

Robust No-Regret Learning in Min-Max Stackelberg Games

Denizalp Goktas
Brown University
Computer Science
Providence, Rhode Island, USA
denizalp_goktas@brown.edu

Jiayi Zhao
Pomona College
Computer Science
Claremont, CA, USA
jz2019@mymail.pomona.edu

Amy Greenwald
Brown University
Computer Science
Providence, Rhode Island, USA
amy_greenwald@brown.edu

ABSTRACT

The behavior of no-regret learning algorithms is well understood in two-player min-max (i.e., zero-sum) games. In this paper, we investigate the behavior of no-regret learning in min-max games *with dependent strategy sets*, where the strategy of the first player constrains the behavior of the second. Such games are best understood as sequential, i.e., min-max Stackelberg, games. We consider two settings, one in which only the first player chooses their actions using a no-regret algorithm while the second player best responds, and one in which both players use no-regret algorithms. For the former case, we show that no-regret dynamics converge to a Stackelberg equilibrium. For the latter case, we introduce a new type of regret, which we call Lagrangian regret, and show that if both players minimize their Lagrangian regrets, then play converges to a Stackelberg equilibrium. We then observe that online mirror descent (OMD) dynamics in these two settings correspond respectively to a known nested (i.e., sequential) gradient descent-ascent (GDA) algorithm and a new simultaneous GDA-like algorithm, thereby establishing convergence of these algorithms to Stackelberg equilibrium. Finally, we analyze the robustness of OMD dynamics to perturbations by investigating online min-max Stackelberg games. We prove that OMD dynamics are robust for a large class of online min-max games with independent strategy sets. In the dependent case, we demonstrate the robustness of OMD dynamics experimentally by simulating them in online Fisher markets, a canonical example of a min-max Stackelberg game with dependent strategy sets.

CCS CONCEPTS

• **Mathematics of computing** → **Convex optimization**; • **Applied computing** → **Economics**; • **Computing methodologies** → **Multi-agent systems**.

KEYWORDS

Equilibrium Computation; Learning in Games; Market Dynamics

ACM Reference Format:

Denizalp Goktas, Jiayi Zhao, and Amy Greenwald. 2022. Robust No-Regret Learning in Min-Max Stackelberg Games. In *Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), Online, May 9–13, 2022, IFAAMAS*, 15 pages.

1 INTRODUCTION

Min-max optimization problems (i.e., zero-sum games) have been attracting a great deal of attention recently because of their applicability to problems in fairness in machine learning [10, 19, 40, 57],

generative adversarial imitation learning [8, 29], reinforcement learning [11], generative adversarial learning [55], adversarial learning [60], and statistical learning, e.g., learning parameters of exponential families [10]. These problems are often modelled as **min-max games**, i.e., constrained min-max optimization problems of the form: $\min_{\mathbf{x} \in X} \max_{\mathbf{y} \in Y} f(\mathbf{x}, \mathbf{y})$, where $f : X \times Y \rightarrow \mathbb{R}$ is continuous, and $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ are non-empty and compact. In **convex-concave min-max games**, where f is convex in \mathbf{x} and concave in \mathbf{y} , von Neumann and Morgenstern’s seminal minimax theorem holds [48]: i.e., $\min_{\mathbf{x} \in X} \max_{\mathbf{y} \in Y} f(\mathbf{x}, \mathbf{y}) = \max_{\mathbf{y} \in Y} \min_{\mathbf{x} \in X} f(\mathbf{x}, \mathbf{y})$, guaranteeing the existence of a saddle point, i.e., a point that is simultaneously a minimum of f in the \mathbf{x} -direction and a maximum of f in the \mathbf{y} -direction. Because of the minimax theorem, we can interpret the constrained optimization problem as a simultaneous-move, zero-sum game, where \mathbf{y}^* (resp. \mathbf{x}^*) is a best-response of the outer (resp. inner) player to the other’s action \mathbf{x}^* (resp. \mathbf{y}^*), in which case a saddle point is also called a minimax point or a Nash equilibrium.

In this paper, we study **min-max Stackelberg games** [25], i.e., constrained min-max optimization problems *with dependent feasible sets* of the form: $\min_{\mathbf{x} \in X} \max_{\mathbf{y} \in Y: g(\mathbf{x}, \mathbf{y}) \geq 0} f(\mathbf{x}, \mathbf{y})$, where $f : X \times Y \rightarrow \mathbb{R}$ is continuous, $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ are non-empty and compact, and $g(\mathbf{x}, \mathbf{y}) = (g_1(\mathbf{x}, \mathbf{y}), \dots, g_K(\mathbf{x}, \mathbf{y}))^T$ with $g_k : X \times Y \rightarrow \mathbb{R}$. Goktas and Greenwald observe that the minimax theorem does not hold in these games [25]. As a result, such games are more appropriately viewed as sequential, i.e., Stackelberg, games for which the relevant solution concept is the Stackelberg equilibrium,¹ where the outer player chooses $\hat{\mathbf{x}} \in X$ before the inner player responds with their choice of $\mathbf{y}(\hat{\mathbf{x}}) \in Y$ s.t. $g(\hat{\mathbf{x}}, \mathbf{y}(\hat{\mathbf{x}})) \geq 0$. The outer player’s objective, which is referred to as their **value function** in the economics literature [41] and which they seek to minimize, is defined as $V_X(\hat{\mathbf{x}}) = \max_{\mathbf{y} \in Y: g(\hat{\mathbf{x}}, \mathbf{y}) \geq 0} f(\hat{\mathbf{x}}, \mathbf{y})$. The inner player’s value function, $V_Y : X \rightarrow \mathbb{R}$, which they seek to maximize, is simply the objective function of the game, given the outer player’s action $\hat{\mathbf{x}}$: i.e., $V_Y(\hat{\mathbf{x}}) = f(\hat{\mathbf{x}}, \mathbf{y})$.

Goktas and Greenwald [25] proposed a polynomial-time first-order method by which to compute Stackelberg equilibria, which they called **nested gradient descent ascent (GDA)**. This method can be understood as an algorithm a third party might run to find an equilibrium, or as a game dynamic that the players might employ if their long-run goal were to reach an equilibrium. Rather than assume that players are jointly working towards the goal of reaching an equilibrium, it is often more reasonable to assume that they play so as to not regret their decisions: i.e., that they employ

Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), P. Faliszewski, V. Mascardi, C. Pelachaud, M.E. Taylor (eds.), May 9–13, 2022, Online. © 2022 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

¹Alternatively, one could view such games as pseudo-games (also known as abstract economies) [3], in which players move simultaneously under the unreasonable assumption that the moves they make will satisfy the game’s dependency constraints. Under this view, the relevant solution concept is generalized Nash equilibrium [20, 21].

a **no-regret learning algorithm**, which minimizes their loss in hindsight. It is well known that when both players in a repeated min-max game are no-regret learners, the players’ strategy profile over time converges to a Nash equilibrium in average iterates: i.e., empirical play converges to a Nash equilibrium (e.g., [23]).

In this paper, we investigate no-regret learning dynamics in repeated min-max Stackelberg games. We assume both an asymmetric and a symmetric setting. In the asymmetric setting, the outer player is a no-regret learner while the inner player best responds; in the symmetric setting, both players are no-regret learners. In the asymmetric case, we show that if the outer player uses a no-regret algorithm that achieves ε -asymmetric regret, then the outer player’s empirical play converges to their ε -Stackelberg equilibrium strategy. In the symmetric case, we introduce a new type of regret, which we call Lagrangian regret,² which assumes access to a solution oracle for the optimal KKT multipliers of the game’s constraints. We then show that if both players use no-regret algorithms that achieve ε -Lagrangian regrets, then the players’ empirical play converges to an ε -Stackelberg equilibrium.

Next, we restrict our attention to a specific no-regret dynamic, namely online mirror descent (OMD) [45]. Doing so yields two algorithms, max-oracle mirror descent (max-oracle MD) and nested mirror descent ascent (nested MDA) in the asymmetric setting, and a new simultaneous GDA-like algorithm [44] in the symmetric setting, which we call Lagrangian mirror descent ascent (LMDA). The first two algorithms converge to ε -Stackelberg equilibrium in $O(1/\varepsilon^2)$ and $O(1/\varepsilon^3)$ iterations, respectively, and the third, in $O(1/\varepsilon^2)$, when a Lagrangian solution oracle exists. As max-oracle gradient [25, 33] and nested GDA [25] are special cases of max-oracle MD and nested MDA, respectively, our convergence bounds complement Goktas and Greenwald’s best iterate convergence results, now proving average iterate convergence for both algorithms. Furthermore, our result on LMDA’s convergence rate suggests the computational superiority of LMDA over nested GDA, when a Lagrangian solution oracle exists. We also note that even when such an oracle does not exist, the Lagrangian solution can be treated as a hyperparameter of the algorithm allowing for a significant speed up in computation.

