})$ are close to each other, i.e., $F(z_{k+\frac{1}{2}}) - F(z_{k-\frac{1}{2}}) \approx F(z_k) - F(z_{k-1})$, then the update in (2.5.4) behaves like the proximal point update with. respect to the mid-point iterates, i.e.,

$$z_{k+\frac{1}{2}} = z_{k-\frac{1}{2}} - \eta F(z_{k+\frac{1}{2}}). \tag{2.5.5}$$

We first derive a result similar to Proposition 2.3.2 for the specific case of EG iterates (Lemma 2.5.1(a)). We then show the the boundedness of the EG iterates in Lemma 2.5.1(b) (Note that the boundedness of the EG updates can also be deduced from the convergence results of [70] and [91]).

Lemma 2.5.1. Let $\{z_k\}, \{z_{k+\frac{1}{2}}\}$ be the iterates generated by the extra-gradient (EG) method introduced in (2.5.3). If Assumptions 2.2.4, 2.2.5 and 2.2.6 hold and the stepsize η satisfies the condition $0 < \eta < 1/L$, then:
(a) The iterates $\{z_k\}, \{z_{k+\frac{1}{2}}\}$ satisfy the following relation:

$$\begin{aligned}
&F(z_{k+\frac{1}{2}})^\top (z_{k+\frac{1}{2}} - z) \\
&\leq \frac{1}{2\eta}\|z_{k-\frac{1}{2}} - z\|^2 - \frac{1}{2\eta}\|z_{k+\frac{1}{2}} - z\|^2 + \frac{L}{2}\|z_{k-\frac{1}{2}} - z_{k-1}\|^2 \\
&\quad + (F(z_{k+\frac{1}{2}}) - F(z_k))^\top (z_{k+\frac{1}{2}} - z) - (F(z_{k-\frac{1}{2}}) - F(z_{k-1}))^\top (z_{k-\frac{1}{2}} - z). \qquad (2.5.6)
\end{aligned}$$

(b) The iterates $\{z_k\}, \{z_{k+\frac{1}{2}}\}$ stay within the compact set $\mathcal{D}$ defined as

$$\mathcal{D} := \{(x, y) \mid \|x - x^*\|^2 + \|y - y^*\|^2 \leq \left(2 + \frac{2}{1 - \eta^2 L^2}\right) (\|x_0 - x^*\|^2 + \|y_0 - y^*\|^2)\}, \qquad (2.5.7)$$

where $(x^*, y^*) = z^* \in \mathcal{Z}^*$ is a saddle point of the problem defined in (2.1.1). Moreover, the sum $\sum_{k=0}^{\infty} \|z_{k+\frac{1}{2}} - z_k\|^2$ is bounded above by

$$\sum_{k=0}^{\infty} \|z_{k+\frac{1}{2}} - z_k\|^2 \leq \frac{\|z_0 - z^*\|^2}{1 - \eta^2 L^2}. \qquad (2.5.8)$$

The result in Lemma 2.5.1 shows that the iterates generated by the update of EG belong to a bounded and closed set. Now we use this result to show that the function value of the averaged iterates converges at a sublinear rate of $\mathcal{O}(1/k)$ to the function value at a saddle point for the EG method in the following theorem.

Theorem 2.5.2. Suppose Assumptions 2.2.4, 2.2.5 and 2.2.6 hold. Let $\{x_{k+1/2}, y_{k+1/2}\}$ be the iterates generated by the EG updates in (2.5.1)-(2.5.2). Let the initial conditions satisfy $x_0 = x_{-1/2}$ and $y_0 = y_{-1/2}$ Consider the definition of the averaged iterates $\widehat{x}_N, \widehat{y}_N$ in (2.2.2) and the compact convex set $\mathcal{D}$ in (2.5.7). If the stepsize η satisfies the condition $\eta = \frac{\sigma}{L}$ for any $\sigma \in (0,1)$, then for all $N \geq 1$, we have

$$\left[\max_{y:(\widehat{x}_N,y)\in\mathcal{D}} f(\widehat{x}_N, y) - f^\star\right] + \left[f^\star - \min_{x:(x,\widehat{y}_N)\in\mathcal{D}} f(x, \widehat{y}_N)\right] \leq \frac{DL\left(16 + \frac{33}{2(1-\sigma^2)}\right)}{N}, \tag{2.5.9}$$

where $f^\star = f(x^*, y^*)$ and $D = \|x_0 - x^*\|^2 + \|y_0 - y^*\|^2$.

Now, similar to Corollary 2.4.3, we have:

Corollary 2.5.3. Suppose Assumptions 2.2.4, 2.2.5 and 2.2.6 hold. Let $\{x_{k+1/2}, y_{k+1/2}\}$ be the iterates generated by the EG updates in (2.5.1)-(2.5.2). Let the initial conditions satisfy $x_0 = x_{-1/2}$ and $y_0 = y_{-1/2}$ Consider the definition of the averaged iterates $\widehat{x}_N, \widehat{y}_N$ in (2.2.2). If the stepsize η satisfies the condition $\eta = \frac{\sigma}{L}$ for any $\sigma \in (0,1)$, then for all $N \geq 1$, we have

$$|f(\widehat{x}_N, \widehat{y}_N) - f^\star| \leq \frac{DL\left(16 + \frac{33}{2(1-\sigma^2)}\right)}{N},$$

where $f^\star = f(x^*, y^*)$.

2.6 Discussion and Numerical Experiments

The main message of this work is that the OGDA algorithm obtains the same convergence rate of $\mathcal{O}(1/k)$, the best achievable rate (see [94]), also achieved by EG. However, the advantage of OGDA is that we need only one gradient computation at each step, as opposed to two gradient computations needed in EG. This shows the computational advantage that OGDA has over EG.

We compare the performance of OGDA and EG in terms of gradient computations, on the bilinear minimax games considered in [94], without any constraint. In particular, we consider the following minimax problem:

$$\min_{x\in\mathbb{R}^n} \max_{y\in\mathbb{R}^n} \; x^\top B y, \tag{2.6.1}$$

where $B \in \mathbb{R}^{n\times n}$ is a sparse random matrix generated as follows. Each element is nonzero independently with probability p. If an element is chosen to be non-zero, it is chosen uniformly

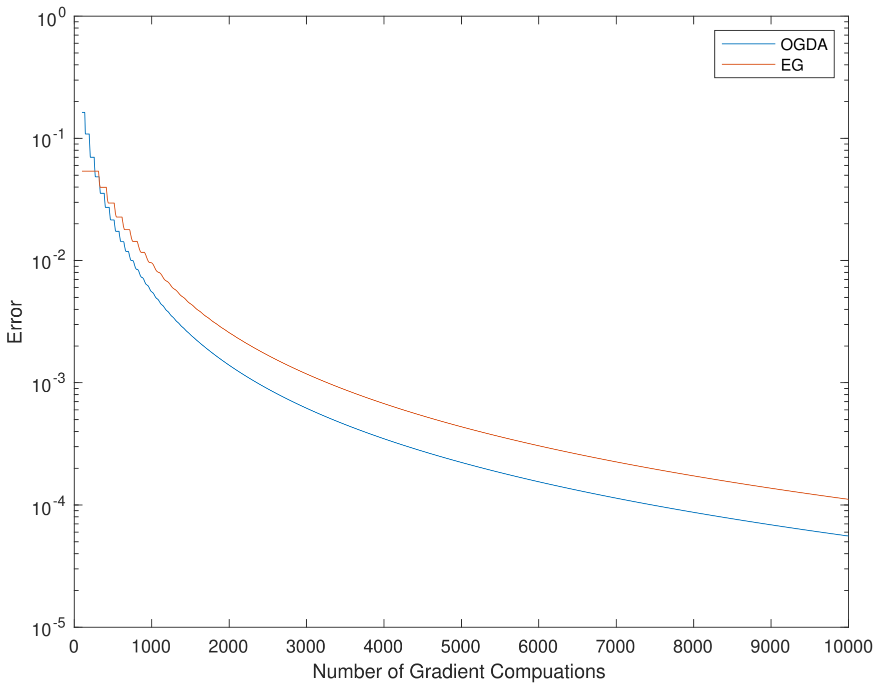


Figure 2-1: Number of Gradient computations required (x-axis) to reach any error level (y-axis) for both OGDA and EG for the problem in Equation (2.6.1)

from $[-1, 1]$. We compare the number of gradient computations required to reach a desired accuracy level for this problem in Figure 2-1. As we observe, both EG and OGDA converge to the saddle point of the bilinear problem at a sublinear rate of $\mathcal{O}(1/k)$, but OGDA slightly outperforms EG in terms of number of gradient evaluations. Once again, this is due to the fact that for both descent and ascent updates of OGDA requires only one gradient computation each, while EG requires two gradient computations for both updates at each iteration.

Note that the Lipschitz constants for the considered problem can be estimated from data using standard line search techniques. In particular, [14] discuss a backward tracking algorithm (ISTA with backtracking) which can be used to estimate the Lipschitz constants, in particular L_{xx} and L_{yy}. Several variants of this algorithm, including the Lipschitz line-search algorithm (Algorithm 2) in [114], can also be used to estimate the Lipschitz constants L_{xx} and L_{yy}. For the specific case of saddle point problems, a recent paper [60] proposes a line search algorithm, to estimate the Lipschitz constant L_{xx}, L_{xy}, L_{yx} and L_{yy}. They propose an algorithm - Accelerated Primal Dual with backtracking (Algorithm 2.3) which uses a backtracking procedure, similar to [81], to locally estimate the Lipschitz constants of the problem. Also, regarding the initial error, we would like to highlight that in the analysis of convex minimization problems or convex-concave saddle point problems, we often have a term of the form $\|x_0 - x^*\|^2$ in the upper bound (for instance see [96, 91]) which shows the effect of initial error. This parameter is hard to estimate in general but can be upper bounded in specific cases. For example, if we are looking at for mixed strategies in zero-sum games, we know that we are looking for a solution lies in the probability simplex, so we can bound the initial error simply by the diameter of the simplex. In general, if we know that our iterates of the algorithm are going to lie in some compact set, we can upper bound the initial distance to the solution simply by the diameter of the compact set.

2.7 Conclusions

In this chapter, we established convergence guarantees of the optimistic gradient ascent-descent (OGDA) and Extra-gradient (EG) methods for unconstrained, smooth, and convex-concave saddle point problems. In particular, we showed a sublinear convergence rate of $\mathcal{O}(1/k)$ in terms of function value error for both OGDA and EG by interpreting them as approximate variants of the proximal point method. This result leads to the first theoretical guarantee for OGDA in convex-concave saddle point problems. Moreover, it provides a simple and short proof for the convergence rate of EG in convex-concave saddle point problems when we measure optimality gap in terms of function value.

2.8 Appendix

We present omitted proofs of this chapter in this section.

2.8.1 Proof of Theorem 2.3.1

The update of the proximal point method can be written as:

$$z_{k+1} = z_k - \eta F(z_{k+1}) \tag{2.8.1}$$

According to this update we can show that

$$\|z_{k+1} - z\|^2 = \|z_k - z\|^2 - 2\eta(z_k - z)^\top F(z_{k+1}) + \eta^2\|F(z_{k+1})\|^2 \tag{2.8.2}$$

Now add and subtract the inner product $2\eta z_{k+1}^\top F(z_{k+1})$ to the right hand side and regroup the terms to obtain

$$\begin{aligned}\|z_{k+1} - z\|^2 = \|z_k - z\|^2 &- 2\eta(z_{k+1} - z)^\top F(z_{k+1}) - 2\eta(x_k - x_{k+1})^\top F(z_{k+1}) \\ &+ \eta^2\|F(z_{k+1})\|^2.\end{aligned} \tag{2.8.3}$$

Replace $F(z_{k+1})$ with $(1/\eta)(-z_{k+1} + z_k)$ to obtain

$$\begin{aligned}&\|z_{k+1} - z\|^2 \\ &= \|z_k - z\|^2 - 2\eta(z_{k+1} - z)^\top F(z_{k+1}) + 2(z_k - z_{k+1})^\top(z_{k+1} - z_k) \\ &\quad + \|z_{k+1} - z_k\|^2 \\ &= \|z_k - z\|^2 - 2\eta(z_{k+1} - z)^\top F(z_{k+1}) - \|z_{k+1} - z_k\|^2.\end{aligned} \tag{2.8.4}$$

On rearranging the terms, we get the following

$$F(z_{k+1})^\top(z_{k+1} - z) = \frac{1}{2\eta}\|z_k - z\|^2 - \frac{1}{2\eta}\|z_{k+1} - z\|^2 - \frac{1}{2\eta}\|z_{k+1} - z_k\|^2, \tag{2.8.5}$$

Now, on substituting $z = z^*$, and noting that $F(z_{k+1})^\top(z_{k+1} - z^*) \geq 0$, we have:

$$\|z_{k+1} - z^*\|^2 \leq \|z_k - z^*\|^2 - \|z_{k+1} - z_k\|^2 \tag{2.8.6}$$

and the proof of boundedness is complete.

On adding Equation (2.8.5) from $k = 0, \cdots N - 1$ and diving by N, we get:

$$\frac{1}{N}\sum_{k=1}^{N} F(z_k)^\top(z_k - z) \leq \frac{\|z_0 - z\|^2}{\eta N} \tag{2.8.7}$$

Now, using Proposition 2.2.8 we can write

$$|f(\widehat{x}_N, \widehat{y}_N) - f^\star| \leq \frac{\|x_0 - x\|^2 + \|y_0 - y\|^2}{\eta N}, \tag{2.8.8}$$

and the proof is complete.

2.8.2 Proof of Lemma 2.5.1

(a) Considering the updates in (2.5.4) and (2.5.5) we can write the update of mid-points in EG as

$$z_{k+\frac{1}{2}} = z_{k-\frac{1}{2}} - \eta F(z_{k+\frac{1}{2}}) + \varepsilon_k, \tag{2.8.9}$$

where,

$$\varepsilon_k = \eta \left[(F(z_{k+\frac{1}{2}}) - F(z_{k-\frac{1}{2}})) - (F(z_k) - F(z_{k-1})) \right]. \tag{2.8.10}$$

Therefore, we can simplify the last term in Equation (2.3.8) of Proposition 2.3.2 as follows:

$$\begin{aligned}
&\frac{1}{\eta}{\varepsilon_k}^\top (z_{k+\frac{1}{2}} - z) \\
&\quad = \frac{1}{\eta} \times [(\eta F(z_{k+\frac{1}{2}}) - \eta F(z_k)) - (\eta F(z_{k-\frac{1}{2}}) - \eta F(z_{k-1}))]^\top (z_{k+\frac{1}{2}} - z) \\
&\quad = (F(z_{k+\frac{1}{2}}) - F(z_k))^\top (z_{k+\frac{1}{2}} - z) - (F(z_{k-\frac{1}{2}}) - F(z_{k-1}))^\top (z_{k-\frac{1}{2}} - z) \\
&\qquad\qquad - (F(z_{k-\frac{1}{2}}) - F(z_{k-1}))^\top (z_{k+\frac{1}{2}} - z_{k-\frac{1}{2}}).
\end{aligned} \tag{2.8.11}$$

Using Lipschitz continuity of the operator F (Lemma 2.2.7(b)) and Young's inequality, we have

$$\begin{aligned}
&\frac{1}{\eta}{\varepsilon_k}^\top (z_{k+\frac{1}{2}} - z) \\
&\quad \leq F(z_{k+\frac{1}{2}}) - F(z_k))^\top (z_{k+\frac{1}{2}} - z) - (F(z_{k-\frac{1}{2}}) - F(z_{k-1}))^\top (z_{k-\frac{1}{2}} - z) \\
&\qquad\qquad + L\|z_{k-\frac{1}{2}} - z_{k-1}\| \|z_{k+\frac{1}{2}} - z_{k-\frac{1}{2}}\| \\
&\quad \leq F(z_{k+\frac{1}{2}}) - F(z_k))^\top (z_{k+\frac{1}{2}} - z) - (F(z_{k-\frac{1}{2}}) - F(z_{k-1}))^\top (z_{k-\frac{1}{2}} - z) \\
&\qquad\qquad + \frac{L}{2}\|z_{k-\frac{1}{2}} - z_{k-1}\|^2 + \frac{L}{2}\|z_{k+\frac{1}{2}} - z_{k-\frac{1}{2}}\|^2
\end{aligned} \tag{2.8.12}$$

Substituting the upper bound in (2.8.12) into Equation (2.3.8) of Proposition 2.3.2, implies that

$$\begin{aligned}
&F(z_{k+\frac{1}{2}})^\top (z_{k+\frac{1}{2}} - z) \\
&\qquad \le \frac{1}{2\eta}\|z_{k-\frac{1}{2}} - z\|^2 - \frac{1}{2\eta}\|z_{k+\frac{1}{2}} - z\|^2 - \frac{1}{2\eta}\|z_{k+\frac{1}{2}} - z_{k-\frac{1}{2}}\|^2 \\
&\qquad\quad + (F(z_{k+\frac{1}{2}}) - F(z_k))^\top (z_{k+\frac{1}{2}} - z) - (F(z_{k-\frac{1}{2}}) - F(z_{k-1}))^\top (z_{k-\frac{1}{2}} - z) \\
&\qquad\quad + \frac{L}{2}\|z_{k-\frac{1}{2}} - z_{k-1}\|^2 + \frac{L}{2}\|z_{k+\frac{1}{2}} - z_{k-\frac{1}{2}}\|^2. \qquad (2.8.13)
\end{aligned}$$

Since $\eta < 1/L$, we have $-\frac{1}{2\eta}\|z_{k+\frac{1}{2}} - z_{k-\frac{1}{2}}\|^2 + \frac{L}{2}\|z_{k+\frac{1}{2}} - z_{k-\frac{1}{2}}\|^2 \le 0$ and therefore

$$\begin{aligned}
&F(z_{k+\frac{1}{2}})^\top (z_{k+\frac{1}{2}} - z) \\
&\le \frac{1}{2\eta}\|z_{k-\frac{1}{2}} - z\|^2 - \frac{1}{2\eta}\|z_{k+\frac{1}{2}} - z\|^2 + \frac{L}{2}\|z_{k-\frac{1}{2}} - z_{k-1}\|^2 \\
&\quad + (F(z_{k+\frac{1}{2}}) - F(z_k))^\top (z_{k+\frac{1}{2}} - z) - (F(z_{k-\frac{1}{2}}) - F(z_{k-1}))^\top (z_{k-\frac{1}{2}} - z). \qquad (2.8.14)
\end{aligned}$$

which completes the proof of Part (a).

(b) Based on the update of EG in (2.5.3), we can write

$$\begin{aligned}
&\|z_k - z\|^2 \\
&= \|z_k - z_{k+1} + z_{k+1} - z\|^2 \\
&= \|z_{k+1} - z\|^2 + 2(z - z_{k+1})^\top (z_{k+1} - z_k) + \|z_{k+1} - z_k\|^2 \\
&= \|z_{k+1} - z\|^2 + 2(z - z_{k+\frac{1}{2}})^\top (z_{k+1} - z_k) \\
&\qquad + 2(z_{k+\frac{1}{2}} - z_{k+1})^\top (z_{k+1} - z_k) + \|z_{k+1} - z_k\|^2 \\
&= \|z_{k+1} - z\|^2 + 2(z - z_{k+\frac{1}{2}})^\top (z_{k+1} - z_k) + \|z_{k+\frac{1}{2}} - z_k\|^2 - \|z_{k+\frac{1}{2}} - z_{k+1}\|^2. \qquad (2.8.15)
\end{aligned}$$

Now we proceed to bound the difference $\|z_{k+\frac{1}{2}} - z_{k+1}\|^2$. Using the fact that the operator F is L-Lipschitz (Lemma 2.2.7(b)), we have

$$\begin{aligned}
\|z_{k+\frac{1}{2}} - z_{k+1}\|^2 &= \eta^2 \|F(z_{k+\frac{1}{2}} - F(z_k)\|^2 \\
&\le \eta^2 L^2 \|z_{k+\frac{1}{2}} - z_k\|^2. \qquad (2.8.16)
\end{aligned}$$

Substituting this upper bound back into (2.8.15) and taking $z = z^*$ implies

$$\begin{aligned} &\|z_k - z^*\|^2 \\ &\quad \geq \|z_{k+1} - z^*\|^2 + 2(z^* - z_{k+\frac{1}{2}})^\top (z_{k+1} - z_k) + (1 - \eta^2 L^2)\|z_{k+\frac{1}{2}} - z_k\|^2. \end{aligned} \tag{2.8.17}$$

Further, since the operator F is monotone, we have

$$\begin{aligned} (z^* - z_{k+\frac{1}{2}})^\top (z_{k+1} - z_k) &= \eta (F(z_{k+\frac{1}{2}}))^\top (z_{k+\frac{1}{2}} - z^*) \\ &\geq \eta (F(z_{k+\frac{1}{2}}) - F(z^*))^\top (z_{k+\frac{1}{2}} - z^*) \\ &\geq 0, \end{aligned} \tag{2.8.18}$$

where in the first inequality we used the fact that $F(z^*) = 0$ (Lemma 2.2.7(c)), and the last inequality holds due to monotonicity of F (Lemma 2.2.7(a)). Therefore, we can replace the inner product $2(z^* - z_{k+\frac{1}{2}})^\top (z_{k+1} - z_k)$ in (2.8.17) by its lower bound 0 to obtain

$$\|z_k - z^*\|^2 \geq \|z_{k+1} - z^*\|^2 + (1 - \eta^2 L^2)\|z_{k+\frac{1}{2}} - z_k\|^2 \tag{2.8.19}$$

The result in (2.8.19) shows that the seqeunce $\|z_k - z^*\|^2$ is non-increasing. Therefore, for any iterate k, it holds that

$$\|z_k - z^*\|^2 \leq \|z_0 - z^*\|^2. \tag{2.8.20}$$

Now, for all $k \geq 0$, we have:

$$\begin{aligned} \|z_{k+\frac{1}{2}} - z^*\|^2 &\leq 2\|z_k - z^*\|^2 + 2\|z_{k+\frac{1}{2}} - z_k\|^2 \\ &\leq \left(2 + \frac{2}{1 - \eta^2 L^2}\right) \|z_k - z^*\|^2 \\ &\leq \left(2 + \frac{2}{1 - \eta^2 L^2}\right) \|z_0 - z^*\|^2 \end{aligned} \tag{2.8.21}$$

where the first inequality follows from the fact that $\forall\ a, b \in \mathbb{R}^d$, $\|a + b\|^2 \leq 2\|a\|^2 + 2\|b\|^2$, the second inequality follows from (2.8.19) and the third inequality follows from (2.8.20).

Therefore from (2.8.20) and (2.8.21), since $0 < 1 - \eta^2 L^2 < 1$, we see that the iterates $\{z_k\}, \{z_{k+\frac{1}{2}}\}$ belong to the compact set $\mathcal{D}$ defined in (2.5.7).

Now by summing both sides of (2.8.19) for $k = 0, \ldots, \infty$, we obtain

$$(1 - \eta^2 L^2) \sum_{k=0}^{\infty} \|z_{k+\frac{1}{2}} - z_k\|^2 \leq \|z_0 - z^*\|^2 \tag{2.8.22}$$

Therefore, by regrouping the terms we obtain

$$\sum_{k=0}^{\infty} \|z_{k+\frac{1}{2}} - z_k\|^2 \leq \frac{\|z_0 - z^*\|^2}{1 - \eta^2 L^2}, \tag{2.8.23}$$

and the claim in (2.5.8) follows.

2.8.3 Proof of Theorem 2.5.2

Using Equation (2.5.6) of Lemma 2.5.1(a), summing it from $k = 0, \cdots, N-1$ and dividing by N, we obtain

$$\begin{aligned}
&\frac{1}{N} \sum_{k=0}^{N-1} F(z_{k+\frac{1}{2}})^\top (z_{k+\frac{1}{2}} - z) \\
&\leq \frac{\frac{1}{2\eta}\|z_0 - z\|^2 + (F(z_{N-\frac{1}{2}}) - F(z_{N-1}))^\top (z_{N-\frac{1}{2}} - z)}{N} + \frac{L}{2N} \sum_{k=0}^{N-1} \|z_{k-\frac{1}{2}} - z_{k-1}\|^2.
\end{aligned} \tag{2.8.24}$$

The bound in Equation (2.5.8) from Lemma 2.5.1(b) yields

$$\begin{aligned}
&\frac{1}{N} \sum_{k=0}^{N-1} F(z_{k+\frac{1}{2}})^\top (z_{k+\frac{1}{2}} - z) \\
&\qquad \leq \frac{L\|z_0 - z\|^2 + (F(z_{N-\frac{1}{2}}) - F(z_{N-1}))^\top (z_{N-\frac{1}{2}} - z)}{N} + \frac{L\|z_0 - z^*\|^2}{2(1 - \eta^2 L^2)N} \\
&\qquad \leq \frac{L\|z_0 - z\|^2 + L\|z_{N-\frac{1}{2}} - z_{N-1}\|\|z_{N-\frac{1}{2}} - z\| + \frac{L}{2(1-\sigma^2)}\|z_0 - z^*\|^2}{N},
\end{aligned} \tag{2.8.25}$$

where in the last inequality we use Lipschitz continuity of the operator F (Lemma 2.2.7(b)) and the fact that $\eta = \frac{\sigma}{L}$. Note that for any $z_1, z_2 \in \mathcal{D}$, we have:

$$
\begin{aligned}
\|z_1 - z_2\| &\leq \|z_1 - z^*\| + \|z_2 - z^*\| \\
&\leq \sqrt{\left(2 + \frac{2}{1 - \eta^2 L^2}\right) \|z_0 - z^*\|} + \sqrt{\left(2 + \frac{2}{1 - \eta^2 L^2}\right) \|z_0 - z^*\|} \\
&\leq 2\sqrt{D\left(2 + \frac{2}{1 - \sigma^2}\right)}. \qquad (2.8.26)
\end{aligned}
$$

Therefore, for any point z in the set $\mathcal{D}$, we can substitute the preceding relation in Equation (2.8.25) to get

$$
\frac{1}{N} \sum_{k=0}^{N-1} F(z_{k+\frac{1}{2}})^\top (z_{k+\frac{1}{2}} - z) \leq \frac{DL\left(16 + \frac{33}{2(1-\sigma^2)}\right)}{N}. \qquad (2.8.27)
$$

Now, using Proposition 2.2.8 we have that for all $x, y \in \mathcal{D}$:

$$
f(\widehat{x}_N, y) - f(x, \widehat{y}_N) \leq \frac{DL\left(16 + \frac{33}{2(1-\sigma^2)}\right)}{N}, \qquad (2.8.28)
$$

where $\widehat{x}_N = \frac{1}{N} \sum_{k=0}^{N-1} x_{k+1/2}$ and $\widehat{y}_N = \frac{1}{N} \sum_{k=0}^{N-1} y_{k+1/2}$ which gives us the following convergence result:

$$
\left[\max_{y:(\widehat{x}_N, y) \in \mathcal{D}} f(\widehat{x}_N, y) - f^\star\right] + \left[f^\star - \min_{x:(x, \widehat{y}_N) \in \mathcal{D}} f(x, \widehat{y}_N)\right] \leq \frac{DL\left(16 + \frac{33}{2(1-\sigma^2)}\right)}{N},
$$

where $f^\star = f(x^*, y^*)$.

Chapter 3

Structured Nonconvex-Nonconcave Problems

3.1 Introduction

In this chapter, we move our focus onto certain structured nonconvex-nonconcave minimax problems. The main motivating example for this setting is multi-agent Reinforcement learning (RL).

Policy gradient (PG) methods have served as the workhorse of modern RL [115, 116, 58], and enjoy the desired properties of being scalable to large state-action spaces, stability with function approximation, as well as sample efficiency. In fact, policy gradient methods have achieved impressive empirical performance in multi-agent RL [78, 139], the regime where many RL's recent successes are pertinent to [120, 100, 118].

Despite the tremendous empirical successes, theoretical foundations of PG methods, even for the single-agent setting, have not been uncovered until recently [45, 1, 145, 129, 85, 22]. The theoretical understanding of PG methods for multi-agent RL remains largely elusive, except for several recent attempts [32, 151, 131, 23]. The key challenge is that in the policy parameter space, even for the basic two-player zero-sum matrix game, the problem becomes *nonconvex-nonconcave* and is computationally intractable in general [36].

In this chapter, we aim to fill in the gap by studying the global convergence of natural PG (NPG) [68], which forms the basis for many popular PG algorithms (e.g., Proximal Policy

Optimization (PPO)/Trust Region Policy Optimization (TRPO)), in the parameter space and for multi-agent learning. We are interested in the setting where the agents take *symmetric* roles and operate *independently*, as it does not require a central coordinator and it scales favorably with the number of agents. Analysis of this setting is challenging precisely because the concurrent updates of the agents makes the learning environment *non-stationary* from one agent's perspective. With *asymmetric* update rules among agents, the non-stationarity issue can be mitigated, and the global convergence of PG methods has been established lately in [146, 32, 151]. However, though being valid as an optimization scheme, asymmetric update-rules might be hard to justify in game-theoretic multi-agent learning with symmetric players. It is thus desirable to develop provably convergent PG methods with symmetric update rules.

We focus on the *last-iterate* convergence of the *policy parameters*, which is critical to establish in order to avoid stability issues during learning. For example, if the norm of the parameters blow up, we might end up with precision issues in computing the gradient and updating the parameters. This is particularly relevant to the setting with function approximation, where we can only operate on *low-dimensional* parameters of the policy, instead of the high-dimensional policy per se. Indeed, we aim to explore the convergence property in the function approximation setting to handle large state-action spaces. Finally, our results are also motivated from the study of *nonconvex-nonconcave minimax* optimization problems, especially those with certain structures that yield global convergence of gradient-based methods. We aim to explore such structures in multi-agent learning with *parameterized* policies.

Contributions. Our contributions are three-fold. First, we identify the non-convergence issue in the policy parameter space of natural PG methods for RL. We show that this issue persists even with entropy regularized rewards. Second, we develop symmetric variants of the natural PG method, i.e., both without and with the optimistic updates ([108]) and establish the last-iterate global convergence to the Nash equilibrium in the policy parameter space. Third, we generalize the scope of symmetric PG methods in game-theoretic multi-agent learning, including two-player zero-sum matrix and Markov games (MGs), multi-player monotone games, and the corresponding linear function approximation settings under certain assumptions, in order to handle enormously large state-action spaces, all with last-iterate parameter convergence rate guarantees. We have also provided numerical experiments to validate the effectiveness of our algorithms.

3.1.1 Related work

Policy gradient RL methods for games. Gradient-descent-ascent (GDA) with projection on simplexes can be viewed as symmetric policy gradient methods for solving matrix games with *direct policy parameterization* [1], which enjoys an average-iterate convergence [24]. Such a guarantee is shared with multiplicative weight update (MWU), which is especially suitable for repeated matrix games, and equivalent to natural PG method with tabular softmax policy parameterization [1]. To the best of our knowledge, however, neither parameter convergence nor function approximation has been studied in this context. For Markov games, [146, 20, 59] have studied global convergence of PG methods for those with a linear quadratic structure; for zero-sum Markov games, [32] established global convergence of independent PG with two-timescale stepsizes for the tabular setting; [151] studied a double-loop natural PG algorithm with function approximation; more recently, [2] proposed a framework of natural actor-critic algorithms. No last-iterate convergence to the Nash equilibrium was established in these works, and these update-rules were all *asymmetric*. [104] developed symmetric policy optimization methods for certain zero-sum Markov games with structured transitions. Concurrently, [148] proposed a policy optimization framework with fast average-iterate convergence guarantees for finite-horizon Markov games. Finally, [74, 149, 37] have studied global convergence of symmetric PG methods in Markov potential games recently, not focused on last-iterate or parameter convergence.

Last-iterate convergence in constrained multi-agent learning. Several papers including [125, 7, 66, 42] and references therein studied the *last-iterate* behavior of strongly monotone games. Furthermore, [52, 53] extended this analysis to the monotone game setting. However, these papers did not consider *parametrized policies*. More specifically, in the matrix game setting with a simplex constraint, papers including [31, 130, 23] showed the last-iterate policy convergence of optimistic methods. However, these papers did not consider policy parameterization or the function approximation settings, and some papers required the assumption that the NE is unique [31, 130]. For Markov games, [151] established last-iterate convergence, but not to NE due to asymmetric update; [131, 23] were, to the best of our knowledge, the only last-iterate policy convergence results in Markov games with symmetric updates. However, these works did not study the function approximation setting, or monotone games beyond the two-player zero-sum case. Also, though having greatly inspired our work, the regularization idea in [23] alone cannot prevent the non-convergence issue of the policy parameters from happening (see §4.2). Our goal, in contrast, is to study the (last-iterate) convergence of the actual policy parameters, and for more general multi-agent learning settings beyond the tabular zero-sum one.

Nonconvex-nonconcave minimax optimization. It is shown that for general nonconvex-nonconcave minimax optimization, even *local* solution concepts [67] may not exist, and finding them can be intractable [36]. Thus, specific structural properties have to be exploited to design efficient algorithms with *global* convergence. [77, 123, 99, 135] have studied the nonconvex-(strongly)-concave or the nonconvex-Polyak-Lojasiewicz (PL) or PL-PL settings, with global convergence rate guarantees. The algorithms in these papers are all *asymmetric* in that they run the inner loop (which solves the maximization problem) multiple times (or on a faster timescale with larger stepsizes) to reach an approximate solution of the inner optimization problem, and then run one step of descent on the outer problem (or on a slower timescale with smaller stepsizes). Closely related to one motivation of our work, [128, 48, 87] studied nonconvex-nonconcave minimax problem with *hidden convexity* structures, and show that GDA can fail to converge globally even so. Interestingly, the benefit of regularization (more generally, strict convexity), and natural gradient flow under Fisher information geometry, were also examined in [48, 87] to establish some positive convergence results. Different from our work, the dynamics there are in *continuous-time*, and the parameterization in [48] is decoupled by dimension, and the convergence rate in the perturbed game is not global. These conditions prevent the application of their results and proof techniques to our setting directly. Also, last-iterate finite-time rates of the iterates were not established in [87].

Notation. For vector $v \in \mathbb{R}^d$, we use $[v]_a$ with $a \in \{1, 2, \cdots, d\}$ to denote the a-th element of v. We use $\|v\|$ to denote the ℓ_2-Euclidean norm of a vector v and $\|Q\|$ to denote the ℓ_2-induced norm of matrix Q. We also use $\|Q\|_\infty$ to denote the infinity norm and $\|Q\|_F$ to denote the Frobenius norm of matrix Q. For a finite-set $\mathcal{S}$, we use $\Delta(\mathcal{S})$ to denote the simplex over $\mathcal{S}$. We use $\mathbb{1}$ to denote the matrix of all ones of appropriate dimension. For any positive integer n, we use $[n]$ to denote the set $\{1, \cdots, n\}$. We use the subscript $-i$ to denote the quantities of all players other than player i. $\mathrm{KL}(p\|q)$ denotes the KL divergence between two probability distributions p and q. For a matrix C, we use $C = [A \mid B]$ to denote the concatenation of the component matrices A and B. For two vectors $x, y \in \mathbb{R}^d$, $x \cdot y$ denotes their inner-product, i.e., $x^\top y$. We use I_d to denote an identity matrix of dimension d.

3.2 Motivation & Background

In this section, we introduce the background of the natural PG methods we study, with two-player[1] zero-sum matrix games being a motivating example.

Zero-sum matrix games. Two-player zero-sum matrix games are characterized by a tuple $(\mathcal{A}, \mathcal{B}, Q)$, where $Q \in \mathbb{R}^{n\times n}$ denotes the cost[2] matrix, $\mathcal{A}$ and $\mathcal{B}$ denote the action spaces of players 1 and 2, respectively. For notational simplicity, we assume both action spaces have cardinality n, i.e., $|\mathcal{A}| = |\mathcal{B}| = n$. Note that our results can be readily generalized to the setting with different action-space cardinalities. For convenience, we use *indices* of the actions to denote the actions, i.e., $\mathcal{A} = \mathcal{B} = \{1, 2, \cdots, n\}$, without loss of generality. Note that the *actual actions* of both players for the same index need not to be the same, and the cost matrix Q needs not to be symmetric. The problem can thus be formulated as a minimax (i.e., saddle-point optimization) problem

$$\min_{g\in\Delta(\mathcal{A})} \max_{h\in\Delta(\mathcal{B})} \quad f(g,h) := g^\top Q h, \tag{3.2.1}$$

[1] Hereafter, we use *player* and *agent* interchangeably.

[2] Note that we can also model it as a payoff, with a negative sign.

where g and h are referred to as the policies/strategies of the players. By Minimax Theorem [98], the min and max operators in (3.2.1) can be interchanged, and the solution concept of *Nash equilibrium* (NE), which is defined as a pair of policies $(g^\star, h^\star)$ such that

$$f(g, h^\star) \geq f(g^\star, h^\star) \geq f(g^\star, h), \quad \text{for any } (g, h) \in \Delta(\mathcal{A}) \times \Delta(\mathcal{B})$$

always holds. In particular, at the Nash equilibrium, the players execute the best-response policies of each other, and have no incentive to deviate from it.

Policy parameterization. To develop policy gradient methods for multi-player learning, the policies $(g, h) \in \Delta(\mathcal{A}) \times \Delta(\mathcal{B})$ are parameterized by some parameters θ and ν. Specifically, consider the following softmax parameterization that is common in practice: for any $a \in \mathcal{A}$ and $b \in \mathcal{B}$

$$g_\theta(a) = \frac{e^{p_\theta(a)}}{\sum_{a' \in \mathcal{A}} e^{p_\theta(a')}}, \qquad h_\nu(b) = \frac{e^{q_\nu(b)}}{\sum_{b' \in \mathcal{B}} e^{q_\nu(b')}}, \tag{3.2.2}$$

where $\theta, \nu \in \mathbb{R}^d$ for some integer $d > 0$, $p_\theta, q_\nu : \mathbb{R}^d \to \mathbb{R}^n$ are two differentiable functions. Note that $g_\theta(a), h_\nu(b) > 0$ for any bounded p_θ, q_ν, and $\sum_{a \in \mathcal{A}} g_\theta(a) = \sum_{b \in \mathcal{B}} h_\nu(b) = 1$. This parameterization gives the following minimax problem for the zero-sum matrix game:

$$\min_{\theta \in \mathbb{R}^n} \max_{\nu \in \mathbb{R}^n} \quad f(\theta, \nu) := g_\theta^\top Q h_\nu, \tag{3.2.3}$$

where by a slight abuse of notation, we use $f(\theta, \nu)$ to denote $f(g_\theta, h_\nu)$. In this section and §3.3, we consider the tabular softmax parameterization where $p_\theta = \theta \in \mathbb{R}^n$ and $q_\nu = \nu \in \mathbb{R}^n$. In §3.4, we consider the setting with function approximation where $d < n$.

The benefits of softmax parameterization are that: 1) it transforms a *constrained* problem over simplexes to an *unconstrained* one, making it easier to implement; 2) it readily incorporates function approximation to deal with large spaces (see §3.4). On the other hand, this policy parameterization makes the optimization problem (3.2.3) more challenging to solve. Indeed, the minimax problem (3.2.3) becomes a nonconvex-nonconcave problem in θ and ν, even with the tabular parameterization as we will show later in Lemma 3.2.2.

Remark 3.2.1 (Hidden bilinear problem)**.** Note that Problem (3.2.3), which resembles a bilinear zero-sum game, in fact falls into the class of *hidden bilinear* minimax problems discussed in [128] (or more generally the *hidden convex-concave* games studied in [48, 87]). It was shown in [48] that for general smooth functions of g_θ and h_ν, vanilla gradient descent-ascent exhibits a variety of behaviors antithetical to convergence to the solutions. We here instead, show that for the specific *softmax parameterization*, and for certain variants of the vanilla gradient-descent-ascent method, the last-iterate convergence rate of the parameters θ and ν can be established.

Natural PG & Non-convergence pitfall. Before proceeding further, we first introduce the *regularized* game:

$$\min_{\theta\in\mathbb{R}^n}\max_{\nu\in\mathbb{R}^n} \quad f_\tau(\theta,\nu) := g_\theta^\top Q h_\nu - \tau\mathcal{H}(g_\theta) + \tau\mathcal{H}(h_\nu), \tag{3.2.4}$$

where the cost of both players is regularized by the Shannon entropy of the policies, with $\tau > 0$ being the regularization parameter, and $\mathcal{H}(\pi) = -\sum_{a\in\mathcal{A}} \pi(a)\log(\pi(a))$ for π on a simplex. The entropy regularization, which is commonly used in single-player RL, enjoys the benefits of both encouraging exploration and accelerating convergence [97, 85]. Our hope is also to exploit the benefits of entropy regularization in the multi-player setting. Indeed, the regularized cost traces its source in the game theory literature [84], to model the imperfect knowledge of the cost matrix Q. In the next lemma, we show that the problem in Equation 3.2.4 can be of the nonconvex-nonconcave type:

Lemma 3.2.2. The minimax problem (3.2.4) is nonconvex in θ and nonconcave in ν, even if $p_\theta = \theta$ and $q_\nu = \nu$.

Note that the nonconvexity in the parameters remains even when we regularize with the entropy of the policy, i.e., $\tau > 0$.

Motivated by the successes of *natural policy gradient* [68] and its variants, as PPO/TRPO [115, 116], in RL practice, we consider the natural PG descent-ascent update for (3.2.4), which is given by

$$\theta_{t+1} = \theta_t - \eta \cdot F_\theta^\dagger(\theta_t) \cdot \frac{\partial f_\tau(\theta_t, \nu_t)}{\partial \theta} = (1 - \eta\tau)\theta_t - \eta Q h_{\nu_t} + \eta\tau\bigg(\log \sum_{a' \in \mathcal{A}} e^{\theta_t(a')} - 1\bigg), \qquad (3.2.5)$$

$$\nu_{t+1} = \nu_t + \eta \cdot F_\nu^\dagger(\nu_t) \cdot \frac{\partial f_\tau(\theta_t, \nu_t)}{\partial \nu} = (1 - \eta\tau)\nu_t + \eta Q^\top g_{\theta_t} - \eta\tau\bigg(\log \sum_{b' \in \mathcal{B}} e^{\nu_t(b')} - 1\bigg), \qquad (3.2.6)$$

where $F_\theta(\theta) = \mathbb{E}_{a\sim g_\theta}[(\nabla_\theta \log g_\theta(a))(\nabla_\theta \log g_\theta(a))^\top]$ and $F_\nu(\nu) = \mathbb{E}_{b\sim h_\nu}[(\nabla_\nu \log h_\nu(b))(\nabla_\nu \log h_\nu(b))^\top]$ are the Fisher information matrices, $M^\dagger$ denotes the pseudo-inverse of the matrix M, and $\eta > 0$ is the stepsize. The derivations for natural policy gradient can be found in §3.9.1 for completeness.

Unfortunately, the vanilla NPG update (3.2.5)-(3.2.6) may fail to converge in the parameter space for *any* stepsize $\eta > 0$. The key reason for the failure is that the mappings represented in (3.2.5)-(3.2.6) may not have a fixed point for a general Q and τ (which could be the only limit point for this dynamics). In fact, this issue persists even when the regularization parameter $\tau = 0$. We formalize this pitfall in the following lemma, with its proof deferred to appendix.

Lemma 3.2.3 (Pitfall of vanilla NPG)**.** There exists a game (3.2.4) with $\tau \geq 0$ and $|\mathcal{B}| = 1$, such that the updates (3.2.5)-(3.2.6) do not converge for any $\eta > 0$.

Remarkably, we emphasize that our Lemma 3.2.3, by construction, also even applies to the single-agent setting, with a regularized cost and the NPG update, as studied recently in [22, 72, 140]. These works only focus on the convergence in the policy space, which does not imply the desired convergence in the policy parameter space. The later becomes especially relevant in the function approximation setting, as we will study later. Finally, we remark that, the non-convergence here also should not be confused with the *last-iterate non-convergence* of no-regret learning algorithms for solving bilinear zero-sum games [35, 9], as our example is essentially a single-agent case. We summarize the importance and motivation of establishing parameter convergence as follows.

Importance of Parameter Convergence:

Numerical instability: Prior works were only able to show the convergence of *values* and/or *policies*, and the convergence behavior of policy parameters was *unclear* (or overlooked). Arguably, having (last-iterate) parameter convergence is the *strongest* type of convergence among the three. In practice, having parameters blow-up to infinity can cause numerical issues. For example, once the size of the parameter crosses a threshold (say 2^{64} for an integer in a 64-bit operating system), there would be overflow issues, and the stored parameter would be void, and NaN (not a number) would be returned by the program. This blow-up would then cause trouble in recovering the policy, or approximating the policy with arbitrary accuracy. In order to circumvent this issue, a common practice in Neural Network training is to do *Clipping/Projection.* In fact, ensuring the *stability* of the model is very important in deep learning. Specifically, consider $\max_{\theta \in \mathbb{R}^n} q^\top g_\theta + \tau \mathcal{H}(g_\theta)$, where we know that under the NPG updates, $g_\theta \to g^\star$ while θ could blow up to infinity. One could clip the parameter θ to some large constant $\theta_{\max}$, i.e., solving $\max_{\|\theta\|_\infty \leq \theta_{\max}} q^\top g_\theta + \tau \mathcal{H}(g_\theta)$ instead. For concreteness, let $n = 2$, $\tau = 1$, $\theta_{\max} = 80$ and $q = [-2, -3]$. The optimal solution is then given by $g_i^\star \propto \exp(q_i)$ for $i = 1, 2$. On running the vanilla NPG algorithm, since we do weight clipping, the algorithm converges to $\theta = [\theta_{\max}, \theta_{\max}]$ corresponding to the distribution $[1/2, 1/2] \neq g^\star$. Meanwhile, the modified NPG we propose converges to $\theta = [-2, -3]$ (see Theorem 3.3.4) which exactly corresponds to the optimal solution $g^\star$. Hence, in practice where the norm of the (neural network) parameters is bounded, one might not obtain policy convergence using vanilla NPG as desired, while our proposed algorithm works.

Nonconvex-nonconcave minimax optimization: The second reason comes from a minimax optimization perspective of solving (3.2.3). We view optimization over the parameter space as an interesting *nonconvex-nonconcave* minimax optimization problem with a *hidden structure* (See Lemma 3.2.2). To the best of our knowledge, we are the first to provide a *symmetric* discrete-time algorithm to solve certain nonconvex-nonconcave problem (and more generally non-monotone variational inequalities) with last-iterate convergence rates, even including the specialized settings (like the ones with Polyak-Łojasiewicz condition [99, 135]).