Finally, we analyze the robustness of OMD dynamics by investigating online min-max Stackelberg games, i.e., min-max Stackelberg games with arbitrary objective and constraint functions from one time step to the next. We prove that OMD dynamics are robust, in that even when the game changes, OMD dynamics track the changing equilibria closely, in a large class of online min-max games with independent strategy sets. In the dependent strategy set case, we demonstrate the robustness of OMD dynamics experimentally by simulating online Fisher markets, a canonical example of an (online) min-max Stackelberg game (with dependent strategy sets) [25]. Even when the Fisher market changes every time step, our OMD dynamics track the changing equilibria closely. These results are somewhat surprising, because optimization problems can be highly sensitive to perturbations of their inputs [5].

Our findings can be summarized as follows:

- In repeated min-max Stackelberg games, when the outer player is a no-regret learner and the inner-player best-responds,

the average of the outer player’s strategies converges to their Stackelberg equilibrium strategy.

- We introduce a new type of regret we call Lagrangian regret and show that in repeated min-max Stackelberg games when both players minimize Lagrangian regret, the average of the players’ strategies converge to a Stackelberg equilibrium.
- We provide convergence guarantees for max-oracle MD and nested MDA to an ε -Stackelberg equilibrium in $O(1/\varepsilon^2)$ and $O(1/\varepsilon^3)$ in average iterates, respectively.
- We introduce a simultaneous GDA-like algorithm, which we call LMDA, and prove that its average iterates converge to an ε -Stackelberg equilibrium in $O(1/\varepsilon^2)$ iterations.
- We prove that max-oracle MD and LMDA are robust to perturbations in a large class of online min-max games (with independent strategy sets).
- We run experiments with Fisher markets which suggest that max-oracle MD and LMDA are robust to perturbations in these online min-max Stackelberg games.

Related Work. Stackelberg games [67] have found important applications in the domain of security (e.g., [49, 59]) and environmental protection (e.g., [22]). These applications have thus far been modelled as Stackelberg games with independent strategy sets. Yet, the increased expressiveness of Stackelberg games with dependent strategy sets may make them a better model of the real world, as they provide the leader with more power to achieve a better outcome by constraining the follower’s choices.

The study of algorithms that compute competitive equilibria in Fisher markets was initiated by Devanur et al. [16], who provided a polynomial-time method for solving these markets assuming linear utilities. More recently, Cheung et al. [9] studied two price adjustment processes, tâtonnement and proportional response dynamics, in dynamic Fisher markets and showed that these price adjustment processes track the equilibrium of Fisher markets closely even when the market is subject to change.

The computation and learning of Nash and generalized Nash equilibrium in min-max games (with independent strategy sets) has been attracting a great deal of attention recently, because of the relevance of these problems to machine learning [1, 13–15, 17], specifically generative adversarial learning [27].

2 MATHEMATICAL PRELIMINARIES

Notation. We use Roman uppercase letters to denote sets (e.g., X), bold uppercase letters to denote matrices (e.g., \mathbf{X}), bold lowercase letters to denote vectors (e.g., \mathbf{p}), and Roman lowercase letters to denote scalar quantities, (e.g., c). We denote the i th row vector of a matrix (e.g., \mathbf{X}) by the corresponding bold lowercase letter with subscript i (e.g., \mathbf{x}_i). Similarly, we denote the j th entry of a vector (e.g., \mathbf{p} or \mathbf{x}_i) by the corresponding Roman lowercase letter with subscript j (e.g., p_j or x_{ij}). We denote the vector of ones of size n by $\mathbf{1}_n$. We denote the set of integers $\{1, \dots, n\}$ by $[n]$, the set of natural numbers by \mathbb{N} , the set of positive natural numbers by \mathbb{N}_+ , the set of real numbers by \mathbb{R} , the set of non-negative real numbers by \mathbb{R}_+ , and the set of strictly positive real numbers by \mathbb{R}_{++} . We denote the orthogonal projection operator onto a convex set C by Π_C , i.e., $\Pi_C(\mathbf{x}) = \arg \min_{\mathbf{y} \in C} \|\mathbf{x} - \mathbf{y}\|^2$. Given a sequence of iterates $\{\mathbf{z}^{(t)}\}_{t=1}^T \subset Z$, we denote the average iterate $\bar{\mathbf{z}}^{(T)} = \frac{1}{T} \sum_{t=1}^T \mathbf{z}^{(t)}$.

²We note that similar notions of Lagrangian regret have been used in other online learning settings (e.g., [4]), but to our knowledge, ours is the first game-theoretic analysis of Lagrangian regret minimization.

Game Definitions. A **min-max Stackelberg game**, (X, Y, f, \mathbf{g}) , is a two-player, zero-sum game, where one player, who we call the **outer** player (resp. the **inner** player), is trying to minimize their loss (resp. maximize their gain), defined by a continuous **objective function** $f : X \times Y \rightarrow \mathbb{R}$, by choosing a strategy from their non-empty and compact **strategy set** $X \subset \mathbb{R}^n$, and (resp. $Y \subset \mathbb{R}^m$) s.t. $\mathbf{g}(\mathbf{x}, \mathbf{y}) \geq \mathbf{0}$ where $\mathbf{g}(\mathbf{x}, \mathbf{y}) = (g_1(\mathbf{x}, \mathbf{y}), \dots, g_K(\mathbf{x}, \mathbf{y}))^T$ with $g_k : X \times Y \rightarrow \mathbb{R}$ continuous. A strategy profile $(\mathbf{x}, \mathbf{y}) \in X \times Y$ is said to be **feasible** iff for all $k \in [K]$, $g_k(\mathbf{x}, \mathbf{y}) \geq 0$. The function f maps a pair of strategies taken by the players $(\mathbf{x}, \mathbf{y}) \in X \times Y$ to a real value (i.e., a payoff), which represents the loss (resp. the gain) of the outer player (resp. the inner player). A min-max game is said to be convex-concave if the objective function f is convex-concave and X and Y are convex sets.

The relevant solution concept for Stackelberg games is the **Stackelberg equilibrium (SE)**: A strategy profile $(\mathbf{x}^*, \mathbf{y}^*) \in X \times Y$ s.t. $\mathbf{g}(\mathbf{x}^*, \mathbf{y}^*) \geq \mathbf{0}$ is an (ϵ, δ) -SE if $\max_{\mathbf{y} \in Y: \mathbf{g}(\mathbf{x}^*, \mathbf{y}) \geq \mathbf{0}} f(\mathbf{x}^*, \mathbf{y}) - \delta \leq f(\mathbf{x}^*, \mathbf{y}^*) \leq \min_{\mathbf{x} \in X} \max_{\mathbf{y} \in Y: \mathbf{g}(\mathbf{x}, \mathbf{y}) \geq \mathbf{0}} f(\mathbf{x}, \mathbf{y}) + \epsilon$. Intuitively, a (ϵ, δ) -SE is a point at which the outer player's (resp. inner player's) payoff is no more than ϵ (resp. δ) away from its optimum. A $(0, 0)$ -SE is guaranteed to exist in min-max Stackelberg games [25]. Note that when $\mathbf{g}(\mathbf{x}, \mathbf{y}) \geq \mathbf{0}$, for all $(\mathbf{x}, \mathbf{y}) \in X \times Y$, the game reduces to a min-max game (with independent strategy sets).

In a min-max Stackelberg game, the outer player's **best-response set** $\text{BR}_X \subset X$, defined as $\text{BR}_X = \arg \min_{\mathbf{x} \in X} V_X(\mathbf{x})$, is independent of the inner player's strategy, while the inner player's **best-response correspondence** $\text{BR}_Y : X \rightrightarrows Y$, defined as $\text{BR}_Y(\mathbf{x}) = \arg \max_{\mathbf{y} \in Y: \mathbf{g}(\mathbf{x}, \mathbf{y}) \geq \mathbf{0}} V_Y(\mathbf{y}; \mathbf{x})$, depends on the outer player's strategy. A $(0, 0)$ -Stackelberg equilibrium $(\mathbf{x}^*, \mathbf{y}^*) \in X \times Y$ is then a tuple of strategies such that $(\mathbf{x}^*, \mathbf{y}^*) \in \text{BR}_X \times \text{BR}_Y(\mathbf{x}^*)$.

An **online min-max Stackelberg game**, $\left\{ (X, Y, f^{(t)}, \mathbf{g}^{(t)}) \right\}$, is a sequence of min-max Stackelberg games played for T time periods. We define the players' value functions at time t in an online min-max Stackelberg game in terms of $f^{(t)}$ and $\mathbf{g}^{(t)}$. Note that when $\mathbf{g}^{(t)}(\mathbf{x}, \mathbf{y}) \geq \mathbf{0}$ for all $\mathbf{x} \in X, \mathbf{y} \in Y$ and all time periods $t \in [T]$, the game reduces to an online min-max game (with independent strategy sets). Moreover, if for all $t, t' \in [T]$, $f^{(t)} = f^{(t')}$, and $\mathbf{g}^{(t)} = \mathbf{g}^{(t')}$, then the game reduces to a **repeated min-max Stackelberg game**, which we denote simply by (X, Y, f, \mathbf{g}) .