Function approximation (FA): Parameter convergence becomes crucial in FA settings (used in practice). Here, the policy lies in a high-dimensional space (or even an infinite-dimensional space if the actions are *continuous*), which we simply do not have access to and/or cannot operate on. The way practitioners run PG methods is to just operate on the low-dimensional policy parameter space. Thus, parameter convergence is necessary to design meaningful stopping criteria for optimization algorithms. If parameters explode to infinity, we cannot decide on how close we are to convergence, and the numerical issue mentioned before would cause trouble in recovering the policy.

3.3 Warm-up: (Optimistic) NPG for Matrix Games

To address the pitfall above about parameter convergence, we introduce two variants of the vanilla NPG (3.2.5)-(3.2.6), and show their convergence for solving matrix games.

3.3.1 NPG for Matrix Games

We first introduce the following variant of the vanilla NPG update:

$$\theta_{t+1} = (1-\eta\tau)\theta_t - \eta Q h_{\nu_t}, \tag{3.3.1}$$

$$\nu_{t+1} = (1-\eta\tau)\nu_t + \eta Q^\top g_{\theta_t}, \tag{3.3.2}$$

where we removed the last term in (3.2.5)-(3.2.6), respectively. Note that these updates correspond to the popular Multiplicative Weights Update (MWU) for the regularized game in policy space (we succinctly represent g_{θ_t} and h_{ν_t} as g_t and h_t, respectively), i.e.,

$$\begin{aligned} g_{t+1}(a) &\propto g_t(a)^{(1-\eta\tau)} e^{-\eta[Qh_t]_a}, \\ h_{t+1}(b) &\propto h_t(b)^{(1-\eta\tau)} e^{\eta[Q^\top g_t]_b}. \end{aligned} \tag{3.3.3}$$

First, we provide a convergence result for the updates in Equations (3.3.1)-(3.3.2), the non-optimistic version, both in terms of policy as well as parameters. In order to do so, we need to first show that the iterates of the regularized MWU in Equations (3.3.1)-(3.3.2) ensure that the policies stay bounded away from the boundary of the simplex. We show this in the following lemma:

Lemma 3.3.1. The policies corresponding to the iterates of regularized MWU in Equations (3.3.1)-(3.3.2) with stepsize $\eta < 1/\tau$ stay within a set $\Delta' \subset \Delta$ which is bounded away from the boundary of the simplex, i.e., $\forall x \in \Delta'$ with $x = (x_1, \cdots, x_n)^\top$, and for all $i \in [n]$, $x_i \geq \delta > 0$ for some δ.

Since the iterates of the policies lie within Δ', a closed and bounded set, and the regularized cost is continuously differentiable with respect to the policies, we let L denote the smoothness constant of the regularized cost in the policy space, i.e.,

$$\|(Mz_1 + \tau\nabla\mathcal{H}(z_1)) - (Mz_2 + \tau\nabla\mathcal{H}(z_2))\| \leq L\|z_1 - z_2\|, \qquad \forall z_1, z_2 \in \mathscr{Z}', \tag{3.3.4}$$

where $z = [g; h]$, $\mathscr{Z}' \in \Delta' \times \Delta'$ and with a slight abuse of notation, we define $\nabla\mathcal{H}(z) = [\nabla_g\mathcal{H}(g); \nabla_h\mathcal{H}(h)]$, and also, we define the matrix $M = \begin{pmatrix} 0 & Q \\ -Q^\top & 0 \end{pmatrix}$.

Finally, the entropy regularized optimization problem in the policy space can be formulated as:

$$\min_{g\in\Delta(\mathcal{A})} \max_{h\in\Delta(\mathcal{B})} = g^\top Qh - \tau\mathcal{H}(g) + \tau\mathcal{H}(h). \tag{3.3.5}$$

Now, we use this to derive the policy and parameter convergence of the MWU updates in Equations (3.3.1)-(3.3.2) in the following theorem.

Theorem 3.3.2. The solution $(g^\star, h^\star)$ to the problem (3.3.5) is unique. Furthermore, let $\theta^\star = \frac{-Qh^\star}{\tau}$ and $\nu^\star = \frac{Q^\top g^\star}{\tau}$. Then on running Equations (3.3.1)-(3.3.2) with the stepsize satisfying $0 < \eta \leq \tau/L^2$, we have:

$$\mathrm{KL}(z^*\|z_{t+1}) \leq \left(1 - \frac{\eta\tau}{2}\right)\mathrm{KL}(z^*\|z_t), \tag{3.3.6}$$

and

$$\|\theta_{t+1} - \theta^\star\|^2 + \|\nu_{t+1} - \nu^\star\|^2 \leq (1 - \eta\tau/4)^t\left(\|\theta_0 - \theta^\star\|^2 + \|\nu_0 - \nu^\star\|^2 + \frac{4C}{\eta\tau}\right), \tag{3.3.7}$$

where $C = \left(1 + \frac{1}{\eta\tau}(1 - \eta\tau)^2\right)4\eta^2\|Q\|_\infty^2\mathrm{KL}(z^\star\|z_0)$. Here $z_0 = (g_{\theta_0}, h_{\nu_0})$ and $z^\star = (g^\star, h^\star)$.

To the best of our knowledge, this is the first policy convergence guarantee for NPG (without optimism) for regularized matrix games. As the theorem above shows, parameter convergence can be established as well. In other words, though the vanilla NPG descent-ascent diverges for Problem (3.2.4), the variant we propose (Equations (3.3.1)-(3.3.2)) *implicitly regularizes* the parameter iterates to converge to a particular solution, in last-iterate. Recall from Lemma 3.2.2 that Problem (3.2.4) is a nonconvex-nonconcave minimax problem and there have been no convergence guarantees, to the best of our knowledge, of any symmetric and simultaneous-update algorithms in general (see [135, 105, 77] and references therein as examples for some structured nonconvex-nonconcave problems). From Theorem 3.3.2, the rate we can hope to achieve with NPG is $\mathcal{O}(\kappa^2 \log(1/\epsilon))$ (number of steps to reach a point ϵ close to the solution), where κ is the condition number L/τ of the problem. In the next subsection, we see how adding optimism ([108]) will improve this rate of convergence.

3.3.2 Optimistic NPG (ONPG) for Matrix Games

In this subsection, we study the variant of the NPG updates in Equations (3.3.1)-(3.3.2) along with optimism [103, 108, 35, 31, 90]. In particular, we introduce the intermediate iterates $(\bar{\theta}_t, \bar{\nu}_t)$, and the following optimistic variant of the NPG (here $(\bar{\theta}_0, \bar{\nu}_0)$ is initialized as (θ_0, ν_0)):

$$\begin{aligned}
\bar{\theta}_{t+1} &= (1-\eta\tau)\theta_t - \eta Q h_{\bar{\nu}_t}, \quad \theta_{t+1} = (1-\eta\tau)\theta_t - \eta Q h_{\bar{\nu}_{t+1}} \\
\bar{\nu}_{t+1} &= (1-\eta\tau)\nu_t + \eta Q^\top g_{\bar{\theta}_t}, \quad \nu_{t+1} = (1-\eta\tau)\nu_t + \eta Q^\top g_{\bar{\theta}_{t+1}}.
\end{aligned}$$

This optimistic update is motivated by the success of optimistic gradient methods in saddle-point problems recently analyzed in several papers including [64, 89, 90]. Note that our algorithm is symmetric in θ and ν updates and both players update simultaneously. The update rules are also tabulated in Algorithm 1.

Remark 3.3.3 (Connections to the literature)**.** Note that the natural PG update rule in Equations (3.3.1)-(3.3.2) has a close relationship to the multiplicative weight update rule [49, 5] in the policy space (g_θ, h_ν), see Section C.3 in [1] for a detailed discussion. Similarly, the optimistic NPG update in Algorithm 1 also relates to the optimistic MWU update [31]. In fact, recent works [130, 23] have shown the last-iterate *policy convergence* of OMWU for zero-sum matrix games (with [130] relying on the uniqueness assumption of the NE). Our goal, in contrast, is to study the (last-iterate) convergence behavior of the actual *policy parameters* (θ, ν), and go beyond the tabular zero-sum setting.

Inspired by the results which show that adding optimistic updates improves convergence rates [108], we next explore our modified NPG updates with optimism, and show that the convergence rate does in fact improves (in line with the comparison of the performance of GDA and optimistic GDA, see e.g., [42]).

In the next theorem, we show that our Optimistic NPG Algorithm (Algorithm 1) in fact converges linearly in last-iterate to a unique point in the set of NE in the parameter space, at a faster rate[3] than the non-optimistic counterpart in Equations (3.3.1)-(3.3.2). The results are formally stated below.

Theorem 3.3.4. Let $\theta^\star = \frac{-Qh^\star}{\tau}$ and $\nu^\star = \frac{Q^\top g^\star}{\tau}$, where $g^\star$ and $h^\star$ are the solutions to the regularized game (3.2.4). Then on running Algorithm 1 with the stepsize satisfying $0 < \eta \leq \min\left\{\frac{1}{2\tau+2\|Q\|_\infty}, \frac{1}{4\|Q\|_\infty}\right\}$, we have:

$$\|\theta_{t+1} - \theta^\star\|^2 + \|\nu_{t+1} - \nu^\star\|^2 \leq (1 - \eta\tau/2)^t V_0, \tag{3.3.8}$$

where $V_0 = \|\theta_0 - \theta^\star\|^2 + \|\nu_0 - \nu^\star\|^2 + \frac{2C}{\eta\tau}$ and $C = \left(1 + \frac{1}{\eta\tau}(1-\eta\tau)^2\right) 4\eta^2\|Q\|_\infty^2 \mathrm{KL}(z^\star \| z_0)$. Here $z_0 = (g_{\theta_0}, h_{\nu_0})$ and $z^\star = (g^\star, h^\star)$.

This result shows that the specific hidden bilinear minimax problem we are dealing with does not fall into the spurious categories discussed in [128], if we resort to the (optimistic) natural PG update. Note that achieving parameter convergence is a non-trivial task since we are dealing with a nonconvex-nonconcave minimax problem (see Lemma 3.2.2) The proof relies on the specific structure of the softmax policy parametrization and the construction of a novel Lyapunov function (see §3.9.2 for more details). Next, we show how the optimistic NPG algorithm solves the original matrix game without regularization.

[3] Note that we say that the optimistic version achieves a faster rate, since the range of stepsizes which permits convergence is much larger for the optimistic variant. This can be noted from the fact that L in equation (3.3.4) will be larger than $\|Q\|_\infty + \tau$.

Corollary 3.3.5. If we run the non-optimistic variant of NPG in Equations (3.3.1)-(3.3.2) or the optimistic version in Algorithm 1 for time $T = \mathcal{O}\left(\frac{\log n}{\eta\epsilon} \log\left(\frac{1}{\epsilon}\right)\right)$ and set $\tau = \epsilon/(8 \log n)$, we have that the output (θ_T, ν_T) is an ϵ-NE of the original unregularized Problem (3.2.3).

We extend these results to simple function approximation settings in the next section.

3.4 Matrix Games with Function Approximation

To handle games with excessively large action spaces, we resort to policy parameterization with function approximation. In particular, consider the following problem:

$$\min_{\theta \in \mathbb{R}^d} \max_{\nu \in \mathbb{R}^d} \quad g_\theta^\top Q h_\nu, \tag{3.4.1}$$

where g_θ and h_ν are both parameterized in a softmax way as in (3.2.2), with the linear function class $p_\theta(a) = \phi_a^\top \theta$ (also called *log-linear* policy in [1]), where $\phi_a \in \mathbb{R}^d$ is a low-dimensional feature representation of the action (see [18, 1]) (Note that usually $d < n$). We define:

$$\Phi = [\phi_1, \phi_2, \cdots, \phi_n] \ \in \mathbb{R}^{d \times n}. \tag{3.4.2}$$

Assumption 3.4.1. Φ is a full rank matrix. In particular, assume that $\Phi = [M \mid 0]$, where $M \in \mathbb{R}^{d \times d}$ is an invertible $d \times d$ square matrix.

Note that the full-rankness of Φ is a standard assumption (see Assumption 6.2 in [1]). It essentially requires the features to be the bases of some low-dimensional space. Furthermore, the results also extend to the case where the matrix Φ is of the from $[M \mid c\mathbb{1}]$ where $\mathbb{1}$ is the matrix of all 1s of appropriate dimension, and c is any constant. This particular structure of the feature matrix, though being restrictive, ensures that the constraint set of policies is convex, as shown next, otherwise the minimax theorem of $\min\max = \max\min$ might not hold, i.e., the Nash equilibrium for the parameterized game does not exist. Moreover, the assumption is also not as restrictive as it seems. For example, in applications of self-driving car and robotics, only a subset of actions (steering angles) is essential in controlling the agent, with other actions being insignificant/redundant. Our feature matrix encodes patterns like these. Moreover, as a first step studying policy optimization in multi-agent learning with function approximation, we start with this simpler setting. Extending the ideas to more general FA settings is an interesting direction worth exploring.

Remark 3.4.2. Note that the results in this section are presented for the case where the feature matrix is identical for both players purely for simplification of notation. The results continue to hold for the case with asymmetric features as well (as long as the feature matrix also satisfies Assumption 3.4.1.

As motivated in the previous section, we study the following regularized problem in order to solve (3.4.1) efficiently:

$$\min_{\theta\in\mathbb{R}^d} \max_{\nu\in\mathbb{R}^d} \quad g_\theta^\top Q h_\nu - \tau\mathcal{H}(g_\theta) + \tau\mathcal{H}(h_\nu), \tag{3.4.3}$$

where $\mathcal{H}$ denotes the entropy function and $\tau > 0$ is the regularization parameter. Note that this problem can still be nonconvex-nonconcave in general, given the example in §3.3 as a special case.

We define the solution to this problem next, the Nash equilibrium in the parameterized policy classes, i.e., in-class NE.

Definition 3.4.3 (ϵ-in-class Nash equilibrium)**.** The policy parameter $(\widetilde{\theta}, \widetilde{\nu})$ is an ϵ-*Nash equilibrium* of the matrix game with function approximation (or ϵ-*in-class* NE), if it satisfies that for all $i \in [N]$,

$$g_{\widetilde{\theta}}^\top Q h_\nu - \epsilon \;\le\; g_{\widetilde{\theta}}^\top Q h_{\widetilde{\nu}} \;\le\; g_\theta^\top Q h_{\widetilde{\nu}} + \epsilon, \qquad \forall \theta,\ \nu\ \in \mathbb{R}^d. \tag{3.4.4}$$

Furthermore, when $\epsilon = 0$, we refer to it as the *in-class Nash Equilibrium.*

3.4.1 Equivalent problem characterization

In this subsection, we study the regularized problem (3.4.3) under a log-linear parametrization and find an equivalent problem in the tabular case.

First, in the following lemma, we characterize the set of distributions covered by this parametrization, and study the equivalent problem in the space of probability vectors.

Lemma 3.4.4. Under Assumption 3.4.1, the log-linear parametrization in Equation (3.2.2) covers all distributions in the following *convex* set:

$$\widetilde{\Delta} = \{\mu : \mu \in \Delta, \mu_{d+1} = \mu_{d+2} = \cdots = \mu_n\}, \tag{3.4.5}$$

and Problem (3.4.3) is equivalent to

$$\min_{g_\theta \in \widetilde{\Delta}} \max_{h_\nu \in \widetilde{\Delta}} \quad g_\theta^\top Q h_\nu - \tau \mathcal{H}(g_\theta) + \tau \mathcal{H}(h_\nu). \tag{3.4.6}$$

Lemma 3.4.4 characterizes the set of distributions which can be represented by log-linear parametrization. Therefore, when we try to solve the matrix game with such function approximation, the best we can hope for is to find an equilibrium within the set $\widetilde{\Delta}$.

Next, we characterize the Nash equilibrium of the regularized Problem (3.4.3) in the function approximation setting, and show its equivalence to another problem in the tabular softmax setting.

Theorem 3.4.5. An in-class Nash equilibrium $(\theta^\star, \nu^\star)$ of Problem (3.4.3) under the function approximation setting exists, and any such in-class NE satisfies:

$$g_{\theta^\star}(a) = \frac{e^{-\frac{[\Psi^\top Q \Psi h_{\nu^\star}]_a}{\tau}}}{\sum_{a'} e^{-\frac{[\Psi^\top Q \Psi h_{\nu^\star}]_{a'}}{\tau}}}, \quad h_{\nu^\star}(a) = \frac{e^{\frac{[\Psi^\top Q^\top \Psi g_{\theta^\star}]_a}{\tau}}}{\sum_{a'} e^{\frac{[\Psi^\top Q^\top \Psi g_{\theta^\star}]_{a'}}{\tau}}},$$

where $\Psi \in \mathbb{R}^{n \times n}$ is defined as:

$$\Psi = \begin{pmatrix} I_d & & \mathbf{0} & \\ & \frac{1}{n-d} & \cdots & \frac{1}{n-d} \\ \mathbf{0} & \cdots & \cdots & \cdots \\ & \frac{1}{n-d} & \cdots & \frac{1}{n-d} \end{pmatrix}. \tag{3.4.7}$$

Furthermore, Problem (3.4.3) is equivalent to[4]

$$\min_{g_\theta \in \Delta} \max_{h_\nu \in \Delta} \quad g_\theta^\top \Psi^\top Q \Psi h_\nu - \tau \mathcal{H}(g_\theta) + \tau \mathcal{H}(h_\nu). \tag{3.4.8}$$

Note that the matrix Ψ defined in Theorem 3.4.5 is the invariant matrix for the set $\widetilde{\Delta}$, i.e., $\Psi\mu = \mu$, $\forall \mu \in \widetilde{\Delta}$.

3.4.2 Optimistic NPG algorithm

From Section 3.3, we see that the optimistic version of NPG leads to faster convergence (of both policy and parameters). This motivates us to focus on the optimistic version of the methods. In this subsection, (and the ones that follow), we focus on optimistic methods instead of their non-optimistic counterparts. Note that similar to Section 3.3, we can derive convergence rates for the non-optimistic versions as well, which would be slower than the corresponding optimistic versions.

As Theorem 3.4.5 characterizes the solution to the function approximation setting to that of the tabular softmax setting, we modify the algorithm for function approximation setting as follows:

$$\begin{aligned} \bar{\theta}_{t+1} &= (1 - \eta\tau)\theta_t - \eta[(M^\top)^{-1} \mid 0]\widetilde{P} Q h_{\bar{\nu}_t}, \\ \theta_{t+1} &= (1 - \eta\tau)\theta_t - \eta[(M^\top)^{-1} \mid 0]\widetilde{P} Q h_{\bar{\nu}_{t+1}}, \end{aligned} \tag{3.4.9}$$

and a similar update for ν to reach the solution of the regularized problem under a log-linear parametrization. Here, the matrix $\widetilde{P}$ is defined as

$$\widetilde{P} = \begin{pmatrix} & \frac{-1}{n-d} & \cdots & \frac{-1}{n-d} \\ I_d & \cdots & \cdots & \cdots \\ & \frac{-1}{n-d} & \cdots & \frac{-1}{n-d} \\ \mathbf{0} & & \mathbf{0} & \end{pmatrix}.$$

[4] By equivalent, we mean that the two problems have the same value at the NE. We will also show the relationship between the solutions in Proposition 3.4.6

The additional term involving the inverse of the feature matrix arises due to the nature of the log-linear function approximation. We make this formal in the following proposition.

Proposition 3.4.6. Consider Algorithm 2 used to solve Problem (3.4.3) under the log-linear parametrization in Equation (3.2.2) under Assumption 3.4.1. Then the iterates of Algorithm 2 have the same guarantees provided in Theorem 3.3.4. Here, the NE parameter value to which the algorithm converges to is given by:

$$\theta^\star = \frac{-[(M^\top)^{-1} \mid 0]\widetilde{P}Qh_{\nu^\star}}{\tau}, \qquad \nu^\star = \frac{[(M^\top)^{-1} \mid 0]\widetilde{P}Q^\top g_{\theta^\star}}{\tau}.$$

Next, we show how the optimistic NPG algorithm solves the original matrix game without regularization.

Corollary 3.4.7. If we run Algorithm 2 for time $T = \mathcal{O}\left(\frac{\log n}{\eta\epsilon}\log\left(\frac{1}{\epsilon}\right)\right)$ and set the regularization parameter $\tau = \epsilon/(8\log n)$, we have that the output (θ_T, ν_T) is an ϵ-in-class NE (Definition 3.4.3) of the unregularized Problem (3.4.1).

3.5 Multi-player Monotone Games

Monotone games. Consider a multi-player continuous game over simplexes, which strictly generalizes the zero-sum matrix game in §4.2. The game is characterized by $(\mathcal{N}, \{\mathcal{A}_i\}_{i\in[N]}, \{f_i\}_{i\in[N]})$, where $\mathcal{N} = [N]$ is the set of players. Without loss of generality, we assume $|\mathcal{A}_i| = n$ for all $i \in [N]$. For notational convenience, let Δ denote the simplex over $\mathcal{A}_i$, and $z := (g_1, g_2, \cdots, g_N) \in \Delta^N$ denote the strategy profile of all N players, with each $g_i \in \Delta$. We define the *pseudo-gradient* operator $F : \Delta^N \to \mathbb{R}^{nN}$ as $F(z) := [\nabla_{g_i} f_i(g_i, g_{-i})]_{i=1}^N$. To make the N-player game tractable, we make the following standard assumptions on F [110, 94, 41].

Assumption 3.5.1 (Monotonicity & Smoothness)**.** The operator F is monotone and smooth, i.e., $\forall z, z' \in \Delta^N$

$$\langle F(z) - F(z'), z - z'\rangle \geq 0, \qquad \|F(z) - F(z')\| \leq L \cdot \|z - z'\|,$$

where $L > 0$ is the Lipschitz constant of the operator F.

The goal is to find the NE, given by strategy $z^\star$ such that $f_i(z_i^\star, z_{-i}^\star) \leq f_i(z_i, z_{-i}^\star),\ \forall\ z_i \in \Delta,\ i \in [N]$. Under Assumption 3.5.1, it is known that the NE exists [110].

Policy parameterization & regularized game. To develop policy gradient methods, we parameterize each policy $g_i \in \Delta$ by g_{θ_i} in the softmax form as before, i.e., for any $a_i \in \mathcal{A}_i$, $g_{\theta_i}(a_i) = e^{p_{\theta_i}(a_i)} \cdot \big(\sum_{a_i' \in \mathcal{A}_i} e^{p_{\theta_i}(a_i')}\big)^{-1}$, where $\theta_i \in \mathbb{R}^d$, and we consider both the tabular case with $p_{\theta_i} = \theta_i$ and the linear function approximation case with $p_{\theta_i}(a_i) = \phi_{a_i}^\top \theta_i$. This parameterization leads to the following set of optimization problems:

$$\min_{\theta_i \in \mathbb{R}^n} \quad f_i(g_{\theta_i}, g_{\theta_{-i}}), \qquad \forall\ i \in [N], \tag{3.5.1}$$

whose solution $(\theta_1^\star, \theta_2^\star, \cdots, \theta_N^\star)$, if exists, corresponds to the Nash equilibrium under this parameterization. Note that (3.5.1) can also be viewed as a nonconvex game (Lemma 3.2.2 is a special case) with a “hidden” *monotone variational inequality* structure, which generalizes the class of hidden convex-concave problems discussed in [48, 87].

Motivated by §4.2, we also consider the regularized game in hope of stronger convergence guarantees for solving (3.5.1). Specifically, the players solve

$$\min_{\theta_i \in \mathbb{R}^n} \quad f_i(g_{\theta_i}, g_{\theta_{-i}}) - \tau \mathcal{H}(g_{\theta_i}), \qquad \forall\ i \in [N], \tag{3.5.2}$$

where $\tau > 0$ and $\mathcal{H}$ is the entropy function. With a small enough τ, the solution to (3.5.2) approximates that to (3.5.1).

3.5.1 Softmax parameterization

We first consider the tabular softmax parameterization with $p_{\theta_i} = \theta_i \in \mathbb{R}^n$ for all $i \in [N]$. In this case, the Nash equilibrium $\theta^\star = (\theta_1^\star, \theta_2^\star, \cdots, \theta_N^\star)$ of the regularized monotone game (3.5.2) satisfies the following property.

Lemma 3.5.2. The NE of the game (3.5.2) exists. A vector $\theta^\star = (\theta_1^\star, \theta_2^\star, \cdots, \theta_N^\star)$ is a NE of (3.5.2) if and only if: $g_{\theta_i^\star}(a) \propto \exp\big(\frac{-[\nabla_{g_{\theta_i}} f_i(g_{\theta_i^\star}, g_{\theta_{-i}^\star})]_a}{\tau}\big)$ and the vector $(g_{\theta_1^\star}, g_{\theta_2^\star}, \cdots, g_{\theta_N^\star})$ is unique. We denote $g_i^\star := g_{\theta_i^\star}$.

Note that although the NE policy $(g_{\theta_1^\star}, g_{\theta_2^\star}, \cdots, g_{\theta_N^\star})$ is unique, the NE parameter $(\theta_1^\star, \theta_2^\star, \cdots, \theta_N^\star)$ is not necessarily the case. Motivated by §4.2 and §3.3, we propose the following update-rule for solving (3.5.2): $\forall$ players $i \in [N]$,

$$\bar{\theta}_i^{t+1} = (1-\eta\tau)\theta_i^t - \eta\nabla_{g_{\theta_i}} f_i(g_{\bar{\theta}_i^t}, g_{\bar{\theta}_{-i}^t}), \qquad \theta_i^{t+1} = (1-\eta\tau)\theta_i^t - \eta\nabla_{g_{\theta_i}} f_i(g_{\bar{\theta}_i^{t+1}}, g_{\bar{\theta}_{-i}^{t+1}}).$$

We refer to the update-rule as *optimistic NPG* (as summarized in Algorithm 3), as it corresponds to the optimistic version of the (specific instance of) natural PG direction for the regularized objective (3.5.2). We choose this specific instance of NPG due to the pitfall discussed in §4.2; and the optimistic update is meant to obtain fast last-iterate convergence. See §3.9.4 for a detailed derivation of the update rule.

As shown in §4.2, the problem (3.5.1) is nonconvex in the policy parameter space, and can be challenging in general. Our strategy is to show that our algorithm solves the regularized problem (3.5.2) fast, with last-iterate parameter convergence (see Theorem 3.5.3), which, with small enough τ, also solves the nonconvex game (3.5.1) (see Corollary 3.5.5).

Theorem 3.5.3. Let $z^\star = (g_i^\star)_{i=1}^N$ be the unique Nash equilibrium given in Lemma 3.5.2. Also, we denote $z_t = (g_{\theta_i^t})_{i=1}^N$. Then for Algorithm 3 with stepsize $0 < \eta < \frac{1}{2(N+4)L+2\tau}$, we have:

$$\begin{aligned} \max\{\mathrm{KL}(z^\star \| z_t), \mathrm{KL}(z^\star \| \bar{z}_{t+1})\} &\leq (1-\eta\tau)^t 2\mathrm{KL}(z^\star \| z_0), \\ \|\theta_{t+1} - \theta^\star\|^2 &\leq (1-\eta\tau/2)^t V_0, \end{aligned} \tag{3.5.3}$$

where $\theta_i^\star = \frac{-\nabla_{g_{\theta_i}} f_i(g_i^\star, g_{-i}^\star)}{\tau}$, $V_0 = \|\theta^t - \theta^\star\|^2 + \frac{2NC}{\eta\tau}$, and $C = 4\eta^2 L^2 \left(1 + \frac{1}{\eta\tau}(1-\eta\tau)^2\right) \mathrm{KL}(z^\star \| z_0)$.

The proof follows by first showing convergence in the policy space, in which we are dealing with a strongly convex problem under convex constraints. We then use this result, along with a novel Lyapunov function to demonstrate the convergence in the parameter space, in which it is a nonconvex problem. The proof technique might of independent interest, and might be generalized to showing convergence in other nonconvex games with a hidden monotonicity structure.

Remark 3.5.4. The proof for Theorem 3.5.3 follows by first showing the convergence rate of the Proximal Point method, and then observing that Optimistic methods approximate this method and could potentially achieve the same convergence rates (see [90] for a unified analysis). We provide a convergence analysis for the Proximal Point and Extragradient methods in §3.9.4.

Now, we present the convergence of Algorithm 3 to an ϵ-NE of the un-regularized problem (3.5.1).

Corollary 3.5.5. If we run Algorithm 3 for time $T = \mathcal{O}\left(\frac{N \log n}{\eta \epsilon} \log\left(\frac{1}{\epsilon}\right)\right)$ and set the regularization parameter $\tau = \epsilon/(4N \log n)$, we have that $\theta^\top = [\theta_1^\top, \theta_2^\top, \cdots, \theta_N^\top]$, the iterate at time T, is an ϵ-NE of Problem (3.5.1).

We extend the results to certain function approximation settings in §3.9.4 in the Appendix.

3.6 Optimistic NPG for Markov Games

We now generalize our results to the sequential decision-making case of Markov games.

Model. A two-player zero-sum Markov game is characterized by the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{B}, P, r, \gamma)$, where $\mathcal{S}$ is the state space; $\mathcal{A}, \mathcal{B}$ are the action spaces of players 1 and 2, respectively; $P : \mathcal{S} \times \mathcal{A} \times \mathcal{B} \to \Delta(\mathcal{S})$ denotes the transition probability of states; $r : \mathcal{S} \times \mathcal{A} \times \mathcal{B} \to [0, 1]$ denotes the bounded reward function of player 1 (thus $-r$ is the reward function of player 2); and $\gamma \in [0, 1)$ is the discount factor. The goal of player 1 (player 2) is to minimize (maximize) the long-term accumulated discounted reward.

Specifically, at each time t, player 1 (player 2) has a Markov stationary policy $g : \mathcal{S} \to \Delta(\mathcal{A})$ $(h : \mathcal{S} \to \Delta(\mathcal{B}))$. The state makes a transition from s_t to s_{t+1} following the probability distribution $P(\cdot \mid s_t, a_t, b_t)$, given (a_t, b_t). As in the Markov decision process model, one can define the *state-value function* under a pair of joint policies (g, h) as

$$V^{g,h}(s) := \mathbb{E}_{a_t \sim g(\cdot \mid s_t), b_t \sim h(\cdot \mid s_t)} \left[\sum_{t \geq 0} \gamma^t r(s_t, a_t, b_t) \,\middle|\, s_0 = s \right].$$

Also, the *state-action/Q-value function* under (g, h) is defined as

$$Q^{g,h}(s,a,b) := \mathbb{E}_{a_t \sim g(\cdot\,|\,s_t), b_t \sim h(\cdot\,|\,s_t)}\bigg[\sum_{t\geq 0}\gamma^t r(s_t,a_t,b_t)\,\bigg|\, s_0 = s, a_0 = a, b_0 = b\bigg].$$

Similar as the matrix game setting, a common solution concept in Markov game is also the (Markov perfect) *Nash equilibrium* policy pair $(g^\star, h^\star)$, which satisfies the following saddle-point inequality:

$$V^{g,h^\star}(s) \leq V^{g^\star,h^\star}(s) \leq V^{g^\star,h}(s), \qquad \forall\ s \in \mathcal{S}. \tag{3.6.1}$$

It follows from [119, 47] that there exists a Nash equilibrium $(g^\star, h^\star) \in \Delta(\mathcal{A})^{|\mathcal{S}|} \times \Delta(\mathcal{B})^{|\mathcal{S}|}$ for finite two-player discounted zero-sum MGs. The state-value $V^\star := V^{g^\star,h^\star}$ is referred to as the *value of the game.* The corresponding Q-value function is denoted by $Q^\star$.

We focus on the softmax parameterization g_θ and h_ν of the policies g and h, respectively.

Policy parameterization. Following the matrix game setting, we also use the softmax parameterization of the policies. Specifically, for any $\theta, \nu \in \mathbb{R}^d$, $(s,a,b) \in \mathcal{S} \times \mathcal{A} \times \mathcal{B}$,

$$g_\theta(a\,|\,s) = \frac{e^{p_\theta(s,a)}}{\sum\limits_{a' \in \mathcal{A}} e^{p_\theta(s,a')}}, \qquad h_\nu(b\,|\,s) = \frac{e^{q_\nu(s,b)}}{\sum\limits_{b' \in \mathcal{B}} e^{q_\nu(s,b')}}. \tag{3.6.2}$$

Note that for any $s \in \mathcal{S}$, $\sum_a g_\theta(a\,|\,s) = \sum_b h_\nu(b\,|\,s) = 1$. We will consider both

1) the tabular case with $p_\theta = \theta$ and $q_\nu = \nu$, where $\theta \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{S}|}$ and $\nu \in \mathbb{R}^{|\mathcal{B}|\times|\mathcal{S}|}$. We will use θ_s and ν_s to denote the parameters corresponding to state s, i.e., $\theta_s = [\theta_{s,1}, \theta_{s,2}, \cdots, \theta_{s,|A|}]$ and similarly ν_s;

2) the linear function approximation case with $p_\theta(s,a) = \theta \cdot \phi(s,a)$ and $q_\nu(s,a) = \nu \cdot \phi(s,a)$ (See §3.9.5 for more details)[5].

The parameterization thus leads to the following definition of the solution concept.

Definition 3.6.1 (ϵ-in-class-NE for Markov games)**.** The policy parameter pair $(\widetilde{\theta}, \widetilde{\nu})$ is an

[5] Once again, note that we use the same features for both players for notational convenience. Our results continue to hold when the two players have different features ϕ.

ϵ-in-class Nash equilibrium if it satisfies that for all $s \in \mathcal{S}$,

$$V^{\widetilde{\theta},\nu}(s) - \epsilon \geq V^{\widetilde{\theta},\widetilde{\nu}}(s) \geq V^{\theta,\widetilde{\nu}}(s) + \epsilon, \qquad \forall \theta, \nu \in \mathbb{R}^d, \tag{3.6.3}$$

where $V^{\theta,\nu}(s) = V^{g_\theta,h_\nu}(s)$ denotes the value of the parameterized policy pair (g_θ, h_ν). Note that if we are in the tabular setting, we will have $d = n$, and the definition covers that of the standard ϵ-NE for Markov games. We also define the ϵ-in-class NE Q-value accordingly.

Given that matrix games considered in §4.2 is a special case of the Markov games with $|\mathcal{S}| = 1$ and $\gamma = 0$, Lemma 3.2.2 implies that finding the NE in Definition 3.6.1 is nonconvex-nonconcave in general, and can be challenging to solve. We show in the following lemma that, for the tabular and linear function approximation settings we consider, such a parameterized NE exists.

Lemma 3.6.2 (Existence of parameterized/in-class NE)**.** Under policy parameterization (3.6.2) with tabular parameterization, the in-class NE defined in Definition 3.6.1 exists.

Motivated by §3.3 and §3.4, we consider the modified version of NPG with optimism to solve this problem.

Optimistic NPG. Following §4.2, we also consider the *regularized* Markov games [50, 144, 23], in hope of favorable convergence guarantees. Define the regularized value functions as

$$V_\tau^{\theta,\nu}(s) := \mathbb{E}_{a_t\sim g_\theta(\cdot\,|\,s_t), b_t\sim h_\nu(\cdot\,|\,s_t)}\left[\sum_{t=0}^{\infty}\gamma^t\big(r_t - \tau\log g_\theta(a_t|s_t) + \tau\log h_\nu(b_t|s_t)\big)\,\bigg|\, s_0 = s\right], \tag{3.6.4}$$

where $r_t = r(s_t, a_t, b_t)$, $\tau < 1$, and

$$Q_\tau^{\theta,\nu}(s,a,b) := r(s,a,b) + \gamma\mathbb{E}_{s'\sim\mathbb{P}(\cdot|s,a,b)}[V_\tau^{\theta,\nu}(s')]. \tag{3.6.5}$$

We denote by $V_\tau^\star$ and $Q_\tau^\star$, the NE value and Q-functions respectively, for the regularized Markov game ("regularized NE"), i.e., $V_\tau^\star = \min_\theta \max_\nu\ V_\tau^{\theta,\nu}$ and $Q_\tau^\star$ is the corresponding Q-function. Note that their existence follows along similar lines as Lemma 3.6.2. As a generalization of regularized matrix games in §4.2, the non-convergence pitfall of vanilla NPG also occurs. We also define the following notation

$$f_\tau\big(Q(s); g_\theta(\cdot \mid s), h_\nu(\cdot \mid s)\big) := -\ g_\theta(\cdot \mid s)^\top Q(s) h_\nu(\cdot \mid s) - \tau\mathcal{H}(g_\theta(\cdot \mid s)) + \tau\mathcal{H}(h_\nu(\cdot \mid s)). \quad (3.6.6)$$

3.6.1 Convergence guarantees

To stabilize the algorithm, we propose the update rule where the parameters (θ, ν) for all states are updated at a faster time scale, and the Q matrix is updated at a slower time scale. To be more precise, at every time t of the outer loop, we solve the matrix game [6]

$$\min_{\theta_s \in \mathbb{R}^n} \max_{\nu_s \in \mathbb{R}^n} \quad f_\tau(Q(s); g_\theta(\cdot \mid s), h_\nu(\cdot \mid s)), \quad (3.6.7)$$

for each state $s \in \mathcal{S}$ by running T_{inner} iterations of the Optimistic NPG algorithm (Algorithm 1). At the end of each inner loop, the outer loop updates the Q matrix for each state $s \in \mathcal{S}$ as $Q_{t+1}(s,a,b) = r(s,a,b) + \gamma\mathbb{E}_{s' \sim \mathbb{P}(\cdot|s,a,b)}[f_\tau(Q_t(s'); g_{\theta_{T_{inner}}}(\cdot \mid s'), h_{\nu_{T_{inner}}}(\cdot \mid s'))]$.

The complete algorithm is presented in Algorithm 6. Note that we use the name ONPG for Markov games because the inner matrix game is solved using the ONPG updates. The two-timescale-type update rule (between the policy and value updates) for solving infinite-horizon Markov games has also been used before in [112, 23, 131].

Next, we provide a convergence result for the performance of Algorithm 6 for the regularized Markov game.

Theorem 3.6.3. *Let $Q_\tau^\star$ be the NE Q-value of the regularized Markov Game under the tabular parametrization. Choose the stepsize $\eta = \frac{1-\gamma}{2(1+\tau(\log n + 1 - \gamma))}$ for the inner loop in Algorithm 6. Let T denote the total number of iterations $(T_{outer} \cdot T_{inner})$. Then, after*

$$\begin{aligned} T_{inner} &= \mathcal{O}\left(\frac{1}{\eta\tau}\left(\log\frac{1}{\epsilon} + \log\frac{1}{1-\gamma} + \log\log n + \log\frac{1}{\eta}\right)\right), \\ T_{outer} &= \mathcal{O}\left(\frac{1}{1-\gamma}\left(\log\frac{d}{\epsilon} + \log\left(\frac{8}{\tau}\left(1 + C^2\|Q^\star\|_F^2\right)\right) + \log\frac{1+\tau\log n}{1-\gamma}\right)\right), \end{aligned} \quad (3.6.8)$$

iterations, we have $\|Q_T - Q_\tau^\star\|_\infty \le \epsilon$ and $\max\{\|\theta_T - \theta^\star\|, \|\nu_T - \nu^\star\|\} \le \epsilon$ where (Q_T, θ_T, ν_T)

[6] Note that here we use the fact that $\min_{g_\theta(\cdot \mid s)} \max_{h_\nu(\cdot \mid s)}[f_\tau \cdots]$ is equivalent to $\min_{\theta_s \in \mathbb{R}^{|A|}} \max_{\nu_s \in \mathbb{R}^{|A|}}[f_\tau \cdots]$ for each s.

is the output of Algorithm 6 after T iterations, and $(\theta^\star, \nu^\star)$ are defined in Equation (3.9.185).

Finally, we show how the optimistic NPG algorithm solves the original Markov game without regularization.

Corollary 3.6.4. If we run Algorithm 6 for time $T_{inner} = \mathcal{O}\left(\frac{\log n}{(1-\gamma)^2\epsilon}\log\left(\frac{1}{\epsilon}\right)\right)$, $T_{outer} = \mathcal{O}\left(\frac{1}{(1-\gamma)}\log\left(\frac{1}{\epsilon}\right)\right)$, and setting $\tau = \mathcal{O}((1-\gamma)\epsilon/\log n)$, the output (θ_T, ν_T) will be an ϵ-in-class NE (Definition 3.6.1) of the original unregularized Markov game.

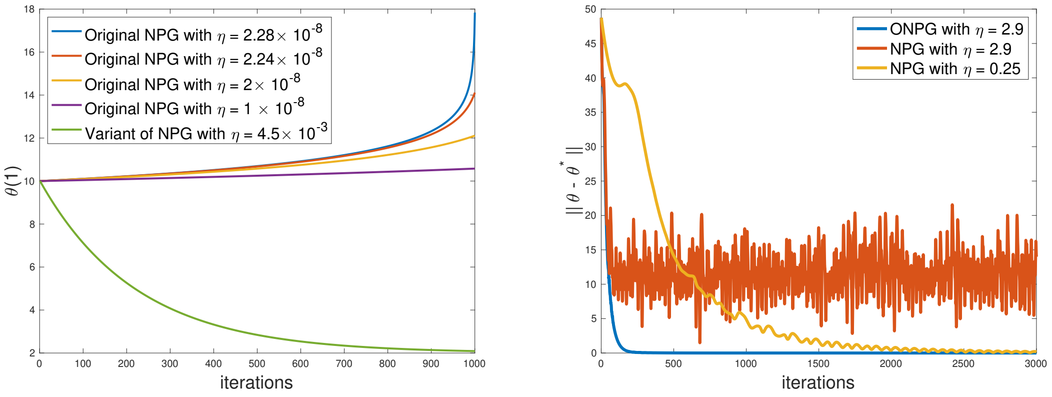
Remark 3.6.5. We remark that Theorem 3.6.3, to the best of our knowledge, is the first to show policy parameter convergence in Markov games with policy parametrization, and is different from the policy convergence results of several recent works [151, 131, 23] (see §3.1.1 for a detailed comparison).

We extend these results to simple function approximation settings in §3.9.5 in the Appendix.

3.7 Simulations

We now provide simulation results to corroborate our theoretical results. First, we study matrix games under the tabular setting in Figure 3-1a. Here, we show the behavior of vanilla NPG and our proposed variant (Equations (3.3.1)-(3.3.2)). We plot the first element of the iterate, i.e., $\theta(1)$ on the y-axis. It is shown that even for vanishingly small stepsizes, vanilla NPG diverges, whereas the proposed variant converges even with reasonable step-size choices. The cost matrix Q is taken to be an identity matrix of dimension 5.

Next, we confirm the convergence of our variant of NPG in Figure 3-1b. In this figure, we compare the behavior of ONPG and NPG, and show that ONPG admits convergence for larger stepsizes. Smaller stepsizes that enables NPG convergence would lead to a slower convergence rate than ONPG. This is in line with our results in Theorems 3.3.2 and 3.3.4.



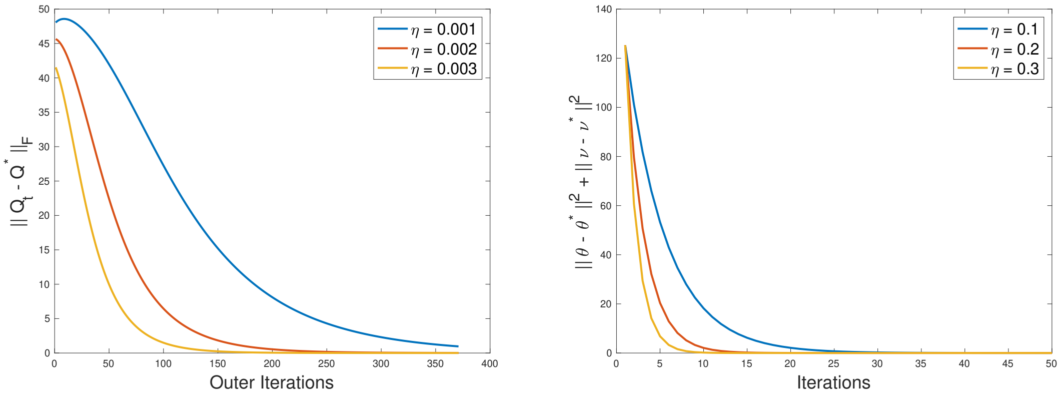
(a) Vanilla NPG vs proposed variant of NPG.

(b) NPG vs ONPG.

Figure 3-1: Comparison of vanilla NPG, proposed variant of NPG and ONPG in matrix game under the tabular setting, in terms of parameter convergence.

3.7.1 ONPG in Markov games with function approximation

Figures 3-2a and 3-2b study the behavior of Algorithm 7 in Markov games with log-linear function approximation, and corroborate the results of Theorem 3.9.22. Here we take the feature matrix $\Phi \in \mathbb{R}^{10\times 100}$, and $|\mathcal{S}| = 10$, i.e., there are 10 states. The first 10 columns correspond to the first action of each of the 10 states. This means that $\Phi_{(s,1)} = e_s$ for $s = \{1, 2, \cdots, 10\}$ where e_s is a standard basis vector with element 1 at position s. We take the discount factor $\gamma = 0.8$. We take the transition probability to be uniform for each state action pair, i.e., $\mathbb{P}(\cdot|s, a, b) = 1/10$ for all (s, a, b), i.e., $\mathbb{P}(s'|s, a, b) = 1/10$ for all state $s' \in \mathcal{S}$. Finally, we take the regularization parameter $\tau = 0.1$.



(a) Q-matrix to in-class NE Q-matrix. (b) Parameters to the in-class NE parameters.

Figure 3-2: Convergence in Markov Games with linear function approximation.

3.8 Concluding Remarks

In this chapter we study the global last-iterate parameter convergence of symmetric policy gradient methods for multi-agent learning. We identified the non-convergence issue of vanilla natural PG in policy parameters, even in presence of regularized reward function, and developed variants of natural PG methods that enjoy last-iterate parameter convergence. We have then expanded the scope of the symmetric PG methods for multi-agent learning, and incorporated function approximation to handle large state-action spaces. Future work includes embracing more general function approximation in policy parameterization, and exploring the power of our approach in nonconvex-nonconcave minimax optimization with other hidden convex structures.

3.9 Appendix

In this section, we provide missing details and proofs from the main part of the chapter.

3.9.1 Missing Details and Proofs in §3.2

Proof of Lemma 3.2.2

We show that the problem

$$\min_{\theta \in \mathbb{R}^n} \max_{\nu \in \mathbb{R}^n} g_\theta^\top Q h_\nu, \tag{3.9.1}$$

is nonconvex-nonconcave.

Let

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{3.9.2}$$

Consider $\theta_1 = (0, 0)$ and $\theta_2 = (\log 4, \log 9)$. This implies $g_{\theta_1} = (1/2, 1/2)^\top$ and $g_{\theta_2} = (4/13, 9/13)^\top$. Also, from the form of Q, we have $Qh_\nu = [h_\nu(1), h_\nu(2)]^\top, \ \ \forall \nu$. Now, for $[h_\nu(1), h_\nu(2)] = [1/3, 2/3]$, we have

$$\frac{1}{2}(g_{\theta_1}^\top Q h_\nu + g_{\theta_2}^\top Q h_\nu) < g_{(\theta_1+\theta_2)/2} Q h_\nu, \tag{3.9.3}$$

which implies nonconvexity in θ.

Similarly, taking $\nu_1 = (0, 0)$ and $\nu_2 = (\log 4, \log 9)$ (which implies $h_{\nu_1} = (1/2, 1/2)^\top$ and $h_{\nu_2} = (4/13, 9/13)^\top$, and taking $g_\theta = (2/3, 1/3)^\top$, we have

$$\frac{1}{2}(g_\theta^\top Q h_{\nu_1} + g_\theta^\top Q h_{\nu_2}) > g_\theta Q h_{(\nu_1+\nu_2)/2}, \tag{3.9.4}$$

which implies nonconcavity in ν.

Note that adding regularization does not get rid of this convexity. For example consider the specific case when h_ν is a constant policy and the matrix Q is 0. We show the nonconvexity of the function $-\tau \mathcal{H}(g_\theta)$ in θ next. Consider $\theta_1 = (0, 0)$ and $\theta_2 = (\log 4, \log 9)$. We have $g_{\theta_1} = (1/2, 1/2)$ and $g_{\theta_2} = (4/13, 9/13)$. Furthermore, we have $g_{(\theta_1+\theta_2)/2} = (1/3, 2/3)$, and we can see that:

$$\frac{-\tau}{2}(\mathcal{H}(g_{\theta_1}) + \mathcal{H}(g_{\theta_2})) < -\tau\mathcal{H}(g_{(\theta_1+\theta_2)/2}) \tag{3.9.5}$$

which shows nonconvexity of $-\tau\mathcal{H}(g_\theta)$. This completes the proof of the lemma. □

Vanilla NPG for matrix games

Next, we compute the Fisher Information Matrix $F_\theta(\theta) = \mathbb{E}_{a\sim g_\theta}\left[\left(\nabla_\theta \log g_\theta(a)\right)\left(\nabla_\theta \log g_\theta(a)\right)^\top\right]$. For the softmax parametrization, we have:

$$\begin{aligned}\nabla_\theta \log g_\theta(a) &= \nabla_\theta \left(\theta(a) - \log\Big(\sum_{a'\in\mathcal{A}} e^{\theta(a')}\Big)\right)\\ &= [-g_\theta(1), -g_\theta(2), \cdots, 1 - g_\theta(a), \cdots, -g_\theta(n)]^\top.\end{aligned} \tag{3.9.6}$$

Now, consider the $(i,i)^{th}$ element of the Fisher information matrix. We have:

$$[F_\theta(\theta)]_{ii} = g_\theta(i)(1 - g_\theta(i))(1 - g_\theta(i)) + \sum_{j\neq i} g_\theta(j)g_\theta(i)^2 = g_\theta(i)(1 - g_\theta(i)). \tag{3.9.7}$$

Similarly, we have the $(i,j)^{th}$ element, where $i \neq j$ is given by:

$$\begin{aligned}&[F_\theta(\theta)]_{ij}\\ &= (1 - g_\theta(i) - g_\theta(j))g_\theta(i)g_\theta(j) - g_\theta(i)(1 - g_\theta(i))g_\theta(j) - g_\theta(j)(1 - g_\theta(j))g_\theta(i)\\ &= -g_\theta(i)g_\theta(j).\end{aligned} \tag{3.9.8}$$

Therefore, the matrix $F_\theta(\theta)$ can be succinctly written as:

$$F_\theta(\theta) = \mathrm{diag}(g_\theta) - g_\theta g_\theta^\top, \tag{3.9.9}$$

where $\text{diag}(g_\theta)$ is a diagonal matrix with entries g_θ. Note that this is in fact $\nabla_\theta g_\theta$ (see [85]), i.e.,

$$\nabla_\theta g_\theta = \text{diag}(g_\theta) - g_\theta g_\theta^\top. \tag{3.9.10}$$

Therefore, we have:

$$F_\theta^\dagger(\theta)\nabla_\theta g_\theta = I. \tag{3.9.11}$$

The update of the vanilla NPG thus simplifies to the following:

$$\begin{aligned}\theta_{t+1} &= \theta_t - \eta \cdot F_\theta^\dagger(\theta_t) \cdot \frac{\partial f_\tau(\theta_t, \nu_t)}{\partial \theta} = \theta_t - \eta \frac{\partial f_\tau(\theta_t, \nu_t)}{\partial g_\theta} \\ &= \theta_t - \eta \left(Qh_{\nu_t} + \tau(\mathbb{1} + \log g_{\theta_t})\right). \end{aligned} \tag{3.9.12}$$

However, since $g_\theta(a) = \frac{e^{\theta(a)}}{\sum_{a' \in \mathcal{A}} e^{\theta(a')}}$, we have

$$\theta_{t+1}(a) = (1 - \eta\tau)\theta_t(a) - \eta\bigg([Qh_{\nu_t}]_a + \tau - \tau \log \sum_{a' \in \mathcal{A}} e^{\theta_t(a')}\bigg). \tag{3.9.13}$$

A similar update for ν leads to the updates in Equations (3.2.5)-(3.2.6). Note that when we write a constant in the update, we mean a constant vector with all elements being the same.

Proof of Lemma 3.2.3

We restate the lemma here first for convenience:

Lemma 3.9.1 (Pitfall of vanilla NPG)**.** There exists a game (3.2.4) with $\tau \geq 0$ (we allow for unregularized games as well) and a dummy player 2, i.e., $|\mathcal{B}| = 1$, for which the updates (3.2.5)-(3.2.6) do not converge for any $\eta > 0$.

Proof. Consider the θ update under NPG:

$$\theta_{t+1}(a) = (1 - \eta\tau)\theta_t(a) - \eta\bigg([Qh_{\nu_t}]_a + \tau - \tau \log \sum_{a' \in \mathcal{A}} e^{\theta_t(a')}\bigg). \tag{3.9.14}$$

From here, it is easy to see that it need not converge for the case where $\tau = 0$, since this would require $Qh_{\nu_t} = 0$ which need not be the case (For example consider $Q = [1 \mid 1]^\top$, and $h_\nu = [1]$. In this case, $h_{\nu_t} = 1$ for any parameter ν_t).

Next, we consider the case where $\tau > 0$. Suppose θ converges to some point $\theta^\star$. Since $|\mathcal{B}| = 1$, we have $h_{\nu_t} = [1]$. Substituting the point $\theta^\star$ into the update we have:

$$\theta^\star(a) = (1 - \eta\tau)\theta^\star(a) - \eta\Bigg([Q]_a + \tau - \tau \log \sum_{a' \in \mathcal{A}} e^{\theta^\star(a')}\Bigg). \tag{3.9.15}$$

This implies:

$$\eta\tau\theta^\star(a) = -\eta\tau\left(\frac{[Q]_a}{\tau} + 1\right) + \eta\tau \log \sum_{a' \in \mathcal{A}} e^{\theta^\star(a')}. \tag{3.9.16}$$

This leads to:

$$\log e^{\theta^\star(a)} - \log \sum_{a' \in \mathcal{A}} e^{\theta^\star(a')} = -\frac{[Q]_a}{\tau} - 1. \tag{3.9.17}$$

However,

$$\log e^{\theta^\star(a)} - \log \sum_{a' \in \mathcal{A}} e^{\theta^\star(a')} = \log \frac{e^{\theta^\star(a)}}{\sum_{a' \in \mathcal{A}} e^{\theta^\star(a')}} = \log g_{\theta^\star}(a). \tag{3.9.18}$$

Substituting this in Equation (3.9.17), we have:

$$g_{\theta^\star}(a) = \exp\left(-\frac{[Q]_a}{\tau} - 1\right). \tag{3.9.19}$$

This need not be a valid probability measure. For example, consider $Q = [-2 \mid 2]^\top$ and $\tau = 1$, we have:

$$g_{\theta^\star}(1) = e > 1, \tag{3.9.20}$$

which contradicts the fact that g_θ is a probability measure. This implies that the original NPG updates cannot have a fixed point, and therefore does not converge for any stepsize

Algorithm 1 Optimistic NPG

```
**Initialize:** $\theta_0 = 0$ and $\nu_0 = 0$.
**for** $t = 1, 2, \cdots$ **do**
    $\bar{\theta}_{t+1} = (1 - \eta\tau)\theta_t - \eta Q h_{\bar{\nu}_t}$
    $\bar{\nu}_{t+1} = (1 - \eta\tau)\nu_t + \eta Q^\top g_{\bar{\theta}_t}$

    $\theta_{t+1} = (1 - \eta\tau)\theta_t - \eta Q h_{\bar{\nu}_{t+1}}$
    $\nu_{t+1} = (1 - \eta\tau)\nu_t + \eta Q^\top g_{\bar{\theta}_{t+1}}$
**end for**
```

$\eta > 0$. □

3.9.2 Missing Details and Proofs in §3.3

Remark 3.9.2. We note that all results presented in this section also follow for the case where the action spaces for both players are asymmetric. However, we stick to the case where the number of actions is the same for both players, for ease of exposition.

Proof of Theorem 3.3.2

Policy convergence: Consider the following modified NPG updates for the regularized game:

$$\theta_{t+1} = (1 - \eta\tau)\theta_t - \eta Q h_{\nu_t}, \tag{3.9.21}$$

$$\nu_{t+1} = (1 - \eta\tau)\nu_t + \eta Q^\top g_{\theta_t}. \tag{3.9.22}$$

Note that these updates correspond to the popular Multiplicative Weights Update [49, 5] for the regularized game in policy space (we succinctly represent g_{θ_t} and h_{ν_t} as g_t and h_t, respectively), i.e.,

$$\begin{aligned} g_{t+1}(a) &\propto g_t(a)^{(1-\eta\tau)} e^{-\eta[Qh_t]_a}, \\ h_{t+1}(b) &\propto h_t(b)^{(1-\eta\tau)} e^{\eta[Q^\top g_t]_b}. \end{aligned} \tag{3.9.23}$$

We can write these updates as a mirror descent update with Bregman function given by the negative entropy (i.e., the corresponding Bregman distance is the KL divergence) as follows:

$$
\begin{aligned}
g_{t+1} &= \underset{g\in\Delta}{\operatorname{argmin}} \,\{\langle Qh_t + \tau\nabla_g\mathcal{H}(g_t), g\rangle + \mathrm{KL}(g\|g_t)\},\\
h_{t+1} &= \underset{h\in\Delta}{\operatorname{argmax}} \,\{\langle Q^\top g_t - \tau\nabla_h\mathcal{H}(h_t), h\rangle - \mathrm{KL}(h\|h_t)\}.
\end{aligned} \tag{3.9.24}
$$

Note that we can write these updates succinctly as one Mirror Descent update in the following form:

$$
z_{t+1} = \underset{z\in\mathcal{Z}}{\operatorname{argmin}} \,\{\langle Mz_t + \tau\nabla\mathcal{H}(z_t), z\rangle + \mathrm{KL}(z\|z_t)\}, \tag{3.9.25}
$$

where $z = [g; h]$, $\mathcal{Z} \in \Delta \times \Delta$ and with slight abuse of notation, we define $\nabla\mathcal{H}(z) = [\nabla_g\mathcal{H}(g); \nabla_h\mathcal{H}(h)]$. Also, we define the matrix

$$
M = \begin{pmatrix} 0 & Q \\ -Q^\top & 0 \end{pmatrix}. \tag{3.9.26}
$$

We can now use properties of mirror decent to analyze the iterates of MWU.

First, we have the following two lemmas which follow from [11], Proposition 2.3, and Lemma D.4 in [121] which will be used to derive the final convergence rate:

Lemma 3.9.3. For all $z \in \mathcal{Z}$, we have

$$
\eta\langle Mz_t + \tau\nabla\mathcal{H}(z_t), z_{t+1} - z\rangle \le \mathrm{KL}(z\|z_t) - \mathrm{KL}(z\|z_{t+1}) - \mathrm{KL}(z_{t+1}\|z_t). \tag{3.9.27}
$$

Lemma 3.9.4. For all $z \in \mathcal{Z}$, we have

$$
\eta\langle Mz + \tau\nabla\mathcal{H}(z), z^* - z\rangle \le -\eta\tau(\mathrm{KL}(z\|z^*) + \mathrm{KL}(z^*\|z)). \tag{3.9.28}
$$

In the next lemma, we show that the iterates of MWU on the regularized problem will be bounded away from the boundary of the simplex.

Lemma 3.9.5. For $\eta < 1/\tau$, the iterates of regularized MWU stay within a set $\Delta' \subset \Delta$ which is bounded away from the boundary of the simplex, i.e., $x_i \ge \delta > 0$ for some $\delta > 0$,

$\forall x \in \Delta'$.

Proof. Consider the update of g. We have the following property from Mirror Descent (see [13])

$$\begin{aligned}\mathrm{KL}(g^*\|g_{t+1}) &\le (1-\eta\tau)\mathrm{KL}(g^*\|g_t) - (1-\eta\tau)\mathrm{KL}(g_{t+1}\|g_t) - \eta\tau\mathcal{H}(g_{t+1}) + \eta\tau\mathcal{H}(g^*) \\ &\qquad - \eta\langle g_{t+1} - g^*, Qh_t\rangle \\ &\le (1-\eta\tau)\mathrm{KL}(g^*\|g_t) + 2\eta\tau\log n + 2\eta\|Q\|_\infty. \end{aligned} \tag{3.9.29}$$

This implies that

$$\mathrm{KL}(g^*\|g_t) \le 2\log n + \frac{2\|Q\|_\infty}{\tau} + \mathrm{KL}(g^*\|g_0). \tag{3.9.30}$$

From the definition of the KL divergence and Equation (3.9.30), we have:

$$g_{t,i} \ge \exp\left(\frac{-1}{g^*_{\min}}\left(2\log n + \frac{2\|Q\|_\infty}{\tau} + \mathrm{KL}(g^*\|g_0) + \mathcal{H}(g^*)\right)\right) > 0, \qquad \forall\ i, \quad \text{and} \quad \forall\ t. \tag{3.9.31}$$

Here $g^*_{\min}$ is the smallest value of the Nash Equilibrium policy (which is greater than 0 since the NE policy of the regularized game is in the interior of the simplex). This completes the proof of the Lemma. □

Since the iterates lie within Δ', we let L denote the Lipschitz constant of $Mz + \tau\nabla\mathcal{H}(z)$ (a continuous function over $\Delta \times \Delta$ whose norm approaches infinity as z approaches the boundary) in the set $\mathcal{Z}'$ (where $\mathcal{Z}' = \Delta' \times \Delta'$), i.e.,

$$\|(Mz_1 + \tau\nabla\mathcal{H}(z_1)) - (Mz_2 + \tau\nabla\mathcal{H}(z_2))\| \le L\|z_1 - z_2\|, \qquad \forall z_1, z_2 \in \mathcal{Z}'. \tag{3.9.32}$$

Now, we use these lemmas to derive the convergence rate of MWU for the regularized problem.

Theorem 3.9.6. Consider the modified NPG updates in Equation (3.9.21)-(3.9.22) with stepsize satisfying $0 \leq \eta \leq \tau/L^2$. We then have:

$$\mathrm{KL}(z^*\|z_{t+1}) \leq \left(1 - \frac{\eta\tau}{2}\right) \mathrm{KL}(z^*\|z_t). \tag{3.9.33}$$

Proof. Note that the constraint $\eta \leq \tau/L^2$ will automatically satisfy $\eta < 1/\tau$ (as needed by Lemma 3.9.5) since $\tau \leq L$.

We have the following string of inequalities:

$$\begin{aligned}
\mathrm{KL}(z^*\|z_{t+1}) &\leq^{*1} KL(z^*\|z_t) - \mathrm{KL}(z_{t+1}\|z_t) + \eta\langle F(z_t) + \tau\nabla\mathcal{H}(z_t), z^* - z_{t+1}\rangle \\
&= \mathrm{KL}(z^*\|z_t) - \mathrm{KL}(z_{t+1}\|z_t) + \eta\langle F(z_{t+1}) + \tau\nabla g(z_{t+1}), z^* - z_{t+1}\rangle \\
&\quad + \eta\langle (F(z_t) + \tau\nabla\mathcal{H}(z_t)) - (F(z_{t+1}) + \tau\nabla\mathcal{H}(z_{t+1})), z^* - z_{t+1}\rangle \\
&\leq^{*2} \mathrm{KL}(z^*\|z_t) - \mathrm{KL}(z_{t+1}\|z_t) - \eta\tau\left(\mathrm{KL}(z_{t+1}\|z^*) + \mathrm{KL}(z^*\|z_{t+1})\right) \\
&\quad + \eta\langle (F(z_t) + \tau\nabla\mathcal{H}(z_t)) - (F(z_{t+1}) + \tau\nabla\mathcal{H}(z_{t+1})), z^* - z_{t+1}\rangle \\
&\leq^{*3} \mathrm{KL}(z^*\|z_t) - \mathrm{KL}(z_{t+1}\|z_t) - \eta\tau\left(\mathrm{KL}(z_{t+1}\|z^*) + \mathrm{KL}(z^*\|z_{t+1})\right) \\
&\qquad + \eta L\|z_{t+1} - z_t\|\|z^* - z_{t+1}\| \\
&\leq^{*4} \mathrm{KL}(z^*\|z_t) - \mathrm{KL}(z_{t+1}\|z_t) - \eta\tau\left(\mathrm{KL}(z_{t+1}\|z^*) + \mathrm{KL}(z^*\|z_{t+1})\right) \\
&\qquad + \frac{1}{2}\|z_{t+1} - z_t\|^2 + \frac{\eta^2 L^2}{2}\|z^* - z_{t+1}\|^2 \\
&\leq^{*5} \mathrm{KL}(z^*\|z_t) - \mathrm{KL}(z_{t+1}\|z_t) - \mathrm{KL}(z_{t+1}\|z^*) + \eta^2 L^2 \mathrm{KL}(z_{t+1}\|z^*) \\
&\qquad\quad - \eta\tau\mathrm{KL}(z_{t+1}\|z^*) - \eta\tau\mathrm{KL}(z^*\|z_{t+1}) \\
&\leq^{*6} \mathrm{KL}(z^*\|z_t) - \eta\tau\mathrm{KL}(z^*\|z_{t+1}).
\end{aligned} \tag{3.9.34}$$

Here (∗1) follows from Lemma 3.9.3, (∗2) follows from Lemma 3.9.4, (∗3) follows from Equation (3.9.32), (∗4) follows from Young's inequality, (∗5) follows from Pinskers inequality and (∗6) follows from $\eta \leq \tau/L^2$. Therefore, we have

$$\mathrm{KL}(z^*\|z_{t+1}) \leq \frac{1}{1+\eta\tau}\mathrm{KL}(z^*\|z_t) \leq \left(1 - \frac{\eta\tau}{2}\right) \mathrm{KL}(z^*\|z_t), \tag{3.9.35}$$

which completes the proof. □

Parameter convergence: Now, we show convergence of the policy parameters. We have the following theorem.

Theorem 3.9.7. Consider the modified NPG updates in Equation (3.9.21)-(3.9.22) with stepsize satisfying $0 \leq \eta \leq \tau/L^2$. We then have:

$$\|\theta_{t+1} - \theta^\star\|^2 + \|\nu_{t+1} - \nu^\star\|^2 \leq (1 - \eta\tau/4)^t \left(\|\theta_0 - \theta^\star\|^2 + \|\nu_0 - \nu^\star\|^2 + \frac{4C}{\eta\tau} \right), \tag{3.9.36}$$

where $\theta^\star = \frac{-Qh^\star}{\tau}$, $\nu^\star = \frac{Q^\top g^\star}{\tau}$ and $C = \left(1 + \frac{1}{\eta\tau}(1 - \eta\tau)^2\right) 4\eta^2 \|Q\|_\infty^2 \mathrm{KL}(z^\star \| z_0)$.

Proof. We begin by first providing the intuition of the proof, when the opponent is playing the NE strategy. We denote the NE strategies of the players as $g^\star$ and $h^\star$. Then, the NPG update has the following form:

$$\theta_{t+1} = (1 - \eta\tau)\theta_t - \eta Q h^\star. \tag{3.9.37}$$

We know that the NE satisfy:

$$g^\star(a) = \frac{e^{-[Qh^\star]_a/\tau}}{\sum_{a' \in \mathcal{A}} e^{-[Qh^\star]_{a'}/\tau}} = \frac{e^{-[Qh^\star]_a/\tau}}{K}, \tag{3.9.38}$$

where we define $K := \sum_{a' \in \mathcal{A}} e^{-[Qh^\star]_{a'}/\tau}$. Taking log on both sides, we have:

$$-\eta Q h^\star = \eta\tau \log g^\star + \eta\tau \log K. \tag{3.9.39}$$

Substituting this back into the θ update in (3.9.37), we have:

$$\begin{aligned}
\theta_{t+1} &= (1 - \eta\tau)\theta_t + \eta\tau \log g^\star + \eta\tau \log K \\
&= \theta_t - \eta\tau\theta_t + \eta\tau \log g^\star + \eta\tau \log K - \eta\tau \log Z_{\theta_t} + \eta\tau \log Z_{\theta_t} \\
&= \theta_t + \eta\tau \log g^\star - \eta\tau \log \left(\frac{e^{\theta_t}}{Z_{\theta_t}} \right) + \eta\tau \log \left(\frac{K}{Z_{\theta_t}} \right) \\
&= \theta_t + \eta\tau \log g^\star - \eta\tau \log g_{\theta_t} + \eta\tau \log \left(\frac{K}{Z_{\theta_t}} \right) = \theta_t + \eta\tau \log \left(\frac{g^\star}{g_{\theta_t}} \right) + \eta\tau \log \left(\frac{K}{Z_{\theta_t}} \right) \\
&= \theta_t + \eta\tau \log \left(\frac{g^\star K}{g_{\theta_t} Z_{\theta_t}} \right).
\end{aligned} \tag{3.9.40}$$

We can further simplify this as:

$$\theta_{t+1} = \theta_t + \eta\tau \log\left(\frac{e^{-[Qh^\star]/\tau}}{e^{\theta_t}}\right), \tag{3.9.41}$$

which is nothing but:

$$\theta_{t+1} = \theta_t + \eta\tau \left(\frac{-Qh^\star}{\tau} - \theta_t\right). \tag{3.9.42}$$

Note that this is the Gradient Descent update on the strongly convex function $\frac{1}{2}\left\|\frac{-Qh^\star}{\tau} - \theta\right\|^2$ with stepsize $\eta\tau$. This update leads to the following convergence guarantees:

$$\|\theta_{t+1} - \theta^\star\|^2 \leq (1-\eta\tau)^2 \|\theta_t - \theta^\star\|^2, \tag{3.9.43}$$

where $\theta^\star = \frac{-Qh^\star}{\tau}$.

The analysis above shows that if one of the players is already at the NE strategy, the parameters of the second player converges to the NE at a linear rate. However, the original NPG update for θ is given by

$$\theta_{t+1} = (1-\eta\tau)\theta_t - \eta Q h^\star + \eta Q(h^\star - h_{\nu_t}). \tag{3.9.44}$$

Since $h_{\bar{\nu}_{t+1}}$ converges to $h^\star$ at a linear rate (from Theorem 3.9.6), we expect the term

$$\varepsilon_t = \eta Q(h^\star - h_{\nu_t}), \tag{3.9.45}$$

to be small, and goes to 0. This is formalized in what follows.

The NPG update for θ can be re-written using ε_t as:

$$\theta_{t+1} = \theta_t + \eta\tau\left(\frac{-Qh^\star}{\tau} - \theta_t\right) + \varepsilon_t. \tag{3.9.46}$$

Once again, defining $\theta^\star = \frac{-Qh^\star}{\tau}$, we have:

$$\begin{aligned}\|\theta_{t+1} - \theta^\star\|^2 = \|\theta_t + \eta\tau\left(\theta^\star - \theta_t\right) + \varepsilon_t - \theta^\star\|^2 &= \|(1-\eta\tau)(\theta_t - \theta^\star) + \varepsilon_t\|^2 \\ &= (1-\eta\tau)^2\|\theta_t - \theta^\star\|^2 + 2(1-\eta\tau)(\theta_t - \theta^\star)^\top\varepsilon_t + \|\varepsilon_t\|^2 \\ &\leq^{*1} (1-\eta\tau)^2\|\theta_t - \theta^\star\|^2 + \eta\tau\|\theta_t - \theta^\star\|^2 + \frac{1}{\eta\tau}(1-\eta\tau)^2\|\varepsilon_t\|^2 + \|\varepsilon_t\|^2 \\ &= (1-\eta\tau+\eta^2\tau^2)\|\theta_t - \theta^\star\|^2 + \left(1 + \frac{1}{\eta\tau}(1-\eta\tau)^2\right)\|\varepsilon_t\|^2, \end{aligned} \tag{3.9.47}$$

where $*1$ follows from Young's inequality.

Next, we analyze the error term $\|\varepsilon_t\|$. We have:

$$\|\varepsilon_t\|^2 = \|\eta Q(h^\star - h_{\nu_t})\|^2 \leq \eta^2\|Q\|_\infty^2\|(h^\star - h_{\nu_t})\|_1^2 \leq^{*1} 2\eta^2\|Q\|_\infty^2(\mathrm{KL}(h^\star\|h_{\nu_t})). \tag{3.9.48}$$

Here $*1$ follows from Pinsker's Inequality. Now, writing the same inequality for ν, we have:

$$\begin{aligned}&\|\theta_{t+1} - \theta^\star\|^2 + \|\nu_{t+1} - \nu^\star\|^2 \\ &\leq (1-\eta\tau+\eta^2\tau^2)(\|\theta_t - \theta^\star\|^2 + \|\nu_t - \nu^\star\|^2) \\ &\qquad\qquad + \left(1 + \frac{1}{\eta\tau}(1-\eta\tau)^2\right)2\eta^2\|Q\|_\infty^2\mathrm{KL}(z^\star\|z_t) \\ &\leq (1-\eta\tau+\eta^2\tau^2)(\|\theta_t - \theta^\star\|^2 + \|\nu_t - \nu^\star\|^2) \\ &\qquad\qquad + \left(1 + \frac{1}{\eta\tau}(1-\eta\tau)^2\right)4\eta^2 C\|Q\|_\infty^2\left(1 - \frac{\eta\tau}{2}\right)^t\mathrm{KL}(z^\star\|z_0).\end{aligned}$$

Define:

$$C = \left(1 + \frac{1}{\eta\tau}(1-\eta\tau)^2\right)4\eta^2\|Q\|_\infty^2\mathrm{KL}(z^\star\|z_0). \tag{3.9.49}$$

Substituting back in Equation (3.9.70), along with the corresponding expression for ν, we have:

$$\|\theta_{t+1} - \theta^\star\|^2 + \|\nu_{t+1} - \nu^\star\|^2 \leq (1-\eta\tau+\eta^2\tau^2)(\|\theta_t - \theta^\star\|^2 + \|\nu_t - \nu^\star\|^2) + C\left(1 - \frac{\eta\tau}{2}\right)^t. \tag{3.9.50}$$

For $\eta\tau < 1/2$ we have:

$$\|\theta_{t+1} - \theta^\star\|^2 + \|\nu_{t+1} - \nu^\star\|^2 \leq (1 - \eta\tau/4)(\|\theta_t - \theta^\star\|^2 + \|\nu_t - \nu^\star\|^2) + C(1 - \eta\tau/2)^t. \tag{3.9.51}$$

Consider the Lyapunov function:

$$V_{t+1} = \|\theta_{t+1} - \theta^\star\|^2 + \|\nu_{t+1} - \nu^\star\|^2 + \frac{4C}{\eta\tau}(1 - \eta\tau/2)^{t+1}. \tag{3.9.52}$$

We have:

$$\begin{aligned} V_{t+1} &\leq (1 - \eta\tau/4)(\|\theta_t - \theta^\star\|^2 + \|\nu_t - \nu^\star\|^2) + C(1 - \eta\tau/2)^t + \frac{4C}{\eta\tau}(1 - \eta\tau/2)(1 - \eta\tau/2)^t \\ &= (1 - \eta\tau/4)(\|\theta_t - \theta^\star\|^2 + \|\nu_t - \nu^\star\|^2) + \frac{4C}{\eta\tau}(1 - \eta\tau)^t(1 - \eta\tau/4) \\ &= (1 - \eta\tau/4)\left(\|\theta_t - \theta^\star\|^2 + \|\nu_t - \nu^\star\|^2 + \frac{4C}{\eta\tau}(1 - \eta\tau)^t\right) = (1 - \eta\tau/4)V_t. \end{aligned} \tag{3.9.53}$$

This shows linear convergence of the parameter θ to $\theta^\star$ since:

$$\|\theta_{t+1} - \theta^\star\|^2 + \|\nu_{t+1} - \nu^\star\|^2 \leq V_{t+1} \leq (1 - \eta\tau/4)^t V_0. \tag{3.9.54}$$

This completes the proof. □

Proof of Theorem 3.3.4

We first prove the following result which follows from Theorem 1 in [23].

Lemma 3.9.8. [Theorem 1, [23]] Consider Algorithm 1. Suppose that the learning rates satisfy:

$$0 < \eta \leq \min\left\{\frac{1}{2\tau + 2\|Q\|_\infty}, \frac{1}{4\|Q\|_\infty}\right\}. \tag{3.9.55}$$

Let $z_t = (g_{\theta_t}, h_{\nu_t})$ and $\bar{z}_t = (g_{\bar{\theta}_t}, h_{\bar{\nu}_t})$. Then:

$$\max\left\{\mathrm{KL}(z^\star\|z_t), \frac{1}{2}\mathrm{KL}(z^\star\|\bar{z}_{t+1})\right\} \leq (1 - \eta\tau)^t \mathrm{KL}(z^\star\|z_0). \tag{3.9.56}$$

Proof. Let g_t and $\bar{g}_t$ denote g_{θ_t} and $g_{\bar{\theta}_t}$ respectively. Also, for any parameter θ, we denote Z_θ to be the normalizing constant $\sum_{a' \in \mathcal{A}} e^{\theta(a')}$ (Define Z_ν similarly). We have:

$$\begin{aligned}
\bar{g}_{t+1}(a) &= \frac{e^{\bar{\theta}_{t+1}(a)}}{\sum_{a' \in \mathcal{A}} e^{\bar{\theta}_{t+1}(a')}} \\
&\propto e^{\bar{\theta}_{t+1}(a)} \;=\; e^{(1-\eta\tau)\theta_t(a) - \eta[Qh_{\bar{\nu}_t}]_a} \;=\; e^{(1-\eta\tau)\theta_t(a) + \log Z_{\theta_t} - \log Z_{\theta_t} - \eta[Qh_{\bar{\nu}_t}]_a} \\
&= e^{(1-\eta\tau)\log e^{\theta_t(a)} + \log Z_{\theta_t} - \log Z_{\theta_t} - \eta[Qh_{\bar{\nu}_t}]_a} \;=\; e^{(1-\eta\tau)\log\left(\frac{e^{\theta_t(a)}}{Z_{\theta_t}}\right) + \log Z_{\theta_t} - \eta[Qh_{\bar{\nu}_t}]_a} \\
&\propto e^{(1-\eta\tau)\log\left(\frac{e^{\theta_t(a)}}{Z_{\theta_t}}\right) - \eta[Qh_{\bar{\nu}_t}]_a} \;=\; e^{(1-\eta\tau)\log g_t(a) - \eta[Qh_{\bar{\nu}_t}]_a} \;=\; e^{\log g_t(a)^{(1-\eta\tau)} - \eta[Qh_{\bar{\nu}_t}]_a} \\
&= g_t(a)^{(1-\eta\tau)} e^{-\eta[Qh_{\bar{\nu}_t}]_a}.
\end{aligned} \tag{3.9.57}$$

Therefore:

$$\bar{g}_{t+1}(a) \propto g_t(a)^{(1-\eta\tau)} e^{-\eta[Qh_{\bar{\nu}_t}]_a}. \tag{3.9.58}$$

Similarly, we have:

$$\begin{aligned}
g_{t+1}(a) &\propto g_t(a)^{(1-\eta\tau)} e^{-\eta[Qh_{\bar{\nu}_{t+1}}]_a}, \\
\bar{h}_{t+1}(a) &\propto h_t(a)^{(1-\eta\tau)} e^{\eta[Q^\top g_{\bar{\theta}_t}]_a}, \\
h_{t+1}(a) &\propto h_t(a)^{(1-\eta\tau)} e^{\eta[Q^\top g_{\bar{\theta}_{t+1}}]_a},
\end{aligned} \tag{3.9.59}$$

which is the same as the OMW updates for the regularized problem in [23]. Therefore, by Theorem 1 in [23], we have convergence of g_θ and h_ν to the solution of the regularized min-max problem. □

We begin by first providing the intuition of the proof, when the opponent is playing the NE strategy. We denote the NE strategies of the players as $g^\star$ and $h^\star$. Then, the optimistic NPG update has the following form:

$$\theta_{t+1} = (1 - \eta\tau)\theta_t - \eta Q h^\star. \tag{3.9.60}$$

From [86], we know that the NE satisfy:

$$g^\star(a) = \frac{e^{-[Qh^\star]_a/\tau}}{\sum_{a' \in \mathcal{A}} e^{-[Qh^\star]_{a'}/\tau}} = \frac{e^{-[Qh^\star]_a/\tau}}{K}, \tag{3.9.61}$$

where we define $K := \sum_{a' \in \mathcal{A}} e^{-[Qh^\star]_{a'}/\tau}$. Taking log on both sides, we have:

$$-\eta Qh^\star = \eta\tau \log g^\star + \eta\tau \log K. \tag{3.9.62}$$

Substituting this back into the θ update in (3.9.60), we have:

$$\begin{aligned}
\theta_{t+1} &= (1-\eta\tau)\theta_t + \eta\tau \log g^\star + \eta\tau \log K \\
&= \theta_t - \eta\tau\theta_t + \eta\tau \log g^\star + \eta\tau \log K - \eta\tau \log Z_{\theta_t} + \eta\tau \log Z_{\theta_t} \\
&= \theta_t + \eta\tau \log g^\star - \eta\tau \log\left(\frac{e^{\theta_t}}{Z_{\theta_t}}\right) + \eta\tau \log\left(\frac{K}{Z_{\theta_t}}\right) \\
&= \theta_t + \eta\tau \log g^\star - \eta\tau \log g_{\theta_t} + \eta\tau \log\left(\frac{K}{Z_{\theta_t}}\right) = \theta_t + \eta\tau \log\left(\frac{g^\star}{g_{\theta_t}}\right) + \eta\tau \log\left(\frac{K}{Z_{\theta_t}}\right) \\
&= \theta_t + \eta\tau \log\left(\frac{g^\star K}{g_{\theta_t} Z_{\theta_t}}\right).
\end{aligned} \tag{3.9.63}$$

We can further simplify this as:

$$\theta_{t+1} = \theta_t + \eta\tau \log\left(\frac{e^{-[Qh^\star]/\tau}}{e^{\theta_t}}\right), \tag{3.9.64}$$

which is nothing but:

$$\theta_{t+1} = \theta_t + \eta\tau \left(\frac{-Qh^\star}{\tau} - \theta_t\right). \tag{3.9.65}$$

Note that this is the Gradient Descent update on the strongly convex function $\frac{1}{2}\left\|\frac{-Qh^\star}{\tau} - \theta\right\|^2$ with stepsize $\eta\tau$. Note that this update leads to the following convergence guarantees:

$$\|\theta_{t+1} - \theta^\star\|^2 \leq (1-\eta\tau)^2 \|\theta_t - \theta^\star\|^2, \tag{3.9.66}$$

where $\theta^\star = \frac{-Qh^\star}{\tau}$.

The analysis above shows that if one of the players is already at the NE strategy, the parameters of the second player converges to the NE at a linear rate. However, the original OGDA update for θ is given by

$$\theta_{t+1} = (1-\eta\tau)\theta_t - \eta Q h^\star + \eta Q(h^\star - h_{\bar{\nu}_{t+1}}). \tag{3.9.67}$$

Since $h_{\bar{\nu}_{t+1}}$ converges to $h^\star$ at a linear rate (from Lemma 3.9.8), we expect the term

$$\varepsilon_t = \eta Q(h^\star - h_{\bar{\nu}_{t+1}}), \tag{3.9.68}$$

to be small, and goes to 0. This is formalized in what follows.

The OGDA update for θ can be re-written using ε_t as:

$$\theta_{t+1} = \theta_t + \eta\tau\left(\frac{-Qh^\star}{\tau} - \theta_t\right) + \varepsilon_t. \tag{3.9.69}$$

Once again, defining $\theta^\star = \frac{-Qh^\star}{\tau}$, we have:

$$\begin{aligned}
\|\theta_{t+1} - \theta^\star\|^2 &= \|\theta_t + \eta\tau\left(\theta^\star - \theta_t\right) + \varepsilon_t - \theta^\star\|^2 = \|(1-\eta\tau)(\theta_t - \theta^\star) + \varepsilon_t\|^2 \\
&= (1-\eta\tau)^2\|\theta_t - \theta^\star\|^2 + 2(1-\eta\tau)(\theta_t - \theta^\star)^\top \varepsilon_t + \|\varepsilon_t\|^2 \\
&\leq^{*1} (1-\eta\tau)^2\|\theta_t - \theta^\star\|^2 + \eta\tau\|\theta_t - \theta^\star\|^2 + \frac{1}{\eta\tau}(1-\eta\tau)^2\|\varepsilon_t\|^2 + \|\varepsilon_t\|^2 \\
&= (1-\eta\tau+\eta^2\tau^2)\|\theta_t - \theta^\star\|^2 + \left(1 + \frac{1}{\eta\tau}(1-\eta\tau)^2\right)\|\varepsilon_t\|^2,
\end{aligned} \tag{3.9.70}$$

where $*1$ follows from Young's inequality.

Next, we analyze the error term $\|\varepsilon_t\|$. We have:

$$\|\varepsilon_t\|^2 = \|\eta Q(h^\star - h_{\bar{\nu}_{t+1}})\|^2 \leq \eta^2\|Q\|_\infty^2\|(h^\star - h_{\bar{\nu}_{t+1}})\|_1^2 \leq^{*1} 2\eta^2\|Q\|_\infty^2(\mathrm{KL}(h^\star\|h_{\bar{\nu}_{t+1}})). \tag{3.9.71}$$

Here $*1$ follows from Pinsker's Inequality. Now, writing the same inequality for ν, we have:

$$\|\theta_{t+1} - \theta^\star\|^2 + \|\nu_{t+1} - \nu^\star\|^2 \leq (1-\eta\tau+\eta^2\tau^2)(\|\theta_t - \theta^\star\|^2 + \|\nu_t - \nu^\star\|^2)$$

$$
\begin{aligned}
&\qquad + \left(1 + \frac{1}{\eta\tau}(1-\eta\tau)^2\right) 2\eta^2 \|Q\|_\infty^2 \mathrm{KL}(z^\star \| \bar{z}_{t+1}) \\
&\le (1 - \eta\tau + \eta^2\tau^2)(\|\theta_t - \theta^\star\|^2 + \|\nu_t - \nu^\star\|^2) \\
&\qquad + \left(1 + \frac{1}{\eta\tau}(1-\eta\tau)^2\right) 4\eta^2 C \|Q\|_\infty^2 (1-\eta\tau)^t \mathrm{KL}(z^\star \| z_0).
\end{aligned}
$$

Define:

$$
C = \left(1 + \frac{1}{\eta\tau}(1-\eta\tau)^2\right) 4\eta^2 \|Q\|_\infty^2 \mathrm{KL}(z^\star \| z_0). \tag{3.9.72}
$$

This gives us (Using Lemma 3.9.8):

$$
\|\theta_{t+1} - \theta^\star\|^2 + \|\nu_{t+1} - \nu^\star\|^2 \le (1 - \eta\tau + \eta^2\tau^2)(\|\theta_t - \theta^\star\|^2 + \|\nu_t - \nu^\star\|^2) + C(1-\eta\tau)^t. \tag{3.9.73}
$$

For $\eta\tau < 1/2$ we have:

$$
\|\theta_{t+1} - \theta^\star\|^2 + \|\nu_{t+1} - \nu^\star\|^2 \le (1 - \eta\tau/2)(\|\theta_t - \theta^\star\|^2 + \|\nu_t - \nu^\star\|^2) + C(1-\eta\tau)^t. \tag{3.9.74}
$$

Consider the Lyapunov function:

$$
V_{t+1} = \|\theta_{t+1} - \theta^\star\|^2 + \|\nu_{t+1} - \nu^\star\|^2 + \frac{2C}{\eta\tau}(1-\eta\tau)^{t+1}. \tag{3.9.75}
$$

We have:

$$
\begin{aligned}
V_{t+1} &\le (1 - \eta\tau/2)(\|\theta_t - \theta^\star\|^2 + \|\nu_t - \nu^\star\|^2) + C(1-\eta\tau)^t + \frac{2C}{\eta\tau}(1-\eta\tau)(1-\eta\tau)^t \\
&= (1 - \eta\tau/2)(\|\theta_t - \theta^\star\|^2 + \|\nu_t - \nu^\star\|^2) + \frac{2C}{\eta\tau}(1-\eta\tau)^t(1-\eta\tau/2) \\
&= (1 - \eta\tau/2)\left(\|\theta_t - \theta^\star\|^2 + \|\nu_t - \nu^\star\|^2 + \frac{2C}{\eta\tau}(1-\eta\tau)^t\right) \\
&= (1 - \eta\tau/2) V_t.
\end{aligned} \tag{3.9.76}
$$

Algorithm 2 Optimistic NPG (Function Approximation)

```
**Initialize:** $\theta_0 = 0$ and $\nu_0 = 0$.
**for** $t = 1, 2, \cdots$ **do**
    $\bar{\theta}_{t+1} = (1 - \eta\tau)\theta_t - \eta[(M^\top)^{-1}|0]\widetilde{P}Qh_{\bar{\nu}_t}$
    $\bar{\nu}_{t+1} = (1 - \eta\tau)\nu_t + \eta[(M^\top)^{-1}|0]\widetilde{P}Q^\top g_{\bar{\theta}_t}$

    $\theta_{t+1} = (1 - \eta\tau)\theta_t - \eta[(M^\top)^{-1}|0]\widetilde{P}Qh_{\bar{\nu}_{t+1}}$
    $\nu_{t+1} = (1 - \eta\tau)\nu_t + \eta[(M^\top)^{-1}|0]\widetilde{P}Q^\top g_{\bar{\theta}_{t+1}}$
**end for**
```

This shows linear convergence of the parameter θ to $\theta^\star$ since:

$$\|\theta_{t+1} - \theta^\star\|^2 + \|\nu_{t+1} - \nu^\star\|^2 \leq V_{t+1} \leq (1 - \eta\tau/2)^t V_0. \tag{3.9.77}$$

This completes the proof. □

Note that proof of Corollary 3.3.5 follows from Remark 4 and the preceding discussion in [23].

3.9.3 Missing Details and Proofs in §3.4

Remark 3.9.9. We note that all results present in this section also follow for the case where the cardinality of the action spaces for both players are unequal. However, we stick to the case where the number of actions is the same for both players for ease of exposition.

Proof of Lemma 3.4.4

The first part of the lemma describes the set of distributions in the m-dimensional simplex covered by this parametrization. Since the set of distributions covered by the log-linear parametrization would be the same for all invertible M[7], for simplicity, we study the case where $M = \mathrm{I}$. Note that this would imply:

$$g_\theta(a) \propto e^{\theta(a)} \qquad \forall a \in \{1, 2, \cdots, d\},$$

[7] To see this, consider $\widetilde{\theta} = M\theta$. Since M is invertible, there is a 1-1 correspondence between $\widetilde{\theta}$ and θ. The distribution parametrized by θ under the function approximation matrix M, is the same as the distribution parametrized by $\widetilde{\theta}$ under the function approximation matrix I. Therefore it is enough to consider the special case of $M = \mathrm{I}$.

$$g_\theta(a) \propto 1 \qquad \forall a \in \{d+1, d+2, \cdots, n\}. \tag{3.9.78}$$

Similarly for h_ν. Therefore, according to this parametrization the first d elements can be chosen freely, and the rest $n-d$ parameters have to be equal. In other words, this parametrization covers the following set of distributions:

$$\widetilde{\Delta} = \{\mu : \mu \in \Delta, \mu_{d+1} = \mu_{d+2} = \cdots = \mu_n\}, \tag{3.9.79}$$

which is a closed convex subset of the n-dimensional simplex. To see that any element of $\widetilde{\Delta}$ can be represented by the log-linear parametrization, we can take the parameters $\theta(a) = \log \mu(a)$ for $a = 1, 2, \cdots, d$[8]. This would be a valid parametrization under the log-linear function approximation setting, and therefore all elements of $\widetilde{\Delta}$ can be represented in this manner. Therefore, since the two sets are equivalent, we can rewrite the problem in terms of the policy vectors lying in the constraint set $\widetilde{\Delta}$ which completes the proof of the lemma.

Proof of Theorem 3.4.5

For the regularized game (3.4.6), let player 2 play the NE strategy $h_\nu^\star$. Then player 1's optimization problem is given by:

$$\min_{g_\theta \in \widetilde{\Delta}} \; g_\theta^\top Q h_\nu^\star - \tau \mathcal{H}(g_\theta). \tag{3.9.80}$$

Define the following Lagrange multipliers (and associated constraints):

$$\begin{aligned}
\lambda \;&:\; \sum_{a=1}^{n} g_\theta(a) = 1, \\
\beta_1 \;&:\; g_\theta(d+1) = g_\theta(d+2), \\
\beta_2 \;&:\; g_\theta(d+2) = g_\theta(d+3), \\
&\cdots \\
\beta_{n-d-1} \;&:\; g_\theta(n-1) = g_\theta(n).
\end{aligned} \tag{3.9.81}$$

[8] Note that if $\mu(a) = 0$, the corresponding paramter would be $-\infty$.