Assumptions. All the theoretical results on min-max Stackelberg games in this paper rely on the following assumption(s):

ASSUMPTION 2.1. 1. (Slater's condition [61, 62]) $\forall \mathbf{x} \in X, \exists \widehat{\mathbf{y}} \in Y$ s.t. $g_k(\mathbf{x}, \widehat{\mathbf{y}}) > 0$, for all $k = 1, \dots, K$; 2. f, g_1, \dots, g_K are continuous and convex-concave; and 3. $\nabla_{\mathbf{x}} f, \nabla_{\mathbf{x}} g_1, \dots, \nabla_{\mathbf{x}} g_K$ are well-defined for all $(\mathbf{x}, \mathbf{y}) \in X \times Y$ and continuous in (\mathbf{x}, \mathbf{y}) ; 4. X and Y are convex.

We note that these assumptions are in line with previous work geared towards solving min-max Stackelberg games [25]. Part 1 of Assumption 2.1, Slater's condition, is a standard constraint qualification condition [6], which is needed to derive the optimality conditions for the inner player's maximization problem; without it the problem becomes analytically intractable. Part 2 of Assumption 2.1 ensures that the value function of the outer player is continuous and convex ([25], Proposition A1), so that the problem affords an efficient solution. Part 3 of Assumption 2.1 can be replaced by a weaker, subgradient boundedness assumption; however, for simplicity, we

assume this stronger condition. Finally, Part 4 of Assumption 2.1 guarantees that projections are polynomial-time operations.

Under Assumption 2.1, the following property holds of the outer player's value function.

PROPOSITION 2.2 ([25], PROPOSITION A.1). *Consider a min-max Stackelberg game (X, Y, f, \mathbf{g}) and suppose that Assumption 2.1 holds, then the outer player's value function $V(\mathbf{x}) = \max_{\mathbf{y} \in Y: \mathbf{g}(\mathbf{x}, \mathbf{y}) \geq \mathbf{0}} f(\mathbf{x}, \mathbf{y})$ is continuous and convex.*

Additional Definitions. Given two normed spaces $(X, \|\cdot\|)$ and $(Y, \|\cdot\|)$, the function $f : X \rightarrow Y$ is L_f -**Lipschitz-continuous** iff $\forall \mathbf{x}_1, \mathbf{x}_2 \in X, \|f(\mathbf{x}_1) - f(\mathbf{x}_2)\| \leq L_f \|\mathbf{x}_1 - \mathbf{x}_2\|$. If the gradient of f , ∇f , is $L_{\nabla f}$ -Lipschitz-continuous, we refer to f as $L_{\nabla f}$ -**Lipschitz-smooth**. A function $f : A \rightarrow \mathbb{R}$ is μ -**strongly convex** if $f(\mathbf{x}_1) \geq f(\mathbf{x}_2) + \langle \nabla_{\mathbf{x}} f(\mathbf{x}_2), \mathbf{x}_1 - \mathbf{x}_2 \rangle + \mu/2 \|\mathbf{x}_1 - \mathbf{x}_2\|^2$, and μ -**strongly concave** if $-f$ is μ -strongly convex.

Online Convex Optimization. An **online convex optimization problem (OCP)** is a decision problem in a dynamic environment which comprises a finite time horizon T , a compact, convex feasible set X , and a sequence of convex differentiable loss functions $\{\ell^{(t)}\}_{t=1}^T$, where $\ell^{(t)} : X \rightarrow \mathbb{R}$ for all $t \in [T]$. A solution to an OCP is a sequence $\{\mathbf{x}^{(t)}\}_{t=1}^T$ with each $\mathbf{x}^{(t)} \in X$. A preferred solution is one that minimizes **average regret**, given by $\text{Regret}^T(\{\mathbf{x}^t\}, \mathbf{x}) = \sum_{t=1}^T \frac{1}{T} \ell^{(t)}(\mathbf{x}^{(t)}) - \sum_{t=1}^T \frac{1}{T} \ell^{(t)}(\mathbf{x})$, for all $\mathbf{x} \in X$. Overloading notation, we also write $\text{Regret}^T(\{\mathbf{x}^t\}) = \max_{\mathbf{x} \in X} \text{Regret}^T(\{\mathbf{x}^t\}, \mathbf{x})$. An algorithm \mathcal{A} that takes as input a sequence of loss functions and outputs decisions such that $\text{Regret}^T(\mathcal{A}(\{\ell^{(t)}\})) \rightarrow 0$ as $T \rightarrow \infty$ is called a **no-regret algorithm**.

For any differentiable convex function $R : X \rightarrow \mathbb{R}$, the **Bregman divergence** between two vectors $\mathbf{w}, \mathbf{u} \in X$ is defined as follows: $\delta_R(\mathbf{w}|\mathbf{u}) = R(\mathbf{w}) - (R(\mathbf{u}) + \langle \nabla R(\mathbf{u}), \mathbf{w} - \mathbf{u} \rangle)$. One first-order no-regret learning algorithm is **Online Mirror Descent (OMD)**, defined as follows for some initial iterate $\mathbf{x}^{(0)} \in X$, a fixed learning rate $\eta > 0$, and a strongly convex regularizer $R: \mathbf{x}^{(t+1)} = \arg \min_{\mathbf{x} \in X} \left\langle \nabla_{\mathbf{x}} \ell^{(t)}(\mathbf{x}^{(t)}), \mathbf{x} \right\rangle + \frac{1}{2\eta} \delta_R(\mathbf{x}|\mathbf{x}^{(t)})$. When $R(\mathbf{x}) = \frac{1}{2} \|\mathbf{x}\|_2^2$, OMD reduces to **projected online gradient descent (OGD)**, given by the update rule: $\mathbf{x}^{(t+1)} = \Pi_X \left(\mathbf{x}^{(t)} - \eta \nabla_{\mathbf{x}} \ell^{(t)}(\mathbf{x}^{(t)}) \right)$. The next theorem bounds the **average regret** of OMD [35]:

THEOREM 2.3. *Suppose that the OMD algorithm generates a sequence of iterates $\{\mathbf{x}^{(t)}\}$ when run with a 1-strongly convex regularizer R^3 . Let $c = \max_{\mathbf{x} \in X, t \in [T]} \delta_R(\mathbf{x}|\mathbf{x}^{(t)})$, and let $\{\ell^{(t)}\}$ be a sequence of functions s.t. for all $t \in \mathbb{N}_+$, $\ell^{(t)} : \mathbb{R}^n \rightarrow \mathbb{R}$ is L -Lipschitz w.r.t. the dual norm $\|\cdot\|_*$. Then, if $\eta = \frac{c}{L\sqrt{2T}}$, OMD achieves average regret bounded as follows: $\text{Regret}^T(\{\mathbf{x}^t\}) \leq cL\sqrt{2T}$.*

3 NO-REGRET LEARNING DYNAMICS

In Stackelberg games, the outer player chooses their strategy assuming the inner player will best respond. When both players' choices are optimal, the outcome is a Stackelberg equilibrium.

In this section, we study no-regret learning dynamics in repeated min-max Stackelberg games in two settings: an **asymmetric** one in

³This assumption is without loss of generality, since any m -strongly-convex regularizer can be transformed into a 1-strongly-convex regularizer

which the outer player is a no-regret learner while the inner player best-responds, and a **symmetric** one in which both players are no-regret learners. Our main results are: 1. In the asymmetric setting, if the outer player employs an asymmetric-regret-minimizing algorithm, play converges to a Stackelberg equilibrium, and 2. in the symmetric setting, if both players employ a no-Lagrangian-regret algorithm, play converges to a Stackelberg equilibrium.

3.1 Asymmetric Learning Setting

We first consider an asymmetric setting in which the inner player best responds to the strategy picked by the outer player, while the outer player employs a no-regret learning algorithm. In min-max Stackelberg games, the two players are adversaries, so this best-response assumption corresponds to the worst case. In many real-world applications, we seek optimal strategies for the outer player, e.g., in security games we are interested in an optimal strategy for the defender/outer player, not the attacker/inner player [36]. Assuming a strong inner player allows us to learn more robust strategies for the outer player.

Given $\mathbf{x} \in X$, let $\mathbf{y}^*(\mathbf{x}) \in \text{BR}_Y(\mathbf{x})$, and consider an online min-max Stackelberg game $\left\{ \left(X, Y, f^{(t)}, \mathbf{g}^{(t)} \right) \right\}$. In an asymmetric setting, the outer player's regret is the difference between the cumulative loss of their sequence of strategies $\{\mathbf{x}^{(t)}\}$ (to which the inner player best responds), and the smallest cumulative loss that the outer player could have achieved by playing a fixed strategy $\mathbf{x} \in X$ (again, to which the inner player best responds), i.e., $\frac{1}{T} \sum_{t=1}^T f^{(t)}(\mathbf{x}^{(t)}, \mathbf{y}^*(\mathbf{x}^{(t)})) - \sum_{t=1}^T \frac{1}{T} f^{(t)}(\mathbf{x}, \mathbf{y}^*(\mathbf{x}))$. We call this regret the **asymmetric regret**, and express it in terms of the outer player's value function $V_X: \text{AsymRegret}_X^T(\{\mathbf{x}^{(t)}\}, \mathbf{x}) = \frac{1}{T} \sum_{t=1}^T V_X^{(t)}(\mathbf{x}^{(t)}) - \sum_{t=1}^T \frac{1}{T} V_X^{(t)}(\mathbf{x})$. As above, we overload notation and write $\text{AsymRegret}_X^T(\{\mathbf{x}^{(t)}\}) = \max_{\mathbf{x} \in X} \text{AsymRegret}_X^T(\{\mathbf{x}^{(t)}\}, \mathbf{x})$.

The main theorem⁴ in this section states the following: assuming the inner player best responds to the strategies of the outer player, if the outer player employs a no-regret algorithm, then the outer player's average strategy converges to their part of a Stackelberg equilibrium strategy.

THEOREM3.1. *Consider a repeated min-max Stackelberg game (X, Y, f, \mathbf{g}) , and suppose the outer player plays a sequence of strategies $\{\mathbf{x}^{(t)}\}$. If, after T iterations, the outer player's asymmetric regret is bounded by ϵ , i.e., $\text{AsymRegret}_X^T(\{\mathbf{x}^{(t)}\}) \leq \epsilon$, then $(\bar{\mathbf{x}}^{(T)}, \mathbf{y}^*(\bar{\mathbf{x}}^{(T)}))$ is a $(\epsilon, 0)$ -Stackelberg equilibrium, where $\mathbf{y}^*(\bar{\mathbf{x}}^{(T)}) \in \text{BR}_Y(\bar{\mathbf{x}}^{(T)})$.*

We remark that although the definition of asymmetric regret looks similar to the standard definition of regret, its structure is very different. Proposition 2.2 is required to ensure that the time-averaged value function $\sum_{t=1}^T V^{(t)}(\mathbf{x})$ is convex in \mathbf{x} .

3.2 Symmetric Learning Setting

We now turn our attention to a setting in which both players are no-regret learners. The most straightforward way to define regret is by considering the outer and inner players' "vanilla" regrets, respectively: $\text{Regret}_X^T(\{\mathbf{x}^{(t)}\}, \mathbf{x}) = \frac{1}{T} \sum_{t=1}^T f^{(t)}(\mathbf{x}^{(t)}, \mathbf{y}^{(t)}) -$

⁴The proofs of all mathematical claims in this section can be found in Appendix B.

$\frac{1}{T} \sum_{t=1}^T f^{(t)}(\mathbf{x}, \mathbf{y}^{(t)})$ and $\text{Regret}_Y^T(\{\mathbf{y}^{(t)}\}, \mathbf{y}) = \frac{1}{T} \sum_{t=1}^T f^{(t)}(\mathbf{x}^{(t)}, \mathbf{y}) - \frac{1}{T} \sum_{t=1}^T f^{(t)}(\mathbf{x}^{(t)}, \mathbf{y}^{(t)})$. In convex-concave min-max games (with independent strategy sets), when both players minimize these regrets, the players' average strategies converge to Nash equilibrium. In min-max Stackelberg games (with dependent strategy sets), however, convergence to a Stackelberg equilibrium is not guaranteed.

EXAMPLE3.2. *Consider the min-max Stackelberg game $\min_{\mathbf{x} \in [-1,1]} \max_{\mathbf{y} \in [-1,1]: 0 \leq 1 - (\mathbf{x} + \mathbf{y})} \mathbf{x}^2 + \mathbf{y} + 1$. The Stackelberg equilibrium of this game is given by $\mathbf{x}^* = 1/2, \mathbf{y}^* = 1/2$.*

If both players employ no-regret algorithms that generate strategies $\{\mathbf{x}^{(t)}, \mathbf{y}^{(t)}\}_{t \in \mathbb{N}_+}$, then at time $T \in \mathbb{N}_+$, there exists $\epsilon > 0$, s.t.

$$\begin{cases} \frac{1}{T} \sum_{t=1}^T \left[\mathbf{x}^{(t)^2} + \mathbf{y}^{(t)} + 1 \right] - \frac{1}{T} \min_{\mathbf{x} \in [-1,1]} \sum_{t=1}^T \left[\mathbf{x}^2 + \mathbf{y}^{(t)} + 1 \right] \leq \epsilon \\ \frac{1}{T} \max_{\mathbf{y} \in [-1,1]} \sum_{t=1}^T \left[\mathbf{x}^{(t)^2} + \mathbf{y} + 1 \right] - \frac{1}{T} \sum_{t=1}^T \left[\mathbf{x}^{(t)^2} + \mathbf{y}^{(t)} + 1 \right] \leq \epsilon \end{cases}$$

Simplifying yields:

$$\begin{cases} \frac{1}{T} \sum_{t=1}^T \mathbf{x}^{(t)^2} - \min_{\mathbf{x} \in [-1,1]} \mathbf{x}^2 \leq \epsilon \\ \max_{\mathbf{y} \in [-1,1]} \mathbf{y} - \frac{1}{T} \sum_{t=1}^T \mathbf{y}^{(t)} \leq \epsilon \end{cases} = \begin{cases} \frac{1}{T} \sum_{t=1}^T \mathbf{x}^{(t)^2} \leq \epsilon \\ 1 - \epsilon \leq \frac{1}{T} \sum_{t=1}^T \mathbf{y}^{(t)} \end{cases}$$

In other words, the average iterates converge to $\mathbf{x} = 0, \mathbf{y} = 1$, which is not the Stackelberg equilibrium of this game.

If the inner player minimizes their vanilla regret without regard to the game's constraints, then their strategies are not guaranteed to be feasible, and thus cannot converge to a Stackelberg equilibrium. To remedy this infeasibility, we introduce a new type of regret we call **Lagrangian regret**, and show that assuming access to a solution oracle for the optimal KKT multipliers of the game's constraints, if both players minimize their Lagrangian regret, then no-regret learning dynamics converge to a Stackelberg equilibrium.

Let $\mathcal{L}_x(\mathbf{y}, \boldsymbol{\lambda}) = f(\mathbf{x}, \mathbf{y}) + \sum_{k=1}^K \lambda_k g_k(\mathbf{x}, \mathbf{y})$ denote the Lagrangian associated with the outer player's value function, or equivalently, the inner player's maximization problem, given the outer player's strategy $\mathbf{x} \in X$. Using this notation, we can re-express the Stackelberg game as $\min_{\mathbf{x} \in X} \max_{\mathbf{y} \in Y: \mathbf{g}(\mathbf{x}, \mathbf{y}) \geq 0} f(\mathbf{x}, \mathbf{y}) = \min_{\mathbf{x} \in X} \max_{\mathbf{y} \in Y} \min_{\boldsymbol{\lambda} \geq 0} \mathcal{L}_x(\mathbf{y}, \boldsymbol{\lambda})$. If the optimal KKT multipliers $\boldsymbol{\lambda}^* \in \mathbb{R}^K$, which are guaranteed to exist by Slater's condition [61], were known, then one could plug them back into the Lagrangian to obtain a convex-concave saddle point problem given by $\min_{\mathbf{x} \in X} \max_{\mathbf{y} \in Y} \mathcal{L}_x(\mathbf{y}, \boldsymbol{\lambda}^*)$. Note that a saddle point of this problem is guaranteed to exist by the min-max theorem [48], since $\mathcal{L}_x(\mathbf{y}, \boldsymbol{\lambda}^*)$ is convex in \mathbf{x} and concave in \mathbf{y} . The next lemma states that the Stackelberg equilibria of a min-max Stackelberg game correspond to the saddle points of $\mathcal{L}_x(\mathbf{y}, \boldsymbol{\lambda}^*)$.

LEMMA3.3. *Any Stackelberg equilibrium $(\mathbf{x}^*, \mathbf{y}^*) \in X \times Y$ of any min-max Stackelberg game (X, Y, f, \mathbf{g}) corresponds to a saddle point of $\mathcal{L}_x(\mathbf{y}, \boldsymbol{\lambda}^*)$, where $\boldsymbol{\lambda}^* \in \arg \min_{\boldsymbol{\lambda} \geq 0} \min_{\mathbf{x} \in X} \max_{\mathbf{y} \in Y} \mathcal{L}_x(\mathbf{y}, \boldsymbol{\lambda})$.*

This lemma tells us that the function $\mathcal{L}_x(\mathbf{y}, \boldsymbol{\lambda}^*)$ represents a new loss function that enforces the game's constraints. Based on this observation, we assume access to a Lagrangian solution oracle that provides us with $\boldsymbol{\lambda}^* \in \arg \min_{\boldsymbol{\lambda} \geq 0} \min_{\mathbf{x} \in X} \max_{\mathbf{y} \in Y} \mathcal{L}_x(\mathbf{y}, \boldsymbol{\lambda}^*)$.

Next, we define a new type of regret which we call **Lagrangian regret**. Given a sequence of strategies $\{\mathbf{x}^{(t)}, \mathbf{y}^{(t)}\}$ played by the outer and inner players in an online min-max Stackelberg game $\left\{ \left(X, Y, f^{(t)}, \mathbf{g}^{(t)} \right) \right\}$, we define their Lagrangian regret, respectively,

as $\text{LagrRegret}_X^T(\{\mathbf{x}^{(t)}\}, \mathbf{x}) = \frac{1}{T} \sum_{t=1}^T \mathcal{L}_{\mathbf{x}^{(t)}}(\mathbf{y}^{(t)}, \boldsymbol{\lambda}^*) - \frac{1}{T} \sum_{t=1}^T \mathcal{L}_{\mathbf{x}^{(t)}}(\mathbf{y}^{(t)}, \boldsymbol{\lambda}^*)$
 and $\text{LagrRegret}_Y^T(\{\mathbf{y}^{(t)}\}, \mathbf{y}) = \frac{1}{T} \sum_{t=1}^T \mathcal{L}_{\mathbf{x}^{(t)}}(\mathbf{y}, \boldsymbol{\lambda}^*) - \frac{1}{T} \sum_{t=1}^T \mathcal{L}_{\mathbf{x}^{(t)}}(\mathbf{y}^{(t)}, \boldsymbol{\lambda}^*)$.
 We further define $\text{LagrRegret}_X^T(\{\mathbf{x}^{(t)}\})$ and $\text{LagrRegret}_Y^T(\{\mathbf{y}^{(t)}\})$
 as expected.