Therefore, for the optimal Lagrange multipliers, taking the first-order optimality conditions (with respect to the variable g_θ), we have:

$$\begin{aligned}
&[Qh^\star_\nu]_a + \tau(\log g_\theta(a) + 1) + \lambda = 0 && \forall a \in \{1, 2, \cdots, d\}, \\
&[Qh^\star_\nu]_a + \tau(\log g_\theta(a) + 1) + \lambda + \beta_{a-d} - \beta_{a-d-1} = 0 && \forall a \in \{d+1, d+2, \cdots, n\},
\end{aligned} \tag{3.9.82}$$

where β_0 and β_{n-d} are defined to be equal to 0. This gives us:

$$g^\star_\theta(a) \propto e^{-[Qh^\star_\nu]_a/\tau} \qquad \forall a \in \{1, 2, \cdots, d\}. \tag{3.9.83}$$

For actions with indices $a > d$, the equality constraints give us:

$$[Qh^\star_\nu]_a + \beta_{a-d} - \beta_{a-d-1} = C \qquad \forall a \in \{d+1, d+2, \cdots, n\}, \tag{3.9.84}$$

for some constant C. On solving these equations for C, we have:

$$C = \frac{1}{n-d} \sum_{a=d+1}^{n} [Qh^\star_\nu]_a, \tag{3.9.85}$$

which gives us

$$g^\star_\theta(a) \propto e^{-C/\tau} \qquad \forall a \in \{d+1, \cdots, n\}. \tag{3.9.86}$$

Now, consider the symmetric matrix $\Psi \in \mathbb{R}^{n \times n}$ defined as:

$$\Psi = \begin{pmatrix} I_d & & \mathbf{0} & \\ & \frac{1}{n-d} & \cdots & \frac{1}{n-d} \\ \mathbf{0} & \cdots & \cdots & \cdots \\ & \frac{1}{n-d} & \cdots & \frac{1}{n-d} \end{pmatrix}. \tag{3.9.87}$$

Note that

$$[\Psi^\top Qh^\star_\nu]_a = [Qh^\star_\nu]_a \quad \forall a \in \{1, 2, \cdots, d\},$$

$$[\Psi^\top Q h_\nu^\star]_a = C \qquad \forall a \in \{d+1, d+2, \cdots, n\}. \tag{3.9.88}$$

Therefore, we can succinctly write the optimal distribution as:

$$g_\theta^\star(a) \propto e^{-[\Psi^\top Q h_\nu^\star]_a/\tau} \qquad \forall a \in \{1, 2, \cdots, n\}. \tag{3.9.89}$$

Also, we have $\Psi\mu = \mu, \ \forall \mu \in \widetilde{\Delta}$.

Now, since $h_\nu^\star$ is the optimal distribution for Player 2, we must have $h_\nu^\star \in \widetilde{\Delta}$, which implies:

$$\Psi h_\nu^\star = h_\nu^\star. \tag{3.9.90}$$

Doing a similar calculation for $g_\nu^\star$, we have that the NE satisfy:

$$g_\theta^\star(a) = \frac{e^{-[\Psi^\top Q \Psi h_\nu^\star]_a/\tau}}{\sum_{a'} e^{-[\Psi^\top Q \Psi h_\nu^\star]_{a'}/\tau}}, \qquad h_\nu^\star(a) = \frac{e^{[\Psi^\top Q^\top \Psi g_\theta^\star]_a/\tau}}{\sum_{a'} e^{[\Psi^\top Q^\top \Psi g_\theta^\star]_{a'}/\tau}}. \tag{3.9.91}$$

Now, from this characterization of the NE, we see that these solutions are also the same as those of the following problem:

$$\min_{g \in \Delta} \max_{h \in \Delta} \ g^\top \Psi^\top Q \Psi h - \tau \mathcal{H}(g) + \tau \mathcal{H}(h). \tag{3.9.92}$$

Note that in the analysis above we have found a solution in the (relative) interior of the constraint set. For example, if for some action a, we have $f_\theta(a) = 0$, then the term $\log f_\theta(a)$ is not defined and the Lagrangian would be different. However, since this is a strongly convex strongly concave minimax problem over a convex compact set, there is a unique solution (see [41]). As we have already found a solution in the interior of the constraint set, by the argument above we have that this is the unique solution. This allows us to solve the first-order KKT optimality conditions (see [17]) to find the solution. Also, since all terms satisfy $f_\theta(a) > 0$ (similarly for $g_\nu(a)$) we do not need to explicitly write down the Lagrange multipliers for the non-negativity constraint. This completes the proof. □

Proof of Proposition 3.4.6

The algorithm is similar to the one proposed for the tabular case. However, since the exponents are elements of $\Phi^\top\theta$ instead of just θ as was in the case of tabular softmax, we have the updates modified as well (we write the updates here for the combined update instead of the two step update for ease of presentation):

$$\begin{aligned}
\Phi^\top\theta_{t+1} &= (1-\eta\tau)\Phi^\top\theta_t - (2-\eta\tau)\eta\Psi^\top\widetilde{P}Q\Psi h_{\nu_t} + (1-\eta\tau)\eta\Psi^\top\widetilde{P}Q\Psi h_{\nu_{t-1}}, \\
\Phi^\top\nu_{t+1} &= (1-\eta\tau)\Phi^\top\nu_t + (2-\eta\tau)\eta\Psi^\top\widetilde{P}Q^\top\Psi g_{\theta_t} - (1-\eta\tau)\eta\Psi^\top\widetilde{P}Q^\top\Psi g_{\theta_{t-1}},
\end{aligned} \tag{3.9.93}$$

where Φ is a full-rank feature matrix. Note that the additional matrix $\widetilde{P}$ is to ensure that the probability vector at each step of the algorithm satisfies the function approximation constraint[9]. Since Φ is full-rank, we can explicitly write this update for θ and ν as follows:

$$\begin{aligned}
\theta_{t+1} &= [\Phi\Phi^\top]^{-1}\Phi\left((1-\eta\tau)\Phi^\top\theta_t - (2-\eta\tau)\eta\Psi^\top\widetilde{P}Q\Psi h_{\nu_t} + (1-\eta\tau)\eta\Psi^\top\widetilde{P}Q\Psi h_{\nu_{t-1}}\right), \\
\nu_{t+1} &= [\Phi\Phi^\top]^{-1}\Phi\left((1-\eta\tau)\Phi^\top\nu_t + (2-\eta\tau)\eta\Psi^\top\widetilde{P}Q^\top\Psi g_{\theta_t} - (1-\eta\tau)\eta\Psi^\top\widetilde{P}Q^\top\Psi g_{\theta_{t-1}}\right),
\end{aligned}$$

which can be simplified to:

$$\begin{aligned}
\theta_{t+1} &= (1-\eta\tau)\theta_t - (2-\eta\tau)\eta[\Phi\Phi^\top]^{-1}\Phi\Psi^\top\widetilde{P}Q\Psi h_{\nu_t} + (1-\eta\tau)\eta[\Phi\Phi^\top]^{-1}\Phi\Psi^\top\widetilde{P}Q\Psi h_{\nu_{t-1}}, \\
\nu_{t+1} &= (1-\eta\tau)\nu_t + (2-\eta\tau)\eta[\Phi\Phi^\top]^{-1}\Phi\Psi^\top\widetilde{P}Q^\top\Psi g_{\theta_t} - (1-\eta\tau)\eta[\Phi\Phi^\top]^{-1}\Phi\Psi^\top\widetilde{P}Q^\top\Psi g_{\theta_{t-1}}.
\end{aligned} \tag{3.9.94}$$

Note that for $\Phi = [M \mid 0]$, we have

$$[\Phi\Phi^\top]^{-1}\Phi = ([M|0][M|0]^\top)^{-1}[M|0] = (MM^\top)^{-1}[M|0] = [(M^\top)^{-1}|0]. \tag{3.9.95}$$

From the previous discussion, since all the terms g_{θ_t} and h_{ν_t} lie in the set $\widetilde{\Delta}$, we have $\Psi g_{\theta_t} = g_{\theta_t}$ and $\Psi h_{\nu_t} = h_{\nu_t}$. Also, we have:

$$[\Phi\Phi^\top]^{-1}\Phi\Psi^\top = [(M^\top)^{-1}|0]\Psi^\top = [(M^\top)^{-1}|0], \tag{3.9.96}$$

[9] This is based on the fact that for a probability vector $x_i \propto e^{y_i}$, we will also have $x_i \propto e^{y_i - k}$ for any constant k independent of the index i.

from the structure of Ψ. Therefore, the update can be written as:

$$\theta_{t+1} = (1-\eta\tau)\theta_t - (2-\eta\tau)\eta[(M^\top)^{-1}|0]\widetilde{P}Qh_{\nu_t} + (1-\eta\tau)\eta[(M^\top)^{-1}|0]\widetilde{P}Qh_{\nu_{t-1}},$$
$$\nu_{t+1} = (1-\eta\tau)\nu_t + (2-\eta\tau)\eta[(M^\top)^{-1}|0]\widetilde{P}Q^\top g_{\theta_t} - (1-\eta\tau)\eta[(M^\top)^{-1}|0]\widetilde{P}Q^\top g_{\theta_{t-1}}, \quad (3.9.97)$$

which completes the proof. □

Simulation to show convergence of ONPG in the function approximation setting

In this section, we show the performance of the ONPG algorithm under Function Approximation (Algorithm 2). The policies have a log linear parametrization, with the feature matrix $\Phi = [I \mid 0] \in \mathbb{R}^{10\times 100}$, and the cost matrix Q is chosen to be a random matrix of dimension 100×100.

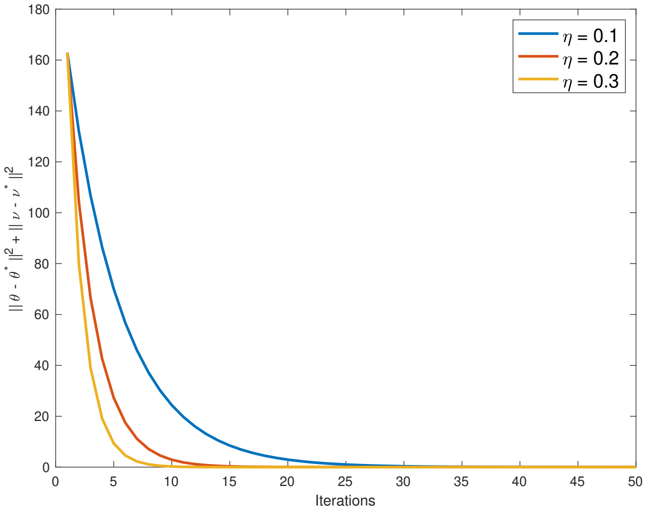


Figure 3-3: Behavior of Algorithm 2 in the matrix game under the log-linear function approximation setting.

3.9.4 Missing Details and Proofs in §3.5

Remark 3.9.10. We note that all results presented in this section also follow for the case where the number of possible actions for each player can be different. However, we stick to the case where the number of action is the same for both players for ease of exposition. Note that the actual action spaces need not be identical, but only their cardinalities.

Algorithm 3 Optimistic NPG for monotone games

```
**Initialize:** $\theta_i^0 = 0$ for all players $i$.
**for** $t = 1, 2, \cdots$ **do**
    $\bar{\theta}_i^{t+1} = (1-\eta\tau)\theta_i^t - \eta\nabla_{g_{\theta_i}} f_i(g_{\bar{\theta}_i^t}, g_{\bar{\theta}_{-i}^t})\ \ \forall i \in [N]$.
    $\theta_i^{t+1} = (1-\eta\tau)\theta_i^t - \eta\nabla_{g_{\theta_i}} f_i(g_{\bar{\theta}_i^{t+1}}, g_{\bar{\theta}_{-i}^{t+1}})\ \ \forall i \in [N]$.
**end for**
```

Definition 3.9.11 (In-class Nash equilibrium for a monotone game)**.** The policy parameter $\theta^\star = [\theta_1^\star, \theta_2^\star, \cdots, \theta_N^\star]$ is an NE under function approximation, i.e., (in-class NE) of the monotone game, if it satisfies that for all $i \in [N]$,

$$f_i(g_{\theta_i^\star}, g_{\theta_{-i}^\star}) \le f_i(g_{\theta_i}, g_{\theta_{-i}^\star}), \qquad \forall \theta_i \in \mathbb{R}^d. \tag{3.9.98}$$

Note that if we are in the tabular setting, we will have $d = n$.

Definition 3.9.12 (ϵ-in-class Nash equilibrium for a monotone game)**.** The policy parameter $(\widetilde{\theta}_1, \cdots, \widetilde{\theta}_N)$ is an ϵ-*Nash equilibrium* under function approximation (or *in-class* ϵ-NE) of the monotone game if it satisfies that for all $i \in [N]$,

$$f_i(g_{\widetilde{\theta}_i}, g_{\widetilde{\theta}_{-i}}) - \epsilon \le f_i(g_{\theta_i}, g_{\widetilde{\theta}_{-i}}), \qquad \forall \theta_i \in \mathbb{R}^d. \tag{3.9.99}$$

Note that if we are in the tabular setting, we will have $d = n$.

We will also use the notation $f_i^\tau(g_i, g_{-i}) = f_i(g_i, g_{-i}) - \tau\mathcal{H}(g_i)$.

Proof of Lemma 3.5.2

We can follow the analysis for a two player game from [86] to write down the solution form for the N-player monotone setting. We let g_i and g_{-i} denote g_{θ_i} and $g_{\theta_{-i}}$ respectively. First, note that the solutions in the policy space exists for the unregularized game, since we are solving a monotone VI over a convex compact set, and this solution is unique if we regularize the game, since in this case we are solving a stringly monotone VI over a convex compact set. See [41].

Consider player $i's$ optimization problem when other players play the equilibrium strategies:

$$\min_{g_i \in \Delta} f_i(g_i, g_{-i}^\star) - \tau \mathcal{H}(g_i). \tag{3.9.100}$$

Since we are in the monotone setting with a strongly convex regularizer, the first order Karush–Kuhn–Tucker (KKT) conditions are both necessary and sufficient. The first order KKT conditions are:

$$[\nabla_{g_i} f_i(g_i^\star, g_{-i}^\star)]_a + \tau(\log g_i^\star + 1) + \lambda = 0,$$

where λ is the Lagrange multiplier corresponding to the simplex constraint. This implies:

$$g_i^\star(a) \propto e^{\frac{-[\nabla_{g_i} f_i(g_i^\star, g_{-i}^\star)]_a}{\tau}}, \tag{3.9.101}$$

which shows the NE in the policy space. Now, to complete the proof of the lemma, we need to find parameters $\theta^\star$ which leads to this distribution. This can be easily seen by setting $\theta_i^\star = \frac{-[\nabla_{g_i} f_i(g_i^\star, g_{-i}^\star)]}{\tau}$, thereby completing the proof of the lemma. □

(Optimistic) NPG for monotone games

As noted in §3.9.1, we have that the Fisher Information matrix $F_\theta(\theta) = \nabla_\theta g_\theta$. Therefore, the NPG update for player i can be simplified as:

$$\begin{aligned}\theta_i^{t+1} &= \theta_i^t - \eta \cdot F_\theta^\dagger(\theta_i^t) \cdot \frac{\partial f_i^\tau(g_{\theta_i^t}, g_{\theta_{-i}^t})}{\partial \theta} = \theta_t - \eta \frac{\partial f_i^\tau(g_{\theta_i^t}, g_{\theta_{-i}^t})}{\partial g_{\theta_i}} \\ &= \theta_t - \eta \left(\nabla_{g_{\theta_i}} f_i(g_{\theta_i^t}, g_{\theta_{-i}^t}) + \tau(\mathbb{1} + \log g_{\theta_i^t}) \right). \end{aligned} \tag{3.9.102}$$

This can be simplified as:

$$\theta_i^{t+1}(a) = (1 - \eta\tau)\theta_i^t(a) - \eta[\nabla_{g_{\theta_i}} f_i(g_{\theta_i^t}, g_{\theta_{-i}^t})]_a + \eta\tau(\log Z_{\theta_i^t} - 1), \tag{3.9.103}$$

where $Z_{\theta_i^t} = \sum_{a' \in \mathcal{A}} e^{\theta_i^t(a')}$. Note that this update will have the same pitfall of parameter divergence as the NPG update for the matrix game, since a matrix game is a special case of the monotone game. Therefore, we propose the following modified version of NPG, as done for the matrix game:

$$\theta_i^{t+1}(a) = (1-\eta\tau)\theta_i^t(a) - \eta[\nabla_{g_{\theta_i}} f_i(g_{\theta_i^t}, g_{\theta_{-i}^t})]_a. \tag{3.9.104}$$

This leads to the modified NPG dynamics for the monotone game. Now, similar to the matrix game, we analyze the optimistic version of this algorithm with updates:

$$\bar{\theta}_i^{t+1} = (1-\eta\tau)\theta_i^t - \eta\nabla_{g_{\theta_i}} f_i(g_{\bar{\theta}_i^t}, g_{\bar{\theta}_{-i}^t}), \qquad \theta_i^{t+1} = (1-\eta\tau)\theta_i^t - \eta\nabla_{g_{\theta_i}} f_i(g_{\bar{\theta}_i^{t+1}}, g_{\bar{\theta}_{-i}^{t+1}}),$$

in §3.5.

Proof of Theorem 3.5.3

We prove the following Lemma first:

Lemma 3.9.13. For any $z = (g_{\theta_i}, g_{\theta_{-i}}) \in \Delta^N$, consider an update of the form:

$$\theta_i^{t+1} = (1-\eta\tau)\theta_i^t - \eta\nabla_{g_{\theta_i}} f_i(g_{\theta_i}, g_{\theta_{-i}}) \quad \forall i \in [N]. \tag{3.9.105}$$

We have:

$$\langle \log z_{t+1} - (1-\eta\tau)\log z_t - \eta\tau \log z^\star, z - z^\star \rangle \le 0, \tag{3.9.106}$$

where $z^\star = (g_i^\star)_{i=1}^N$.

Proof. In the proof below, we define $g_i := g_{\theta_i}$ and $g_i^t := g_{\theta_i^t}$. From the update sequence in Equation (3.9.105), we have

$$\log g_i^{t+1} = (1-\eta\tau)\log g_i^t - \eta\nabla_{g_i} f_i(g_i, g_{-i}) + c \cdot \mathbb{1}, \tag{3.9.107}$$

where c is the normalization constant. This implies:

$$\langle \log g_i^{t+1} - (1-\eta\tau)\log g_i^t, g_i - g_i^\star \rangle = \langle -\eta\nabla_{g_i} f_i(g_i, g_{-i}) + c \cdot \mathbb{1}, g_i - g_i^\star \rangle$$

$$= \langle -\eta \nabla_{g_i} f_i(g_i, g_{-i}), g_i - g_i^\star \rangle. \tag{3.9.108}$$

Note that $\langle c \cdot \mathbb{1}, g_i - g_i^\star \rangle = 0$, since $g_i, g_i^\star \in \Delta$. Since this is true for all players i, we have:

$$\langle \log z_{t+1} - (1-\eta\tau) \log z_t, z - z^\star \rangle = -\eta \sum_i \langle \nabla_{g_i} f_i(g_i, g_{-i}), g_i - g_i^\star \rangle = -\eta \langle F(z), z - z^\star \rangle. \tag{3.9.109}$$

Now, from the properties of the NE, we have:

$$\eta\tau \log g_i^\star = -\eta \nabla_{g_i} f_i(g_i^\star, g_{-i}^\star) + c \cdot \mathbb{1}, \tag{3.9.110}$$

which gives us:

$$\langle \eta\tau \log g_i^\star, g_i - g_i^\star \rangle = \langle -\eta \nabla_{g_i} f_i(g_i^\star, g_{-i}^\star), g_i - g_i^\star \rangle. \tag{3.9.111}$$

Since this is true for all players i, we have

$$\langle \eta\tau \log z^\star, z - z^\star \rangle = -\eta \langle F(z^\star), z - z^\star \rangle. \tag{3.9.112}$$

From Equations (3.9.109) and (3.9.112), we have:

$$\langle \log z_{t+1} - (1-\eta\tau) \log z_t - \eta\tau \log z^\star, z - z^\star \rangle = -\eta \langle F(z) - F(z^\star), z - z^\star \rangle \leq 0, \tag{3.9.113}$$

where the last step uses the monotonicity assumption of F. This completes the proof of the lemma. □

Lemma 3.9.14. For updates of the form:

$$\theta_i^{t+1} = (1-\eta\tau)\theta_i^t - \eta \nabla_{g_{\theta_i}} f_i(g_{\bar\theta_i^{t+1}}, g_{\bar\theta_{-i}^{t+1}}), \quad \forall i \in [N], \tag{3.9.114}$$

we have

$$\begin{aligned}(1-\eta\tau)\mathrm{KL}(z^\star \| z_t) \geq{}& (1-\eta\tau)\mathrm{KL}(\bar z_{t+1} \| z_t) + \eta\tau \mathrm{KL}(\bar z_{t+1} \| z^\star) + \mathrm{KL}(z_{t+1} \| \bar z_{t+1}) \\ &- \langle \log \bar z_{t+1} - \log z_{t+1}, \bar z_{t+1} - z_{t+1} \rangle + \mathrm{KL}(z^\star \| z_{t+1}),\end{aligned} \tag{3.9.115}$$

and

$$\mathrm{KL}(z^\star \| \bar{z}_{t+1}) = \mathrm{KL}(z^\star \| z_{t+1}) - \mathrm{KL}(\bar{z}_{t+1} \| z_{t+1}) - \langle z^\star - \bar{z}_{t+1}, \log \bar{z}_{t+1} - \log z_{t+1} \rangle. \tag{3.9.116}$$

Proof. From the definition of KL divergence we have:

$$-\langle \log z_{t+1} - (1-\eta\tau)\log z_t - \eta\tau \log z^\star, z^\star \rangle = -(1-\eta\tau)\mathrm{KL}(z^\star \| z_t) + \mathrm{KL}(z^\star \| z_{t+1}). \tag{3.9.117}$$

Next, note that:

$$\begin{aligned}
&\langle \log z_{t+1} - (1-\eta\tau)\log z_t - \eta\tau \log z^\star, \bar{z}_{t+1} \rangle \\
&= \langle \log \bar{z}_{t+1} - (1-\eta\tau)\log z_t - \eta\tau \log z^\star, \bar{z}_{t+1} \rangle + \langle \log \bar{z}_{t+1} - \log z_{t+1}, z_{t+1} \rangle \\
&\qquad\qquad - \langle \log \bar{z}_{t+1} - \log z_{t+1}, \bar{z}_{t+1} - z_{t+1} \rangle \\
&= (1-\eta\tau)\mathrm{KL}(\bar{z}_{t+1} \| z_t) + \eta\tau \mathrm{KL}(\bar{z}_{t+1} \| z^\star) + \mathrm{KL}(z_{t+1} \| \bar{z}_{t+1}) \\
&\qquad\qquad - \langle \log \bar{z}_{t+1} - \log z_{t+1}, \bar{z}_{t+1} - z_{t+1} \rangle.
\end{aligned} \tag{3.9.118}$$

Now, substituting $z = \bar{z}_{t+1}$ in Lemma 3.9.13 we have:

$$\langle \log z_{t+1} - (1-\eta\tau)\log z_t - \eta\tau \log z^\star, \bar{z}_{t+1} - z^\star \rangle \leq 0. \tag{3.9.119}$$

Substituting Equations (3.9.117) and (3.9.118) in (3.9.119), we get the Inequality (3.9.115). Inequality (3.9.116) follows from the properties of KL divergence. □

From the ONPG updates in Algorithm (3), we have:

$$\log \bar{g}_i^{t+1} - \log g_i^{t+1} = -\eta \left(\nabla_{g_i} f_i(\bar{g}_i^t, \bar{g}_{-i}^t) - \nabla_{g_i} f_i(\bar{g}_i^{t+1}, \bar{g}_{-i}^{t+1}) \right) + c \cdot \mathbb{1}. \tag{3.9.120}$$

This implies:

$$\begin{aligned}
&\langle \log \bar{g}_i^{t+1} - \log g_i^{t+1}, \bar{g}_i^{t+1} - g_i^{t+1} \rangle \\
&= -\eta \langle \nabla_{g_i} f_i(\bar{g}_i^t, \bar{g}_{-i}^t) - \nabla_{g_i} f_i(g_i^t, g_{-i}^t) + \nabla_{g_i} f_i(g_i^t, g_{-i}^t) - \nabla_{g_i} f_i(\bar{g}_i^{t+1}, \bar{g}_{-i}^{t+1}), \bar{g}_i^{t+1} - g_i^{t+1} \rangle
\end{aligned}$$

$$
\begin{aligned}
&\leq^{*1} \eta \left(\|\nabla_{g_i} f_i(\bar{g}_i^t, \bar{g}_{-i}^t) - \nabla_{g_i} f_i(g_i^t, g_{-i}^t)\| + \|\nabla_{g_i} f_i(g_i^t, g_{-i}^t) - \nabla_{g_i} f_i(\bar{g}_i^{t+1}, \bar{g}_{-i}^{t+1})\| \right) \|\bar{g}_i^{t+1} - g_i^{t+1}\| \\
&\leq^{*2} \eta L \left(\|\bar{g}_i^t - g_i^t\| + \|\bar{g}_{-i}^t - g_{-i}^t\| + \|\bar{g}_i^{t+1} - g_i^t\| + \|\bar{g}_{-i}^{t+1} - g_{-i}^t\| \right) \|\bar{g}_i^{t+1} - g_i^{t+1}\| \\
&\leq^{*3} \frac{1}{2}\eta L \left(\|\bar{g}_i^t - g_i^t\|^2 + \|\bar{g}_{-i}^t - g_{-i}^t\|^2 + \|\bar{g}_i^{t+1} - g_i^t\|^2 + \|\bar{g}_{-i}^{t+1} - g_{-i}^t\|^2 + 4\|\bar{g}_i^{t+1} - g_i^{t+1}\|^2 \right),
\end{aligned}
\tag{3.9.121}
$$

where $(*1)$ follows from the fact that $a^\top b \leq \|a\|\|b\|$, $(*2)$ follows from Assumption 3.5.1, and $(*3)$ follows from $x \cdot y \leq \frac{1}{2}(x^2 + y^2)$. Since this is true for all players i, we have:

$$
\begin{aligned}
&\langle \log \bar{z}_{t+1} - \log z_{t+1}, \bar{z}_{t+1} - z_{t+1} \rangle \\
&\leq \frac{1}{2}\eta L \sum_{i=1}^{N} \left(\|\bar{g}_i^t - g_i^t\|^2 + \|\bar{g}_{-i}^t - g_{-i}^t\|^2 + \|\bar{g}_i^{t+1} - g_i^t\|^2 + \|\bar{g}_{-i}^{t+1} - g_{-i}^t\|^2 + 4\|\bar{g}_i^{t+1} - g_i^{t+1}\|^2 \right) \\
&\leq \frac{1}{2}\eta L \left(N\|\bar{z}_t - z_t\|^2 + N\|\bar{z}_{t+1} - z_t\|^2 + 4\|\bar{z}_{t+1} - z_{t+1}\|^2 \right) \\
&\leq^{*1} \eta L \left(N\mathrm{KL}(z_t\|\bar{z}_t) + N\mathrm{KL}(\bar{z}_{t+1}\|z_t) + 4\mathrm{KL}(z_{t+1}\|\bar{z}_{t+1}) \right),
\end{aligned}
\tag{3.9.122}
$$

where $(*1)$ follows from Pinsker's Inequality and the fact that the l_1 norm is an upper bound for the l_2 norm. Substituting this in Equation (3.9.115), we have

$$
\begin{aligned}
\mathrm{KL}(z^\star\|z_{t+1}) \leq (1 - \eta\tau)\mathrm{KL}(z^\star\|z_t) - (1 - \eta\tau - N\eta L)\mathrm{KL}(\bar{z}_{t+1}\|z_t) - \eta\tau\mathrm{KL}(\bar{z}_{t+1}\|z^\star) \\
- (1 - 4\eta L)\mathrm{KL}(z_{t+1}\|\bar{z}_{t+1}) + N\eta L\mathrm{KL}(z_t\|\bar{z}_t).
\end{aligned}
\tag{3.9.123}
$$

For $\eta < \frac{1}{2(N+4)L+2\tau}$, we have: $N\eta L \leq (1 - \eta\tau)(1 - 4\eta L)$. This gives us:

$$
\begin{aligned}
\mathrm{KL}(z^\star\|z_{t+1}) + (1 - 4\eta L)\mathrm{KL}(z_{t+1}\|\bar{z}_{t+1}) &\leq (1 - \eta\tau)\mathrm{KL}(z^\star\|z_t) + N\eta L\mathrm{KL}(z_t\|\bar{z}_t) \\
&\leq (1 - \eta\tau)\left(\mathrm{KL}(z^\star\|z_t) + (1 - 4\eta L)\mathrm{KL}(z_t\|\bar{z}_t)\right).
\end{aligned}
$$

Define:

$$
V_t = \mathrm{KL}(z^\star\|z_t) + (1 - 4\eta L)\mathrm{KL}(z_t\|\bar{z}_t). \tag{3.9.124}
$$

Then we have:

$$V_{t+1} \leq (1-\eta\tau)V_t, \tag{3.9.125}$$

and therefore:

$$\mathrm{KL}(z^\star \| z_{t+1}) \leq V_{t+1} \leq (1-\eta\tau)^t V_0 = (1-\eta\tau)^t \mathrm{KL}(z^\star \| z_0), \tag{3.9.126}$$

which shows convergence of $\mathrm{KL}(z^\star \| z_{t+1})$. Next we show convergence of $\mathrm{KL}(z^\star \| \bar{z}_{t+1})$.

Similar to derivation of Equation (3.9.122), we have

$$-\langle z^\star - \bar{z}_{t+1}, \log \bar{z}_{t+1} - \log z_{t+1} \rangle \leq \eta L \left(N\mathrm{KL}(z_t \| \bar{z}_t) + N\mathrm{KL}(\bar{z}_{t+1} \| z_t) + 4\mathrm{KL}(z^\star \| \bar{z}_{t+1}) \right). \tag{3.9.127}$$

Substituting this in Equation (3.9.116), we have:

$$(1-4\eta L)\mathrm{KL}(z^\star \| \bar{z}_{t+1}) \leq \mathrm{KL}(z^\star \| z_{t+1}) + \eta L \left(N\mathrm{KL}(z_t \| \bar{z}_t) + N\mathrm{KL}(\bar{z}_{t+1} \| z_t) \right). \tag{3.9.128}$$

Now, using Equation (3.9.123), we have:

$$\begin{aligned}
(1-&4\eta L)\mathrm{KL}(z^\star \| \bar{z}_{t+1}) \\
&\leq (1-\eta\tau)\mathrm{KL}(z^\star \| z_t) - (1-\eta\tau-2N\eta L)\mathrm{KL}(\bar{z}_{t+1} \| z_t) - \eta\tau\mathrm{KL}(\bar{z}_{t+1} \| z^\star) \\
&\quad - (1-4\eta L)\mathrm{KL}(z_{t+1} \| \bar{z}_{t+1}) + 2N\eta L\mathrm{KL}(z_t \| \bar{z}_t) \\
&\leq (1-\eta\tau)\mathrm{KL}(z^\star \| z_t) + 2N\eta L\mathrm{KL}(z_t \| \bar{z}_t) \\
&\leq \mathrm{KL}(z^\star \| z_t) + (1-4\eta L)\mathrm{KL}(z_t \| \bar{z}_t) := V_t,
\end{aligned} \tag{3.9.129}$$

which gives us:

$$\mathrm{KL}(z^\star \| \bar{z}_{t+1}) \leq 2V_t \leq 2(1-\eta\tau)^t V_0 = 2(1-\eta\tau)^t \mathrm{KL}(z^\star \| z_0). \tag{3.9.130}$$

This completes the proof of the first part of the theorem.

Next, we prove parameter convergence. We have from Equation (3.9.70):

$$\|\theta_i^{t+1} - \theta_i^\star\|^2 = (1 - \eta\tau + \eta^2\tau^2)\|\theta_i^t - \theta_i^\star\|^2 + \left(1 + \frac{1}{\eta\tau}(1 - \eta\tau)^2\right)\|\varepsilon_t\|^2, \tag{3.9.131}$$

where $\theta_i^\star = \frac{-\nabla_{g_i} f_i(g_i^\star, g_{-i}^\star)}{\tau}$ (Note that the term ε_t however is different here from definition of ε_t in Equation (3.9.70)). Now, for ONPG, we have:

$$\begin{aligned}\|\varepsilon_t\|^2 &= \eta^2\|\nabla_{g_i} f_i(g_i^\star, g_{-i}^\star) - \nabla_{g_i} f_i(\bar{g}_i^{t+1}, \bar{g}_{-i}^{t+1})\|^2 \\ &\leq \eta^2 L^2\|\bar{z}_{t+1} - z^\star\|^2 \leq 2\eta^2 L^2 \mathrm{KL}(z^\star\|z_{t+1}).\end{aligned} \tag{3.9.132}$$

This gives us:

$$\|\theta_i^{t+1} - \theta_i^\star\|^2 \leq (1 - \eta\tau + \eta^2\tau^2)\|\theta_i^t - \theta_i^\star\|^2 + C(1 - \eta\tau)^t, \tag{3.9.133}$$

where

$$C = 4\eta^2 L^2\left(1 + \frac{1}{\eta\tau}(1 - \eta\tau)^2\right)\mathrm{KL}(z^\star\|z_0), \tag{3.9.134}$$

from the first part of the Theorem proved above. For $\eta\tau < 1/2$, this reduces to:

$$\|\theta_i^{t+1} - \theta_i^\star\|^2 \leq (1 - \eta\tau/2)\|\theta_i^t - \theta_i^\star\|^2 + C(1 - \eta\tau)^t. \tag{3.9.135}$$

Now, consider the Lyapunov function:

$$V_{t+1} = \|\theta_i^{t+1} - \theta_i^\star\|^2 + \frac{2C}{\eta\tau}(1 - \eta\tau)^{t+1}. \tag{3.9.136}$$

We have:

$$\begin{aligned}V_{t+1} &\leq (1 - \eta\tau/2)\|\theta_i^t - \theta_i^\star\|^2 + C(1 - \eta\tau)^t + \frac{2C}{\eta\tau}(1 - \eta\tau)(1 - \eta\tau)^t \\ &= (1 - \eta\tau/2)\|\theta_i^t - \theta_i^\star\|^2 + \frac{2C}{\eta\tau}(1 - \eta\tau)^t(1 - \eta\tau + \eta\tau/2) \\ &= (1 - \eta\tau/2)\|\theta_i^t - \theta_i^\star\|^2 + \frac{2C}{\eta\tau}(1 - \eta\tau)^t(1 - \eta\tau/2)\end{aligned}$$

$$= (1 - \eta\tau/2)\left(\|\theta_i^t - \theta_i^\star\|^2 + \frac{2C}{\eta\tau}(1-\eta\tau)^t\right) = (1-\eta\tau/2)V_t. \tag{3.9.137}$$

This shows linear convergence of the parameter θ_i to $\theta_i^\star$ since: $\|\theta_i^{t+1} - \theta_i^\star\|^2 \leq V_{t+1} \leq (1-\eta\tau/2)^t V_0$. Merging these inequalities for all players i completes the proof of the Theorem.

□

Proof of Corollary 3.5.5

We define $g_i := g_{\theta_i}$. The duality gap for the regularized game is given by:

$$\begin{aligned}
\mathrm{DG}_\tau(g_1, g_2, \cdots, g_N) &= \sum_{i=1}^N \left[f_i^\tau(g_i, g_{-i}) - \min_{\widetilde{g}_i} f_i^\tau(\widetilde{g}_i, g_{-i})\right] = \max_{\widetilde{g}_1, \widetilde{g}_2, \cdots, \widetilde{g}_N} \sum_{i=1}^N \left[f_i^\tau(g_i, g_{-i}) - f_i^\tau(\widetilde{g}_i, g_{-i})\right] \\
&= \max_{\widetilde{g}_1, \widetilde{g}_2, \cdots, \widetilde{g}_N} \sum_{i=1}^N [f_i^\tau(g_i, g_{-i}) - f_i^\tau(g_i, g_{-i}^\star) + f_i^\tau(g_i, g_{-i}^\star) - f_i^\tau(\widetilde{g}_i, g_{-i}^\star) + f_i^\tau(\widetilde{g}_i, g_{-i}^\star) - f_i^\tau(\widetilde{g}_i, g_{-i})] \\
&\leq \max_{\widetilde{g}_1, \widetilde{g}_2, \cdots, \widetilde{g}_N} \sum_{i=1}^N [f_i^\tau(g_i, g_{-i}) - f_i^\tau(g_i, g_{-i}^\star) + f_i^\tau(g_i, g_{-i}^\star) - f_i^\tau(g_i^\star, g_{-i}^\star) + f_i^\tau(\widetilde{g}_i, g_{-i}^\star) - f_i^\tau(\widetilde{g}_i, g_{-i})].
\end{aligned} \tag{3.9.138}$$

Next, we note that:

$$\begin{aligned}
&\sum_{i=1}^N f_i^\tau(g_i, g_{-i}) - f_i^\tau(g_i, g_{-i}^\star) + f_i^\tau(g_i, g_{-i}^\star) - f_i^\tau(g_i^\star, g_{-i}^\star) + f_i^\tau(\widetilde{g}_i, g_{-i}^\star) - f_i^\tau(\widetilde{g}_i, g_{-i}) \\
&\leq \sum_{i=1}^N \left[\|f_i^\tau(g_i, g_{-i}) - f_i^\tau(g_i, g_{-i}^\star)\| + \|f_i^\tau(g_i, g_{-i}^\star) - f_i^\tau(g_i^\star, g_{-i}^\star)\| + \|f_i^\tau(\widetilde{g}_i, g_{-i}^\star) - f_i^\tau(\widetilde{g}_i, g_{-i})\|\right] \\
&\leq \sum_{i=1}^N \left[\|f_i^\tau(g_i, g_{-i}) - f_i^\tau(g_i, g_{-i}^\star)\| + \|f_i^\tau(g_i, g_{-i}^\star) - f_i^\tau(g_i^\star, g_{-i}^\star)\| + \|f_i^\tau(\widetilde{g}_i, g_{-i}^\star) - f_i^\tau(\widetilde{g}_i, g_{-i})\|\right] \\
&\leq C_1 \sum_{i=1}^N \left[\|g_i - g_i^\star\| + \|g_{-i} - g_{-i}^\star\|\right] \leq C_2 N \sqrt{\mathrm{KL}(z^\star \| z_t)}.
\end{aligned} \tag{3.9.139}$$

This follows by noting that the functions f_i^τ are Lipschitz since they are continuous functions defined on a compact domain. The last step follows from Pinsker's inequality and the fact that the l_1 norm is an upper bound for the l_2 norm.

Algorithm 4 Proximal Point Method

```
**Initialize:** $\theta_i^0 = 0$ for all players $i$.
**for** $t = 1, 2, \cdots$ **do**
    $\theta_i^{t+1} = (1 - \eta\tau)\theta_i^t - \eta\nabla_{g_{\theta_i}} f_i(g_{\theta_i^{t+1}}, g_{\theta_i^{t+1}})\ \ \forall i \in [N]$.
**end for**
```

Combining the two inequalities (3.9.138) and (3.9.139) we have:

$$\mathrm{DG}_\tau(g_1, g_2, \cdots, g_N) \leq C_2 N \sqrt{\mathrm{KL}(z^\star \| z_t)}. \tag{3.9.140}$$

Let DG denote the Duality gap of the unregularized problem. Then we have:

$$\mathrm{DG}(g_1, g_2, \cdots, g_N) \leq \mathrm{DG}_\tau(g_1, g_2, \cdots, g_N) + 2N\tau \log n. \tag{3.9.141}$$

Therefore, setting $\tau = \frac{\epsilon}{4N \log n}$ and solving the regularized problem to an accuracy of $\frac{\epsilon^2}{4C_2^2 N^2}$ in terms of KL divergence, we have that:

$$\mathrm{DG}(g_1, g_2, \cdots, g_N) \leq \epsilon, \tag{3.9.142}$$

completing the proof. □

Proximal point and Extragradient methods for multi-player monotone games

We define $g_i := g_{\theta_i}$ and $g_i^t := g_{\theta_i^t}$ for simplicity.

Proximal-point updates

In this subsection, we show the convergence of the Proximal Point (PP) updates to the NE of the regularized N-player monotone game. The PP algorithm is presented in Algorithm 4

Theorem 3.9.15. Let $z^\star = (g_i^\star)_{i=1}^N$ be the Nash equilibrium of Problem (3.5.2). Also, we denote $z_t = (g_i^t)_{i=1}^N$. Define $\theta_i^\star := \frac{-\nabla_{g_i} f_i(g_i^\star, g_{-i}^\star)}{\tau}$. Then for updates in Algorithm 4, we have for $\eta\tau < 1/2$:

- $\mathrm{KL}(z^\star \| z_{t+1}) \leq (1 - \eta\tau)^t \mathrm{KL}(z^\star \| z_0)$.

- $\|\theta_{t+1} - \theta^\star\|^2 \leq (1 - \eta\tau/2)^t V_0$, where,

$$V_0 = \|\theta^t - \theta^\star\|^2 + \frac{2NC}{\eta\tau}, \qquad\qquad C = 2\eta^2 L^2 \left(1 + \frac{4}{\eta\tau}(1-\eta\tau)^2\right) \mathrm{KL}(z^\star \| z_0).$$

Proof. From the definiton of the KL divergence we have:

$$-\langle \log z_{t+1} - (1-\eta\tau)\log z_t - \eta\tau \log z^\star, z^\star \rangle = -(1-\eta\tau)\mathrm{KL}(z^\star\|z_t) + \mathrm{KL}(z^\star\|z_{t+1}), \tag{3.9.143}$$

and

$$\langle \log z_{t+1} - (1-\eta\tau)\log z_t - \eta\tau \log z^\star, z_{t+1} \rangle = (1-\eta\tau)\mathrm{KL}(z_{t+1}\|z_t) + \mathrm{KL}(z_{t+1}\|z^\star). \tag{3.9.144}$$

Substituting in Lemma 3.9.13 with $z = z_{t+1}$, we have:

$$\langle \log z_{t+1} - (1-\eta\tau)\log z_t - \eta\tau \log z^\star, z_{t+1} - z^\star \rangle \leq 0, \tag{3.9.145}$$

and using the two equalities, we get:

$$-(1-\eta\tau)\mathrm{KL}(z^\star\|z_t) + \mathrm{KL}(z^\star\|z_{t+1}) + (1-\eta\tau)\mathrm{KL}(z_{t+1}\|z_t) + \mathrm{KL}(z_{t+1}\|z^\star) \leq 0. \tag{3.9.146}$$

This implies:

$$\mathrm{KL}(z^\star\|z_{t+1}) \leq (1-\eta\tau)\mathrm{KL}(z^\star\|z_t). \tag{3.9.147}$$

This shows linear convergence of the KL divergence to the Nash equilibrium for the proximal point method, which completes the proof of the first part of the Theorem.

When the agents are playing the NE strategy, the updates reduce to

$$\theta_i^{t+1} = (1-\eta\tau)\theta_i^t - \eta\nabla_{g_i} f_i(g_i^\star, g_{-i}^\star). \tag{3.9.148}$$

From Lemma 3.9.13, we know that the NE satisfy:

$$g_i^\star(a) = \frac{e^{\frac{-[\nabla_{g_i} f_i(g_i^\star, g_{-i}^\star)]_a}{\tau}}}{K}, \tag{3.9.149}$$

where $K = \sum_{a' \in \mathcal{A}_i} e^{\frac{-[\nabla_{g_i} f_i(g_i^\star, g_{-i}^\star)]_{a'}}{\tau}}$. Taking log on both sides, we have:

$$-\eta \nabla_{g_i} f_i(g_i^\star, g_{-i}^\star) = \eta\tau \log g_i^\star + \eta\tau \log K. \tag{3.9.150}$$

Substituting this back into the θ update, we have:

$$\begin{aligned}
\theta_i^{t+1} &= (1-\eta\tau)\theta_i^t + \eta\tau \log g_i^\star + \eta\tau \log K \\
&= \theta_i^t - \eta\tau\theta_i^t + \eta\tau \log g_i^\star + \eta\tau \log K - \eta\tau \log Z_{\theta_i^t} + \eta\tau \log Z_{\theta_i^t} \\
&= \theta_i^t + \eta\tau \log g_i^\star - \eta\tau \log\left(\frac{e^{\theta_i^t}}{Z_{\theta_i^t}}\right) + \eta\tau \log\left(\frac{K}{Z_{\theta_i^t}}\right) \\
&= \theta_i^t + \eta\tau \log g_i^\star - \eta\tau \log g_{\theta_i^t} + \eta\tau \log\left(\frac{K}{Z_{\theta_i^t}}\right) \\
&= \theta_i^t + \eta\tau \log\left(\frac{g_i^\star}{g_{\theta_i^t}}\right) + \eta\tau \log\left(\frac{K}{Z_{\theta_i^t}}\right) = \theta_i^t + \eta\tau \log\left(\frac{g_i^\star K}{g_{\theta_i^t} Z_{\theta_i^t}}\right).
\end{aligned} \tag{3.9.151}$$

We can further simplify this as:

$$\theta_i^{t+1} = \theta_i^t + \eta\tau \log\left(\frac{e^{-[\nabla_{g_i} f_i(g_i^\star, g_{-i}^\star)]/\tau}}{e^{\theta_i^t}}\right). \tag{3.9.152}$$

This is nothing but:

$$\theta_i^{t+1} = \theta_i^t + \eta\tau \left(\frac{-[\nabla_{g_i} f_i(g_i^\star, g_{-i}^\star)]}{\tau} - \theta_i^t\right). \tag{3.9.153}$$

Now the Proximal Point updates can be written as:

$$\theta_i^{t+1} = \theta_i^t + \eta\tau \left(\frac{-\nabla_{g_i} f_i(g_i^\star, g_{-i}^\star)}{\tau} - \theta_i^t\right) + \varepsilon_t, \tag{3.9.154}$$

where $\varepsilon_t = \eta \nabla_{g_i} f_i(g_i^\star, g_{-i}^\star) - \eta \nabla_{g_i} f_i(g_i^{t+1}, g_{-i}^{t+1})$.