The **saddle point residual** of a point $(\mathbf{x}^*, \mathbf{y}^*) \in X \times Y$ w.r.t. a convex-concave function $h : X \times Y \rightarrow \mathbb{R}$ is given by $\max_{\mathbf{y} \in Y} h(\mathbf{x}^*, \mathbf{y}) - \min_{\mathbf{x} \in X} h(\mathbf{x}, \mathbf{y}^*)$. When the saddle point residual of (\mathbf{x}, \mathbf{y}) w.r.t. $\mathcal{L}_{\mathbf{x}}(\mathbf{y}, \boldsymbol{\lambda}^*)$ is 0, the saddle point is a $(0, 0)$ -Stackelberg equilibrium.

The main theorem of this section now follows: if both players play so as to minimize their Lagrangian regret, then their average strategies converge to a Stackelberg equilibrium. The bound is given in terms of the saddle point residual of the iterates generated.

THEOREM 3.4. *Consider a repeated min-max Stackelberg game (X, Y, f, \mathbf{g}) , and suppose the outer and the players generate sequences of strategies $\{(\mathbf{x}^{(t)}, \mathbf{y}^{(t)})\}$ using a no-Lagrangian-regret algorithm. If after T iterations, the Lagrangian regret of both players is bounded by ε , i.e., $\text{LagrRegret}_X^T(\{\mathbf{x}^{(t)}\}) \leq \varepsilon$ and $\text{LagrRegret}_Y^T(\{\mathbf{y}^{(t)}\}) \leq \varepsilon$, then the following convergence bound holds on the saddle point residual of $(\bar{\mathbf{x}}^{(T)}, \hat{\mathbf{y}}^{(T)})$ w.r.t. the Lagrangian: $0 \leq \max_{\mathbf{y} \in Y} \mathcal{L}_{\bar{\mathbf{x}}^{(T)}}(\mathbf{y}, \boldsymbol{\lambda}^*) - \min_{\mathbf{x} \in X} \mathcal{L}_{\mathbf{x}}(\hat{\mathbf{y}}^{(T)}, \boldsymbol{\lambda}^*) \leq 2\varepsilon$.*

Having provided convergence to Stackelberg equilibrium of general no-regret learning dynamics in repeated min-max Stackelberg games, we now proceed to investigate the convergence and robustness properties of a specific example of a no-regret learning dynamic, namely online mirror descent (OMD).

4 ONLINE MIRROR DESCENT

In this section, we apply the results we have derived for general no-regret learning dynamics to Online Mirror Descent (OMD) specifically [46, 58]. We then study the robustness properties of OMD in min-max Stackelberg games.

4.1 Convergence Analysis

When the outer player is an OMD learner minimizing its asymmetric regret and the inner player best responds, we obtain the max-oracle mirror descent (MD) algorithm (Algorithm 1), a special case of which was first proposed by Jin et al. [33] for min-max games (with independent strategy sets) under the name of max-oracle GD. Goktas and Greenwald [25] extended their algorithm from min-max games (with independent strategy sets) to min-max Stackelberg games and proved its convergence in best iterates. Max-oracle MD (Algorithm 1) is a further generalization of both algorithms.

The following corollary of Theorem 3.1, which concerns convergence of the more general max-oracle MD algorithm in average iterates, complements Goktas and Greenwald's result on the convergence of max-oracle GD (Algorithm 3, Appendix C) in best iterates: if the outer player employs a strategy that achieves ε -asymmetric regret, then the max-oracle MD algorithm is guaranteed to converge to the outer player's $(\varepsilon, 0)$ -Stackelberg equilibrium strategy in average iterates after $O(1/\varepsilon^2)$ iterations, assuming the inner player best responds.

We note that since V_X is convex, by Proposition 2.2, V_X is sub-differentiable. Moreover, for all $\hat{\mathbf{x}} \in X$, $\hat{\mathbf{y}} \in \text{BR}_Y(\hat{\mathbf{x}})$, $\nabla_{\mathbf{x}} f(\hat{\mathbf{x}}, \hat{\mathbf{y}}) +$

Algorithm 1 Max-Oracle Mirror Descent (MD)

Inputs: $X, Y, f, \mathbf{g}, \boldsymbol{\eta}, T, \mathbf{x}^{(0)}, R$ **Output:** $\mathbf{x}^*, \mathbf{y}^*$

- 1: **for** $t = 1, \dots, T$ **do**
- 2: Find $\mathbf{y}^*(\mathbf{x}^{(t-1)}) \in \text{BR}_Y(\mathbf{x}^{(t-1)})$
- 3: Set $\mathbf{y}^{(t-1)} = \mathbf{y}^*(\mathbf{x}^{(t-1)})$
- 4: Set $\boldsymbol{\lambda}^{(t-1)} = \boldsymbol{\lambda}^*(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t-1)})$
- 5: Set $\mathbf{x}^{(t)} = \arg \min_{\mathbf{x} \in X} \left(\nabla_{\mathbf{x}} \mathcal{L}_{\mathbf{x}^{(t-1)}}(\mathbf{y}^{(t-1)}, \boldsymbol{\lambda}^{(t-1)}), \mathbf{x} \right) + \frac{1}{2\boldsymbol{\eta}_t} \delta_R(\mathbf{x} | \mathbf{x}^{(t-1)})$
- 6: **end for**
- 7: Set $\bar{\mathbf{x}}^{(T)} = \frac{1}{T} \sum_{t=1}^T \mathbf{x}^{(t)}$
- 8: Set $\hat{\mathbf{y}}^{(T)} \in \text{BR}_Y(\bar{\mathbf{x}}^{(T)})$
- 9: **return** $(\bar{\mathbf{x}}^{(T)}, \hat{\mathbf{y}}^{(T)})$

$\sum_{k=1}^K \lambda_k^* g_k(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ is an arbitrary subgradient of the value function at $\hat{\mathbf{x}}$ by Goktas and Greenwald's subdifferential envelope theorem [25]. We add that similar to Goktas and Greenwald, we assume that the optimal KKT multipliers $\boldsymbol{\lambda}^*(\mathbf{x}^{(t)}, \hat{\mathbf{y}}(\mathbf{x}^{(t)}))$ associated with a solution $\hat{\mathbf{y}}(\mathbf{x}^{(t)})$ can be computed in constant time.

COROLLARY 4.1. *Let $c = \max_{\mathbf{x} \in X} \|\mathbf{x}\|$ and let $L_f = \max_{(\hat{\mathbf{x}}, \hat{\mathbf{y}}) \in X \times Y} \|\nabla_{\mathbf{x}} f(\hat{\mathbf{x}}, \hat{\mathbf{y}})\|$. If Algorithm 1 is run on a repeated min-max Stackelberg game (X, Y, f, \mathbf{g}) , with $\boldsymbol{\eta}_t = \frac{c}{L_f \sqrt{2T}}$, for all iteration $t \in [T]$ and any $\mathbf{x}^{(0)} \in X$, then $(\bar{\mathbf{x}}^{(T)}, \hat{\mathbf{y}}^{(T)})$ is a $(cL_f \sqrt{2}/\sqrt{T}, 0)$ -Stackelberg equilibrium. Furthermore, for any $\varepsilon \in (0, 1)$, there exists $N(\varepsilon) \in O(1/\varepsilon^2)$ s.t. for all $T \geq N(\varepsilon)$, there exists an iteration $T^* \leq T$ s.t. $(\bar{\mathbf{x}}^{(T^*)}, \hat{\mathbf{y}}^{(T^*)})$ is an $(\varepsilon, 0)$ -Stackelberg equilibrium.*

Note that we can relax Theorem 3.1 to instead work with an approximate best response of the inner player, i.e., given the strategy of the outer player $\hat{\mathbf{x}}$, instead of playing an exact best-response, the inner player could compute a $\hat{\mathbf{y}}$ s.t. $f(\hat{\mathbf{x}}, \hat{\mathbf{y}}) \geq \max_{\mathbf{y} \in Y: \mathbf{g}(\hat{\mathbf{x}}, \mathbf{y}) \geq 0} f(\hat{\mathbf{x}}, \mathbf{y}) - \varepsilon$. Moreover, the inner player could run gradient (or mirror) ascent on $f(\hat{\mathbf{x}}, \mathbf{y})$ to find $\hat{\mathbf{y}}$, instead of assuming a best-response oracle in Algorithm 1. We can combine the fact that gradient ascent on Lipschitz smooth functions converges in $O(1/\varepsilon)$ iterations [46] with our novel convergence rate for max-oracle MD to conclude that the average iterates computed by nested GDA [25] converge to an $(\varepsilon, \varepsilon)$ -Stackelberg equilibrium in $O(1/\varepsilon^3)$ iterations. If additionally, f is strongly convex in \mathbf{y} , then the iteration complexity can be reduced to $O(1/\varepsilon^2 \log(1/\varepsilon))$.