Defining $\theta_i^\star = \frac{-\nabla_{g_i} f_i(g_i^\star, g_{-i}^\star)}{\tau}$, we have:

$$\begin{aligned}
\|\theta_i^{t+1} - \theta_i^\star\|^2 &= \|\theta_i^t + \eta\tau\left(\theta_i^\star - \theta_i^t\right) + \varepsilon_t - \theta_i^\star\|^2 = \|(1-\eta\tau)(\theta_i^t - \theta_i^\star) + \varepsilon_t\|^2 \\
&= (1-\eta\tau)^2\|\theta_i^t - \theta_i^\star\|^2 + 2(1-\eta\tau)(\theta_i^t - \theta_i^\star)^\top \varepsilon_t + \|\varepsilon_t\|^2 \\
&\leq^{*1} (1-\eta\tau)^2\|\theta_i^t - \theta_i^\star\|^2 + \eta\tau\|\theta_i^t - \theta_i^\star\|^2 + \frac{4}{\eta\tau}(1-\eta\tau)^2\|\varepsilon_t\|^2 + \|\varepsilon_t\|^2 \\
&= (1-\eta\tau+\eta^2\tau^2)\|\theta_i^t - \theta_i^\star\|^2 + \left(1 + \frac{4}{\eta\tau}(1-\eta\tau)^2\right)\|\varepsilon_t\|^2, \qquad (3.9.155)
\end{aligned}$$

where $*1$ follows from Young's inequality.

Now, for the proximal point methods, we have:

$$\begin{aligned}
\|\varepsilon_t\|^2 &= \eta^2\|\nabla_{g_i} f_i(g_i^\star, g_{-i}^\star) - \nabla_{g_i} f_i(g_i^{t+1}, g_{-i}^{t+1})\|^2 \\
&\leq \eta^2 L^2 \|z_{t+1} - z^\star\|^2 \leq 2\eta^2 L^2 \mathrm{KL}(z^\star \| z_{t+1}), \qquad (3.9.156)
\end{aligned}$$

where it follows from Pinsker's inequality and the fact that the l_1 norm is an upper bound for the l_2 norm.

This gives us:

$$\|\theta_i^{t+1} - \theta_i^\star\|^2 \leq (1-\eta\tau+\eta^2\tau^2)\|\theta_i^t - \theta_i^\star\|^2 + C(1-\eta\tau)^t, \qquad (3.9.157)$$

where

$$C = 2\eta^2 L^2 \left(1 + \frac{4}{\eta\tau}(1-\eta\tau)^2\right)\mathrm{KL}(z^\star \| z_0). \qquad (3.9.158)$$

For $\eta\tau < 1/2$, this reduces to:

$$\|\theta_i^{t+1} - \theta_i^\star\|^2 \leq (1-\eta\tau/2)\|\theta_i^t - \theta_i^\star\|^2 + C(1-\eta\tau)^t. \qquad (3.9.159)$$

Algorithm 5 Extragradient Method

```
**Initialize:** $g_0$ and $h_0$.
**for** $t = 1, 2, \cdots$ **do**
    $\bar{\theta}_i^{t+1} = (1-\eta\tau)\theta_i^t - \eta\nabla_{g_{\theta_i}} f_i(g_{\theta_i^t}, g_{\theta_{-i}^t}) \quad \forall i \in [N].$
    $\theta_i^{t+1} = (1-\eta\tau)\theta_i^t - \eta\nabla_{g_{\theta_i}} f_i(g_{\bar{\theta}_i^{t+1}}, g_{\bar{\theta}_{-i}^{t+1}}) \quad \forall i \in [N].$
**end for**
```

Now, consider the Lyapunov function:

$$V_{t+1} = \|\theta_i^{t+1} - \theta_i^\star\|^2 + \frac{2C}{\eta\tau}(1-\eta\tau)^{t+1}. \tag{3.9.160}$$

We have:

$$\begin{aligned}
V_{t+1} &\leq (1-\eta\tau/2)\|\theta_i^t - \theta_i^\star\|^2 + C(1-\eta\tau)^t + \frac{2C}{\eta\tau}(1-\eta\tau)(1-\eta\tau)^t \\
&= (1-\eta\tau/2)\|\theta_i^t - \theta_i^\star\|^2 + \frac{2C}{\eta\tau}(1-\eta\tau)^t(1-\eta\tau+\eta\tau/2) \\
&= (1-\eta\tau/2)\|\theta_i^t - \theta_i^\star\|^2 + \frac{2C}{\eta\tau}(1-\eta\tau)^t(1-\eta\tau/2) \\
&= (1-\eta\tau/2)\left(\|\theta_i^t - \theta_i^\star\|^2 + \frac{2C}{\eta\tau}(1-\eta\tau)^t\right) \\
&= (1-\eta\tau/2)V_t.
\end{aligned} \tag{3.9.161}$$

This shows linear convergence of the parameter θ_i to $\theta_i^\star$ since:

$$\|\theta_i^{t+1} - \theta_i^\star\|^2 \leq V_{t+1} \leq (1-\eta\tau/2)^t V_0. \tag{3.9.162}$$

Merging these inequalities for all players i completes the proof of the theorem. □

Extragradient updates

In this subsection, we show the convergence of the Extragradient method to the NE of the regularized N-player monotone game. The EG algorithm is presented in Algorithm 5.

Theorem 3.9.16. Let $z^\star = (g_i^\star)_{i=1}^N$ be the unique Nash equilibrium of Problem (3.5.2). Also, we denote $z_t = (g_{\theta_i^t})_{i=1}^N$. Then for updates in Algorithm 5, we have for stepsize satisfying $0 < \eta < \frac{1}{2NL+\tau}$:

- Convergence of distributions:

$$\max\left\{\mathrm{KL}(z^\star\|z_t), \mathrm{KL}(z^\star\|\bar{z}_{t+1})\right\} \le (1-\eta\tau)^t 2\mathrm{KL}(z^\star\|z_0). \tag{3.9.163}$$

- Convergence of parameters:

$$\|\theta_{t+1}-\theta^\star\|^2 \le (1-\eta\tau/2)^t V_0, \tag{3.9.164}$$

where,

$$V_0 = \|\theta^t-\theta^\star\|^2 + \frac{2NC}{\eta\tau}, \qquad C = 4\eta^2L^2\left(1+\frac{4}{\eta\tau}(1-\eta\tau)^2\right)\mathrm{KL}(z^\star\|z_0).$$

Proof. From the EG updates in Algorithm 5, we have:

$$\log \bar{g}_i^{t+1} - \log g_i^{t+1} = -\eta\left(\nabla_{g_i} f_i(g_i^t, g_{-i}^t) - \nabla_{g_i} f_i(\bar{g}_i^{t+1}, \bar{g}_{-i}^{t+1})\right) + c\cdot\mathbb{1}. \tag{3.9.165}$$

This implies:

$$\begin{aligned}
\langle \log \bar{g}_i^{t+1} - \log g_i^{t+1}, \bar{g}_i^{t+1} - g^i_{t+1}\rangle &= -\eta\langle \nabla_{g_i} f_i(g_i^t, g_{-i}^t) - \nabla_{g_i} f_i(\bar{g}_i^{t+1}, \bar{g}_{-i}^{t+1}), \bar{g}_i^{t+1} - g_i^{t+1}\rangle \\
&\le^{*1} \eta\|\nabla_{g_i} f_i(g_i^t, g_{-i}^t) - \nabla_{g_i} f_i(\bar{g}_i^{t+1}, \bar{g}_{-i}^{t+1})\| \cdot \|\bar{g}_i^{t+1} - g_i^{t+1}\| \\
&\le^{*2} \eta L\left(\|\bar{g}_i^{t+1} - g_i^t\| + \|\bar{g}_{-i}^{t+1} - g_{-i}^t\|\right)\|\bar{g}_i^{t+1} - g_i^{t+1}\| \\
&\le^{*3} \frac{1}{2}\eta L\left(\|\bar{g}_i^{t+1} - g_i^t\|^2 + 2\|\bar{g}_i^{t+1} - g_i^{t+1}\|^2 + \|\bar{g}_{-i}^{t+1} - g_{-i}^t\|^2\right),
\end{aligned} \tag{3.9.166}$$

where $(*1)$ follows from the fact that $a^\top b \le \|a\|\|b\|$, $(*2)$ follows from Assumption 3.5.1, and $(*3)$ follows from $x\cdot y \le \frac{1}{2}(x^2+y^2)$. Since this is true for all players, we have:

$$\begin{aligned}
&\langle \log \bar{z}_{t+1} - \log z_{t+1}, \bar{z}_{t+1} - z_{t+1}\rangle \\
&\qquad \le \frac{1}{2}\eta L \sum_{i=1}^{N}\left(\|\bar{g}_i^{t+1} - g_i^t\|^2 + 2\|\bar{g}_i^{t+1} - g_i^{t+1}\|^2 + \|\bar{g}_{-i}^{t+1} - g_{-i}^t\|^2\right) \\
&\qquad \le \frac{1}{2}\eta L\left(N\|\bar{z}_{t+1} - z_t\|^2 + 2\|\bar{z}_{t+1} - z_{t+1}\|^2\right)
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{N}{2}\eta L \left(\|\bar{z}_{t+1} - z_t\|^2 + \|\bar{z}_{t+1} - z_{t+1}\|^2 \right) \\
&\leq^{*1} N\eta L \left(\mathrm{KL}(\bar{z}_{t+1}\|z_t) + \mathrm{KL}(z_{t+1}\|\bar{z}_{t+1}) \right), \qquad (3.9.167)
\end{aligned}
$$

where $(*1)$ follows from Pinsker's Inequality and the fact that the l_1 norm is an upper bound for the l_2 norm. Substituting this in Equation (3.9.115), we have

$$
\begin{aligned}
\mathrm{KL}(z^\star\|z_{t+1}) \leq (1-\eta\tau)\mathrm{KL}(z^\star\|z_t) &- (1-\eta\tau-N\eta L)\mathrm{KL}(\bar{z}_{t+1}\|z_t) \\
&- \eta\tau\mathrm{KL}(\bar{z}_{t+1}\|z^\star) - (1-N\eta L)\mathrm{KL}(z_{t+1}\|\bar{z}_{t+1}). \qquad (3.9.168)
\end{aligned}
$$

If $\eta \leq \frac{1}{\tau+NL}$, we have:

$$
\mathrm{KL}(z^\star\|z_{t+1}) \leq (1-\eta\tau)\mathrm{KL}(z^\star\|z_t). \qquad (3.9.169)
$$

Similar to the derivation of Equation (3.9.167), we have:

$$
-\langle \log \bar{z}_{t+1} - \log z_{t+1}, z^\star - \bar{z}_{t+1} \rangle \leq N\eta L \left(\mathrm{KL}(\bar{z}_{t+1}\|z_t) + \mathrm{KL}(z^\star\|\bar{z}_{t+1}) \right). \qquad (3.9.170)
$$

Substituting this in Equation (3.9.116) we have:

$$
(1-N\eta L)\mathrm{KL}(z^\star\|\bar{z}_{t+1}) \leq \mathrm{KL}(z^\star\|z_{t+1}) + N\eta L\mathrm{KL}(\bar{z}_{t+1}\|z_t). \qquad (3.9.171)
$$

Now, plugging Inequality (3.9.168) in (3.9.171) we have:

$$
\begin{aligned}
(1-N\eta L)\mathrm{KL}(z^\star\|\bar{z}_{t+1}) &\leq (1-\eta\tau)\mathrm{KL}(z^\star\|z_t) \\
&- (1-\eta\tau-2N\eta L)\mathrm{KL}(\bar{z}_{t+1}\|z_t) - \eta\tau\mathrm{KL}(\bar{z}_{t+1}\|z^\star) - (1-N\eta L)\mathrm{KL}(z_{t+1}\|\bar{z}_{t+1}). \\
&\qquad (3.9.172)
\end{aligned}
$$

With stepsize $\eta < \frac{1}{\tau+2NL}$ we have:

$$
\mathrm{KL}(z^\star\|\bar{z}_{t+1}) \leq 2\mathrm{KL}(z^\star\|z_t) \leq 2(1-\eta\tau)^t\mathrm{KL}(z^\star\|z_0), \qquad (3.9.173)
$$

which completes the first part of the proof.

The proof of parameter convergence follows exactly from the proof of Theorem 3.5.3 and we avoid rewriting it here. □

Parameterization with function approximation

In this section, we discuss the monotone game setting with function approximation for policy parameterization, as discussed for matrix games in §3.4. In this setting, the regularized problem each player i faces is:

$$\min_{\theta_i \in \mathbb{R}^d} \quad f_i(g_{\theta_i}, g_{\theta_{-i}}) - \tau \mathcal{H}(g_{\theta_i}), \tag{3.9.174}$$

where g_{θ_i} is a log-linear policy parametrization. In the next lemma, we show the existence of a NE in this setting as well as the equivalence of this problem to one on the entire simplex, in the same spirit as in §3.4 for matrix games:

Lemma 3.9.17. The in-class NE (Definition 3.9.11) for the unregularized (and regularized) monotone game under the log-linear policy parametrization exists. Also, under Assumption 3.4.1, solving Problem (3.9.174) for all i is equivalent to:

$$\min_{g_{\theta_i} \in \Delta} \quad f_i(\Psi g_{\theta_i}, \Psi g_{\theta_{-i}}) - \tau \mathcal{H}(g_{\theta_i}), \tag{3.9.175}$$

where Ψ is defined in Equation (3.4.7).

Now, using Lemma 3.9.17, and Proposition 3.4.6, we describe an algorithm to solve this problem in the following corollary.

Corollary 3.9.18. The update rule:

$$\begin{aligned}
\bar{\theta}_i^{t+1} &= (1 - \eta\tau)\theta_i^t - [(M^\top)^{-1}|0]\eta\widetilde{P}\nabla_{g_{\theta_i}} f_i(g_{\bar{\theta}_i^t}, g_{\bar{\theta}_{-i}^t}) \\
\theta_i^{t+1} &= (1 - \eta\tau)\theta_i^t - [(M^\top)^{-1}|0]\eta\widetilde{P}\nabla_{g_{\theta_i}} f_i(g_{\bar{\theta}_i^{t+1}}, g_{\bar{\theta}_{-i}^{t+1}}),
\end{aligned}$$

solves Problem (3.9.174) with similar guarantees given by Theorem 3.5.3. Here, the NE parameter value to which the algorithm converges to is given by $\theta_i^\star = \frac{-[(M^\top)^{-1}|0]\widetilde{P}\nabla_{g_{\theta_i}} f_i(g_i^\star, g_{-i}^\star)}{\tau}$. Furthermore, by choosing the regularization parameter τ small enough, like in Corollary 3.5.5,

Algorithm 6 Optimistic NPG for Markov Games

```
Initialize: $Q_0 = 0$
for $t = 1, 2, \cdots, T_{outer}$ do
    for $s = 1, 2, \cdots, |\mathcal{S}|$ do
        Let $Q_t(s, a, b) := r(s, a, b) + \gamma \mathbb{E}_{s' \sim \mathbb{P}(\cdot|s,a,b)}[V_t(s')]$
        Solve $\min_{\theta \in \mathbb{R}^n} \max_{\nu \in \mathbb{R}^n} f_\tau(Q(s); g_\theta, h_\nu)$ by running the Optimistic NPG algorithm (Algorithm 1) for $T_{inner}$ iterations and return the last iterates $(\theta_{T_{inner}}, \nu_{T_{inner}})$.
        Set $V_{t+1}(s) = f_\tau(Q_t(s); g_{\theta_{T_{inner}}}(\cdot \,|\, s), h_{\nu_{T_{inner}}}(\cdot \,|\, s))$
    end for
end for
```

we reach an ϵ-in-class NE (Definition 3.9.12) of the unregularized monotone game under the function approximation setting.

Proof of Lemma 3.9.17

The proof of this Lemma follows along the lines of Lemma 3.4.4 and Theorem 3.4.5. The key here is to notice that:

$$\nabla_{g_{\theta_i}} f_i(\Psi g_{\theta_i}, \Psi g_{\theta_{-1}}) = \Psi^\top \nabla_{\Psi g_{\theta_i}} f_i(\Psi g_{\theta_i}, \Psi g_{\theta_{-1}}), \tag{3.9.176}$$

and that: $\Psi \mu = \mu, \quad \forall \mu \in \widetilde{\Delta}$. The rest of the proof is identical to the proof of Lemma 3.4.4.

□

3.9.5 Missing Details and Proofs in §3.6

Remark 3.9.19. We note that all results presented in this section also follow for the case where the cardinality of the action spaces for both players are asymmetric. However, we stick to the case where the number of action is the same for both players in all states for ease of exposition.

Proof of Theorem 3.6.3

We first have the following lemma which shows the smoothness property of the NE policy with respect to the game matrix Q.

Lemma 3.9.20. Consider the following entropy regularized game:

$$\min_{x\in\widetilde{\Delta}}\max_{y\in\widetilde{\Delta}} \quad x^\top Qy - \tau\mathcal{H}(x) + \tau\mathcal{H}(y), \tag{3.9.177}$$

where $\widetilde{\Delta} \subseteq \Delta$ is a convex compact subset of the simplex given by Equation (3.4.5). Let $(x_Q^\star, y_Q^\star)$ denote the unique solution to this problem (note that this is unique since we have a strongly convex-strongly concave objective over a compact convex set). Then, we have:

$$\max\left\{\|x_{Q_1}^\star - x_{Q_2}^\star\|, \|y_{Q_1}^\star - y_{Q_2}^\star\|\right\} \le C\cdot\|Q_1 - Q_2\|_F, \tag{3.9.178}$$

for some constant $C > 0$ and for any $Q_1, Q_2 \in \mathbb{R}^{n\times n}$.

Proof. First notice that by the proof of Theorem 3.4.5, solving (3.9.177) is equivalent to solving

$$\min_{x\in\Delta}\max_{y\in\Delta} \quad x^\top\Psi^\top Q\Psi y - \tau\mathcal{H}(x) + \tau\mathcal{H}(y), \tag{3.9.179}$$

with Ψ being defined in Equation (3.4.7), which admits a unique solution. In other words, the solution $(x_Q^\star, y_Q^\star)$ also solves (3.9.179). Also, notice that the solution to (3.9.179) always lies in the relative interior of Δ, given by (3.9.91). In other words, $[x_Q^\star]_a > 0$ and $[y_Q^\star]_b > 0$ for all a and b. Due to the simplex constraint, the free variable is of dimension $n-1$, and the last dimension of x can be represented as $1 - \sum_{a=1}^{n-1} x_a > 0$ (similarly for y). Let $\widetilde{x} = (x_1, x_2, \cdots, x_{n-1})^\top$, $\widetilde{y} = (y_1, y_2, \cdots, y_{n-1})^\top$, and $f(x,y) := x^\top\Psi^\top Q\Psi y - \tau\mathcal{H}(x) + \tau\mathcal{H}(y)$. Recall that $f(x,y)$ is strongly convex in x and strongly concave in y. Note that $f(x,y) = f\big(\Lambda(\widetilde{x}), \Lambda(\widetilde{y})\big) = \widetilde{f}(\widetilde{x},\widetilde{y}) := \Lambda(\widetilde{x})^\top\Psi^\top Q\Psi\Lambda(\widetilde{y}) - \tau\mathcal{H}\big(\Lambda(\widetilde{x})\big) + \tau\mathcal{H}\big(\Lambda(\widetilde{y})\big)$, where $x = \Lambda(\widetilde{x}) = \begin{bmatrix} I \\ -\mathbb{1}^\top \end{bmatrix}\widetilde{x} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$, and $\mathbb{1}$ denotes an all-one vector of proper dimension. Note that $\widetilde{f}(\widetilde{x},\widetilde{y})$ is also strongly convex in $\widetilde{x}$ and strongly concave in $\widetilde{y}$, as for any $\widetilde{y}$, the Hessian $\nabla^2_{\widetilde{x}}\widetilde{f}(\widetilde{x},\widetilde{y}) = \begin{bmatrix} I \mid -\mathbb{1} \end{bmatrix}\nabla^2_x f(\Lambda(\widetilde{x}), \Lambda(\widetilde{y}))\begin{bmatrix} I \\ -\mathbb{1}^\top \end{bmatrix} \succ 0$, so is the Hessian with respect to $\widetilde{y}$ for any $\widetilde{x}$. Hence, the solution to the minimax problem

$$\min_{\left\{\widetilde{x}\,\middle|\,\widetilde{x}_a\geq 0, 1-\sum_{a=1}^{n-1}\widetilde{x}_a\geq 0\right\}}\ \max_{\left\{\widetilde{y}\,\middle|\,\widetilde{y}_b\geq 0, 1-\sum_{b=1}^{n-1}\widetilde{y}_b\geq 0\right\}}\ \widetilde{f}(\widetilde{x},\widetilde{y}) \tag{3.9.180}$$

is given by $(\widetilde{x}_Q^\star, \widetilde{y}_Q^\star)$, where $\widetilde{x}_Q^\star$ and $\widetilde{y}_Q^\star$ are just the first $n-1$ dimensions of $x_Q^\star$ and $y_Q^\star$, satisfying $[\widetilde{x}_Q^\star]_a > 0$, $[\widetilde{y}_Q^\star]_b > 0$, and $1-\sum_{a=1}^{n-1}[\widetilde{x}_Q^\star]_a > 0$, $1-\sum_{b=1}^{n-1}[\widetilde{y}_Q^\star]_b > 0$, i.e., the constraints in (3.9.180) are not violated at $(\widetilde{x}_Q^\star, \widetilde{y}_Q^\star)$. By KKT conditions, it holds that at $(\widetilde{x}_Q^\star, \widetilde{y}_Q^\star)$

$$\tau\nabla_{\widetilde{x}}\mathcal{H}(\Lambda(\widetilde{x}_Q^\star)) - \left[I \mid -\mathbb{1}\right]\Psi^\top Q\Psi\Lambda(\widetilde{y}_Q^\star) = 0 \tag{3.9.181}$$

$$\tau\nabla_{\widetilde{y}}\mathcal{H}(\Lambda(\widetilde{y}_Q^\star)) + \left[I \mid -\mathbb{1}\right]\Psi^\top Q^\top\Psi\Lambda(\widetilde{x}_Q^\star) = 0. \tag{3.9.182}$$

Define operator $\mathcal{G}\big(\widetilde{x},\widetilde{y},\mathrm{vec}(Q)\big)$ as

$$\mathcal{G}\big(\widetilde{x},\widetilde{y},\mathrm{vec}(Q)\big) := \begin{bmatrix} \tau\nabla_{\widetilde{x}}\mathcal{H}(\Lambda(\widetilde{x})) - \left[I \mid -\mathbb{1}\right]\Psi^\top Q\Psi\Lambda(\widetilde{y}) \\ \tau\nabla_{\widetilde{y}}\mathcal{H}(\Lambda(\widetilde{y})) + \left[I \mid -\mathbb{1}\right]\Psi^\top Q^\top\Psi\Lambda(\widetilde{x}) \end{bmatrix},$$

where $\mathrm{vec}(Q)$ is the vectorization of Q. Then, $(\widetilde{x}_Q^\star, \widetilde{y}_Q^\star)$ is given by the solution to $\mathcal{G}\big(\widetilde{x},\widetilde{y},\mathrm{vec}(Q)\big) = 0$. Note that the Jacobian of $\mathcal{G}$ with respect to $[\widetilde{x}^\top, \widetilde{y}^\top]^\top$ is

$$\mathcal{M}\big(\widetilde{x},\widetilde{y},\mathrm{vec}(Q)\big) := \left[\frac{\partial\mathcal{G}}{\partial\widetilde{x}} \mid \frac{\partial\mathcal{G}}{\partial\widetilde{y}}\right] = \begin{bmatrix} \tau\nabla_{\widetilde{x}}^2\mathcal{H}(\Lambda(\widetilde{x})) & -\left[I \mid -\mathbb{1}\right]\Psi^\top Q\Psi\begin{bmatrix} I \\ -\mathbb{1}^\top\end{bmatrix} \\ \left[I \mid -\mathbb{1}\right]\Psi^\top Q^\top\Psi\begin{bmatrix} I \\ -\mathbb{1}^\top\end{bmatrix} & \tau\nabla_{\widetilde{y}}^2\mathcal{H}(\Lambda(\widetilde{y})) \end{bmatrix},$$

which is always invertible for any $\widetilde{x}$ and $\widetilde{y}$ belonging to the constraints in (3.9.180), due to the fact that $\tau\nabla_{\widetilde{x}}^2\mathcal{H}(\Lambda(\widetilde{x})), \tau\nabla_{\widetilde{y}}^2\mathcal{H}(\Lambda(\widetilde{y})) \succ 0$, and $\mathcal{M}\big(\widetilde{x},\widetilde{y},\mathrm{vec}(Q)\big)$ is skew-symmetric, yielding the fact that the real parts of the eigenvalues of $\mathcal{M}\big(\widetilde{x},\widetilde{y},\mathrm{vec}(Q)\big)$, which are the eigenvalues of $(\mathcal{M}^\top + \mathcal{M})/2$, are always positive. In fact, the real parts are uniformly lower bounded by some constant $\eta > 0$ for any $\widetilde{x}$ and $\widetilde{y}$ belong to the constraints in (3.9.180), due to the strong convexity of $\mathcal{H}(\Lambda(\widetilde{x}))$ and $\mathcal{H}(\Lambda(\widetilde{y}))$. Hence, we have

$$\left\|\mathcal{M}\big(\widetilde{x},\widetilde{y},\mathrm{vec}(Q)\big)^{-1}\right\|_2 \leq \frac{2}{\lambda_{\min}\Big(\mathcal{M}\big(\widetilde{x},\widetilde{y},\mathrm{vec}(Q)\big)+\mathcal{M}\big(\widetilde{x},\widetilde{y},\mathrm{vec}(Q)\big)^\top\Big)} \leq \frac{1}{\eta}.$$

Due to the invertibility of $\mathcal{M}\big(\widetilde{x},\widetilde{y},\mathrm{vec}(Q)\big)$, we can apply implicit function theorem [71] for any solution to $\mathcal{G}\big(\widetilde{x},\widetilde{y},\mathrm{vec}(Q)\big)=0$, and obtain that for any such a solution $(\widetilde{x}_Q^\star,\widetilde{y}_Q^\star,\mathrm{vec}(Q))$, there exists a neighborhood of it such that for any $(\widetilde{x},\widetilde{y},\mathrm{vec}(\widetilde{Q}))$ in the neighborhood

$$\frac{\partial[\widetilde{x}^\top,\widetilde{y}^\top]^\top}{\partial\mathrm{vec}(\widetilde{Q})} = -\mathcal{M}\big(\widetilde{x},\widetilde{y},\mathrm{vec}(\widetilde{Q})\big)^{-1}\cdot\frac{\partial\mathcal{G}\big(\widetilde{x},\widetilde{y},\mathrm{vec}(\widetilde{Q})\big)}{\partial\mathrm{vec}(\widetilde{Q})}.$$

Notice that $\frac{\partial\mathcal{G}(\widetilde{x},\widetilde{y},\mathrm{vec}(\widetilde{Q}))}{\partial\mathrm{vec}(\widetilde{Q})}$ is uniformly bounded in norm on the constrained sets in (3.9.180), due to the boundedness of the sets. Hence, there exists a uniform constant $C'>0$ such that

$$\left\|\frac{\partial[(\widetilde{x}_Q^\star)^\top,\ (\widetilde{y}_Q^\star)^\top]^\top}{\partial\mathrm{vec}(Q)}\right\|_2 \leq \left\|\mathcal{M}\big(\widetilde{x}_Q^\star,\widetilde{y}_Q^\star,\mathrm{vec}(Q)\big)^{-1}\right\|_2\cdot\left\|\frac{\partial\mathcal{G}\big(\widetilde{x}_Q^\star,\widetilde{y}_Q^\star,\mathrm{vec}(Q)\big)}{\partial\mathrm{vec}(\widetilde{Q})}\right\|_2 \leq C'$$

for any $(\widetilde{x}_Q^\star,\widetilde{y}_Q^\star,\mathrm{vec}(Q))$. By the mean-value theorem, we know that

$$\left\|[(\widetilde{x}_{Q_1}^\star)^\top,\ (\widetilde{y}_{Q_1}^\star)^\top]-[(\widetilde{x}_{Q_2}^\star)^\top,\ (\widetilde{y}_{Q_2}^\star)^\top]\right\|_2 \leq C'\cdot\left\|\mathrm{vec}(Q_1)-\mathrm{vec}(Q_2)\right\|_2.$$

Finally, notice that

$$\begin{aligned}&\left\|[(x_{Q_1}^\star)^\top,\ (y_{Q_1}^\star)^\top]-[(x_{Q_2}^\star)^\top,\ (y_{Q_2}^\star)^\top]\right\|_2\\ &\qquad\leq\left\|\begin{bmatrix}I\\-\mathbb{1}^\top\end{bmatrix}\right\|_2\cdot\left\|[(\widetilde{x}_{Q_1}^\star)^\top,\ (\widetilde{y}_{Q_1}^\star)^\top]-[(\widetilde{x}_{Q_2}^\star)^\top,\ (\widetilde{y}_{Q_2}^\star)^\top]\right\|_2,\end{aligned}$$

which completes the proof by the equivalence of norms. □

Tabular case and proof of Theorem 3.6.3

We set the stepsize to be:

$$\eta=\frac{1-\gamma}{2(1+\tau(\log n+1-\gamma))}. \tag{3.9.183}$$

The convergence of the Q-function follows from Lemma 3.9.8, along with Theorem 2 in [23]. In particular, it is shown that when the inner problem is solved upto an accuracy of ϵ, we have:

$$\|Q_t - Q^\star\|_\infty \leq \epsilon + \gamma^t \|Q_0 - Q^\star\|_\infty \tag{3.9.184}$$

We show the convergence of the parameter next. We define:

$$\theta^\star(s) = \frac{-Q^\star_\tau h_{\nu^\star}(\cdot \,|\, s)}{\tau}, \qquad \nu^\star(s) = \frac{{Q^\star_\tau}^\top g_{\theta^\star}(\cdot \,|\, s)}{\tau}. \tag{3.9.185}$$

Similarly, we also define $\theta^\star_{Q_t}$ and $\nu^\star_{Q_t}$ as:

$$\theta^\star_{Q_t} = \frac{-Q_t h_{\nu^\star_{Q_t}}(\cdot \,|\, s)}{\tau}, \qquad \nu^\star(s) = \frac{{Q_t}^\top g_{\theta^\star_{Q_t}}(\cdot \,|\, s)}{\tau}. \tag{3.9.186}$$

We have:

$$\|\theta_t - \theta^\star_{Q^\star}\|^2 = \|\theta_t - \theta^\star_{Q_t} + \theta^\star_{Q_t} - \theta^\star\|^2 \leq 2\|\theta_t - \theta^\star_{Q_t}\|^2 + 2\|\theta^\star_{Q_t} - \theta^\star\|^2. \tag{3.9.187}$$

Now, the first term converges approximately after the inner loop terminates. We can analyze the second term as follows:

$$\begin{aligned}
&\|\theta^\star_{Q_t} - \theta^\star\|^2 =^{*1} \frac{1}{\tau}\|Q_t h_{\nu^\star_{Q_t}} - Q^\star h_{\nu^\star_{Q^\star}}\|^2 = \frac{1}{\tau}\|Q_t h_{\nu^\star_{Q_t}} - Q^\star h_{\nu^\star_{Q_t}} + Q^\star h_{\nu^\star_{Q_t}} - Q^\star h_{\nu^\star_{Q^\star}}\|^2 \\
&\leq \frac{2}{\tau}\left(\|Q_t h_{\nu^\star_{Q_t}} - Q^\star h_{\nu^\star_{Q_t}}\|^2 + \|Q^\star h_{\nu^\star_{Q_t}} - Q^\star h_{\nu^\star_{Q^\star}}\|^2\right) \\
&\leq \frac{2}{\tau}\left(\|Q_t - Q^\star\|_F^2 + \|Q^\star\|_F^2 \|h_{\nu^\star_{Q_t}} - h_{\nu^\star_{Q^\star}}\|^2\right) \\
&\leq^{*2} \frac{2}{\tau}\left(\|Q_t - Q^\star\|_F^2 + C\|Q^\star\|_F^2\|Q_t - Q^\star\|_F^2\right) = \frac{2}{\tau}\left(1 + C^2\|Q^\star\|_F^2\right)\|Q_t - Q^\star\|_F^2,
\end{aligned} \tag{3.9.188}$$

where $(*1)$ follows from the definition of $\theta^\star$ and $(*2)$ follows from Lemma 3.9.20. Substituting this in Equation (3.9.187) we have:

$$\|\theta_t - \theta^\star_{Q^\star}\|^2 \leq 2\|\theta_t - \theta^\star_{Q_t}\|^2 + \frac{4}{\tau}\left(1 + C^2\|Q^\star\|_F^2\right)\|Q_t - Q^\star\|_F^2. \tag{3.9.189}$$

Therefore, we have $\|\theta_t - \theta^\star_{Q^\star}\|^2 \leq \epsilon$ if

$$\|\theta_t - \theta^\star_{Q_t}\|^2 \leq \frac{\epsilon}{4}, \qquad \|Q_t - Q^\star\|_F^2 \leq \frac{\epsilon}{\frac{8}{\tau}\left(1 + C^2\|Q^\star\|_F^2\right)}. \tag{3.9.190}$$

Note that the first term can be achieved by setting the inner-loop iterations T_{inner} to be the following (from Theorem 3.3.4):

$$T_{inner} = \mathcal{O}\left(\frac{1}{\eta\tau}\left(\log\frac{1}{\epsilon} + \log\frac{1}{1-\gamma} + \log\log n + \log\frac{1}{\eta}\right)\right), \tag{3.9.191}$$

and the second term can be achieved by noting that:

$$\|Q_t - Q^\star\|_F \leq d\|Q_t - Q^\star\|_\infty, \tag{3.9.192}$$

for $d = |\mathcal{S}| \times |A|$. Now, using Inequality (3.9.184), we can set the outer-loop iterations T_{outer} to be:

$$T_{outer} = \mathcal{O}\left(\frac{1}{1-\gamma}\left(\log\frac{d}{\epsilon} + \log\left(\frac{8}{\tau}\left(1 + C^2\|Q^\star\|_F^2\right)\right) + \log\frac{1+\tau\log n}{1-\gamma}\right)\right), \tag{3.9.193}$$

to get the desired convergence result. This completes the proof. □

Function approximation setting

In this subsection, we discuss Markov games where the policies have a log-linear parametrization. The basic idea is to follow the tabular setting, but only for those states for which there is an action for which the feature vector corresponding to the state-action pair is non-zero. We first make the following assumption on the feature matrix ϕ.

Assumption 3.9.21. The feature matrix Φ is full rank, Moreover, it is of the form $\Phi = [\phi_1, \phi_2, \cdots, \phi_{|\mathcal{S}|\times|\mathcal{A}|}] = [I \mid 0]$.

Note that this assumption is similar to Assumption 3.4.1 for the matrix game. This The full rank assumption is standard in the literature. Furthermore, this particular structure of the feature matrix, though being restrictive, ensures that the constraint set of policies is convex (similar to the case of matrix games), otherwise the minimax theorem of $\min\max = \max\min$

might not hold, i.e., the Nash equilibrium for the parameterized game does not exist. See the paragraph below Assumption 3.4.1 for further discussion on the structure of the feature matrix. We next describe the detailed description of the setup.

Setup

Each column of Φ is a feature vector corresponding to some state action pair (s, a). Note that for each state, there could be 0 to $\min\{n, d\}$ actions for which the feature vector is non-zero.

Now, consider a state $s \in S$. Define $A_s = \{a \in \mathcal{A} : \Phi_{s,a} \neq 0\}$ where $\Phi_{s,a}$ corresponds to the column in the feature matrix for state s, and action a, and here 0 denotes the zero vector. Therefore A_s is the set of actions in state s for which the feature vector is non-zero. For sake of notational simplicity, let these be the actions $1, 2, \cdots, |A_s|$. Note that A_s can be an empty set. We further assume that the first $|A_1|$ columns of Φ are corresponding to state 1, the next $|A_2|$ columns correspond to state 2 and so on. Note that we have $\sum_{s\in\mathcal{S}} |A_s| = d$.

For state s, if A_s is nonempty, define the feature matrix $\Phi_s = [I_{|A_s|} \mid 0] \in \mathbb{R}^{|A_s|\times n}$. Note that this would be the feature matrix corresponding to each state for the original feature matrix Φ. Now, define $\widetilde{\Delta}_s$ corresponding to each state s using the feature matrix Φ_s, as in Equation (3.4.5). This corresponds to the set of admissible distributions under the function approximation setting for state s. If the set A_s is empty, we take $\widetilde{\Delta}_s$ to be the singleton set with the uniform distribution. Furthermore, we define $\widetilde{\Delta} = \times_{s\in\mathcal{S}} \widetilde{\Delta}_s$.

Next, for notational convenience, we let the first d_1 columns of the Matrix Φ correspond to state s_1, i.e., $d_1 = |A_{s_1}|$, the next d_2 columns correspond to actions in state s_2 and so on till finally the columns from $d - d_D + 1$ to column d correspond to state s_D, i.e., we partition the columns for which the feature vector is non-zero into the different states they correspond to. Therefore, D corresponds to the number of states for which there is at least one action for which the feature vector corresponding to the state action pair is non-zero. This means that the states $s_1, s_2, \cdots, s_D$ are the only states for which there is at least one action with a nonzero feature vector, and therefore $\widetilde{\Delta}_s$ is not a singleton set for these states. For all other states $s \in \mathcal{S}\backslash\{s_1, s_2, \cdots, s_D\}$, we have that $\widetilde{\Delta}_s$ is a singleton set containing the uniform distribution. We will also separate the parameters as follows: θ_{s_1} denotes the first d_1 elements of θ, θ_{s_2} denotes the next d_2 elements of θ and so on. Similarly for ν.

Now, the algorithm used to solve the Markov game in this function approximation setting, is similar to the tabular setting, except that we only have to run the inner iteration on the states $s_i,\ i = \{1, 2, \cdots, D\}$. We describe the algorithm in detail in Algorithm 7 in §3.9.5.

Theorem 3.9.22. Let $Q_\tau^\star$ be the in-class NE (Definition 3.6.1) Q-value defined in Equation (3.6.5) of the regularized problem, under the log-linear parametrization satisfying Assumption 3.9.21. Note that we have existence of the in-class NE from Lemma 3.9.25. Choose the stepsize $\eta = \frac{1-\gamma}{2(1+\tau(\log n+1-\gamma))}$ for the inner loop in Algorithm 7.Then, after running Algorithm 7 for

$$\begin{aligned} T_{inner} &= \mathcal{O}\left(\frac{1}{\eta\tau}\left(\log\frac{1}{\epsilon} + \log\frac{1}{1-\gamma} + \log\log n + \log\frac{1}{\eta}\right)\right), \\ T_{outer} &= \mathcal{O}\left(\frac{1}{1-\gamma}\left(\log\frac{D}{\epsilon} + \log\left(\frac{8}{\tau}\left(1 + C^2\|Q^\star\|_F^2\right)\right) + \log\frac{1+\tau\log n}{1-\gamma}\right)\right), \end{aligned} \tag{3.9.194}$$

iterations, we have $\|Q_T - Q_\tau^\star\|_\infty \le \epsilon$ and $\max\{\|\theta_T - \theta^\star\|, \|\nu_T - \nu^\star\|\} \le \epsilon$ where (Q_T, θ_T, ν_T) is the output of Algorithm 7 after T iterations and $(\theta^\star, \nu^\star)$ are defined in Equation (3.9.195).

Remark 3.9.23. Note that Theorem 3.9.22 provides the convergence rate for a two player Markov game under the function approximation setting. This covers the tabular case by setting the feature matrix, Φ (See Equation (3.4.2) in §3.4), to be equal to the identity matrix. In particular, making this substitution recovers the results of Theorem 3.6.3 .

Remark 3.9.24. We remark that Theorem 3.9.22, to the best of our knowledge, provides the first symmetric algorithm with convergence rate guarantees for Markov games under the function approximation setting. The only other existing result in this setting is [151], where the update is asymmetric, and one of the players performs multiple updates while the other player updates once. An asymmetric update-rule also appears in [32], without function approximation. Our results also improve over [131, 23] by generalizing the results to the case of certain function approximation, as well as showing parameter convergence.

Proof of Theorem 3.9.22

Algorithm 7 Optimistic NPG for Markov Games with Function Approximation

```
Initialize: $Q_0 = 0$
for $t = 1, 2, \cdots, T_{outer}$ do
    Let $Q_t(s,a,b) := r(s,a,b) + \gamma \mathbb{E}_{s' \sim \mathbb{P}(\cdot|s,a,b)}[-f_\tau(Q_t(s'); g_{\theta_{T_{inner}}}(\cdot \mid s'), h_{\nu_{T_{inner}}}(\cdot \mid s'))]$
    for $i = 1, 2, \cdots, D$ do
        Solve $\min_{\theta \in \mathbb{R}^{|A_{s_i}|}} \max_{\nu \in \mathbb{R}^{|B_{s_i}|}} f_\tau(Q(s_i); g_\theta, h_\nu)$ by running the Optimistic NPG algorithm (Algorithm 2) with feature matrix $\Phi_{s_i}$ for $T_{inner}$ iterations and return the last iterates $(\theta_{s_i}^{T_{inner}}, \nu_{s_i}^{T_{inner}})$.
    end for
    Set $(\theta_{T_{inner}}, \nu_{T_{inner}}) = \big([\theta_{s_1}^{T_{inner}}, \theta_{s_2}^{T_{inner}}, \cdots, \theta_{s_D}^{T_{inner}}], [\nu_{s_1}^{T_{inner}}, \nu_{s_2}^{T_{inner}}, \cdots, \nu_{s_D}^{T_{inner}}]\big)$.
end for
```

In order to characterize the point where the parameter converges to, we define:

$$\theta^\star = [(\theta^\star_{s_1})^\top, (\theta^\star_{s_2})^\top, \cdots, (\theta^\star_{s_D})^\top]^\top, \ \nu^\star = [(\nu^\star_{s_1})^\top, (\nu^\star_{s_2})^\top, \cdots, (\nu^\star_{s_D})^\top]^\top, \tag{3.9.195}$$

where

$$\theta^\star_{s_i} = -[I_{|A_{s_i}|}|0]\frac{Q^\star_\tau(s_i) h_{\nu^\star}(\cdot \mid s_i)}{\tau} \in \mathbb{R}^{A_{s_i}}, \qquad \nu^\star_{s_i} = [I_{|A_{s_i}|}|0]\frac{Q^\star_\tau(s_i)^\top g_{\theta^\star}(\cdot \mid s_i)}{\tau} \in \mathbb{R}^{A_{s_i}},$$

and $Q^\star_\tau$ is the in-class NE Q-value (see Definition 3.6.1). The existence of this in-class NE follows from Lemma 3.9.25.

The in-class Nash equilibrium $(\theta^\star, \nu^\star)$ under the function approximation setting satisfies:

$$V^{\theta^\star,\nu}(s) \geq V^{\theta^\star,\nu^\star}(s) \geq V^{\theta,\nu^\star}(s) \qquad \forall \theta, \nu \in \mathbb{R}^d, \ \forall s \in \mathcal{S}. \tag{3.9.196}$$

Note that this is equivalent to (from Lemma 3.4.4):

$$V^{g_{\theta^\star},h_\nu}(s) \geq V^{g_{\theta^\star},h_{\nu^\star}}(s) \geq V^{g_\theta,h_{\nu^\star}}(s) \qquad \forall g_\theta, h_\nu \in \widetilde{\Delta}, \ \forall s \in \mathcal{S}. \tag{3.9.197}$$

We denote the NE V (and Q) as $V^\star$ (and $Q^\star$) (We use this notation for the general regularized version with $\tau \geq 0$, i.e., we do not explicitly state the dependence on τ), i.e.,

$$Q^\star(s,a,b) = r(s,a,b) + \gamma \mathbb{E}_{s' \sim \mathbb{P}(\cdot|s,a,b)}[V^\star(s')]. \tag{3.9.198}$$

Next, we define the soft-Bellman operator as [10]:

$$\mathcal{T}_\tau(Q)(s,a,b) := r(s,a,b) + \gamma \mathbb{E}_{s' \sim \mathbb{P}(\cdot | s,a,b)}[- \min_{\theta \in \mathbb{R}^{|A_s|}} \max_{\nu \in \mathbb{R}^{|A_s|}} f_\tau(Q(s'); g_\theta(\cdot \,|\, s'), h_\nu(\cdot \,|\, s'))], \tag{3.9.199}$$

where

$$f_\tau(Q(s); g_\theta(\cdot \,|\, s), h_\nu(\cdot \,|\, s) := -\, g_\theta(\cdot \,|\, s)^\top Q(s) h_\nu(\cdot \,|\, s) - \tau \mathcal{H}(g_\theta(\cdot \,|\, s)) + \tau \mathcal{H}(h_\nu(\cdot \,|\, s)), \tag{3.9.200}$$

and $Q(s)$ is the Q-value matrix at state s. Let θ_s, ν_s be the NE parameters at state s. Note that by NE we mean the solution to the min-max problem (which is the in-class NE in the function approximation setting). Then the concatenation $\theta = [\theta_{s_1}^\top, \theta_{s_2}^\top, \cdots, \theta_{s_D}^\top]^\top$ (and similarly for ν) denotes the parameters.

Note that the inner min max problem is equivalent to (from §3.4):

$$\min_{g_\theta(\cdot|s') \in \widetilde{\Delta}_{s'}} \max_{h_\nu(\cdot|s') \in \widetilde{\Delta}_{s'}} \quad f_\tau(Q(s'); g_\theta(\cdot \,|\, s'), h_\nu(\cdot \,|\, s')). \tag{3.9.201}$$

Consider the value iteration:

$$Q_{t+1} = \mathcal{T}_\tau(Q_t). \tag{3.9.202}$$

We have $\|Q_t - Q^\star\|_\infty \leq \gamma^t \|Q_0 - Q^\star\|_\infty$, due to the non-expansiveness property of the min max operator and the contracting factor $\gamma < 1$. Note from Lemma 3.9.25, this fixed point in fact corresponds to the in-class NE $Q-$value matrix of the regularized Markov game.

The inner problem, which solves the saddle-point problem, is solved for each state with the input feature matrix Φ_s using Algorithm 2. The iteration complexity follows from a similar analysis to the tabular case. This completes the proof. □

[10] Note that we use the structure of the feature matrix Φ, along with Theorem 3.4.5, to show that $\min_{g_\theta(\cdot \,|\, s) \in \widetilde{\Delta}} \max_{h_\nu(\cdot \,|\, s) \in \widetilde{\Delta}}$ is equivalent to $\min_{\theta_s \in \mathbb{R}^{|A_s|}} \max_{\nu_s \in \mathbb{R}^{|A_s|}}$

Proof of Lemma 3.6.2

Consider the operator [11]

$$\mathcal{T}(V)(s) := \max_{\theta_s \in \mathbb{R}^{|A|}} \min_{\nu_s \in \mathbb{R}^{|A|}} \mathbb{E}_{a \sim g_\theta(\cdot|s), b \sim h_\nu(\cdot|s)} \left[r(s,a,b) + \gamma \mathbb{E}_{s' \sim \mathbb{P}(\cdot|s,a,b)}[V(s')] \right]. \tag{3.9.203}$$

The proof here is for the regularized game, i.e., with $\tau \geq 0$[12].Note that this is the operator for the value function V corresponding to the Q-value operator in Equation (3.9.199) with $\tau = 0$. We have, from the nonexpansive property of the min-max operator (see for example [47]):

$$\|\mathcal{T}(V_1) - \mathcal{T}(V_2)\|_\infty \leq \gamma \|V_1 - V_2\|_\infty, \tag{3.9.204}$$

which shows that this is a contracting operator and therefore has a unique fixed point by the Banach Fixed Point theorem. We show that this fixed point will lead to the in-class Nash equilibrium policy defined in Definition 3.6.1. Let the fixed point be denoted by $V^\star$, and let $(\theta^\star, \nu^\star)$ be the maxmin policy parameters in Equation (3.9.203) when plugging in $V^\star$. Note that $\theta^\star = (\theta^\star_s)_{s=1}^{|S|}$ [13] and similarly $\nu^\star$. We will show that $(\theta^\star, \nu^\star)$ is in fact the NE policy parameters. We have:

$$\begin{aligned} V^\star(s) &= \mathbb{E}_{a \sim g_{\theta^\star}(\cdot|s), b \sim h_{\nu^\star}(\cdot|s)} \left[r(s,a,b) + \gamma \mathbb{E}_{s' \sim \mathbb{P}(\cdot|s,a,b)}[V^\star(s')] \right] \\ &= \min_{\nu_s \in \mathbb{R}^{|A|}} \mathbb{E}_{a \sim g_{\theta^\star}(\cdot|s), b \sim h_{\nu}(\cdot|s)} \left[r(s,a,b) + \gamma \mathbb{E}_{s' \sim \mathbb{P}(\cdot|s,a,b)}[V^\star(s')] \right] \\ &\leq \mathbb{E}_{a \sim g_{\theta^\star}(\cdot|s), b \sim h_{\nu}(\cdot|s)} \left[r(s,a,b) + \gamma \mathbb{E}_{s' \sim \mathbb{P}(\cdot|s,a,b)}[V^\star(s')] \right] \qquad \forall\, \nu \in \mathbb{R}^d. \end{aligned} \tag{3.9.205}$$

Now applying Lemma 4.3.3 in [47] to Inequality (3.9.205) for all states s, we have:

$$V^\star \leq V^{\theta^\star, \nu} \qquad \forall\, \nu \in \mathbb{R}^d, \tag{3.9.206}$$

[11] Note that we here we use the structure of the tabular parametrization, by noting that $\max_{g_\theta(\cdot\,|\,s) \in \Delta} \min_{h_\nu(\cdot\,|\,s) \in \Delta}$ is equivalent to $\max_{\theta_s \in \mathbb{R}^{|A|}} \min_{\nu_s \in \mathbb{R}^{|A|}}$ using Theorem 3.3.2.

[12] We suppress the dependency on τ in the notation, by dropping the subscript τ for Q, V, and the operator $\mathcal{T}$.

[13] Since we are in the tabular setting, each element $\theta^\star_s$ is of length $|A|$.

where $d = |\mathcal{S}| \times |A|$. Applying the same inequality for $\nu^\star$, we have:

$$V^{\theta^\star,\nu} \geq V^\star \geq V^{\theta,\nu^\star} \qquad \forall\ \theta, \nu \in \mathbb{R}^d. \tag{3.9.207}$$

Furthermore, since we have:

$$V^\star = \mathbb{E}_{a\sim g_{\theta^\star}(\cdot|s), b\sim h_{\nu^\star}(\cdot|s)} \left[r(s,a,b) + \gamma \mathbb{E}_{s'\sim \mathbb{P}(\cdot|s,a,b)}[V^\star(s')] \right]. \tag{3.9.208}$$

Lemma 4.3.3 in [47] gives us that $V^\star = V^{\theta^\star,\nu^\star}$. Combining this with Inequality (3.9.207), we have:

$$V^{\theta^\star,\nu} \geq V^{\theta^\star,\nu^\star} \geq V^{\theta,\nu^\star} \qquad \forall \theta, \nu \in \mathbb{R}^d, \tag{3.9.209}$$

which shows that $(\theta^\star, \nu^\star)$ is the required NE. Finally, using Theorem 4.3.2 (iii) in [47], we can find the NE $(\theta^\star, \nu^\star)$ for each state $s \in \mathcal{S}$ by solving the following matrix game (for which the solution is guaranteed to exist, from Theorem 3.3.2) :

$$\max_{\theta_s \in \mathbb{R}^{|A|}} \min_{\nu_s \in \mathbb{R}^{|A|}} \mathbb{E}_{a\sim g_\theta(\cdot|s), b\sim h_\nu(\cdot|s)} \left[r(s,a,b) + \gamma \mathbb{E}_{s'\sim \mathbb{P}(\cdot|s,a,b)}[V^\star(s')] \right].$$

This completes the proof.

□

Lemma 3.9.25 (Existence of parameterized/in-class NE under Linear FA)**.** Under policy parameterization (3.6.2) with log linear policy parametrization and Assumption 3.9.21, the in-class NE defined in Definition 3.6.1 exists.

Proof. The proof follows exactly along the lines of Lemma 3.6.2, along with the fact that the matrix game under linear function approximation and Assumption 3.9.21 has a solution, as shown in Theorem 3.4.5. □

Chapter 4

Generalization of Minimax Learners

4.1 Introduction

In this chapter, we move our attention from the problem of optimization, to the problem of generalization. Stochastic minimax optimization, a classical and fundamental problem in operations research and game theory, involves solving the following problem:

$$\min_{w\in W}\max_{\theta\in\Theta} E_{z\sim P_z}[f(w,\theta;z)].$$

Such minimax formulations have recently received increasing attention in machine learning. Most existing works (including the first two chapters) have focused on the *optimization* aspect of the problem, i.e., studying the rates of convergence, robustness, and optimality of algorithms for solving an empirical version of the problem where it approximates the expectation by an average over a sampled dataset, in various minimax settings including convex-concave [95, 91], nonconvex-concave [76, 106], and certain special nonconvex-nonconcave [99, 135] problems.

However, the optimization aspect is not sufficient to achieve the success of stochastic minimax optimization in machine learning. In particular, as in classical supervised learning, which is usually studied as a *minimization* problem [63], the out-of-sample *generalization* performance is a key metric for evaluating the learned models. The study of generalization guarantees in minimax optimization (and related machine learning problems) has not received significant attention until recently [4, 46, 138, 73, 43, 143]. Specifically, existing works along this line have investigated two types of generalization guarantees: *uniform* convergence generalization bounds, and *algorithm-dependent* generalization bounds. The former is more general and irrespective of the optimization algorithms being used, while the latter is usually finer and really explains what happens in practice, when optimization algorithms play an indispensable role. In fact, the former might not be able to explain generalization performance in deep learning, e.g., these bounds can increase with the training dataset size and easily become vacuous in practice [92], making the latter a more favorable metric for understanding the success of minimax optimization in machine learning.

Algorithm-dependent generalization for minimax optimization has been studied recently in [43, 73, 134, 137]. These papers build on the algorithmic stability framework developed in [16], which are further investigated in [61]. In particular, these works have studied *primal risk* and/or (variants of) *primal-dual risk* under different convexity and smoothness assumptions of the objective. Primal risk (see formal definition in §4.2) is a natural extension of the definition of risk from minimization problems. Primal-dual risk, on the other hand, is defined similarly but based on the duality gap of the solution. It is know that it is well-defined and can be optimized to zero only when the global saddle-point exists (i.e., min and max can be interchanged). Based on these metrics, [43, 73] compare the performance of specific algorithms, e.g., gradient descent-ascent (GDA) and gradient descent-max (GDMax).

Although these metrics are natural extensions of generalization metrics from the *minimization* setting, they might not be the most suitable ones for studying generalization in stochastic *minimax* optimization, especially in the *nonconvex* settings that is pervasive in machine/deep learning applications, where the global saddle-point might not exist. In particular, we are interested in the following fundamental question:

What is a good metric to study generalization of minimax learners[1]*?*

In this final chapter, we answer this question, by first identifying the inadequacies of the existing metric, and proposing a new metric, the *primal gap* that overcomes these inadequacies. We then provide generalization error bounds for the newly proposed metric, and discuss how it captures information not included in the other existing metrics.

[1] We use *learner* and *learning algorithm* interchangeably.

Reference	Assumption	Metric	Rate
[43]	NC-μ-SC	PR	$L\sqrt{\kappa^2+1}\epsilon$
[73]	NC-μ-SC	PR	$L(1+\kappa)\epsilon$
[73]	μ-SC-SC	PD	$\sqrt{2}L(1+\kappa)\epsilon$
This work (Theorem 3)	NC-C	PG	$\sqrt{4L\ell C_p^2 \cdot \sqrt{\epsilon} + \epsilon L} + 4L_\theta^* C_e/\sqrt{n}$
This work (Lemma 1)	NC-C	PR	$\sqrt{4L\ell C_p^2 \cdot \sqrt{\epsilon} + \epsilon L}$
This work (Theorem 7)	C-C	PD	$\left(\sqrt{4L\ell C_p^2} + \sqrt{4L\ell (C_p^w)^2}\right)\sqrt{\epsilon} + 2\epsilon L$

Table 4.1: Generalization bounds for ϵ-stable algorithms. PR stands for Primal Risk, PD stands for the primal-dual risk and PG stands for the primal gap. NC-μ-SC stands for nonconvex-μ strongly concave. μ-SC-SC stands for μ strongly convex-μ strongly concave. NC-C stands for nonconvex-concave. C-C convex-concave. L is the Lipschitz constant of the function f. κ stands for the condition number L/μ. The constants in the theorems have been defined in the appropriate sections. Note that there are other results in [43, 73] for cases where the expectation and max operator can be interchanged. This case is almost identical to the minimization problem and we thus do not include it in the table.

Contributions. First, we introduce an example through which we identify the inadequacies of *primal risk*, a well-studied metric for generalization in stochastic minimax optimization, in capturing the generalization behavior of *nonconvex-concave* minimax problems. Second, to address the issue, we propose a new metric – the *primal gap*, which provably avoids the issue in the example, and derive its generalization error bounds. Next, we leverage this new metric to compare the generalization behavior of GDA and GDMax, two popular algorithms for minimax optimization and GAN training, and answer the question of *when does GDA generalize better than GDMax?* Moreover, we also address two open questions in the literature: establishing generalization error bounds for primal risk and primal-dual risk without strong concavity or assuming that the maximization and expectation can be interchanged, while at least one of these assumptions was needed in the literature [43, 73, 134, 137]. Finally, under certain assumptions of the max learner, our results also generalize to the nonconvex-nonconcave setting.

4.1.1 Related work

Algorithms for minimax optimization. There is a vast literature on algorithms for minimax optimization. The most popular algorithms include the Extragradient (EG), the

Optimistic Gradient Descent Ascent (OGDA) and the Gradient Descent Ascent and their variants. The EG algorithms introduced in [70], has been analyzed in several papers including [91, 89, 90, 53] for (strongly)convex-(strongly)concave problems. Another popular algorithm is OGDA introduced in [103] and has been analyzed in several recent works including [33, 64, 52]. Once again, all these works focus on the (strongly)convex-(strongly)concave setting. Stochastic versions of these algorithms in similar settings have also been analyzed in several papers including [95, 64, 42]. A few papers including [76, 142, 65, 150, 102, 69, 142] analyze gradient based algorithms in the nonconvex-(strongly)concave cases. Some papers including [106, 136, 101, 55] analyze special cases of nonconvex-nonconcave (like nonconvex-PL) for algorithms like GDA and its variants. However, in this chapter, we are interested in the generalization performance of these algorithms. We summarize below the most related literature that studies the generalization behavior in minimax optimization problems.

Algorithm-independent generalization. Specific to the machine learning problems of GAN and adversarial training, there have been several papers studying the uniform convergence generalization bounds. [4] establish a uniform convergence generalization bound which depends on the number of discriminator parameters. [133] connect the stability-based theory to differential privacy ([117]) in GANs and numerically study the generalization behavior in GANs. [147, 8] analyze the Rademacher complexity of the players to show the uniform convergence bounds for GANs. In the simpler Gaussian setting, [46] and [113] derive bounds for GANs and adversarial training, respectively. The uniform convergence bounds for adversarial training have also been studied under several statistical learning frameworks, e.g., PAC-Bayes [44], Rademacher complexity [138], margin-based [132], and VC analysis [6]. Recently, [143] investigate the generalization of empirical saddle point (ESP) solution in strongly-convex-concave problems using a stability-based approach. Note that these results are not specific to the optimization algorithms being used.

Algorithm-dependent generalization. Algorithm specific generalization bounds for minimax optimization have attracted increasing attention. Based on the algorithmic stability framework in [16], [43] have established generalization bounds of standard gradient descent-ascent and proximal point algorithms under the convex-concave setting, and those of stochastic GDA and GDMax under the nonconvex-strongly concave setting. Concurrently, [73] derive high-probability generalization bounds for both convex-concave and weakly convex-weakly concave settings, with possibly nonsmooth objectives, also through the lens of algorithmic stability. Both works hinged on the metrics of *primal risk* and *primal-dual risk*. As shown in the present work, the former is not necessarily suitable to characterize the generalization behavior of minimax optimization, while the latter is known to be appropriate only when the saddle point exists, which is usually not the case in the nonconvex settings that are common in machine learning. Following this line of work, [134] provide generalization bounds specifically for adversarial training, which is essentially the primal risk, also using the algorithmic stability framework. Recently, [137] study the generalization of stochastic GDA under differential privacy constraints.

4.2 Preliminaries

4.2.1 Problem formulation

In this chapter, we consider the following (stochastic) minimax problem:

$$\min_{w \in W} \max_{\theta \in \Theta} E_{z \sim P_z} f(w, \theta; z). \tag{4.2.1}$$

We make the following assumption on the sets W and Θ throughout the chapter.

Assumption 1. W and Θ are convex, closed sets, and we further assume that W is compact with $\|w\| \leq M(W), \forall w \in W$. Here $M(W)$ is a constant dependent on the set W.

Let $r(w, \theta) = E_{z \sim P_z} f(w, \theta; z)$. For a training dataset $S = \{z_1, \cdots, z_n\}$ with n i.i.d. variables drawn from P_z, we define $r_S(w, \theta) = \frac{1}{n} \sum_{i=1}^{n} f(w, \theta; z_i)$. Next, we define the following quantity:

Definition 1 (Primal risk (empirical/population))**. Primal population risk** is given by[2]

$$r(w) = \max_{\theta \in \Theta} E_{z \sim P_z} f(w, \theta; z),$$

and the **primal empirical risk** is given by:

$$r_S(w) = \max_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^{n} f(w, \theta; z_i).$$

Throughout this chapter we use (w_S, θ_S) to denote a solution of the minimax problem: $\min_{w \in W} \max_{\theta \in \Theta} r_S(w, \theta)$. Notice that (w_S, θ_S) need not be a global saddle-point of r_S. Furthermore, we use (w^*, θ^*) to denote a solution of $\min_{w \in W} \max_{\theta \in \Theta} r(w, \theta)$. Once again, notice that (w^*, θ^*) may not be a saddle point of r.

The goal in Problem (4.2.1) is to minimize the primal population risk $r(w)$. Note that this function can be decomposed as

$$r(w) = r_S(w) + (r(w) - r_S(w)). \tag{4.2.2}$$

In practice, we only have access to $r_S(w, \theta)$, and our goal is to design algorithms for minimizing $r(w)$ using dataset S. Suppose A is a learning algorithm initialized at $(w, \theta) = (0, 0)$. We define (w_S^A, θ_S^A) to be the output of Algorithm A using dataset S.

From Equation (4.2.2), it is clear if we ensure $r_S(w_S^A)$ as well as $r(w_S^A) - r_S(w_S^A)$ are small, this would guarantee that $r(w_S^A)$ is small, which is the goal of Problem (4.2.1). Note that we can always ensure that $r_S(w_S^A)$ is small by using a good optimization Algorithm A (if the problem is tractable). The main goal in the study of generalization is therefore to estimate the generalization error of the primal risk, as defined below.

Definition 2. The generalization error for the primal risk is defined as:

$$\zeta_{gen}^P(A) = E_S E_A [r(w_S^A) - r_S(w_S^A)]. \tag{4.2.3}$$

[2] Note that we slightly abuse the notation here by allowing r and r_S to have inputs that can be both w and (w, θ). The distinction will be clear from context.

Here the expectations are taken over the randomness in the dataset S, as well as any randomness used in the Algorithm A.

This metric has been used to study generalization in stochastic minimization problems, i.e., when the maximization set Θ is a singleton, as well as several recent works in stochastic minimax optimization (see [61, 43, 73]).

We are interested in the question of when the solution to the empirical problem w_S^A has good *generalization behavior*, i.e., when $E[r(w_S^A) - \min_{w \in W} r(w)]$ is small – w_S^A is an approximate minimizer of the primal population risk r. In the next subsection, we briefly describe why the generalization error of the primal risk $\zeta_{gen}^P(A)$ is a good measure to study the generalization behavior in minimization problems.

$\zeta_{gen}^P(A)$ for minimization problems

Consider a stochastic optimization problem of the form

$$\min_{w \in W} \; E_{z \sim P_z}[g(w; z)]. \tag{4.2.4}$$

We define the (minimization) primal risk (population and empirical version respectively) as: $r(w) = E_{z \sim P_z} g(w; z)$, and $r_S(w) = \frac{1}{n}\sum_{i=1}^n g(w; z_i)$. The generalization error $\zeta_{gen}^{P,min}(A)$ for the (minimization) primal risk is the same as in Definition 2 using the (minimization) primal risk.

Assume that the generalization error of the primal risk for an Algorithm A is small, say $\zeta_{gen}^{P,min}(A) \leq \epsilon$. This implies that (from Definition 2): $E[r(w_S^A)] \leq E[r_S(w_S^A)] + \epsilon$. Note that the expectation is with respect to S and A. Now, in order to show that w_S^A has good generalization behavior, we first see that:

$$E[r(w_S^A) - \min_{w \in W} r(w)] \leq E[r_S(w_S^A)] + \epsilon - \min_{w \in W} r(w). \tag{4.2.5}$$

However, note that for minimization problems, since $E[r_S] = r$, we have that[3] $\min_{w \in W} r(w) \geq$

[3] Here we use the fact that $E_z[\min_x f(x, z)] \leq \min_x E_z[f(x, z)]$.

$E[\min_{w\in W} r_S(w)]$, which gives us:

$$\begin{aligned}&E[r(w_S^A) - \min_{w\in W} r(w)] \\ &\qquad \le E[r_S(w_S^A)] + \epsilon - E[\min_{w\in W} r_S(w)] = E[r_S(w_S^A) - \min_{w\in W} r_S(w)] + \epsilon = \epsilon.\end{aligned}$$

Therefore, for minimization problems, if the generalization error for primal risk is small, the solution to the empirical risk minimization problem has good generalization behavior. Next, we highlight some results in the literature which discusses generalization error bounds of the primal risk. These results depend on the concept of algorithmic stability we use later.

4.2.2 Stability of algorithms

Stability analysis is a powerful tool to analyze the generalization behavior of algorithms (see [16]). In this section, we will review some definitions and theoretical results about stability bounds existing in the current literature. More specifically, in this chapter, we adopt the following definition of stability:

Definition 3 (ϵ-stable Algorithm)**.** Suppose that A is a randomized algorithm for solving the stochastic minimax problem. We define (w_S^A, θ_S^A) as the output of Algorithm A using dataset S. We say S and S' are neighboring dataset if they defer only in one sample. An Algorithm A is defined to be ϵ-stable if $E_A\|w_S^A - w_{S'}^A\| \le \epsilon$ and $E_A\|\theta_S^A - \theta_{S'}^A\| \le \epsilon$ for any neighboring datasets S and S'.

[61] gives the following basic result for the generalization error of $r_S(w)$.

Theorem 1 ([61])**.** Consider the (stochastic) minimization problem defined in 4.2.4. Suppose $g(\cdot; z)$ is $\bar{L}$-Lipschitz continuous, i.e., $\forall z$, it holds that $\|g(w_1; z) - g(w_2; z)\| \le \bar{L}\|w_1 - w_2\|, \forall w_1, w_2 \in W$. Then, for an ϵ-stable Algorithm A, we have $|E_S E_A[r(w_S^A) - r_S(w_S^A)]| \le \bar{L}\epsilon$.

When is primal risk a valid metric for minimax learners?

According to the above discussions for minimization problems, we know that the primal risk is a valid metric to study generalization behavior in these problems, and furthermore, the generalization error bound of the primal risk can be estimated in terms of algorithmic stability. However, Theorem 1 cannot be directly extended to analyze the generalization behavior of minimax learners because we have an additional maximization step before taking expectation.

A natural question emerges: Under what conditions does primal risk serve as a valid metric to study generalization behavior of minimax problems. One sufficient condition is when the maximization step and expectation can be interchanged, i.e., when

$$\max_{\theta\in\Theta} E_{z\sim P_z} f(w,\theta;z) = E_{z\sim P_z}[\max_{\theta\in\Theta} f(w,\theta;z)]$$

for any distribution P_z. Letting $f_{\max}(w;z) = \max_{\theta\in\Theta} f(w,\theta;z)$, we further have

$$r(w) = \max_{\theta\in\Theta} E_{z\sim P_z} f(w,\theta;z) = E_{z\sim P_z}[\max_{\theta\in\Theta} f(w,\theta;z)] = E_{z\sim P_z} f_{\max}(w;z).$$

Therefore, the minimax problem in (4.2.1) is equivalent to the (stochastic) minimization problem with loss function $f_{\max}(w;z)$. Moreover, letting $P(S)$ be the uniform distribution over the dataset $S = \{z_1, \cdots, z_n\}$, we have

$$r_S(w) = \max_{\theta\in\Theta} E_{z\sim P(S)}[f(w,\theta;z)] E_{z\sim P(S)}[\max_{\theta\in\Theta} f(w,\theta;z)] = \frac{1}{n}\sum_{i=1}^{n} f_{\max}(w;z_i).$$

Therefore, $r_S(w)$ is just the empirical primal risk corresponding to the minimization problem with loss function $f_{\max}(w;z)$. Hence, Theorem 1 can be directly used to minimax problems where the maximization and expectation can be interchanged.

Theorem 2. Suppose that $f(w,\theta;z)$ is $\bar{L}$-Lipschitz continuous with respect to w, i.e., $|f(w_1,\theta;z) - f(w_2,\theta;z)| \le \bar{L}\|w_1 - w_2\|$ for any $w_1, w_2 \in W, \theta \in \Theta$ and z. If an Algorithm A is ϵ-stable, we have

$$E_S E_A[r(w_S^A) - r_S(w_S^A)] \le \bar{L}\epsilon.$$

Proof. From the previous analysis along with Theorem 1, it suffices to show that $f_{\max}(\cdot;z)$ is

$\bar{L}$-Lipschitz continuous. In fact, we have

$$\begin{aligned} f_{\max}(w_1;z) - f_{\max}(w_2;z) &= f(w_1,\theta(w_1);z) - f(w_2,\theta(w_2);z) \\ &\leq f(w_1,\theta(w_1);z) - f(w_2,\theta(w_1);z) \leq \bar{L}\|w_1 - w_2\|, \end{aligned}$$

where $\theta(w) \in \arg\max_{\theta\in\Theta} f(w,\theta;z)$, the first inequality is because of the definition of $\theta(w)$ and the second inequality is because of the Lipschitz continuity of f with respect to w. Using the same argument, we can prove

$$f_{\max}(w_2;z) - f_{\max}(w_1;z) \leq \bar{L}\|w_1 - w_2\|.$$

Therefore, we prove the $\bar{L}$-Lipschitz continuity of $f_{\max}(\cdot;z)$ and hence finish the proof. □

By the above discussion, we know that if maximization and expectation can be interchanged, the minimax problem can be reduced to a minimization problem and hence the primal risk is a valid metric for studying the generalization behavior of minimax learners and the generalization error can be estimated using the same method as for minimization problems. In practice, the adversarial-training problems can be such an example of minimax problems.

Example 1 (Adversarial-training)**.** We consider the adversarial training problem [80]. Suppose we have loss function $g(w;z)$ for a supervised learning problem. Here z denotes the training sample and w denotes the model parameter. Due to the noise in the data or due to an adversarial attack, for any sample z, we consider an uncertainty set $B(z,\epsilon_0)$ around it. The goal is to train a model that is robust to the data with possible perturbation in the uncertainty set. Let θ_z be some adversarial sample from the set $B(z,\epsilon_0)$ and let θ be an infinite dimensional vector (functional) with the component θ_z corresponding to the sample z. Define the function $\iota_B(v)$ to be the indicator function of the set B, i.e., $\iota_B(v) = 0$ if $v \in B$ and $\iota_B(v) = \infty$ otherwise. The goal of adversarial training is to solve the following minimax problem:

$$\min_{w}\max_{\theta} \quad E_{z\sim P_z} f(w,\theta;z), \tag{4.2.6}$$

where $f(w, \theta; z) = g(w; \theta_z) + \iota_{B(z,\epsilon_0)}(\theta_z)$. For any distribution P_z over z's, we have

$$\begin{aligned}\max_{\theta} \ E_{z\sim P_z} f(w, \theta; z) &= \max_{\theta} \ E_{z\sim P_z}[g(w; \theta_z) + \iota_{B(z,\epsilon_0)}(\theta_z)] \\ &= E_{z\sim P_z}[\max_{\theta_z} \ (g(w; \theta_z) + \iota_{B(z,\epsilon_0)}(\theta_z))] \\ &= E_{z\sim P_z}[\max_{\theta} \ f(w, \theta; z)],\end{aligned}$$

where the second and the third equalities use the fact that $\theta_{z'}$ does not contribute to $f(w, \theta; z)$ if $z \neq z'$. Therefore, the expectation and maximization can be interchanged in adversarial training problems. This implies that the results of Theorem 2 can be applied and therefore primal risk is a valid metric to study the generalization behavior in such problems.

Unfortunately, maximization and expectation are not necessarily interchangeable for many minimax problems. If they are not interchangeable, it is unclear how to estimate the generalization error bound of the primal risk. In fact, whether primal risk is still a good metric for studying generalization behavior in such problems remains elusive.

In the next section, we will see how to estimate generalization error bound of primal risk for nonconvex-concave and even nonconvex-nonconcave problems. To the best of our knowledge, this is the first result which provides generalization error bounds for the primal risk without assuming the interchangeability or strong concavity of the inner maximization problems (see e.g., [73]). Furthermore, we will see that even in some simple minimax problems, the generalization error bound of the primal risk can fail to capture the generalization behavior of minimax learners. We then propose a new metric and use its generalization error to properly characterize the generalization behavior of minimax learners.

4.3 Primal Gap: A New Metric to Study Generalization

The key idea behind the success of $\zeta^P_{gen}(A)$ as a way to characterize to study generalization for minimization learners is that $E[r_S(w)] = r(w)$ for any w, which is no longer the case in the minimax case. In fact, we first show via example that a good bound for the generalization error of primal risk does not imply good generalization behavior for minimax learners.

4.3.1 Primal risk can fail for minimax learners

We provide an example where the generalization error of the primal risk is small, but the final solution to the empirical problem has poor generalization behavior. In this example, the minimizer of $r_S(w)$ is suboptimal for $r(w)$ with high probability, and $E_S[r(w_S) - r(w^*)]$ is large.

Example 2 (Analytical example)**.** Let $y \sim N(0, 1/\sqrt{n})$ be a Gaussian random variable in $\mathbb{R}$. Define the truncated Gaussian variable $z \sim P_z$ as follows: $z = y$ if $|y| < \lambda \log n/\sqrt{n}$ and $z = \lambda \log n/\sqrt{n}$ if $y \geq \lambda \log n/\sqrt{n}$. Let $f(w, \theta; z) = \frac{1}{2}w^2 - \left(\frac{1}{2n^2}\theta^2 - z\theta + 1\right)w$, where $w \in W = [0, 1]$, $\theta \in \Theta = [-\lambda n, \lambda n]$ with a sufficiently large $\lambda > 0$, and $z_i \sim P_z$ be i.i.d truncated Gaussian variables. Then, we have $r_S(w, \theta) = \frac{1}{2}w^2 - \left(\frac{1}{2n^2}\theta^2 - \frac{\sum_{i=1}^n z_i}{n}\theta + 1\right)w$, and

$$r(w, \theta) = \frac{1}{2}w^2 - \left(\frac{1}{2n^2}\theta^2 + 1\right)w. \tag{4.3.1}$$

Note that this leads to the primal population risk function: $r(w) = \frac{1}{2}w^2 - w$.

It is not hard to see that we always have $r_S(w) \geq r(w)$. Note that this means $\zeta_{gen}^P(A) \leq 0$, and thus we have a small generalization error for primal risk. However, we can prove that for large enough λ,

$$E_S[r(w_S) - r(w^*)] \geq 0.02. \tag{4.3.2}$$

This means that w_S has a constant error compared to w^* in terms of the population risk, despite that its generalization error is small. This phenomenon is due to that $\min_{w \in W} r_S(w) - \min_{w \in W} r(w) > c$ for some $c > 0$, and hence minimizing $r_S(w)$ is very different from minimizing $r(w)$.

This example shows that the generalization error of primal risk is not a good measure to study generalization in minimax learners. The main drawback is that $\min_w r_S(w)$ and $\min_w r(w)$ can be very different. We now introduce another more practical example, from GAN training, to further illustrate this point.

Example 3 (GAN-training example)**.** Suppose that we have a real distribution P_r in $\mathbb{R}^d$ which can be represented as $G^*(y)$ with $y \in \mathbb{R}^k$ drawn from a standard Gaussian distribution

P_0 and a mapping $G^* : \mathbb{R}^k \to \mathbb{R}^d$. For an arbitrary generator G, we define P_G to be the distribution of the random variable $G(y)$ with $y \sim P_0$. So our goal is to find a generator G such that $P_G = P_r$. GAN is a popular tool for solving this problem. Consider a GAN with generator G, parametrized by w and discriminator D parametrized by θ. The goal of GAN training is to find a pair of a generator G and a discriminator D that solves the minimax problem:

$$\begin{aligned}\min_G \max_D \quad & \{E_{x\sim P_r}\phi(D(x)) + E_{x\sim P_G}[\phi(1 - D(x))]\} \\ & = \min_w \max_\theta \quad \{E_{x\sim P_r}\phi(D_\theta(x)) + E_{y\sim P_0}[\phi(1 - D_\theta(G_w(y)))]\},\end{aligned}$$

where $\phi : \mathbb{R} \to \mathbb{R}$ is concave, monotonically increasing and $\phi(u) = -\infty$ for $u \leq 0$. To connect to the minimax formulation in (4.2.1), we note that $z = (x, y)$, and $P_z = P_r \times P_0$. Also, we denote

$$r(w, \theta) = E_{x\sim P_r}\phi(D_\theta(x)) + E_{y\sim P_0}[\phi(1 - D_\theta(G_w(y)))]$$

to be the population risk. We now give the empirical version of this problem. Let $S_1 = \{x_1, \cdots, x_n\}$ and $S_2 = \{y_1, \cdots, y_n\}$. Let $S = S_1 \cup S_2$ and $r_S(w, \theta) = \frac{1}{n}\left(\sum_{i=1}^n \phi(D_\theta(x_i) + \phi(1 - D_\theta(G_w(y_i))))\right)$. We assume that P_{G_w} has the same support set as P_r. Moreover, we assume that $\|w - w^*\| \leq 0.5$ and $G_w(y)$ is 1-Lipschitz w.r.t. w for any y. Here w^* denotes the parameter for which $G_{w^*} = G^*$. Then, combining Theorem B.1 in [4] and the Lipschitz continuity of $G_w(y)$ as well as $\|w - w^*\| \leq 0.5$, we have that the distance between the sets S_1 and $\{G_w(y_1), G_w(y_2), \cdots, G_w(y_n)\}$ will be larger than 0.6 with probability greater than $1 - O(n^2/e^d)$. Now, if n is only of polynomial size of d, the optimal discriminator for disjoint datasets outputs 1 on one dataset, and 0 on the other. On the other hand, when $w = w^*$, the optimal discriminator for the population problem outputs 1/2 for any sample it receives. Combining these two results, we have:

$$E_S[\min_{w\in W} r_S(w) - \min_{w\in W} r(w)] \geq (1 - \delta)\,(2\phi(1) - 2\phi(1/2))$$

which is bounded away from 0.

Note that in this example, we also have $E_S[\min_w r_S(w) - \min_w r(w)] > 0$, implying that using $\zeta^P_{gen}(A)$ might not be a good way to characterize the generalization behavior in GAN training. To address this issue, we next define a new metric, the primal gap, and use its generalization error to study the generalization of minimax learners.

4.3.2 Primal gap to the rescue

The population and empirical versions of the primal gap are defined as follows:

Definition 4 (Primal gap (empirical/population))**.** The **population primal gap** is defined as

$$\Delta(w) = r(w) - \min_{w \in W} r(w),$$

and the **empirical primal gap** is defined as

$$\Delta_S(w) = r_S(w) - \min_{w \in W} r_S(w).$$

Notice that these two primal gaps can always take 0 at $w_S \in \arg\min_{w \in W} r_S(w)$ and $w^* \in \arg\min_{w \in W} r(w)$ respectively even if the saddle point of problem (4.2) does not exist. Next, we define the expected generalization error of this primal gap as follows:

Definition 5. The generalization error for the primal gap is defined as

$$\zeta^{PG}_{gen}(A) = E_S E_A[\Delta(w^A_S) - \Delta_S(w^A_S)].$$

Remark 1. For Example 1, since the maximization and expectation can be interchanged, the minimax problem is equivalent to a minimization problem. Then we have

$$\begin{aligned} E_S[\min_w r_S(w)] &= E_S[\min_w \max_\theta E_{z \sim P_z(S)} f(w, \theta; z)] = E_S[\min_w E_{z \sim P_z(S)}[\max_\theta f(w, \theta; z)]] \\ &= E_S[\min_w E_{z \sim P_z(S)}[f_{\max}(w; z)]] \leq E_S[E_{z \sim P_z(S)}[f_{\max}(w; z)]] \end{aligned}$$

for any w. Therefore, we have $E_S[\min_w r_S(w)] \leq \min_w r(w)$. Consequently, we have $\zeta^P_{gen} \geq \zeta^{PG}_{gen}$, which means that good generalization bounds for the primal risk implies good generalization bounds for the primal gap. Therefore, if the maximization and expectation are interchangeable, primal risk is sufficient to study the generalization behavior because the generalization error of the primal risk is an upper bound of the generalization error of the primal gap in this case.

Now we provide bounds on $\zeta_{gen}^{PG}(A)$ for a stable algorithm A, and show that in Example 2, $\zeta_{gen}^{PG}(A)$ cannot be small (unlike $\zeta_{gen}^{P}(A)$).

4.3.3 Relationship between generalization and stability

We provide bounds for the generalization error of the primal gap (Definition 5) for an ϵ-stable Algorithm A. We will focus on the nonconvex-concave case where the following assumptions are made throughout the rest of the chapter.

Assumption 2. The function f in Problem (4.2.1) is nonconvex-concave, i.e., $f(w, \cdot; z)$ is a concave function for all $w \in W$ and for all z.

Next we define the notion of *capacity*, which will play a key role in the bounds we derive for $\zeta_{gen}^{PG}(A)$.

Definition 6 (Capacity)**.** For any $w \in W$ and any constraint set Θ, we define

$$\Theta(w) = \arg\max_{\theta\in\Theta} r(w, \theta) \qquad \Theta_S(w) = \arg\max_{\theta\in\Theta} r_S(w, \theta).$$

We define the capacities C_p and C_e as:

$$C_p(\Theta) = \max_{w\in W} \operatorname{dist}(0, \Theta(w)), \qquad C_e(\Theta) = \max_{S} \max_{w\in W} \operatorname{dist}(0, \Theta_S(w)),$$

where $\operatorname{dist}(p, \mathcal{S})$ denotes the distance between a point p to a set $\mathcal{S}$ in Euclidean space, i.e.,

$$\operatorname{dist}(p, \mathcal{S}) := \inf_{q\in\mathcal{S}} \|p - q\|_2.$$

For the specific constraint set in Problem (4.2.1), we succinctly denote the capacities as C_p and C_e, respectively.

The norm of the model parameter (its distance to 0) is usually viewed as the metric for the complexity of the model. In fact, the norm of the optimal solution determines the Rademacher complexity of the function class in statistical learning theory [126]. Moreover, in deep learning, minimum-norm solution of overparameterized neural networks is well-known to enjoy better generalization performance [141]. Hence, we view the capacity constant C_e and C_p as natural metrics to capture the model complexity for the best response of the max learner, i.e., the power of the maximizer, when using the empirical data set and population data respectively.

Now, we are ready to discuss the relationship between the stability bound and the generalization error of algorithms in nonconvex-concave minimax problems. All proofs have been deferred to the appendix. We make the following assumptions throughout the chapter:

Assumption 3. The gradient of f is ℓ-Lipschitz-continuous for all z, i.e., for all z

$$\|\nabla f(w_1,\theta_1;z)-\nabla f(w_2,\theta_2;z)\| \le \ell(\|w_1-w_2\|+\|\theta_1-\theta_2\|),\ \forall w_1,w_2\in W,\ \forall \theta_1,\theta_2\in\Theta.$$

Moreover, fixing $w\in W$, the partial gradient $\nabla_\theta f(w,\cdot;z)$ is $\ell_{\theta\theta}$-Lipschitz continuous with respect to θ for all z, i.e., $\|\nabla_\theta f(w,\theta_1;z)-\nabla_\theta f(w,\theta_2;z)\|\le \ell_{\theta\theta}\|\theta_1-\theta_2\|, \forall w\in W, \quad \forall \theta_1,\theta_2\in\Theta$.

Assumption 4. For any $\Theta_1\subseteq\Theta$, we assume that f is $L(\Theta_1)$-Lipschitz-continuous with respect to $w\in W, \theta\in\Theta_1$ for all z, i.e., $\|f(w_1,\theta_1;z)-f(w_2,\theta_2;z)\|\le L(\Theta_1)(\|w_1-w_2\|+\|\theta_1-\theta_2\|), \quad \forall w_1,w_2\in W,\ \forall\theta_1,\theta_2\in\Theta_1$, and the gradient $\nabla f(w,\theta;z)$ is uniformly bounded as $\|\nabla_{w,\theta} f(w,\theta;z)\|\le L(\Theta_1)$ for all z and $w\in W, \theta\in\Theta_1$. Moreover, $f(w^*,\cdot;z)$ is L_θ^*-Lipschitz continuous with respect to θ where $w^*\in\arg\min_{w\in W} r(w)$. We also define $L:=L(B(0,2C_p+1)\cap\Theta)$ and $L_r:=L(B(0,r)\cap\Theta)$, where $B(v,r)$ denotes the l_2-ball with radius r centered at v.

Note that we can decompose the generalization error of the primal gap as follows:

$$\begin{aligned}\zeta_{gen}^{PG}(A) &:= E_S E_A[\Delta(w_S^A)-\Delta_S(w_S^A)]\\ &= E_S E_A[r(w_S^A)-r_S(w_S^A)] + E_S[\min_{w\in W} r_S(w) - \min_{w\in W} r(w)]\\ &= \zeta_{gen}^{P}(A) + E_S\big[\min_{w\in W} r_S(w)-\min_{w\in W} r(w)\big].\end{aligned}$$

Next, we provide a bound on the generalization error for the primal risk $\zeta^{P}_{gen}(A)$. To the best of our knowledge, this is the first bound for $\zeta^{P}_{gen}(A)$ in the nonconvex-concave (without strong concavity) setting.

Lemma 1. The generalization error of the primal risk of an ϵ-stable Algorithm A for a minimax problem with concave maximization problem can be bounded by $\zeta^{P}_{gen}(A) \leq \sqrt{4L\ell C_p^2} \cdot \sqrt{\epsilon} + \epsilon L$.

We show that this dependence on ϵ is tight in the Appendix Section 4.6.3.

Since we already have the generalization error for the primal risk $E_S E_A[r(w_S^A) - r_S(w_S^A)]$ from Lemma 1, we only need to estimate

$$\begin{aligned} E_S E_A \big[\min_{w \in W} r_S(w) - \min_{w \in W} r(w) \big] & \\ = E_S \big[\min_{w \in W} r_S(w) - \min_{w \in W} r(w) \big] & \qquad \text{[Primal Min Error]}. \end{aligned} \tag{4.3.3}$$

The following theorem gives the generalization bound of the primal gap using the upper bound from Lemma 1 and bounding the Primal Min Error in Equation (4.3.3).

Theorem 3. Suppose Algorithm A is ϵ-stable. The generalization error bound of the primal gap is given by

$$\zeta^{PG}_{gen}(A) \leq \sqrt{4L\ell C_p^2} \cdot \sqrt{\epsilon} + \epsilon L + 4L_\theta^* C_e / \sqrt{n}.$$

The first term in the bound above is from the generalization bound of the primal risk, as shown in Lemma 1. Note that the bound in Lemma 1 only involves C_p, as the key in the analysis is to upper-bound the population risk $r(w_S^A)$, which requires bounding the power of the maximizer using the population capacity C_p. This reflects the intuition that the power of the maximizer should affect the generalization behavior of minimax learners, and the stronger the maximizer is, the harder for the learner to generalize. On the other hand, the bound in Theorem 3 additionally involve C_e, the empirical capacity. Technically, C_e (instead of C_p) appears since we need to bound $\min_w r_S(w)$ (defined on the empirical dataset) in the Primal Min Error term in (4.3.3). We show the tightness of this bound in Section 4.6.3. The appearance of C_e reflects the intuition that the difference between the maximizers of the empirical and population risks should make a difference in characterizing the generalization of minimax learners. This intuition cannot be captured by the generalization error of the primal risk, as in Lemma 1. Note that in the minimization case, the Primal Min Error can be upper-bounded directly by zero, and such a distinction disappears, making primal risk a valid metric.

4.3.4 Revisiting Example 2

Recall Example 2 in Section 4.3.1. In this example, we have that the primal risk has a small generalization error, but the solution w_S does not generalize well. In particular, as shown in the appendix (Proposition 4), we have

$$E_S[\min_{w\in W} r_S(w) - \min_{w\in W} r(w)] \geq 0.005. \tag{4.3.4}$$

On the other hand, it is easy to compute that $L_\theta^* = \lambda \log n/\sqrt{n}$ and $C_e = \lambda n$. Therefore, by Theorem 3, we have an upper bound for the Primal Min Error (see Equation (4.3.3)): $E_S[\min_{w\in W} r_S(w) - \min_{w\in W} r(w)] \leq 4L_\theta^* C_e/\sqrt{n} = 4\log n$, which is tight up to a log factor according to (4.3.4). Therefore, the primal gap has a constant generalization error which is consistent with the observation that the solution to the empirical problem does not have good generalization behavior.

4.3.5 Nonconvex-nonconcave case

In this section, we extend our results to the nonconvex-nonconcave setting. We will show that under certain assumptions on the inner maximization problem, we can derive generalization error bounds for the primal risk and primal gap in terms of algorithmic stability.

We make the following assumptions on the inner maximization problem:

Assumption 5. For any $\gamma > 0$, there exists an algorithm which outputs $\theta_P^\gamma(w)$, for the inner maximization problem $\max_{\theta\in\Theta} r(w,\theta)$, satisfying the following conditions:

1. $r(w) - r(w, \theta_P^\gamma(w)) \leq \gamma$.

2. $\|\theta_P^\gamma(w) - \theta_P^\gamma(w')\| \leq \frac{\lambda_p}{\gamma}\|w - w'\|$ with some constant $\lambda_p > 0$ for all $w, w' \in W$.

Assumption 6. For any $\gamma > 0$, there exists an algorithm which outputs $\theta_E^\gamma(S)$, for the inner maximization problem $\max_{\theta\in\Theta} r_S(w^*, \theta)$, satisfying the following conditions:

1. $r_S(w^*) - r_S(w^*, \theta_E^\gamma(S)) \leq \gamma$.

2. For any neighboring dataset S, S', we have $\|\theta_E^\gamma(S) - \theta_E^\gamma(S')\| \leq \frac{\lambda_e}{n\gamma}$ with some constant $\lambda_e > 0$.

The following lemma gives sufficient conditions for these two assumptions to hold.

Lemma 2. Consider constants $D_e \geq \gamma$ and $D_p \geq \gamma$.

1. Suppose that gradient ascent with diminishing stepsizes c_0/t for the problem $\max_{\theta\in\Theta} r(w, \theta)$ has convergence rate $r(w) - r(w, \theta^s) \leq D_p/s$. Then we define $\theta_p^\gamma(w)$ by performing $s = D_p/\gamma$ steps of gradient ascent. Then, $\theta_p^\gamma(w)$ satisfies Assumption 5.

2. Suppose that gradient ascent with constant stepsize c_0 for the problem $\max_{\theta\in\Theta} r(w, \theta)$ has convergence rate $r(w) - r(w, \theta^s) \leq D_p\eta^s$ for some constant $0 < \eta < 1$. Then we define $\theta_p^\gamma(w)$ by $s = \log(D_p/\gamma)/\log(1/\eta)$ steps of gradient ascent. Then, $\theta_p^\gamma(w)$ satisfies Assumption 5.

3. Suppose that gradient ascent with diminishing stepsizes c_0/t for the problem $\max_{\theta\in\Theta} r_S(w, \theta)$ has convergence rate $r_S(w) - r_S(w, \theta^s) \leq D_p/s$. Then we define $\theta_e^\gamma(S)$ by performing $s = D_e/\gamma$ steps of gradient ascent. Then, $\theta_e^\gamma(S)$ satisfies Assumption 6.

4. Suppose that gradient ascent with constant stepsize c_0 for the problem $\max_{\theta\in\Theta} r_S(w, \theta)$ has convergence rate $r_S(w) - r_S(w, \theta^s) \leq D_e\eta^s$ for some constant $0 < \eta < 1$. Then we define $\theta_e^\gamma(w)$ by $s = \log(D_e/\gamma)/\log(1/\eta)$ steps of gradient ascent. Then, $\theta_e^\gamma(S)$ satisfies Assumption 6.

Remark 2. Note that for some practical nonconvex optimization problems in machine learning, gradient descent indeed converges to the global minima at a reasonably fast rate, e.g., in training deep overparametrized neural networks [38], robust least squares problems [40], phase retrieval and matrix completion [79]. Our Assumptions 5 and 6 can be viewed as an abstract summary of some benign properties of gradient descent for certain nonconvex optimization problems.