Similarly, we can also consider the symmetric case, in which both the outer and inner players minimize their Lagrangian regrets, as OMD learners with access to a Lagrangian solution oracle that returns $\boldsymbol{\lambda}^* \in \arg \min_{\boldsymbol{\lambda} \geq 0} \min_{\mathbf{x} \in X} \max_{\mathbf{y} \in Y} \mathcal{L}_{\mathbf{x}}(\mathbf{y}, \boldsymbol{\lambda})$. In this case, we obtain the **Lagrangian mirror descent ascent (LMDA)** algorithm (Algorithm 2). The following corollary of Theorem 3.4 states that LMDA converges in average iterates to an ε -Stackelberg equilibrium in $O(1/\varepsilon^2)$ iterations.

COROLLARY 4.2. *Let $b = \max_{\mathbf{x} \in X} \|\mathbf{x}\|$, $c = \max_{\mathbf{y} \in Y} \|\mathbf{y}\|$, and $L_{\mathcal{L}} = \max_{(\hat{\mathbf{x}}, \hat{\mathbf{y}}) \in X \times Y} \|\nabla_{\mathbf{x}} \mathcal{L}_{\hat{\mathbf{x}}}(\hat{\mathbf{y}}, \boldsymbol{\lambda}^*)\|$. If Algorithm 2 is run on a repeated min-max Stackelberg game (X, Y, f, \mathbf{g}) , with $\boldsymbol{\eta}_t^{\mathbf{x}} = \frac{b}{L_{\mathcal{L}} \sqrt{2T}}$ and $\boldsymbol{\eta}_t^{\mathbf{y}} = \frac{c}{L_{\mathcal{L}} \sqrt{2T}}$, for all iterations $t \in [T]$ and any $\mathbf{x}^{(0)} \in X$, then*

Algorithm 2 Lagrangian Mirror Descent Ascent (LMDA)

Inputs: $\lambda^*, X, Y, f, g, \eta^x, \eta^y, T, \mathbf{x}^{(0)}, \mathbf{y}^{(0)}, R$ **Output:** $\mathbf{x}^*, \mathbf{y}^*$

- 1: **for** $t = 1, \dots, T - 1$ **do**
- 2: Set $\mathbf{x}^{(t)} = \arg \min_{\mathbf{x} \in X} \left(\nabla_{\mathbf{x}} \mathcal{L}_{\mathbf{x}^{(t-1)}}(\mathbf{y}^{(t-1)}, \lambda^*), \mathbf{x} \right) + \frac{1}{2\eta_t^x} \delta_R(\mathbf{x} \parallel \mathbf{x}^{(t)})$
- 3: Set $\mathbf{y}^{(t)} = \arg \max_{\mathbf{y} \in Y} \left(\nabla_{\mathbf{y}} \mathcal{L}_{\mathbf{x}^{(t-1)}}(\mathbf{y}^{(t-1)}, \lambda^*), \mathbf{y} \right) - \frac{1}{2\eta_t^y} \delta_R(\mathbf{y} \parallel \mathbf{y}^{(t-1)})$
- 4: **end for**
- 5: **return** $\{(\mathbf{x}^{(t)}, \mathbf{y}^{(t)})\}_{t=1}^T$

the following convergence bound holds on the saddle point residual of $(\bar{\mathbf{x}}^{(T)}, \bar{\mathbf{y}}^{(T)})$ w.r.t. the Lagrangian: $0 \leq \max_{\mathbf{y} \in Y} \mathcal{L}_{\bar{\mathbf{x}}^{(T)}}(\mathbf{y}, \lambda^*) - \min_{\mathbf{x} \in X} \mathcal{L}_{\bar{\mathbf{y}}^{(T)}}(\mathbf{x}, \lambda^*) \leq \frac{2\sqrt{2}L_{\mathcal{L}}}{\sqrt{T}} \max\{b, c\}$.

We remark that in certain rare cases the Lagrangian can become degenerate in \mathbf{y} , in that the \mathbf{y} terms in the Lagrangian might cancel out when λ^* is plugged back into Lagrangian, leading LMDA to not update the \mathbf{y} variables, as demonstrated by the following example:

EXAMPLE 4.3. Consider the following min-max Stackelberg game: $\min_{\mathbf{x} \in [-1, 1]} \max_{\mathbf{y} \in [-1, 1]: 0 \leq 1 - (x+y)} x^2 + y + 1$. When we plug the optimal KKT multiplier $\lambda^* = 1$ into the Lagrangian associated with the outer player's value function, we obtain $\mathcal{L}_{\mathbf{x}}(\mathbf{y}, \lambda) = x^2 + y + 1 - (x+y) = x^2 - x + 1$, with $\frac{\partial \mathcal{L}}{\partial x} = 2x - 1$ and $\frac{\partial \mathcal{L}}{\partial y} = 0$. It follows that the \mathbf{x} iterate converges to $1/2$, but the \mathbf{y} iterate will never be updated, and hence unless \mathbf{y} is initialized at its Stackelberg equilibrium value, LMDA will not converge to a Stackelberg equilibrium.

In general, this degeneracy issue occurs when $\forall \mathbf{x} \in X, \nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}) = -\sum_{k=1}^K \lambda_k^* \nabla_{\mathbf{y}} g_k(\mathbf{x}, \mathbf{y})$. We can sidestep the issue by restricting our attention to min-max Stackelberg games with convex-strictly-concave objective functions, which is sufficient to ensure that the Lagrangian is not degenerate in \mathbf{y} [6]. However, we observe in our experiments that even for convex-non-strictly-concave min-max Stackelberg games, LMDA, specifically with regularizer $R(\mathbf{x}) = \|\mathbf{x}\|_2^2$ (i.e., LGDA; Algorithm 4, Appendix C), converges to Stackelberg equilibrium.

4.2 Robustness Analysis

Our analysis thus far of min-max Stackelberg games has assumed the same game is played repeatedly. In this section, we expand our consideration to online min-max Stackelberg games more generally, allowing the objective function to change from one time step to the next, as in the OCO framework. Providing dynamics that are robust to ongoing game changes is crucial, as the real world is rarely static.

Online games bring with them a host of interesting issues. Notably, even though the environment might change from one time step to the next, the game still exhibits a Stackelberg equilibrium during each stage of the game. However, one cannot reasonably expect the players to play an equilibrium during each stage, since even in a repeated game setting, known game dynamics require multiple iterations before players can reach an approximate equilibrium. Players cannot immediately best respond, but they can behave like boundedly rational agents who take a step in the direction of their optimal strategy during each iteration. In general online games, equilibria also become dynamic objects, which can never be reached unless the game stops changing.

Corollaries 4.1 and 4.2 tell us that OMD dynamics are effective equilibrium-finding strategies in repeated min-max Stackelberg

games. However, they do not provide any intuition about the robustness of OMD dynamics to perturbations in the game. In this section, we ask whether OMD dynamics can track Stackelberg equilibria when the game changes. Ultimately, our theoretical results only concern online min-max games (with independent strategy sets), for which Nash, not Stackelberg, equilibrium is the relevant solution concept. Nonetheless, we provide experimental evidence that suggests that the results we prove may also apply more broadly to online min-max Stackelberg games (with dependent strategy sets). We note that our robustness analysis focuses on projected OGD dynamics, a special case of OMD dynamics, for ease of analysis.

We first consider the asymmetric setting, in which the outer player is a no-regret learner and the inner player best-responds. In this setting, we show that when the outer player plays according to projected OGD dynamics in an arbitrary online min-max game, the outer player's strategies closely track their Nash equilibrium strategies. The following result states that regardless of the initial strategy of the outer player, projected OGD dynamics are always within a $2d/\delta$ radius of the outer player's Nash equilibrium strategy.

THEOREM 4.4. Consider an online min-max game $\{(X, Y, f^{(t)})\}_{t=1}^T$. Suppose that, for all $t \in [T]$, $f^{(t)}$ is μ -strongly convex in \mathbf{x} and strictly concave in \mathbf{y} , and $f^{(t)}$ is $L_{\nabla f}$ -Lipschitz smooth. Suppose the outer player generates a sequence of actions $\{\mathbf{x}^{(t)}\}_{t=1}^T$ by using projected OGD on the loss functions $\{V^{(t)}\}_{t=1}^T$ with learning rate $\eta \leq \frac{2}{\mu + L_{\nabla f}}$, and further suppose the inner player generates a sequence of best-responses $\{\mathbf{y}^{(t)}\}_{t=1}^T$ to each iterate of the outer player. For all $t \in [T]$, let $\mathbf{x}^{(t)*} \in \arg \min_{\mathbf{x} \in X} V^{(t)}(\mathbf{x})$, $\Delta^{(t)} = \|\mathbf{x}^{(t+1)*} - \mathbf{x}^{(t)*}\|$, and $\delta = \frac{2\eta\mu L_{\nabla f}}{L_{\nabla f} + \mu}$. We then have: $\|\mathbf{x}^{(T)*} - \mathbf{x}^{(T)}\| \leq (1 - \delta)^{T/2} \|\mathbf{x}^{(0)*} - \mathbf{x}^{(0)}\| + \sum_{t=1}^T (1 - \delta)^{\frac{T-t}{2}} \Delta^{(t)}$. If additionally, for all $t \in [T]$, $\Delta^{(t)} \leq d$, then: $\|\mathbf{x}^{(T)*} - \mathbf{x}^{(T)}\| \leq (1 - \delta)^{T/2} \|\mathbf{x}^{(0)*} - \mathbf{x}^{(0)}\| + \frac{d}{\delta}$.