Furthermore, we assume that $f(\cdot,\cdot;z)$ is L-Lipschitz[4] continuous in $W \times \Theta$. This, along with Assumptions 5 and 6, allows us to derive the generalization error bounds of the primal risk and primal gap in terms of algorithmic stability.

Lemma 3. Suppose that Assumption 5 holds. If a minimax learning Algorithm A is an ϵ-stable algorithm, we have

$$\zeta^{P}_{gen}(A) \leq L\epsilon + \sqrt{L\lambda_p}\sqrt{\epsilon}.$$

Similarly, we can derive the generalization bound for the primal gap given the above assumptions.

Theorem 4. Suppose Assumptions 5 and 6 hold. Then we have

$$\zeta^{PG}_{gen}(A) \leq \zeta^{P}_{gen}(A) + \sqrt{L\lambda_e}/\sqrt{n}.$$

The proof of this theorem is similar to the proof of Lemma 3 and Theorem 3 and hence omitted.

4.4 Comparison of GDA and GDMax

In Section 4.3.3, we provide generalization bounds for the primal gap for any ϵ-stable algorithm. In this section, we focus on two algorithms in particular – GDA and GDMax. These two algorithms are described in Algorithms 8 and 9 in Appendix 4.6.4.

We note that though analyzing the *optimization* properties of GDA/stochastic GDA for solving the empirical minimax problem is an important topic, our focus in this chapter is on studying the generalization behavior of these algorithms. We assume that the empirical version of the stochastic minimax problem can be solved by GDA and GDMax, i.e., we assume that GDA and GDMax satisfy the following assumption:

Assumption 7. Let A be a minimax learner, such as GDA or GDMax. Then we assume that A has the following convergence rate: $E_A[r_S(w^t) - \min_{w\in W} r_S(w)] \leq (\phi_A(M(W)) +$

[4] Note that this is different from the L defined for the nonconvex-concave case. Here L captures the Lipschitz constant over the whole constraint set. In the nonconvex-concave case, $L = L(B(0, 2C_p + 1))$.

$\phi_A(C_e))/\psi_A(t)$, where $M(W)$ is the maximum of the norms of w, and $\phi_A(s)$, $\psi_A(s)$ are nonnegative, increasing functions that tend to infinity as $s \to \infty$.

For simplicity, throughout this section, we assume that $\|f(w,\theta;z)\| \leq 1$ for all w, θ, and z. The next theorem provides a bound for the population primal gap $\Delta(w_S^A) := r(w_S^A) - \min_{w\in W} r(w)$. Note that the goal of any algorithm is to make this gap as small as possible.

For an Algorithm A and subsets $W_0 \subseteq W, \Theta_0 \subseteq \Theta$, we define $A(W_0, \Theta_0)$ as the algorithm which restricts A to solve (4.2.1) under constraint sets W_0 and Θ_0. Specifically, $A(W, \Theta)$ is just A.

Theorem 5. Let $w_S^{A,t}, \theta_S^{A,t}$ be the t-th iterate generated by Algorithm A using dataset S. Assume that $\{\theta_S^{A,t}\} \subseteq \Theta_0 = \Theta_\theta^A$ for $t \leq T$ with probability $1-\delta$ (due to the randomness in S) and $B(0, C_p) \subseteq \Theta_\theta^A$. Here $B(v,r)$ denotes the l_2-ball with radius r centered at v. Let $A_0 = A(W, \Theta_0)$. Then after T iterations of Algorithm A, the population primal gap can be bounded as:

$$\begin{aligned}
&E_S[r(w_S^{A,T}) - \min_{w\in W} r(w)] \\
&\leq \underbrace{(\phi_{A_0}(M(W)) + \phi_{A_0}(C_e(\Theta_\theta^A)))/\psi_{A_0}(T) + 4L_\theta^* C_e(\Theta_\theta^A)/\sqrt{n}}_{II} + \underbrace{\zeta_{gen}^P(A_0)}_{I} + \delta,
\end{aligned}$$

where $\zeta_{gen}^P(A_0) = E_S E_A[r(w_S^{A_0,T}) - r_S(w_S^{A_0,T})]$ is the generalization error of the primal risk of Algorithm A_0.

Remark 3. Theorem 5 builds a closer connection between generalization behavior and the dynamics of the minimax learner A. It shows that suitable restriction to the max learner can lead to better minimax learner, in terms of generalization. We make this clear in the comparison of GDA and GDMax by analyzing the three terms in Theorem 5.

4.4.1 Analyzing the term I

First, we study the generalization error bound of the primal risk, i.e., ζ_{gen}^P in Theorem 5. For GDA, we can estimate ζ_{gen}^P by using Lemma 1. Therefore, it suffices to estimate the stability of GDA. We do this in the following lemma:

Lemma 4. Let $c_0 = \max\{\alpha_0, \beta_0\}$, If we use diminishing stepsizes $\alpha_t = \alpha_0/t$ and $\beta_t = \beta_0/t$ for GDA for T iterations, we have the stability bound $\epsilon^{GDA} \leq 2L_{\Theta_\theta^{GDA}} T^{c_0 \ell}/(n\ell)$.

Now, since we have a bound for $\zeta_{gen}^P(A)$ for an ϵ-stable Algorithm A in Lemma 1, we can substitute the stability bound for GDA from Lemma 4 in this expression to get a bound on $\zeta_{gen}^P(GDA)$ for GDA. We do this in the next proposition. We can bound $\zeta_{gen}^P(A_0)$ for GDA by substituting the stability bound in Lemma 4 into Lemma 1 (letting $\epsilon = \epsilon^{GDA}$).

Proposition 1. Let $c_0 = \max\{\alpha_0, \beta_0\}$ and assume that $f(\cdot, \cdot; z)$ is $L_{\Theta_\theta^{GDA}}$-Lipschitz-continuous inside the set $W \times \Theta_\theta^{GDA}$. For GDA with diminishing stepsizes $\alpha_0/t, \beta_0/t$ run for T iterations (denoted by GDA_T), the generalization error of the primal risk can be bounded by:

$$\zeta_{gen}^P(GDA_T) \leq (L_{\Theta_\theta^{GDA}})^{3/2} \sqrt{8C_p^2/\ell} \sqrt{T^{c_0 \ell}/n} + 2L^2_{\Theta_\theta^{GDA}} T^{c_0 \ell}/(n\ell).$$

However, for GDMax, we can not compute a uniform stability bound that vanishes as n goes to infinity. In fact, we can show from the following simple example that ζ_{gen}^P(GDMax) can be a constant that is independent of n, which means that for the case where $r(w, \theta)$ is nonconvex-concave, the generalization error of primal risk of GDMax can be undesirable.

Example 4 (Constant generalization error of primal risk for GDMax)**.** Consider a dataset S with n elements. Define the objective function: $f(w, \theta; z) = \left(\frac{w}{n^2} - z\right)\theta - \frac{\theta^2}{2n}$, where $w \in W = [-n\sqrt{n}, n\sqrt{n}]$, $\theta \in \Theta = \mathbb{R}$ and z is drawn from the uniform distribution over $\{-1/\sqrt{n}, 1/\sqrt{n}\}$. We have

$$r_S(w) = \frac{n^2}{2} \left(\frac{w}{n^2} - \frac{1}{n} \sum_{i=1}^{n} z_i \right)^2,$$

and $r(w) = \frac{w^2}{2n^2}$. Therefore, $\min_{w \in W} r(w) = 0$. From the definition of the function f and the sets W and Θ, we have $\ell = 1/n^2$, $L = \mathcal{O}(1/\sqrt{n})$.

Note that one step of GDMax can attain the minimizer of $r_S(w)$ (since it is a one dimensional quadratic problem), i.e., $w_S = n \sum_{i=1}^n z_i$ and $r_S(w_S) = 0$. Furthermore, we have $E_S r(w_S) = E[\frac{(\sum_{i=1}^n z_i)2}{2}] = 1/2 > 0$. Thus, ζ_{gen}^P(GDMax) $= E[r(w_S) - r_S(w_S)] = 1/2 > 0$ cannot be made small.

Therefore, from Proposition 1 and Example 4, we see that the bound for the expected population primal gap contains the term ζ_{gen}^P which cannot be bounded for GDMax, whereas can be bounded for GDA which leads us to the conclusion that GDA generalizes better than GDMax for such problems. However, it is possible to bound ζ_{gen}^P(GDMax) in certain problems, and in this case the other terms in Theorem 5 become crucial. We analyze them next.

4.4.2 Analyzing the term II

As shown in Example 2, sometimes GDMax can have a good generalization bound for the primal risk. Therefore, we need to analyze the other two terms in Theorem 5, i.e., $(\phi_A(M_w)+\phi_A(C_e(\Theta_\theta^A)))/\psi_A(T)$ and $L_\theta^* C_e(\Theta_\theta^A)/\sqrt{n}$. For these two terms, since L_θ^* is fixed, the constant $C_e(\Theta_\theta^A)$ is the key term which differentiates the performance of different algorithms.

By definition, the constant $C_e(\Theta_\theta^{GDMax})$ for GDMax is nearly C_e (See Definition 6). Therefore, the population primal gap after T steps of GDMax is dominated by C_e if C_e is large. However, the set Θ_θ^{GDA} for GDA can be much smaller than Θ, which implies that $C_e(\Theta_\theta^{GDA})$ can be much smaller than C_e. This phenomenon can be seen from Example 2: If we perform one step of GDMax with primal stepsize 1, we can attain $w^1 = w_S$. Then $E_S[r(w_S^1) - \min_{w\in W} r(w)] \geq 0.005$ from (4.3.2). For GDA, we can see that $w^1 = 1$ after one step of GDA with stepsize 1. Therefore, GDA generalizes better than GDMax. Generally, we have the following estimate of $C_e(\Theta_\theta^{GDA})$.

Lemma 5. Let $L_0 = \max_z \|\nabla f(w_0, \theta_0; z)\|$. Let $c_0 = \max\{\alpha_0, \beta_0\}$. If we use diminishing stepsizes $\alpha_t = \alpha_0/t$ and $\beta_t = \beta_0/t$ for GDA, then after T steps we have $\|\theta^t\| \leq T^{c_0\ell} L_0/\ell$ for $t \in [T]$.

Therefore, if C_e is much larger than C_p, using GDA with $C_p \leq T^{c_0\ell} L_0/\ell \leq C_e$ is better than GDMax. We make this more concrete in the context of GAN training next.

4.4.3 GAN training

We now study the specific case of GAN training to explore why GDA might generalize better than GDMax. This is numerically verified in the literature, such as [43]. Specifically, we revisit Example 3, and consider a special case: D is restricted to be a overparametrized linear function with respect to θ. Define the descriminator $D(x) = \Phi^T(x)v + b_0$, where $\Phi(x) = [\Phi_1(x), \cdots, \Phi_m(x)]^T \in \mathbb{R}^m$ is the feature matrix and $b_0 \in \mathbb{R}$. Also suppose that G is parametrized by w and $G^* = G_{w^*}$. Then the GAN problem can be written as $\min_{w \in W} \max_{\theta \in \Theta} \; r(w, \theta)$, where

$$r(w, \theta) = E_{x \sim P_r}[\phi(v^T \Phi(x) + b_0)] + E_{y \sim P_0}[\phi(1 - v^T \Phi(G_w(y)) - b_0)].$$

Here $\theta = (v, b_0)$. Assume that $\sqrt{\sigma_{\max}\left(E_{x \sim P_{G_w}} \Phi(x)\Phi^T(x)\right)} \leq \bar{\sigma}_{\max}/\sqrt{m}$, where $\sigma_{\max}(\cdot)$ denotes the largest singular value of a matrix and $\bar{\sigma}_{\max} > 0$ is a constant. Also assume that $E_{x \sim P_{G_w}} \Phi(x)\Phi^T(x)$ is full rank. Also, we assume that $|\phi'(\lambda)| \leq L_\phi$ for any $\lambda \in [0, 1]$. Therefore, we have $E[\|\nabla_\theta f(w, \theta; z)\|^2] \approx L_\phi^2 \bar{\sigma}_{\max}^2$. Then it is reasonable to assume that $\|\nabla f\| \leq \mathcal{O}(1)$.

Lemma 6. Suppose $\Phi(x)$ is sub-Gaussian and the matrix

$$Q_S = \begin{bmatrix} \Phi(x_1) & \Phi(x_2) \cdots & \Phi(x_n) & \Phi(G_w(y_1)) & \cdots & \Phi(G_w(y_n)) \end{bmatrix}$$

is full column rank ($m > n$) with probability 1. Then with probability at least $1 - C\delta$ with some constant C, we have $\|\theta_S(w^*)\| \geq \Omega(\sqrt{n})$, where $\theta_S(w^*) \in \arg\max_{\theta \in \Theta} r_S(w^*, \theta)$.

Now, for $\theta \in \arg\max_{\theta' \in \Theta} r(w^*, \theta')$, it can be easily seen that $v = 0, b_0 = 1/2$ in this case. Therefore, $C_p \approx 1/2$. Finally, combining the previous discussion on GDA in Lemma 5, and using the fact that C_e is large from Lemma 6, we see from Theorem 5 that GDA can generalize better than GDMax. More detailed discuss of the GAN-training example and Lemma 6 can be found in Section 4.6.4.

4.5 Conclusions

In this chapter, we first demonstrate the shortcomings of one popular metric, the primal risk, in terms of characterizing the generalization behavior of minimax learners. We then propose a new metric, the primal gap, whose generalization error overcomes these shortcomings and captures the generalization behavior of algorithms that solve stochastic minimax problems.

Finally, we use this newly proposed metric to study the generalization behavior of two different algorithms – GDA and GDMax, and study cases where GDA has a better generalization behavior than GDMax. Future directions include further investigation of the proposed new metric, the primal gap, and deriving its (tighter) generalization error bounds in other structured stochastic minimax optimization problems in machine learning.

4.6 Appendix

In this section, we present supplementary material and proofs omitted from the main text of the chapter.

4.6.1 Existing Related Results

From [43], we have the following theorem showing the connection between stability and generalization for minimax problems.

Theorem 6 ([43])**.** Consider an Algorithm A which is ϵ-stable. We have the following two claims:

1. If the maximization and the expectation can be swapped when computing $r(w)$, then

$$E_S E_A[\zeta^P_{gen}(A)] \leq \epsilon.$$

2. If $f(\cdot, \cdot; z)$ is nonconvex-strongly-concave and f is μ-strongly-concave with respect to θ, then

$$E_S E_A[\zeta^P_{gen}(A)] \leq L\sqrt{\kappa^2 + 1}\epsilon.$$

Remark 4. In [73], the authors proved a generalization bound in a weak sense, i.e., they consider the weak duality gap:

$$(\max_{\theta \in \Theta} E_S E_A r(w^A_S, \theta) - \min_{w \in W} E_S E_A r(w, \theta^A_S)) - (\max_{\theta \in \Theta} E_S E_A r_S(w^A_S, \theta) - \min_{w \in W} E_S E_A r_S(w, \theta^A_S)).$$

However, notice that the expectation is inside the min and max operators. It does not deal with the coupling of the maximization and expectation.

Remark 5. According to Theorem 6, the generalization bound for ζ_{gen}^P scales with the condition number κ_θ, and therefore cannot give useful bounds in the absence of strong concavity (when $\kappa_\theta \to \infty$).

Remark 6. The generalization bounds for ζ_{gen}^P of algorithms for problems in terms of stability without strong concavity is still open to the best of our knowledge. As mentioned in [73], finding generalization bounds without the strong concavity assumption is an interesting open problem.

4.6.2 Analysis of Example 2

In this section, we analyze the toy example given in Example 2.

Proposition 2. For the risk function and data distribution given in Example 2, we have

$$E_S[r(w) - r_S(w)] \leq 0$$

for any $w \in W$.

Proof. For a fixed w, $r(w) = w^2/2 - w$. On the other hand,

$$\begin{aligned} r_S(w) &= \max_{\theta \in \Theta} r(w, \theta) && (4.6.1) \\ &\geq r_S(w, 0) && (4.6.2) \\ &= r(w). && (4.6.3) \end{aligned}$$

Therefore, we have the desired result. □

Next, we prove that $|\sum_{i=1}^n z_i|$ will stay in the interval $[0.5, \lambda]$ with high probability.

Lemma 7. For large enough $\lambda > 2$, we have

$$\Pr\left(\left|\sum_{i=1}^n z_i\right| \in [0.5, \lambda]\right) > 0.4, \qquad \Pr\left(\left|\sum_{i=1}^n z_i\right| \in [2, \lambda]\right) > 0.01.$$

Proof. Let $y_i \sim N(0, 1/\sqrt{n}), i = 1, \cdots, n$ be n i.i.d. variables. Then $\sum_{i=1}^n y_i \sim N(0,1)$. According to the table of Normal distribution, we have $\Pr(|\sum_{i=1}^n y_i| \in [0.5, \lambda]) \geq 0.41$. By the definition of z_i, we have

$$\begin{aligned}&\Pr(|\sum_{i=1}^n z_i| \in [0.5, \lambda]) \\ &\quad \geq \Pr(|\sum_{i=1}^n y_i| \in [0.5, \lambda], |y_i| < 3\log n/\sqrt{n}) + \Pr(\max_{i \in [n]}(|y_i|) \geq 3\log n/\sqrt{n}).\end{aligned}$$

For the first term, we have

$$\begin{aligned}&\Pr(|\sum_{i=1}^n y_i| \in [0.5, \lambda], |y_i| < 3\log n/\sqrt{n}) \\ &\quad \geq \Pr(|\sum_{i=1}^n y_i| \in [0.5, \lambda]) - \Pr(\max_{i \in [n]}(|y_i|) \geq 3\log n/\sqrt{n}) \\ &\quad \geq 0.41 - \sum_{i=1}^n \Pr(|y_i| \geq 3\log n/\sqrt{n}) \\ &\quad \geq 0.41 - ne^{-\gamma 9 \log^2 n} \geq 0.41 - 1/n^{\lambda\gamma - 1}.\end{aligned}$$

Taking λ sufficiently large yields the desired result, where the first inequality is because of the union bound and the second inequality is due to the tail bound of Normal distribution. Therefore, $\Pr(|\sum_{i=1}^n z_i| \in [0.5, \lambda]) > 0.4$ for sufficiently large n. The second statement follows similarly, noting from the table of Normal distribution that $\Pr(|\sum_{i=1}^n y_i| \in [0.5, \lambda]) \geq 0.046$. □

Proposition 3. For sufficiently large $\lambda > 0$, we have

$$E_S[r(w_S) - \min_{w \in W} r(w)] \geq 0.001.$$

Proof. If $|\sum_{i=1}^n z_i| \in [0.5, \lambda]$, we have

$$w_S = \max(0, 1 - (\sum_{i=1}^n z_i)^2/2) \leq 0.9.$$

In this case, we have

$$r(w_S) - \min_{w \in W} r(w) \geq 0.005, \tag{4.6.4}$$

by direct calculation. Therefore, we have

$$\begin{aligned} & E_S[r(w_S) - \min_{w \in W} r(w)] && (4.6.5) \\ & \geq \Pr(|\sum_{i=1}^{n} z_i| \in [0.5, \lambda]) \cdot 0.05 + \Pr(|\sum_{i=1}^{n} z_i| \notin [0.5, \lambda]) \cdot 0 && (4.6.6) \\ & \geq 0.02, && (4.6.7) \end{aligned}$$

where the first inequality is because of (4.6.4) and the fact that $r(w_S) - \min_{w \in W} r(w) \geq 0$ for any S. □

Proposition 4. For sufficiently large $\lambda > 0$, we have:

$$E_S[\min_{w \in W} r_S(w) - \min_{w \in W} r(w)] \geq 0.005$$

for Example 2.

Proof. If $|\sum_{i=1}^{n} z_i| \geq \lambda > 2$, we have $w_S = 0$ and hence $r_S(w_S) = 0$. If $|\sum_{i=1}^{n} z_i| \leq \lambda$, we have

$$r_S(w_S) - r(w^*) \geq r_S(w_S) - r(w_S) = w_S(\sum_{i=1}^{n} z_i)^2/2 \geq 0.$$

Therefore, $\min_{w \in W} r_S(w) \geq \min_{w \in W} r(w)$ for any S. By Lemma 7, we can prove that $\Pr(|\sum_{i=1}^{n} z_i| \in [2, \lambda]) \geq 0.01$ for sufficiently large λ. Notice that for $|\sum_{i=1}^{n} z_i| \in [2, \lambda]$, $r_S(w_S) - \min_{w \in W} r(w) = 1/2$. Therefore, we have

$$E_S[\min_{w \in W} r_S(w) - \min_{w \in W} r(w)] \geq \Pr(|\sum_{i=1}^{n} z_i| \in [2, \lambda]) \cdot 1/2 \geq 0.005.$$

This completes the proof. □

4.6.3 Proofs in Section 4.3

Proof of Lemma 1

In this subsection, we assume that A is an ϵ-stable algorithm. For any $w \in W$, let $\Theta_S(w) = \arg\max_{\theta\in\Theta} r_S(w,\theta)$ and $\Theta(w) = \arg\max_{\theta\in\Theta} r(w,\theta)$ be the solution sets of the problems. Let $\theta(w)$ be any element in $\Theta(w)$. Then

$$\begin{aligned} E_A E_S[r(w_S^A) - r_S(w_S^A)] &= E_A E_S[r(w_S^A, \theta(w_S^A)) - r_S(w_S^A, \theta_S(w_S^A))] \\ &\le E_A E_S[r(w_S^A, \theta(w_S^A)) - r_S(w_S^A, \theta(w_S^A))], \end{aligned}$$

where the inequality is because $r_S(w_S^A, \theta_S(w_S^A)) \ge r_S(w_S^A, \theta)$ for any θ. Let f be μ-strongly concave with respect to θ. We denote the condition number by $\kappa_\theta = \ell_{\theta\theta}/\mu$.

In the strongly concave case, $\Theta(w)$ has a unique element $\theta(w)$, which is κ_θ-Lipschitz continuous with respect to w (see [76]).

Then, defining $\widetilde{f}(w,z) = f(w,\theta(w);z)$, the minimax problem reduces to the usual minimization problem on the function $\widetilde{f}$. The stability and the Lipschitz continuity of $\theta(w)$ with respect to w yield the generalization bound of $L\sqrt{\kappa^2+1}\epsilon$. This is the result shown in Theorem 1 of [43].

However, if the maximization problem is not strongly concave, we lose the Lipschitz continuity and the uniqueness. To overcome this difficulty, we define an approximate maximizer $\bar{\theta}(w)$ to $r(w,\theta)$. Concretely speaking, we define $\bar{\theta}(w)$ to be the point after s steps of gradient ascent for the function $r(w,\cdot)$ with a stepsize $1/\ell_{\theta\theta}$ and being initialized at 0. Then we have the following lemma:

Lemma 8. For any $w \in W$, we have[5]

1. $\|\bar{\theta}(w) - \bar{\theta}(w')\| \le s\frac{\ell}{\ell_{\theta\theta}}\|w - w'\|$.

2. $r(w) - r(w, \bar{\theta}(w)) \le \ell_{\theta\theta} C_p^2/s$.

[5] For point 2, it holds when $s > 0$. For $s = 0$, we have the bound $r(w) - r(w,\bar{\theta}(w)) \le \ell_{\theta\theta} C_p^2$. We do not separate this degenerate case for ease of presentation.

Proof. To prove the first part, let $\theta_0 = \theta'_0 = 0$. Define θ_t, θ'_t recursively as follows:

$$\theta_{t+1} = \theta_t + \nabla_\theta r(w, \theta_t)/\ell_{\theta\theta}$$

and

$$\theta'_{t+1} = \theta'_t + \nabla_\theta r(w', \theta'_t)/\ell_{\theta\theta}.$$

We prove $\|\theta_t - \theta'_t\| \le t\frac{\ell}{\ell_{\theta\theta}}\|w - w'\|$ by induction. For $t = 0$, $\theta_0 - \theta'_0 = 0$. Assume the induction hypothesis $\|\theta_{t-1} - \theta'_{t-1}\| \le (t-1)\frac{\ell}{\ell_{\theta\theta}}\|w - w'\|$ holds. We have

$$\begin{aligned}
\|\theta_t - \theta'_t\| &= \|(\theta_{t-1} + \nabla_\theta r(w, \theta_{t-1})/\ell_{\theta\theta}) - (\theta'_{t-1} + \nabla_\theta r(w, \theta'_{t-1})/\ell_{\theta\theta}) \\
&\qquad + (\nabla_\theta r(w, \theta'_{t-1}) - \nabla_\theta r(w', \theta'_{t-1}))/\ell_{\theta\theta}\| \\
&\le \|(\theta_{t-1} + \nabla_\theta r(w, \theta_{t-1})/\ell_{\theta\theta}) - (\theta'_{t-1} + \nabla_\theta r(w, \theta'_{t-1})/\ell_{\theta\theta})\| \\
&\qquad + \|(\nabla_\theta r(w, \theta'_{t-1}) - \nabla_\theta r(w', \theta'_{t-1}))/\ell_{\theta\theta}\| \\
&\le \|\theta_{t-1} - \theta'_{t-1}\| + \ell\|w - w'\|/\ell_{\theta\theta} \\
&\le (t-1)\frac{\ell}{\ell_{\theta\theta}}\|w - w'\| + \frac{\ell}{\ell_{\theta\theta}}\|w - w'\| \\
&= t\frac{\ell}{\ell_{\theta\theta}}\|w - w'\|,
\end{aligned}$$

where the first inequality follows from the triangle inequality, the second inequality follows from non-expansiveness of gradient ascent for concave functions and the ℓ-Lipschitz continuity of ∇r, and the third inequality follows from the induction hypothesis.

Therefore, letting $t = s$ completes the proof of the first part. The second part of this lemma is just the convergence result for gradient ascent on smooth concave functions (see e.g., [96]). □

Consider a virtual algorithm $\bar{A}$: for any S, the algorithm returns $w = w_S^A$ and $\theta = \bar{\theta}(w_S^A)$.

Lemma 9. The stability of this virtual algorithm is $\epsilon\sqrt{\left(s\frac{\ell}{\ell_{\theta\theta}}\right)^2 + 1}$.

Proof. It is direct from the first part of Lemma 8. □

Then we have the generalization bound of $r_S(w, \theta)$:

Lemma 10. We have

$$E_S E_A[r(w_S^A, \bar{\theta}(w_S^A)) - r_S(w_S^A, \bar{\theta}(w_S^A))] \le \epsilon L \sqrt{\left(s\frac{\ell}{\ell_{\theta\theta}}\right)^2 + 1}.$$

Proof. For any z, by Assumption 4, we have

$$\|f(w_S^{\bar{A}}, \theta_S^{\bar{A}}; z) - f(w_{S'}^{\bar{A}}, \theta_{S'}^{\bar{A}}; z)\| \le \epsilon L \sqrt{\left(s\frac{\ell}{\ell_{\theta\theta}}\right)^2 + 1}.$$

The result follows directly from the standard stability theory in [61]. □

Now we are ready to derive the generalization error bound of the Primal Risk for an Algorithm A with ϵ-stability. First, we have

$$\begin{aligned}
E_S E_A[r(w_S^A) - r_S(w_S^A)] &\le E_S E_A[r(w_S^A) - r_S(w_S^A, \bar{\theta}(w_S^A))] \\
&\le E_S E_A[(r(w_S^A, \bar{\theta}(w_S^A) + \ell_{\theta\theta} C_p^2/s) - r_S(w_S^A, \bar{\theta}(w_S^A))] \\
&= E_S E_A[r(w_S^A, \bar{\theta}(w_S^A)) - r_S(w_S^A, \bar{\theta}(w_S^A))] + \ell_{\theta\theta} C_p^2/s \\
&\le \epsilon L \sqrt{\left(s\frac{\ell}{\ell_{\theta\theta}}\right)^2 + 1} + \ell_{\theta\theta} C_p^2/s \\
&\le \epsilon L s \frac{\ell}{\ell_{\theta\theta}} + \frac{\ell_{\theta\theta} C_p^2}{s} + \epsilon L
\end{aligned}$$

where the first inequality is because $r_S(w_S^A) = \max_\theta r_S(w_S^A, \theta)$, the second inequality is because of the second part of Lemma 8 and the last inequality is because of Lemma 10. Optimizing over[6] s, the generalization error is bounded by $\zeta_{gen}^P(A) \le \sqrt{4L\ell C_p^2} \cdot \sqrt{\epsilon} + \epsilon L$. This completes the proof. □

Tightness of the bound for Primal Risk

Consider the following risk function: $f(w, \theta; z) = \sqrt{n/\epsilon}((w/(n\sqrt{n}\epsilon) - z)\theta - \theta^2/(2n\sqrt{n}\epsilon))$, where $w \in W = [-\lambda\epsilon\sqrt{n}\log n, \lambda\epsilon\sqrt{n}\log n]$ and $\theta \in \Theta = \mathbb{R}$. The sample z is drawn from the uniform distribution over $\{-1/\sqrt{n}, 1/\sqrt{n}\}$. Then we have $r(w) = \sqrt{n/\epsilon}(w^2/(2\epsilon n\sqrt{n}))$

[6] Here we assume that the optimal s is a real number greater than 0. Constraining s to be an integer and also incorporating 0 does not change the result and we ignore this case here. See also Footnote 5.

and $r_S(w) = \sqrt{n/\epsilon}((\epsilon n\sqrt{n})(w/(\epsilon n\sqrt{n}) - \sum_{i=1}^n z_i/n)^2/2)$. Now we have $\ell = \frac{\sqrt{n/\epsilon}}{(\epsilon n\sqrt{n})}$, $C_p = \lambda\epsilon\sqrt{n}\log n$ and $L = \frac{\sqrt{n/\epsilon}}{\sqrt{n}}$. If we perform one-step of GDMax with stepsize $1/\ell_{r_S}$ where $\ell_{r_S} = \sqrt{n/\epsilon}/(\epsilon n\sqrt{n})$, then we attain $w_S = \arg\min_{w\in W} r_S(w)$. The stability bound of the GDMax is ϵ. Therefore, the generalization error of the primal risk is estimated as:

$$\zeta_{gen}^P(GDMax) \leq \sqrt{8L\ell C_p^2}\sqrt{\epsilon} = 8\lambda\log n\sqrt{\epsilon}.$$

On the other hand, $\sum_{i=1}^n z_i/n \in [-\lambda\log n/n, \lambda\log n/n]$ holds with probability at least $1 - C/n^\lambda$ by Hoefding inequality. Let $\bar{w}_S = \epsilon\sqrt{n}(\sum_{z_i\in S} z_i)$. Then with probability at least $1 - C/n^\lambda$, $w_S = \bar{w}_S$ Notice that $E_S[r(\bar{w}_S) - r_S(\bar{w}_S)] = E_S[\sqrt{n}\cdot(\sqrt{\epsilon}n\sqrt{n})\cdot(\sum_{z_i\in S} z_i/n)^2] = \sqrt{\epsilon}$. It is not hard to show that $|r(\bar{w}_S)| \leq n\sqrt{\epsilon}$, $r_S(\bar{w}_S) = 0$, $|r(w_S)| \leq 2n\sqrt{\epsilon}$ and $|r_S(w_S)| \leq 2n\sqrt{\epsilon}$. Then we have

$$E_S[r(\bar{w}_S) - r_S(\bar{w}_S)] - E_S[r(w_S) - r_S(w_S)] \leq 5Cn\sqrt{\epsilon}/n^\lambda.$$

Therefore, $E[r(w_S) - r_S(w_S)] \geq \sqrt{\epsilon}/2$ for sufficiently large λ and n. Then in this example we have

$$\sqrt{\epsilon}/2 \leq \zeta_{gen}^P(A) \leq 8\lambda\log n\sqrt{\epsilon}.$$

For $\epsilon \leq 1/n^{\tau+1}$, we have

$$\log n \leq \frac{1}{\tau+1}\log(1/\epsilon).$$

Therefore, the estimate $\zeta_{gen}^P \leq \lambda\sqrt{\epsilon}\log(1/\epsilon)/(\tau+1)$ is tight up to a $\log(1/\epsilon)$ factor.

Proof of Theorem 3

Recall that the empirical primal gap is defined as

$$\Delta_S(w) = r_S(w) - \min_{w\in W} r_S(w)$$

and the population primal gap is given by

$$\Delta(w) = r(w) - \min_{w\in W} r(w).$$

Suppose we are given an ϵ-stable Algorithm A. We then want to derive the generalization error

$$\zeta_{gen}^{PG}(A) = E_S E_A[\Delta(w_S^A) - \Delta_S(w_S^A)].$$

Since we already have the generalization error for the primal risk $E_S E_A[r(w_S^A) - r_S(w_S^A)]$ in Theorem 1, we only need to estimate

$$E_S E_A[\min_{w \in W} r_S(w) - \min_{w \in W} r(w)] = E_S[\min_{w \in W} r_S(w) - \min_{w \in W} r(w)]$$

to get a generalization error bound on the primal gap.

Lemma 11. Let $w^* \in \arg\min_{w \in W} r(w)$. Suppose that $f(w^*, \cdot; z)$ is L_θ^* Lipschitz continuous with respect to θ. Then we have

$$E_S[\min_{w \in W} r_S(w) - \min_{w \in W} r(w)] \leq 4L_\theta^* C_e / \sqrt{n}.$$

Proof. We use similar techniques as in the proof of Lemma 1.

Step 1. We define an approximate maximizer $\widetilde{\theta}_S$ of the function $r_S(w^*, \cdot)$. $\widetilde{\theta}_S$ is attained by performing s steps of gradient ascent to $r_S(w^*, \cdot)$ with stepsize $1/\ell_{\theta\theta}$ and being initialized at 0.

Similar to Lemma 8, we have the following lemma:

Lemma 12. We have the following properties:

1. $\|\widetilde{\theta}_S - \widetilde{\theta}_{S'}\| \leq 2sL_\theta^* / (n\ell_{\theta\theta})$.

2. $r_S(w^*) - r_S(w^*, \widetilde{\theta}_S) \leq \ell_{\theta\theta} C_e^2 / s$.

Proof. The proof is similar to the proof of Lemma 8. To prove the first part, let $\widetilde{\theta}_0 = \widetilde{\theta}'_0 = 0$. Define $\widetilde{\theta}_t, \widetilde{\theta}'_t$ recursively as follows:

$$\widetilde{\theta}_{t+1} = \widetilde{\theta}_t + \nabla_\theta r_S(w^*, \widetilde{\theta}_t) / \ell_{\theta\theta}$$

and

$$\widetilde{\theta}'_{t+1} = \widetilde{\theta}'_t + \nabla_\theta r_{S'}(w^*, \widetilde{\theta}'_t) / \ell_{\theta\theta}.$$

We prove $\|\widetilde{\theta}_t - \widetilde{\theta}'_t\| \leq L^*_\theta/(n\ell_{\theta\theta})$ by induction. For $t = 0$, $\widetilde{\theta}_0 - \widetilde{\theta}'_0 = 0$. Assume the induction hypothesis $\|\widetilde{\theta}_{t-1} - \widetilde{\theta}'_{t-1}\| \leq (t-1)L^*_\theta/(n\ell_{\theta\theta})$ holds. We have

$$\begin{aligned}
\|\widetilde{\theta}_t - \widetilde{\theta}'_t\| &= \|(\widetilde{\theta}_{t-1} + \nabla_\theta r_S(w^*, \widetilde{\theta}_{t-1})/\ell_{\theta\theta}) - (\widetilde{\theta}'_{t-1} + \nabla_\theta r_S(w^*, \widetilde{\theta}'_{t-1})/\ell_{\theta\theta}) \\
&\qquad + (\nabla_\theta r_S(w^*, \widetilde{\theta}'_{t-1}) - \nabla_\theta r_{S'}(w^*, \widetilde{\theta}'_{t-1}))/\ell_{\theta\theta}\| \\
&\leq \|(\widetilde{\theta}_{t-1} + \nabla_\theta r_S(w^*, \widetilde{\theta}_{t-1})/\ell_{\theta\theta}) - (\widetilde{\theta}'_{t-1} + \nabla_\theta r_S(w^*, \widetilde{\theta}'_{t-1})/\ell_{\theta\theta})\| \\
&\qquad + \|(\nabla_\theta r_S(w^*, \widetilde{\theta}'_{t-1}) - \nabla_\theta r_{S'}(w^*, \widetilde{\theta}'_{t-1}))/\ell_{\theta\theta}\| \\
&\leq \|\widetilde{\theta}_{t-1} - \widetilde{\theta}'_{t-1}\| + \ell\|w - w'\|/\ell_{\theta\theta} \\
&\leq (t-1)\frac{2L^*_\theta}{n\ell_{\theta\theta}} + \frac{2L^*_\theta}{n\ell_{\theta\theta}} \\
&= t\frac{2L^*_\theta}{n\ell_{\theta\theta}},
\end{aligned}$$

where the first inequality follows from the triangle inequality, the second inequality follows from non-expansiveness of gradient ascent for concave functions and the L^*_θ-Lipschitz continuity of $f(w^*, \cdot; z)$, and the third inequality follows from the induction hypothesis.

Therefore, letting $t = s$ completes the proof of the first part. The second part of this lemma is just the convergence result for gradient ascent on smooth concave functions (see e.g., [96]). □

We then define the virtual algorithm $\widetilde{A}$ given by $w_S^{\widetilde{A}} = w^*$ and $\theta_S^{\widetilde{A}} = \widetilde{\theta}_S$. Since the output argument w of $\widetilde{A}$ is always w^*, the stability of $\widetilde{A}$ only depends on $\widetilde{\theta}_S$. Then the stability bound of this virtual algorithm is given in the following lemma:

Lemma 13. The stability of Algorithm $\widetilde{A}$ is given by $\epsilon_{sta}(\widetilde{A}) = 2s(L^*_\theta)^2/(n\ell_{\theta\theta})$.

Then by the standard stability theory in [61], we have

$$|E_S E_A[r_S(w^*, \widetilde{\theta}_S) - r(w^*, \widetilde{\theta}_S)]| \leq 2s(L^*_\theta)^2/(n\ell_{\theta\theta}). \tag{4.6.8}$$

Step 2. We have

$$E_S[\min_{w\in W} r_S(w) - \min_{w\in W} r(w)] \overset{\text{(i)}}{=} E_S[r_S(w_S) - r(w^*, \theta^*)]$$

$$
\begin{aligned}
&\overset{\text{(ii)}}{\leq} E_S[r_S(w^*) - r(w^*, \theta^*)] \\
&\overset{\text{(iii)}}{\leq} E_S[r_S(w^*, \widetilde{\theta}_S) - r(w^*, \theta^*)] + \ell_{\theta\theta} C_e^2/s \\
&\overset{\text{(iv)}}{\leq} E_S[r_S(w^*, \widetilde{\theta}_S) - r(w^*, \widetilde{\theta}_S)] + \ell_{\theta\theta} C_e^2/s,
\end{aligned}
$$

where (i) follows from the definition of w^*, θ^*, (ii) follows since w_S minimizes $r_S(w)$, (iii) follows from Lemma 12, and (iv) follows from the optimality of θ^* given w^*. Then by (4.6.8), we have

$$
\begin{aligned}
E_S[\min_{w \in W} r_S(w) - \min_{w \in W} r(w)] &\leq E_S[r_S(w^*, \widetilde{\theta}_S) - r(w^*, \widetilde{\theta}_S)] + \ell_{\theta\theta} C_e^2/s && (4.6.9) \\
&\leq 2s(L_\theta^*)^2/(n\ell_{\theta\theta}) + \ell_{\theta\theta} C_e^2/s && (4.6.10) \\
&\leq 4L_\theta^* C_e/\sqrt{n} && (4.6.11)
\end{aligned}
$$

which completes the proof. □

The final statement of the theorem follows from Lemma 11 and Lemma 1. □

Tightness of Bound for Primal Min Error

We can construct an example with C_e,L_θ^* independent of n and the upper bound for ζ_{gen}^{PM} is tight up a log factor.

Consider the following example:

- Let $f(w, \theta; z) = \frac{1}{M}(w^2/2 + w(\log^2 n\theta^2/(2K^2) + n \log nz\theta/K + 1))$, where $w \in W = [-1, 1]$, $\theta \in \Theta = [-\lambda K, \lambda K]$ for some arbitrary constants $K > 0$ and $M > 0$. z is drawn from a truncated Gaussian distribution. Concretely speaking, let $y \sim N(0, 1/\sqrt{n})$. Then $z = y$ if $|y| \leq \lambda \log n/\sqrt{n}$ and $z = \lambda \log n/\sqrt{n}$ if $y \geq \lambda \log n/\sqrt{n}$. Notice that $Mf(w, \theta; z) = h(w, \theta'; z)$, where $\theta' = n \log n\theta/K$, $h(w, \theta'; z) = w^2/2 + w((\theta')^2/(2n^2) + z\theta' + 1)$, $\theta' = n \log n\theta/K \in [-\lambda n \log n, \lambda n \log n]$. Notice that h is just the risk function in Example 1. Therefore, we can estimate the lower bound of the Primal Min Error corresponding to f using the result in Example. The lower bound of Primal Min Error of the problem in Example 1 (corresponding to h) is 0.005. Then the Primal-Min Error (corresponding to *f)* $\zeta_{gen}^{PM}(A) = E_S \min_{w\in W} r_S(w) - \min_{w\in W} r(w) \geq 0.005/M$. On the other hand, it is not hard to have $L_\theta^* = \lambda\sqrt{n}\log^2 n/MK$, $C_e = \lambda K$. Therefore, $L_\theta^* C_e = \lambda^2\sqrt{n}\log^2 n/M$. Let $M = \sqrt{n}\log^2 n$. Then $L_\theta^* = \lambda/K$.

If K does not depend on n, L_θ^*, C_e do not depend on n. Therefore, the Primal-Min Error $\zeta_{gen}^{PM}(A)$ satisfies $0.005\lambda^2(L_\theta^* C_e)/(\log^2 n\sqrt{n}) = 0.005(L_\theta^* C_e)/(ML_\theta^* C_e) \leq \zeta_{gen}^{PM}(A) \leq (L_\theta^* C_e)/\sqrt{n}$, which is tight up to a $\log^2 n$.

- If we let $M = 1$ and $K = 1$, then $\zeta_{gen}^{PM}(A) \leq \lambda^2\log^2 n$ as discussed in the main part of the chapter. This upper bound means that we can not attain an arbitrary accuracy $\delta > 0$. The lower bound, i.e., $\zeta_{gen}^{PM}(A) \geq 0.005\lambda^4$, also implies that we can not attain an arbitrary accuracy.

- If we let $M = 1/(\lambda^2\log^2 n)$ and $K = 1$, $L_\theta^* C_e/\sqrt{n} = 1$. This upper bound implies that we can not let ζ_{gen}^{PM} smaller than arbitrariy required accuracy. Return to the lower bound, i.e., $\zeta_{gen}^{PM}(A) \geq 0.005/(\lambda^2\log^2 n)$. If we want to attain an accuracy δ, the required sample complexity is $2^{\lambda/\sqrt{\delta}}$, which is larger than a polynomial size and hence is still viewed as intractable. In this sense, the upper bound and the lower bound do not make a major difference.

- Combining the two points above, we can conclude that in terms of sample complexity, our bound is tight (up to logarithmic factors).

Proof of Lemma 2

We only prove the first part of this lemma and the others can be proved similarly. Let $s = [D_p/\gamma] + 1$, where $[r]$ denotes the largest integer no more than r. To prove the first part, let $\theta_0 = \theta_0' = 0$. Define θ_t, θ_t' recursively as follows:

$$\theta_{t+1} = \theta_t + c_0\nabla_\theta r(w, \theta_t)/t$$

and

$$\theta_{t+1}' = \theta_t' + c_0\nabla_\theta r(w', \theta_t')/t.$$

We prove $\|\theta_t - \theta'_t\| \leq t\frac{\ell}{\ell_{\theta\theta}}\|w - w'\|$ by induction. For $t = 0$, $\theta_0 - \theta'_0 = 0$. Assume the induction hypothesis $\|\theta_{t-1} - \theta'_{t-1}\| \leq (t-1)\frac{\ell}{\ell_{\theta\theta}}\|w - w'\|$. We have

$$\begin{aligned}
\|\theta_t - \theta'_t\| &= \|(\theta_{t-1} + c_0\nabla_\theta r(w, \theta_{t-1})/t) - (\theta'_{t-1} + c_0\nabla_\theta r(w, \theta'_{t-1})/t) & (4.6.12)\\
&\quad + c_0(\nabla_\theta r(w, \theta'_{t-1}) - \nabla_\theta r(w', \theta'_{t-1}))/t\| & (4.6.13)\\
&\leq \|(\theta_{t-1} + c_0\nabla_\theta r(w, \theta_{t-1})/t) - (\theta'_{t-1} + c_0\nabla_\theta r(w, \theta'_{t-1})/t)\| & (4.6.14)\\
&\quad + c_0\|(\nabla_\theta r(w, \theta'_{t-1}) - \nabla_\theta r(w', \theta'_{t-1}))/t\| & (4.6.15)\\
&\leq (1 + c_0\ell_{\theta\theta}/t)\|\theta_{t-1} - \theta'_{t-1}\| + c_0\ell\|w - w'\|/t. & (4.6.16)
\end{aligned}$$

Here the first inequality follows from the triangle inequality, the second inequality follows from the $\ell_{\theta\theta}-$Lipschitz continuity of $\nabla_\theta r$ and ℓ-Lipschitz continuity of ∇r. Therefore, we have

$$\|\theta_t - \theta'_t\| \leq (1 + c_0\ell_{\theta\theta}/t)\|\theta_{t-1} - \theta'_{t-1}\| + c_0\ell\|w - w'\|/t.$$

Let $\delta_t = \|\theta_t - \theta'_t\|$. Then by the above recursion, we have

$$\delta_t + \ell/\ell_{\theta\theta}\|w - w'\| \leq \prod_{i=1}^{t}(1 + c_0\ell_{\theta\theta}/i)\ell\|w - w'\|/\ell_{\theta\theta}.$$

Using the inequalities $e^a \geq 1 + a$ and $\sum_{i=1}^{t} 1/i \leq \log t$, we have

$$\delta_t \leq \frac{t\ell}{\ell_{\theta\theta}}\|w - w'\|.$$

Letting $t = s$ yields

$$\|\theta_p^\gamma(w) - \theta_p^\gamma(w')\| \leq \frac{s\ell}{\ell_{\theta\theta}}\|w - w'\|.$$

Since $D_p > \gamma$, we have

$$s \leq [D_p/\gamma] + 1 \leq 2D_p/\gamma.$$

Hence,

$$s\frac{\ell}{\ell_{\theta\theta}} \cdot \gamma \leq 2D_p\ell/\ell_{\theta\theta}.$$

Setting $\lambda_p = 2D_p\ell/\ell_{\theta\theta}$ yields the desired result.