We can derive a similar robustness result in the symmetric setting, where the outer and inner players are both projected OGD learners. The following result states that regardless of the initial strategies of the two players, projected OGD dynamics follow the Nash equilibrium of the game, always staying within a $4d/\delta$ radius.

THEOREM 4.5. Consider an online min-max game $\{(X, Y, f^{(t)})\}_{t=1}^T$. Suppose that, for all $t \in [T]$, $f^{(t)}$ is μ_x -strongly convex in \mathbf{x} and μ_y -strongly concave in \mathbf{y} , and $f^{(t)}$ is $L_{\nabla f}$ -Lipschitz smooth. Let $\{(\mathbf{x}^{(t)}, \mathbf{y}^{(t)})\}_{t=1}^T$ be the strategies played by the outer and inner players, assuming that the outer player uses a projected OGD algorithm on the losses $\{f^{(t)}(\cdot, \mathbf{y}^{(t)})\}_{t=1}^T$ with $\eta_x = \frac{2}{\mu_x + L_{\nabla f}}$ and the inner player uses a projected OGD algorithm on the losses $\{-f^{(t)}(\mathbf{x}^{(t)}, \cdot)\}_{t=1}^T$ with $\eta_y = \frac{2}{\mu_y + L_{\nabla f}}$. For all $t \in [T]$, let $\mathbf{x}^{(t)*} \in \arg \min_{\mathbf{x} \in X} f^{(t)}(\mathbf{x}, \mathbf{y}^{(t)})$, $\mathbf{y}^{(t)*} \in \arg \min_{\mathbf{y} \in Y} f^{(t)}(\mathbf{x}^{(t)}, \mathbf{y})$, $\Delta_{\mathbf{x}}^{(t)} = \|\mathbf{x}^{(t+1)*} - \mathbf{x}^{(t)*}\|$, $\Delta_{\mathbf{y}}^{(t)} = \|\mathbf{y}^{(t+1)*} - \mathbf{y}^{(t)*}\|$, $\delta_{\mathbf{x}} = \frac{2\eta_x \mu_x L_{\nabla f}}{L_{\nabla f} + \mu_x}$, and $\delta_{\mathbf{y}} = \frac{2\eta_y \mu_y L_{\nabla f}}{L_{\nabla f} + \mu_y}$. We then have: $\|\mathbf{x}^{(T)*} - \mathbf{x}^{(T)}\| + \|\mathbf{y}^{(T)*} - \mathbf{y}^{(T)}\| \leq (1 - \delta_{\mathbf{x}})^{T/2} \|\mathbf{x}^{(0)*} - \mathbf{x}^{(0)}\| + (1 - \delta_{\mathbf{y}})^{T/2} \|\mathbf{y}^{(0)*} - \mathbf{y}^{(0)}\| + \sum_{t=1}^T (1 - \delta_{\mathbf{x}})^{\frac{T-t}{2}} \Delta_{\mathbf{x}}^{(t)} + \sum_{t=1}^T (1 - \delta_{\mathbf{y}})^{\frac{T-t}{2}} \Delta_{\mathbf{y}}^{(t)}$. If additionally, for all $t \in [T]$, $\Delta_{\mathbf{x}}^{(t)} \leq d$ and $\Delta_{\mathbf{y}}^{(t)} \leq d$, and $\delta =$

$$\min\{\delta_{\mathbf{y}}, \delta_{\mathbf{x}}\}, \text{ then: } \left\| \mathbf{x}^{(T)*} - \mathbf{x}^{(T)} \right\| + \left\| \mathbf{y}^{(T)*} - \mathbf{y}^{(T)} \right\| \leq 2(1 - \delta)^{T/2} \left(\left\| \mathbf{x}^{(0)*} - \mathbf{x}^{(0)} \right\| + \left\| \mathbf{y}^{(0)*} - \mathbf{y}^{(0)} \right\| \right) + \frac{4d}{\delta}.$$

The proofs of the above theorems are relegated to Appendix B. These theorems establish the robustness of projected OGD dynamics for min-max games in both the asymmetric and symmetric settings by showing that the dynamics closely track the Nash equilibria in a large class of min-max games (with independent strategy sets). These results also suggest that general OMD dynamics, e.g., OMD with entropy as a regularizer, are robust to perturbation. As we are not able to extend these theoretical robustness guarantees to min-max Stackelberg games (with dependent strategy sets), we instead ran a series of experiments with online Fisher markets, which are canonical examples of min-max Stackelberg games [25], to investigate the empirical robustness guarantees of projected OGD dynamics for this class of min-max Stackelberg games.

5 ONLINE FISHER MARKETS

The Fisher market model, attributed to Irving Fisher [7], has received a great deal of attention in the literature, especially by computer scientists, as it has proven useful in the design of electronic marketplaces. We now study OMD dynamics in online Fisher markets, which are instances of min-max Stackelberg games [25].

A **Fisher market** consists of n buyers and m divisible goods [7]. Each buyer $i \in [n]$ has a budget $b_i \in \mathbb{R}_+$ and a utility function $u_i : \mathbb{R}_+^m \rightarrow \mathbb{R}$. Each good $j \in [m]$ has supply $s_j \in \mathbb{R}_+$. A Fisher market is thus given by a tuple $(n, m, U, \mathbf{b}, \mathbf{s})$, where $U = \{u_1, \dots, u_n\}$ is a set of utility functions, one per buyer; $\mathbf{b} \in \mathbb{R}_+^n$ is a vector of buyer budgets; and $\mathbf{s} \in \mathbb{R}_+^m$ is a vector of good supplies. We abbreviate as $(U, \mathbf{b}, \mathbf{s})$ when n and m are clear from context. An **online Fisher market** is a sequence of Fisher markets $\left\{ \left(U^{(t)}, \mathbf{b}^{(t)}, \mathbf{s}^{(t)} \right) \right\}_{t=1}^T$.

An **allocation** $X = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T \in \mathbb{R}_+^{n \times m}$ is an assignment of goods to buyers, represented as a matrix s.t. $x_{ij} \geq 0$ denotes the amount of good $j \in [m]$ allocated to buyer $i \in [n]$. Goods are assigned **prices** $\mathbf{p} = (p_1, \dots, p_m)^T \in \mathbb{R}_+^m$. A tuple (\mathbf{p}^*, X^*) is said to be a **competitive equilibrium (CE)** of Fisher market $(U, \mathbf{b}, \mathbf{s})$ if 1. buyers are utility maximizing, constrained by their budget, i.e., $\forall i \in [n], \mathbf{x}_i^* \in \arg \max_{\mathbf{x}: \mathbf{x} \cdot \mathbf{p}^* \leq b_i} u_i(\mathbf{x})$; and 2. the market clears, i.e., $\forall j \in [m], p_j^* > 0 \Rightarrow \sum_{i \in [n]} x_{ij}^* = s_j$ and $p_j^* = 0 \Rightarrow \sum_{i \in [n]} x_{ij}^* \leq s_j$.

Goktas and Greenwald [25] observe that any CE (\mathbf{p}^*, X^*) of a Fisher market (U, \mathbf{b}) corresponds to a Stackelberg equilibrium of the following min-max Stackelberg game:⁵

$$\min_{\mathbf{p} \in \mathbb{R}_+^m} \max_{X \in \mathbb{R}_+^{n \times m}: X\mathbf{p} \leq \mathbf{b}} \sum_{j \in [m]} s_j p_j + \sum_{i \in [n]} b_i \log(u_i(\mathbf{x}_i)) \quad (1)$$

Let $\mathcal{L} : \mathbb{R}_+^m \times \mathbb{R}^{n \times m} \rightarrow \mathbb{R}_+$ be the Lagrangian of the outer player's value function in Equation (1), i.e., $\mathcal{L}_{\mathbf{p}}(X, \boldsymbol{\lambda}) = \sum_{j \in [m]} s_j p_j + \sum_{i \in [n]} b_i \log(u_i(\mathbf{x}_i)) + \sum_{i \in [n]} \lambda_i (b_i - \mathbf{x}_i \cdot \mathbf{p})$. One can show the existence of a Lagrangian solution oracle for the Lagrangian of Equation (1) such that $\boldsymbol{\lambda}^* = \mathbf{1}_m$. We then have: 1. by Goktas and Greenwald's envelope theorem, the subdifferential of the outer player's value function is given by $\nabla_{\mathbf{p}} V(\mathbf{p}) = \mathbf{s} - \sum_{i \in [n]} \mathbf{x}_i^*(\mathbf{p})$, where $\mathbf{x}_i^*(\mathbf{p}) \in \arg \max_{\mathbf{x} \in \mathbb{R}_+^m: \mathbf{x} \cdot \mathbf{p} \leq b_i} u_i(\mathbf{x})$, 2. the gradient of the Lagrangian w.r.t.

the prices, given the Lagrangian solution oracle, is $\nabla_{\mathbf{p}} \mathcal{L}_{\mathbf{p}}(X, \boldsymbol{\lambda}^*) = \mathbf{s} - \sum_{i \in [n]} \mathbf{x}_i$ and $\nabla_{\mathbf{x}_i} \mathcal{L}_{\mathbf{p}}(X, \boldsymbol{\lambda}^*) = \frac{b_i}{u_i(\mathbf{x}_i)} \nabla_{\mathbf{x}_i} u_i(\mathbf{x}_i) - \mathbf{p}$, where $\boldsymbol{\lambda}^* = \mathbf{1}_m$ [26].