Algorithm 8 GDA

```
Require: initial iterate $(w_S^0, \theta_S^0) = (0,0)$, stepsizes $\alpha_t, \beta_t$, projection operators $P_W$ and $P_\Theta$;
for $t = 0, \dots, T-1$ do
    $w_S^{t+1} = P_W\left(w_S^t - \alpha_t \nabla_w r_S(w,\theta)\right)$
    $\theta_S^{t+1} = P_\Theta\left(\theta_S^t + \beta_t \nabla_\theta r_S(w,\theta)\right)$
end for
```

Algorithm 9 GDMax

```
Require: initial iterate $(w_S^0, \theta_S^0) = (0,0)$, stepsizes $\alpha_t$, projection operators $P_W$ and $P_\Theta$;
for $t = 0, \dots, T-1$ do
    $w_S^{t+1} = P_W\left(w_S^t - \alpha_t \nabla_w r_S(w,\theta)\right)$
    $\theta_S^{t+1} = \operatorname*{argmax}_{\theta\in\Theta} r_S(w_S^{t+1}, \theta)$
end for
```

Proof of Lemma 3

This is similar to the proof of Lemma 1. We first define the virtual algorithm $\bar{A}$ which outputs $(w_S^A, \theta_p^\gamma(w_S^A))$. By Assumption 5, it can be easily seen that $\bar{A}$ is $(1+\lambda_p/\gamma)\epsilon$-stable. Then by Theorem 1, we have

$$E_S E_A[r(w_S^A, \theta_p^\gamma(w_S^A)) - r_S(w_S^A, \theta_p^\gamma(w_S^A))] \le L(1+\lambda_p/\gamma)\epsilon.$$

This gives us:

$$\begin{aligned} E_S E_A[r(w_S^A) - r_S(w_S^A)] &\le E_S E_A[r(w_S^A, \theta_p^\gamma(w_S^A)) - r_S(w_S^A, \theta_p^\gamma(w_S^A))] + \gamma \\ &\le L\epsilon + L\lambda_p \epsilon/\gamma + \gamma. \end{aligned}$$

Taking $\gamma = \sqrt{L\lambda_p}\sqrt{\epsilon}$, we have

$$\zeta_{gen}^p(A) \le L\epsilon + \sqrt{L\lambda_p}\sqrt{\epsilon}.$$

4.6.4 Proofs in Section 4.4

Proof of Theorem 5

First, we have

$$
\begin{aligned}
&E_S E_{A_0}[r(w_S^{A_0,T}) - \min_{w\in W} r(w)] \\
&= E_S E_{A_0}[r_S(w_S^{A_0,T}) - \min_{w\in W} r_S(w)] + E_S E_{A_0}[r(w_S^{A_0,T}) - r_S(w_S^{A_0,T})] \\
&\qquad + E_S E_{A_0}[\min_{w\in W} r_S(w) - \min_{w\in W} r(w)]. \qquad (4.6.17)
\end{aligned}
$$

Furthermore, by Assumption 7 and Theorem 3, we have

$$E_S E_{A_0}[r(w_S^{A_0,T}) - \min_{w\in W} r(w)] \leq (\phi_{A_0}(M_w) + \phi_{A_0}(C_e(\Theta_0)))/\psi_{A_0}(T) + \zeta_{gen}^P(A_0) + L_\theta^* C_e(\Theta_0)/\sqrt{n}.$$

Next, notice that the output of A_0 is equal to the output of A with probability at least $1-\delta$ and $\|r(w)\| \leq 1$. Therefore, we have

$$|E_S E_A[r(w_S^{A,T})] - E_S E_{A_0}[r(w_S^{A_0,T})]| \leq \delta,$$

which gives the desired result. □

Proof of Lemma 4

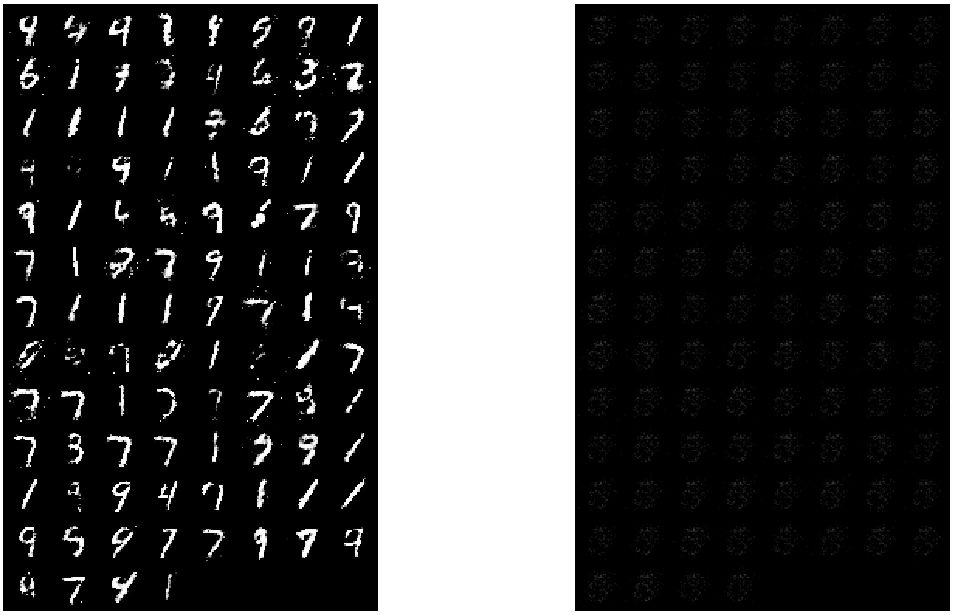
Define $\delta_t = \|(w_S^t, \theta_S^t) - (w_{S'}^t, \theta_{S'}^t)\|$. We have

$$\delta_{t+1} \leq (1 + c_0\ell/t)\delta_t + 2c_0 L_{\Theta_\theta^{GDA}}/nt.$$

Therefore,

$$\delta_{t+1} + \frac{2L_{\Theta_\theta^{GDA}}}{\ell n} \leq (1 + c_0\ell/t)\left(\delta_t + \frac{2L_{\Theta_\theta^{GDA}}}{\ell n}\right) \leq \frac{2L_{\Theta_\theta^{GDA}}}{\ell n} T^{c_0\ell}, \qquad (4.6.18)$$

which completes the proof. □



(a) GDA (b) GDMax

Figure 4-1: Comparison of the results on MNIST generated by GDA and GDMax.

Proof of Lemma 5

For a fixed dataset S, let $g_t = \nabla r_S(w^t, \theta^t)$ and $d_t = \|(w^0, \theta^0) - (w^t, \theta^t)\|$. Then we have $g_t \leq L_0 + d_t \ell$ and $d_{t+1} \leq d_t + c_0 g_t / t$. Substituting the first inequality into the second one, we have

$$d_{t+1} \leq d_t + c_0 d_t / t + L_0 c_0 / t,$$

which gives us

$$d_{t+1} + L/\ell \leq (1 + c_0 \ell / t)(d_t + L_0/\ell).$$

Multiplying this inequality from 0 to $T-1$ yields

$$d_T \leq T^{c_0 \ell} L_0 / \ell,$$

which completes the proof. □

Proof of Lemma 6

Let $u = [1, 1, \cdots, 1, 0, \cdots, 0]^T \in \mathbb{R}^{2n}$. Then $\theta_S(w)$ satisfies $Q_S^T \theta_S(w) = u - b_0 e$, where $e = [1, 1, \cdots, 1]^T \in \mathbb{R}^{2n}$. It can be easily seen that $\|u - b_0 e\| \geq \sqrt{n}/2$.

We can also show that $\sigma_{\max}(Q_S) \leq 2\sigma_{\max} \cdot \sigma_{\max}(P)$, where $P \in \mathbb{R}^{2n \times m}$ is full row-rank and independent rows. Moreover, every row of P has covariance matrix $I_m/\sqrt{m}$. Then by random matrix theory (see [127]), we have $\sigma_{\max}(P) \leq \mathcal{O}(\sqrt{m}/\sqrt{m} - C\sqrt{n}/\sqrt{m} + \log(1/\delta)/\sqrt{m}) = \mathcal{O}(1)$ with probability $1 - C\delta$. Therefore, we have $\theta_S(w) \geq \Omega(\sqrt{n})$. □

4.6.5 Experiments on GAN-training

In this section, we provide some numerical results to corroborate our theoretical findings.

Setup

We train a GAN on MNIST data using two algorithms – GDA and GDMax. Since the stability is improved by using adaptive methods like Adam, we use Adam-descent-ascent (ADA) and Adam-descent-max (ADMax) instead. ADA simultaneously trains the generator and the discriminator, while ADMax trains the optimal discriminator for each generator step. We simulate this by taking 10 steps of ascent for every descent step. Figure 4-1 plots the images generated by GANs trained using these two algorithms. Finally, in Figure 4-2, we plot the norms of the discriminator trained by these two algorithms.

Results

Figure 4-1 plots the images generated by GANs trained using GDA and GDMax (using Adam instead of the simple gradient step). As predicted by the theory in Section 4.6.4, we can see that GDA produces better images than the corresponding GAN trained using GDMax. Furthermore, the claim that $C_e >> C_p$ can be seen from Figure 4-2 where we see that the norm of the discriminator trained using GDMax is much larger than the norm of the discriminator trained using GDA. This follows from the results in Section 4.4.2. GDMax trains the discriminator to exactly distinguish between the empirical data generated by the true and fake distributions. Therefore, when they are nearly the same, their empirical distributions would be close as well. This would imply that the discriminator would need to have a very large slope (Lipschitz constant) to exactly distinguish between the two empirical datasets, and this in turn leads to a large discriminator norm (which captures the Lipschitz

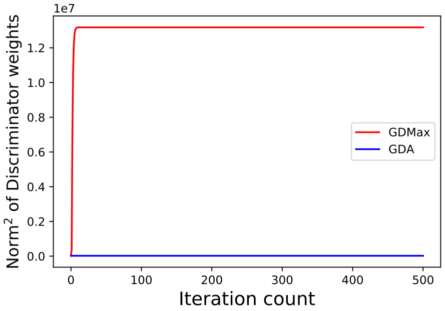


Figure 4-2: Comparison of the norm squares of discriminator weights.

constant of the discriminator).

4.6.6 Generalization Error for Primal-Dual Risk

If the saddle-point exists, the *primal-dual risk* is often a good measure of generalization:

Definition 7. [Primal-dual risk] The population and empirical primal-dual (PD) risks are defined as:

$$\Delta^{PD}(w,\theta) = \max_{\theta' \in \Theta} r(w,\theta') - \min_{w' \in W} r(w',\theta),$$

and

$$\Delta_S^{PD}(w,\theta) = \max_{\theta' \in \Theta} r_S(w,\theta') - \min_{w' \in W} r_S(w',\theta).$$

A point (w,θ) is called a saddle-point of r_S (or r) if $\Delta_S^{PD}(w,\theta) = 0$ (or $\Delta^{PD}(w,\theta) = 0$). Furthermore, if a saddle-point (w_S,θ_S) exists for $r_S(\cdot,\cdot)$, we have $w_S = \min_{w \in W} r_S(w)$. Moreover, if $w_S \in \arg\min_{w \in W} r_S(w)$ and $\theta_S \in \arg\max_{\theta \in \Theta} r_S(w_S,\theta)$, then (w_S,θ_S) is a saddle point of $r_S(\cdot,\cdot)$.

Notice that if we can find an approximate saddle point (w_S,θ_S) of $r_S(w,\theta)$, i.e., $\Delta_S^{PD}(w_S,\theta_S) < \epsilon$ and guarantee that $\Delta^{PD}(w_S,\theta_S) - \Delta_S^{PD}(w_S,\theta_S)$ is small, we can guarantee that $\Delta(w_S,\theta_S)$ is small and therefore (w_S,θ_S) is an approximate saddle point of $r(\cdot,\cdot)$. Hence if the saddle point exists for $r_S(\cdot,\cdot)$, the generalization error of the primal-dual risk can be a good measure for the generalization of the solution to the empirical problem. We define the expected generalization error for the primal-dual risk as follows:

Definition 8. The generalization error for the primal-dual risk is defined as

$$\zeta_{gen}^{PD}(A) = E_S E_A[\Delta^{PD}(w_S^A, \theta_S^A) - \Delta_S^{PD}(w_S^A, \theta_S^A)].$$

The generalization of the primal-dual risk for convex-concave problems

Similar to Definition 6, we define the W-capacity as follows:

Definition 9 (W-Capacity)**.** Let

$$W^*(\theta) = \min_{w \in W} r(w, \theta), \text{ and } \quad W_S(\theta) = \min_{w \in W} r_S(w, \theta).$$

The W-capacities C_e^w and C_p^w are defined as

$$\begin{aligned} C_p^w &= \max_{\theta} \operatorname{dist}(0, W^*(\theta) \\ C_e^w &= \max_{S,\theta} \operatorname{dist}(0, W_S(\theta)). \end{aligned} \tag{4.6.19}$$

Next, we also define the following:

Definition 10. Let $f^-(\theta, w; z) = -f(w, \theta; z)$. We first have

$$r^-(\theta, w) = E_{z \sim P_z}[f^-(\theta, w; z)], \qquad r_S^-(\theta, w) = \frac{1}{n}\sum_{i=1}^{n} f^-(\theta, w; z_i). \tag{4.6.20}$$

Furthermore, we define:

$$\begin{aligned} r^-(\theta) &= \max_{w \in W} r^-(\theta, w) = -(\min_{w \in W} r(w, \theta)) \\ r_S^-(\theta) &= \max_{w \in W} r_S^-(\theta, w) = -(\min_{w \in W} r_S(w, \theta)). \end{aligned} \tag{4.6.21}$$

Now, we have the following bound for the generalization error of the primal-dual risk, $\zeta_{gen}^{PD}(A)$ for an ϵ-stable Algorithm A:

Theorem 7. Suppose that Algorithm A is ϵ-stable. The generalization error $\zeta_{gen}^{PD}(A)$ for convex-concave problem, i.e., when $f(\cdot,\cdot;z)$ is convex-concave for all z, is bounded by:

$$\zeta_{gen}^{PD}(A) \leq \left(\sqrt{4L\ell C_p^2} + \sqrt{4L\ell(C_p^w)^2}\right)\sqrt{\epsilon} + 2\epsilon L.$$

Proof. Notice that

$$\begin{aligned} \zeta_{gen}^{PD}(A) &= E_S E_A[\Delta^{PD}(w_S^A, \theta_S^A) - \Delta_S^{PD}(w_S^A, \theta_S^A)] && (4.6.22) \\ &= E_S E_A[r(w_S^A) - r_S(w_S^A)] + E_S E_A[r^-(\theta_S^A) - r_S^-(\theta_S^A)]. && (4.6.23) \end{aligned}$$

The two terms can be bounded by Lemma 1 respectively. By Lemma 1, we have

$$E_S E_A[r(w_S^A) - r_S(w_S^A)] \leq \sqrt{4L\ell C_p^2}\sqrt{\epsilon} + \epsilon L$$

and

$$E_S E_A[r^-(\theta_S^A) - r_S^-(\theta_S^A)] \leq \sqrt{4L\ell(C_p^w)^2}\sqrt{\epsilon} + \epsilon L.$$

Combining these two inequalities yields the desired result. □

$\zeta_{gen}^{PD}(T)$ for the proximal point algorithm

In this section, we study the generalization behavior of the proximal point algorithm (PPA) ((See Equation (3) in [43])). By [43], the stability of T steps of PPA can be bounded as follows:

Lemma 14 ([43])**.** The stability of T steps of PPA can be bounded by $\epsilon \leq \mathcal{O}(T/n)$.

Therefore, substituting the result of Lemma 14 in Theorem 7, we have the following bound for ζ_{gen} for T steps of PPA:

Theorem 8. After T steps of PPA, the generalization error of the primal-dual risk can be bounded by:

$$\zeta_{gen}^{PD}(T) \leq \mathcal{O}\left(\sqrt{T/n} + T/n\right).$$

The population primal-dual risk of PPA

Finally, we give the population primal-dual risk after T steps of PPA. By [90], we have the following convergence result of PPA.

Lemma 15 ([90])**.** Let (w_S^t, θ_S^t) be the iterates obtained after t iterations of proximal point algorithm on the function $r_S(\cdot,\cdot)$ and $\bar{w}_S^t = \frac{1}{t}\sum_{i=1}^t w_S^i, \bar{\theta}_S^t = \frac{1}{t}\sum_{i=1}^t \theta_S^i$ be the averaged iterates. Then we have

$$\Delta_S^{PD}(\bar{w}_S^T, \bar{\theta}_S^T) \leq \ell(C_e^2 + (C_e^w)^2)/T.$$

Combining Lemma 15 and Theorem 8, we have the following result:

Theorem 9. Let (w_S^t, θ_S^t) be the iterates obtained after t iterations of proximal point algorithm on the function $r_S(\cdot,\cdot)$ and $\bar{w}_S^t = \frac{1}{t}\sum_{i=1}^t w_S^i, \bar{\theta}_S^t = \frac{1}{t}\sum_{i=1}^t \theta_S^i$ be the averaged iterates. Then, the expected population primal-dual risk at the point $(\bar{w}_S^t, \bar{\theta}_S^t)$ can be bounded by:

$$E_S[\Delta^{PD}(w_S^t, \theta_S^t)] \leq \mathcal{O}\left(1/T + \sqrt{T/n} + T/n\right).$$

Chapter 5

Conclusions

In this thesis, we study minimax formulations, i.e., we look at problems of the form

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} f(x, y)$$

We study two aspects of algorithms that solve these problems - convergence and generalization.

In Chapters 2 and 3 we study the convergence properties of algorithms which solve this optimization problem. In Chapter 2, we look at the case where the function f is convex-concave. We show that simple algorithms like Gradient-Descent Ascent (GDA) fail to converge in this setting. Motivated by this, we look at algorithms like the Optimistic Gradient Descent Ascent (OGDA) and the Extra-Gradient (EG) method and show that these algorithms in fact converge (at an optimal rate) to the desired solution. The main insight used is the fact that these algorithms are trying to approximate the Proximal Point (PP) method, which is a conceptual method that converges at an arbitrarily fast rate. Using this connection, we establish a unified analysis for OGDA and EG and establish their convergence in the convex-concave setting.

Next, in Chapter 3 we extend our studies to the nonconvex-nonconcave setting. Finding a solution to this problem is hard in general. We restrict our attention to the case of zero-sum games where the policies are parametrized using the softmax parametrization. In this case, we show that using the ideas of optimism developed in Chapter 2, we, in fact, get optimal convergence rates to the solution. Several open questions remain in this direction. For example, what other ‘non-convexities’ can we handle, i.e, which class of hidden nonconvex-nonconcave games are tractable? Furthermore, for other classes of nonconvexities, cen we develop similar ideas to optimism to get the optimal convergence rates? As in most machine learning applications, the nonconvexties in the loss function arise due to parametrization of the function class (say using neural networks). Therefore, if we have a handle on how to deal with such nonconvexities, while exploiting the convexity of the original loss function, we can hopefully provide tractable algorithms for several machine learning algorithms with provable guarantees.

Finally, in Chapter 4, we study the generalization behavior of these algorithms. Most ML applications use a loss function which involve an expectation over the data distribution. Since we do no have access to this data distribution, we have to settle for the empirical average over the given dataset. The question we study in this chapter is when is the solution to the empirical minimax problem a good proxy for the solution to the population problem. We show that existing metrics like the primal risk and the primal-dual risk are inadequate and does not capture the generalization performance. We then propose a new metric, the primal gap, which encapsulates the generalization behavior of minimax learners. We provide (optimal) generalization bounds for this new metric, and finally use these bounds to compare the generalization performance of several popular algorithms like GDA and GDMax.

Bibliography

[1] Alekh Agarwal, Sham M Kakade, Jason D Lee, and Gaurav Mahajan. Optimality and approximation with policy gradient methods in Markov decision processes. *arXiv preprint arXiv:1908.00261*, 2019.

[2] Ahmet Alacaoglu, Luca Viano, Niao He, and Volkan Cevher. A natural actor-critic framework for zero-sum Markov games. In *International Conference on Machine Learning*, pages 307–366. PMLR, 2022.

[3] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning*, pages 214–223, 2017.

[4] Sanjeev Arora, Rong Ge, Yingyu Liang, Tengyu Ma, and Yi Zhang. Generalization and equilibrium in generative adversarial nets (GANs). In *International Conference on Machine Learning*, pages 224–232. PMLR, 2017.

[5] Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: A meta-algorithm and applications. *Theory of Computing*, 8(1):121–164, 2012.

[6] Idan Attias, Aryeh Kontorovich, and Yishay Mansour. Improved generalization bounds for robust learning. In *Algorithmic Learning Theory*, pages 162–183. PMLR, 2019.

[7] Waïss Azizian, Ioannis Mitliagkas, Simon Lacoste-Julien, and Gauthier Gidel. A tight and unified analysis of gradient-based methods for a whole spectrum of differentiable games. In *International Conference on Artificial Intelligence and Statistics*, pages 2863–2873. PMLR, 2020.

[8] Yu Bai, Tengyu Ma, and Andrej Risteski. Approximability of discriminators implies diversity in GANs. *arXiv preprint arXiv:1806.10586*, 2018.

[9] James P Bailey and Georgios Piliouras. Multiplicative weights update in zero-sum games. In *ACM Conference on Economics and Computation*, pages 321–338, 2018.

[10] Tamer Basar and Geert Jan Olsder. *Dynamic noncooperative game theory*, volume 23. Siam, 1999.

[11] Heinz H Bauschke, Jonathan M Borwein, and Patrick L Combettes. Bregman monotone optimization algorithms. *SIAM Journal on control and optimization*, 42(2):596–636, 2003.

[12] Heinz H Bauschke, Patrick L Combettes, et al. *Convex analysis and monotone operator theory in Hilbert spaces*, volume 408. Springer, 2011.

[13] Amir Beck and Marc Teboulle. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31(3):167–175, 2003.

[14] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

[15] Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust optimization*, volume 28. Princeton University Press, 2009.

[16] Olivier Bousquet and André Elisseeff. Stability and generalization. *Journal of Machine Learning Research*, 2:499–526, 2002.

[17] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

[18] Satchuthananthavale RK Branavan, Harr Chen, Luke S Zettlemoyer, and Regina Barzilay. Reinforcement learning for mapping instructions to actions. Association for Computational Linguistics, 2009.

[19] Ronald E Bruck Jr. On the weak convergence of an ergodic iteration for the solution of variational inequalities for monotone operators in hilbert space. *Journal of Mathematical Analysis and Applications*, 61(1):159–164, 1977.

[20] Jingjing Bu, Lillian J Ratliff, and Mehran Mesbahi. Global convergence of policy gradient for sequential zero-sum linear quadratic dynamic games. *arXiv preprint arXiv:1911.04672*, 2019.

[21] Regina S Burachik, Alfredo N Iusem, and Benar Fux Svaiter. Enlargement of monotone operators with applications to variational inequalities. *Set-Valued Analysis*, 5(2):159–180, 1997.

[22] Shicong Cen, Chen Cheng, Yuxin Chen, Yuting Wei, and Yuejie Chi. Fast global convergence of natural policy gradient methods with entropy regularization. *Operations Research*, 2021.

[23] Shicong Cen, Yuting Wei, and Yuejie Chi. Fast policy extragradient methods for competitive games with entropy regularization. *arXiv preprint arXiv:2105.15186*, 2021.

[24] Nicolo Cesa-Bianchi and Gabor Lugosi. *Prediction, Learning, and Games.* Cambridge University Press, 2006.

[25] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*, 40(1):120–145, 2011.

[26] George HG Chen and R Tyrrell Rockafellar. Convergence rates in forward–backward splitting. *SIAM Journal on Optimization*, 7(2):421–444, 1997.

[27] Yichen Chen and Mengdi Wang. Stochastic primal-dual methods and sample complexity of reinforcement learning. *arXiv preprint arXiv:1612.02516*, 2016.

[28] Yunmei Chen, Guanghui Lan, and Yuyuan Ouyang. Optimal primal-dual methods for a class of saddle point problems. *SIAM Journal on Optimization*, 24(4):1779–1814, 2014.

[29] Chao-Kai Chiang, Tianbao Yang, Chia-Jung Lee, Mehrdad Mahdavi, Chi-Jen Lu, Rong Jin, and Shenghuo Zhu. Online optimization with gradual variations. In *COLT 2012 - The 25th Annual Conference on Learning Theory, June 25-27, 2012, Edinburgh, Scotland*, pages 6.1–6.20, 2012.

[30] 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. PMLR, 2018.

[31] C Daskalakis and Ioannis Panageas. Last-iterate convergence: Zero-sum games and constrained min-max optimization. In *Innovations in Theoretical Computer Science (ITCS)*, 2019.

[32] Constantinos Daskalakis, Dylan J Foster, and Noah Golowich. Independent policy gradient methods for competitive reinforcement learning. In *Advances in Neural Information Processing Systems*, 2020.

[33] Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. Training gans with optimism. *arXiv preprint arXiv:1711.00141*, 2017.

[34] Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. Training gans with optimism. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*, 2018.

[35] Constantinos Daskalakis and Ioannis Panageas. The limit points of (optimistic) gradient descent in min-max optimization. In *Neural Information Processing Systems*, 2018.

[36] Constantinos Daskalakis, Stratis Skoulakis, and Manolis Zampetakis. The complexity of constrained min-max optimization. In *Proceedings of Symposium on Theory of Computing*, pages 1466–1478, 2021.

[37] Dongsheng Ding, Chen-Yu Wei, Kaiqing Zhang, and Mihailo Jovanovic. Independent policy gradient for large-scale Markov potential games: Sharper rates, function approximation, and game-agnostic convergence. In *International Conference on Machine Learning*, pages 5166–5220. PMLR, 2022.

[38] Simon Du, Jason Lee, Haochuan Li, Liwei Wang, and Xiyu Zhai. Gradient descent finds global minima of deep neural networks. In *International conference on machine learning*, pages 1675–1685. PMLR, 2019.

[39] Simon S. Du and Wei Hu. Linear convergence of the primal-dual gradient method for convex-concave saddle point problems without strong convexity. In *The 22nd International Conference on Artificial Intelligence and Statistics, AISTATS 2019, 16-18 April 2019, Naha, Okinawa, Japan*, pages 196–205, 2019.

[40] Laurent El Ghaoui and Hervé Lebret. Robust solutions to least-squares problems with uncertain data. *SIAM Journal on matrix analysis and applications*, 18(4):1035–1064, 1997.

[41] Francisco Facchinei and Jong-Shi Pang. *Finite-dimensional variational inequalities and complementarity problems.* Springer Science & Business Media, 2007.

[42] Alireza Fallah, Asuman Ozdaglar, and Sarath Pattathil. An optimal multistage stochastic gradient method for minimax problems. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 3573–3579. IEEE, 2020.

[43] Farzan Farnia and Asuman Ozdaglar. Train simultaneously, generalize better: Stability of gradient-based minimax learners. In *International Conference on Machine Learning*, pages 3174–3185. PMLR, 2021.

[44] Farzan Farnia, Jesse M Zhang, and David Tse. Generalizable adversarial training via spectral normalization. *arXiv preprint arXiv:1811.07457*, 2018.

[45] Maryam Fazel, Rong Ge, Sham M Kakade, and Mehran Mesbahi. Global convergence of policy gradient methods for the linear quadratic regulator. In *International Conference on Machine Learning*, 2018.

[46] Soheil Feizi, Farzan Farnia, Tony Ginart, and David Tse. Understanding GANs in the LQG setting: Formulation, generalization and stability. *IEEE Journal on Selected Areas in Information Theory*, 1(1):304–311, 2020.

[47] Jerzy Filar and Koos Vrieze. *Competitive Markov Decision Processes.* Springer Science & Business Media, 2012.

[48] 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.

[49] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

[50] Matthieu Geist, Bruno Scherrer, and Olivier Pietquin. A theory of regularized Markov decision processes. In *International Conference on Machine Learning*, pages 2160–2169, 2019.

[51] Gauthier Gidel, Hugo Berard, Pascal Vincent, and Simon Lacoste-Julien. A variational inequality perspective on generative adversarial nets. *arXiv preprint arXiv:1802.10551*, 2018.

[52] Noah Golowich, Sarath Pattathil, and Constantinos Daskalakis. Tight last-iterate convergence rates for no-regret learning in multi-player games. *arXiv preprint arXiv:2010.13724*, 2020.

[53] Noah Golowich, Sarath Pattathil, Constantinos Daskalakis, and Asuman Ozdaglar. Last iterate is slower than averaged iterate in smooth convex-concave saddle point problems. In *Conference on Learning Theory*, pages 1758–1784. PMLR, 2020.

[54] 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.

[55] Benjamin Grimmer, Haihao Lu, Pratik Worah, and Vahab Mirrokni. The landscape of the proximal point method for nonconvex-nonconcave minimax optimization. *arXiv preprint arXiv:2006.08667*, 2020.

[56] Osman Güler. On the convergence of the proximal point algorithm for convex minimization. *SIAM Journal on Control and Optimization*, 29(2):403–419, 1991.

[57] Osman Güler. New proximal point algorithms for convex minimization. *SIAM Journal on Optimization*, 2(4):649–664, 1992.

[58] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. *arXiv preprint arXiv:1801.01290*, 2018.

[59] Ben M Hambly, Renyuan Xu, and Huining Yang. Policy gradient methods find the Nash equilibrium in n-player general-sum linear-quadratic games. *Available at SSRN 3894471*, 2021.

[60] Erfan Yazdandoost Hamedani and Necdet Serhat Aybat. A primal-dual algorithm for general convex-concave saddle point problems. *arXiv preprint arXiv:1803.01401*, 2018.

[61] Moritz Hardt, Benjamin Recht, and Yoram Singer. Train faster, generalize better: Stability of stochastic gradient descent. In *33rd International Conference on Machine Learning, ICML 2016*, pages 1868–1877. International Machine Learning Society (IMLS), 2016.

[62] Martin Hast, KJ Astrom, Bo Bernhardsson, and Stephen Boyd. PID design by convex-concave optimization. In *Control Conference (ECC), 2013 European*, pages 4460–4465. Citeseer, 2013.

[63] Trevor Hastie, Robert Tibshirani, and Jerome H Friedman. *The elements of statistical learning: Data mining, inference, and prediction*, volume 2. Springer, 2009.

[64] Yu-Guan Hsieh, Franck Iutzeler, Jérôme Malick, and Panayotis Mertikopoulos. On the convergence of single-call stochastic extra-gradient methods. *arXiv preprint arXiv:1908.08465*, 2019.

[65] Minhui Huang, Kaiyi Ji, Shiqian Ma, and Lifeng Lai. Efficiently escaping saddle points in bilevel optimization. *arXiv preprint arXiv:2202.03684*, 2022.

[66] Adam Ibrahim, Waıss Azizian, Gauthier Gidel, and Ioannis Mitliagkas. Linear lower bounds and conditioning of differentiable games. In *International Conference on Machine Learning*, pages 4583–4593. PMLR, 2020.

[67] 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. PMLR, 2020.

[68] Sham M Kakade. A natural policy gradient. In *Advances in Neural Information Processing Systems*, pages 1531–1538, 2002.

[69] Weiwei Kong, Jefferson G Melo, and Renato DC Monteiro. Complexity of a quadratic penalty accelerated inexact proximal point method for solving linearly constrained nonconvex composite programs. *SIAM Journal on Optimization*, 29(4):2566–2593, 2019.

[70] GM Korpelevich. The extragradient method for finding saddle points and other problems. *Matecon*, 12:747–756, 1976.

[71] Steven G Krantz and Harold R Parks. *The Implicit Function Theorem: History, Theory, and Applications.* Springer Science & Business Media, 2012.

[72] Guanghui Lan. Policy mirror descent for reinforcement learning: Linear convergence, new sampling complexity, and generalized problem classes. *arXiv preprint arXiv:2102.00135*, 2021.

[73] Yunwen Lei, Zhenhuan Yang, Tianbao Yang, and Yiming Ying. Stability and generalization of stochastic gradient methods for minimax problems. *arXiv preprint arXiv:2105.03793*, 2021.

[74] Stefanos Leonardos, Will Overman, Ioannis Panageas, and Georgios Piliouras. Global convergence of multi-agent policy gradient in Markov potential games. *arXiv preprint arXiv:2106.01969*, 2021.

[75] Tengyuan Liang and James Stokes. Interaction matters: A note on non-asymptotic local convergence of generative adversarial networks. In *The 22nd International Conference on Artificial Intelligence and Statistics, AISTATS 2019, 16-18 April 2019, Naha, Okinawa, Japan*, pages 907–915, 2019.

[76] 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. PMLR, 2020.

[77] Tianyi Lin, Chi Jin, and Michael I Jordan. On gradient descent ascent for nonconvex-concave minimax problems. *arXiv preprint arXiv:1906.00331*, 2019.

[78] Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, OpenAI Pieter Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. In *Advances in Neural Information Processing Systems*, pages 6379–6390, 2017.

[79] Cong Ma, Kaizheng Wang, Yuejie Chi, and Yuxin Chen. Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion, and blind deconvolution. *Foundations of Computational Mathematics*, 2019.

[80] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. *stat*, 1050:9, 2017.

[81] Yura Malitsky and Thomas Pock. A first-order primal-dual algorithm with linesearch. *SIAM Journal on Optimization*, 28(1):411–432, 2018.

[82] Yura Malitsky and Matthew K Tam. A forward-backward splitting method for monotone inclusions without cocoercivity. *arXiv preprint arXiv:1808.04162*, 2018.

[83] Bernard Martinet. Brève communication. régularisation d'inéquations variationnelles par approximations successives. *Revue française d'informatique et de recherche opérationnelle. Série rouge*, 4(R3):154–158, 1970.

[84] Richard D McKelvey and Thomas R Palfrey. Quantal response equilibria for normal form games. *Games and economic behavior*, 10(1):6–38, 1995.

[85] Jincheng Mei, Chenjun Xiao, Csaba Szepesvari, and Dale Schuurmans. On the global convergence rates of softmax policy gradient methods. In *International Conference on Machine Learning*, pages 6820–6829. PMLR, 2020.

[86] Panayotis Mertikopoulos and William H Sandholm. Learning in games via reinforcement and regularization. *Mathematics of Operations Research*, 41(4):1297–1324, 2016.

[87] Andjela Mladenovic, Iosif Sakos, Gauthier Gidel, and Georgios Piliouras. Generalized natural gradient flows in hidden convex-concave games and gans. In *International Conference on Learning Representations*, 2021.

[88] Aryan Mokhtari, Asuman Ozdaglar, and Sarath Pattathil. A unified analysis of extra-gradient and optimistic gradient methods for saddle point problems: Proximal point approach. pages 1497–1507, 2020.

[89] Aryan Mokhtari, Asuman Ozdaglar, and Sarath Pattathil. A unified analysis of extra-gradient and optimistic gradient methods for saddle point problems: Proximal point approach. In *International Conference on Artificial Intelligence and Statistics*, pages 1497–1507. PMLR, 2020.

[90] Aryan Mokhtari, Asuman E Ozdaglar, and Sarath Pattathil. Convergence rate of $\mathcal{O}(1/\mathrm{k})$ for optimistic gradient and extragradient methods in smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 30(4):3230–3251, 2020.

[91] Renato DC Monteiro and Benar Fux Svaiter. On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. *SIAM Journal on Optimization*, 20(6):2755–2787, 2010.

[92] Vaishnavh Nagarajan and J Zico Kolter. Uniform convergence may be unable to explain generalization in deep learning. *Advances in Neural Information Processing Systems*, 32, 2019.

[93] Angelia Nedić and Asuman Ozdaglar. Subgradient methods for saddle-point problems. *Journal of optimization theory and applications*, 142(1):205–228, 2009.

[94] 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.

[95] Arkadi Nemirovski, Anatoli Juditsky, Guanghui Lan, and Alexander Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19(4):1574–1609, 2009.

[96] Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2013.

[97] Gergely Neu, Anders Jonsson, and Vicenç Gómez. A unified view of entropy-regularized Markov decision processes. *arXiv preprint arXiv:1705.07798*, 2017.

[98] J v Neumann. Zur theorie der gesellschaftsspiele. *Mathematische Annalen*, 100(1):295–320, 1928.

[99] 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.

[100] OpenAI. Openai five. `https://blog.openai.com/openai-five/`, 2018.

[101] Dmitrii M Ostrovskii, Babak Barazandeh, and Meisam Razaviyayn. Nonconvex-nonconcave min-max optimization with a small maximization domain. *arXiv preprint arXiv:2110.03950*, 2021.

[102] Dmitrii M Ostrovskii, Andrew Lowy, and Meisam Razaviyayn. Efficient search of first-order nash equilibria in nonconvex-concave smooth min-max problems. *SIAM Journal on Optimization*, 31(4):2508–2538, 2021.

[103] Leonid Denisovich Popov. A modification of the arrow-hurwicz method for search of saddle points. *Mathematical Notes*, 28(5):845–848, 1980.

[104] Shuang Qiu, Xiaohan Wei, Jieping Ye, Zhaoran Wang, and Zhuoran Yang. Provably efficient fictitious play policy optimization for zero-sum Markov games with structured transitions. In *International Conference on Machine Learning*, pages 8715–8725. PMLR, 2021.

[105] Hassan Rafique, Mingrui Liu, Qihang Lin, and Tianbao Yang. Non-convex min-max optimization: Provable algorithms and applications in machine learning. *arXiv preprint arXiv:1810.02060*, 2018.

[106] Hassan Rafique, Mingrui Liu, Qihang Lin, and Tianbao Yang. Weakly-convex concave min-max optimization: Provable algorithms and applications in machine learning. *arXiv preprint arXiv:1810.02060*, 2018.

[107] Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. In *COLT 2013 - The 26th Annual Conference on Learning Theory, June 12-14, 2013, Princeton University, NJ, USA*, pages 993–1019, 2013.

[108] Sasha Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. In *Advances in Neural Information Processing Systems*, pages 3066–3074, 2013.

[109] R Tyrrell Rockafellar. Monotone operators and the proximal point algorithm. *SIAM journal on control and optimization*, 14(5):877–898, 1976.

[110] J Ben Rosen. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica: Journal of the Econometric Society*, pages 520–534, 1965.

[111] Maziar Sanjabi, Jimmy Ba, Meisam Razaviyayn, and Jason D Lee. On the convergence and robustness of training gans with regularized optimal transport. In *Advances in Neural Information Processing Systems*, pages 7091–7101, 2018.

[112] Muhammed Sayin, Kaiqing Zhang, David Leslie, Tamer Basar, and Asuman Ozdaglar. Decentralized q-learning in zero-sum markov games. *Advances in Neural Information Processing Systems*, 34:18320–18334, 2021.

[113] Ludwig Schmidt, Shibani Santurkar, Dimitris Tsipras, Kunal Talwar, and Aleksander Madry. Adversarially robust generalization requires more data. *Advances in Neural Information Processing Systems*, 31, 2018.

[114] Mark Schmidt, Reza Babanezhad, Mohamed Ahmed, Aaron Defazio, Ann Clifton, and Anoop Sarkar. Non-uniform stochastic average gradient method for training conditional random fields. In *artificial intelligence and statistics*, pages 819–828, 2015.

[115] John Schulman, Sergey Levine, Pieter Abbeel, Michael Jordan, and Philipp Moritz. Trust region policy optimization. In *International Conference on Machine Learning*, pages 1889–1897, 2015.

[116] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[117] Shai Shalev-Shwartz, Ohad Shamir, Nathan Srebro, and Karthik Sridharan. Learnability, stability and uniform convergence. *The Journal of Machine Learning Research*, 11:2635–2670, 2010.

[118] Shai Shalev-Shwartz, Shaked Shammah, and Amnon Shashua. Safe, multi-agent, reinforcement learning for autonomous driving. *arXiv preprint arXiv:1610.03295*, 2016.

[119] Lloyd S Shapley. Stochastic games. *Proceedings of the National Academy of Sciences*, 39(10):1095–1100, 1953.

[120] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of Go without human knowledge. *Nature*, 550(7676):354–359, 2017.

[121] Samuel Sokota, Ryan D'Orazio, J Zico Kolter, Nicolas Loizou, Marc Lanctot, Ioannis Mitliagkas, Noam Brown, and Christian Kroer. A unified approach to reinforcement learning, quantal response equilibria, and two-player zero-sum games. *arXiv preprint arXiv:2206.05825*, 2022.

[122] Marc Teboulle. Convergence of proximal-like algorithms. *SIAM Journal on Optimization*, 7(4):1069–1083, 1997.

[123] Kiran Koshy Thekumparampil, Prateek Jain, Praneeth Netrapalli, and Sewoong Oh. Efficient algorithms for smooth minimax optimization. *arXiv preprint arXiv:1907.01543*, 2019.

[124] Paul Tseng. On linear convergence of iterative methods for the variational inequality problem. *Journal of Computational and Applied Mathematics*, 60(1-2):237–252, 1995.

[125] Paul Tseng. On linear convergence of iterative methods for the variational inequality problem. *Journal of Computational and Applied Mathematics*, 60(1-2):237–252, 1995.

[126] Vladimir N Vapnik. An overview of statistical learning theory. *IEEE transactions on neural networks*, 10(5):988–999, 1999.

[127] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.

[128] 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.

[129] Lingxiao Wang, Qi Cai, Zhuoran Yang, and Zhaoran Wang. Neural policy gradient methods: Global optimality and rates of convergence. *arXiv preprint arXiv:1909.01150*, 2019.

[130] Chen-Yu Wei, Chung-Wei Lee, Mengxiao Zhang, and Haipeng Luo. Linear last-iterate convergence in constrained saddle-point optimization. 2020.

[131] Chen-Yu Wei, Chung-Wei Lee, Mengxiao Zhang, and Haipeng Luo. Last-iterate convergence of decentralized optimistic gradient descent/ascent in infinite-horizon competitive Markov games. *arXiv preprint arXiv:2102.04540*, 2021.

[132] Colin Wei and Tengyu Ma. Improved sample complexities for deep neural networks and robust classification via an all-layer margin. In *International Conference on Learning Representations*, 2019.

[133] Bingzhe Wu, Shiwan Zhao, Chaochao Chen, Haoyang Xu, Li Wang, Xiaolu Zhang, Guangyu Sun, and Jun Zhou. Generalization in generative adversarial networks: A novel perspective from privacy protection. *Advances in Neural Information Processing Systems*, 32, 2019.

[134] Yue Xing, Qifan Song, and Guang Cheng. On the algorithmic stability of adversarial training. *Advances in Neural Information Processing Systems*, 34, 2021.

[135] Junchi Yang, Negar Kiyavash, and Niao He. Global convergence and variance-reduced optimization for a class of nonconvex-nonconcave minimax problems. *arXiv preprint arXiv:2002.09621*, 2020.

[136] Junchi Yang, Antonio Orvieto, Aurelien Lucchi, and Niao He. Faster single-loop algorithms for minimax optimization without strong concavity. *arXiv preprint arXiv:2112.05604*, 2021.

[137] Zhenhuan Yang, Shu Hu, Yunwen Lei, Kush R Varshney, Siwei Lyu, and Yiming Ying. Differentially private SGDA for minimax problems. *arXiv preprint arXiv:2201.09046*, 2022.

[138] Dong Yin, Ramchandran Kannan, and Peter Bartlett. Rademacher complexity for adversarially robust generalization. In *International Conference on Machine Learning*, pages 7085–7094. PMLR, 2019.

[139] Chao Yu, Akash Velu, Eugene Vinitsky, Yu Wang, Alexandre Bayen, and Yi Wu. The surprising effectiveness of MAPPO in cooperative, multi-agent games. *arXiv preprint arXiv:2103.01955*, 2021.

[140] Wenhao Zhan, Shicong Cen, Baihe Huang, Yuxin Chen, Jason D Lee, and Yuejie Chi. Policy mirror descent for regularized reinforcement learning: A generalized framework with linear convergence. *arXiv preprint arXiv:2105.11066*, 2021.

[141] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, 64(3):107–115, 2021.

[142] Jiawei Zhang, Peijun Xiao, Ruoyu Sun, and Zhiquan Luo. A single-loop smoothed gradient descent-ascent algorithm for nonconvex-concave min-max problems. *Advances in Neural Information Processing Systems*, 33:7377–7389, 2020.

[143] Junyu Zhang, Mingyi Hong, Mengdi Wang, and Shuzhong Zhang. Generalization bounds for stochastic saddle point problems. In *International Conference on Artificial Intelligence and Statistics*, pages 568–576. PMLR, 2021.

[144] Kaiqing Zhang, Sham M Kakade, Tamer Başar, and Lin F Yang. Model-based multi-agent RL in zero-sum Markov games with near-optimal sample complexity. *arXiv preprint arXiv:2007.07461*, 2020.

[145] Kaiqing Zhang, Alec Koppel, Hao Zhu, and Tamer Başar. Global convergence of policy gradient methods to (almost) locally optimal policies. *arXiv preprint arXiv:1906.08383*, 2019.

[146] Kaiqing Zhang, Zhuoran Yang, and Tamer Başar. Policy optimization provably converges to Nash equilibria in zero-sum linear quadratic games. In *Advances in Neural Information Processing Systems*, 2019.

[147] Pengchuan Zhang, Qiang Liu, Dengyong Zhou, Tao Xu, and Xiaodong He. On the discrimination-generalization tradeoff in GANs. *arXiv preprint arXiv:1711.02771*, 2017.

[148] Runyu Zhang, Qinghua Liu, Huan Wang, Caiming Xiong, Na Li, and Yu Bai. Policy optimization for markov games: Unified framework and faster convergence. *arXiv preprint arXiv:2206.02640*, 2022.

[149] Runyu Zhang, Zhaolin Ren, and Na Li. Gradient play in multi-agent Markov stochastic games: Stationary points and convergence. *arXiv preprint arXiv:2106.00198*, 2021.

[150] Siqi Zhang, Junchi Yang, Cristóbal Guzmán, Negar Kiyavash, and Niao He. The complexity of nonconvex-strongly-concave minimax optimization. In *Uncertainty in Artificial Intelligence*, pages 482–492. PMLR, 2021.

[151] Yulai Zhao, Yuandong Tian, Jason D Lee, and Simon S Du. Provably efficient policy gradient methods for two-player zero-sum markov games. *arXiv preprint arXiv:2102.08903*, 2021.