We first consider OMD dynamics for Fisher markets in the asymmetric setting, in which the outer player determines their strategy via projected OGD and the inner player best-responds. This setup yields a dynamic version of a natural price adjustment process known as tâtonnement [68], this variant of which was first studied by Cheung et al. [9] (Algorithm 5, Appendix C).

We also consider OMD dynamics in the symmetric setting, specifically the case in which both the outer and inner players employ projected OGD simultaneously, which yields myopic best-response dynamics [43] (Algorithm 6, Appendix C). In words, at each time step, the (fictional Walrasian) auctioneer takes a gradient descent step to minimize its regret, and then all the buyers take a gradient ascent step to minimize their Lagrangian regret. These GDA dynamics can be seen as myopic best-response dynamics for boundedly rational sellers and buyers.

Experiments. In order to better understand the robustness properties of Algorithms 5 and 6 in an online min-max Stackelberg game that is subject to perturbation across time, we ran a series of experiments with online Fisher Markets assuming three different classes of utility functions.⁶ Each utility structure endows Equation (1) with different smoothness properties, which allows us to compare the efficiency of the algorithms under varying conditions. Let $\mathbf{v}_i \in \mathbb{R}^m$ be a vector of valuation parameters that describes the utility function of buyer $i \in [n]$. We consider the following utility function classes: 1. linear: $u_i(\mathbf{x}_i) = \sum_{j \in [m]} v_{ij} x_{ij}$; 2. Cobb-Douglas: $u_i(\mathbf{x}_i) = \prod_{j \in [m]} x_{ij}^{v_{ij}}$; and 3. Leontief: $u_i(\mathbf{x}_i) = \min_{j \in [m]} \left\{ \frac{x_{ij}}{v_{ij}} \right\}$.

To simulate an online Fisher market, we fix a range for every market parameter and draw from that range uniformly at random during each iteration. Our goal is to understand how closely OMD dynamics track the CE of the Fisher markets as they vary with time. We compare the iterates $(\mathbf{p}^{(t)}, X^{(t)})$ computed by the algorithms and the CE $(\mathbf{p}^{(t)*}, X^{(t)*})$ of the market $(U^{(t)}, \mathbf{b}^{(t)}, \mathbf{s}^{(t)})$ at each iteration t . The difference between these outcomes is measured as $\left\| \mathbf{p}^{(t)*} - \mathbf{p}^{(t)} \right\|_2 + \left\| X^{(t)*} - X^{(t)} \right\|_2$.

In our experiments, we ran Algorithms 5 and 6 on 100 randomly initialized online Fisher markets. We depict the distance to the CE at each iteration for a single experiment chosen at random in Figures 1 and 2. In these figures, we observe that the OMD dynamics are closely tracking the CE as they vary with time. A more detailed description of our experimental setup can be found in Appendix D.

We observe from Figures 1 and 2 that for both Algorithms 5 and 6, we obtain an empirical convergence rate relatively close to $O(1/\sqrt{T})$ under Cobb-Douglas utilities, and a slightly slower empirical convergence rate under linear utilities. Recall that $O(1/\sqrt{T})$ is the convergence rate guarantee we obtained for both algorithms, assuming a fixed learning rate in a repeated Fisher market (Corollaries 4.1 and 4.2). Our theoretical results assumed fixed learning rates, but since those results apply to repeated games while our experiments apply to online Fisher markets, we selected variable learning

⁵The first term in this program is slightly different than the first term in the program presented by Goktas and Greenwald [25], since supply is assumed to be 1 for each work.

⁶Our code can be found at <https://github.com/Sadie-Zhao/Dynamic-Minmax-Games>.

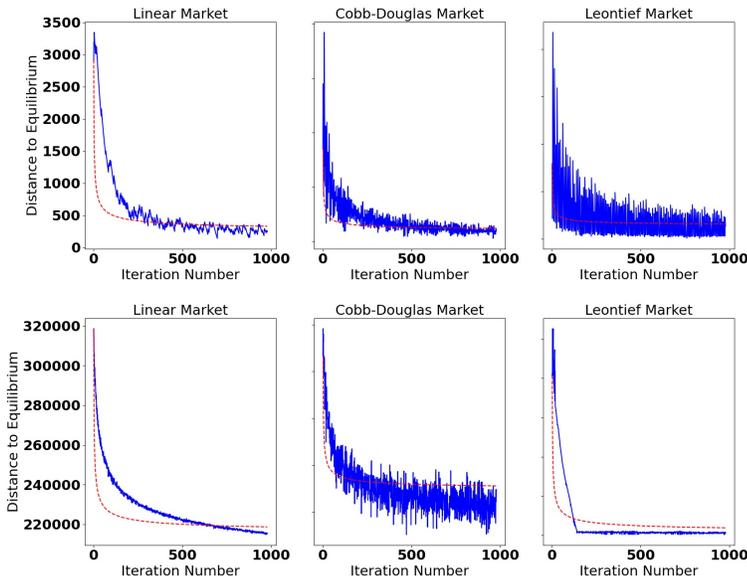


Figure 1: In blue, we depict a trajectory of distances between computed allocation-price pairs and equilibrium allocation-price pairs, when Algorithm 5 is run on randomly initialized online linear, Cobb-Douglas, and Leontief Fisher markets. In red, we plot an arbitrary $O(1/\sqrt{T})$ function.

Figure 2: In blue, we depict a trajectory of distances between computed allocation-price pairs and equilibrium allocation-price pairs, when Algorithm 6 is run on randomly initialized online linear, Cobb-Douglas, and Leontief Fisher markets. In red, we plot an arbitrary $O(1/\sqrt{T})$ function.

rates. After manual hyper-parameter tuning, for Algorithm 5, we chose a dynamic learning rate of $\eta_t = \frac{1}{\sqrt{t}}$, while for Algorithm 6, we chose learning rates of $\eta_t^x = \frac{5}{\sqrt{t}}$ and $\eta_t^y = \frac{0.01}{\sqrt{t}}$, for all $t \in [T]$. For these optimized learning rates, we obtain empirical convergence rates close to what the theory predicts.

In Fisher markets with Leontief utilities, the objective function is not differentiable. Correspondingly, online Fisher markets with Leontief utilities are the hardest markets of the three for our algorithms to solve. Still, we only see a slightly slower than $O(1/\sqrt{T})$ empirical convergence rate. In these experiments, the convergence curve generated by Algorithm 6 has a less erratic behavior than the one generated by Algorithm 5. Due to the non-differentiability of the objective function, the gradient ascent step in Algorithm 6 for buyers with Leontief utilities is very small, effectively dampening any potentially erratic changes in the iterates.

Our experiments suggest that OMD dynamics (Algorithms 5 and 6) are robust enough to closely track the changing CE in online Fisher markets. We note that tâtonnement dynamics (Algorithm 5) seem to be more robust than myopic best response dynamics (Algorithm 6), i.e., the distance to equilibrium allocations is smaller at each iteration of tâtonnement. This result is not surprising, as tâtonnement computes a utility-maximizing allocation for the buyers at each time step. Even though Theorems 4.4 and 4.5 only provide theoretical guarantees on the robustness of OMD dynamics in online min-max games (with independent strategy sets), it seems that similar theoretical robustness results may be attainable in online min-max Stackelberg games (with dependent strategy sets).

6 CONCLUSION

We began this paper by considering no-regret learning dynamics in repeated min-max Stackelberg games in two settings: an asymmetric setting in which the outer player is a no-regret learner and the inner player best responds, and a symmetric setting in which both players are no-regret learners. For both of these settings, we

proved that no-regret learning dynamics converge to a Stackelberg equilibrium of the game. We then specialized the no-regret algorithm employed by the players to online mirror descent (OMD), which yielded two new algorithms, max-oracle MD and nested MDA in the asymmetric setting, and a new simultaneous GDA-like algorithm [44], which we call Lagrangian MDA, in the symmetric setting. As these algorithms are no-regret learning algorithms, our earlier theorems imply convergence to ϵ -Stackelberg equilibria in $O(1/\epsilon^2)$ iterations for max-oracle MD and LMDA, and $O(1/\epsilon^3)$ iterations for nested MDA.

Finally, as many real-world applications involve changing environments, we investigated the robustness of OMD dynamics by analyzing how closely they track Stackelberg equilibria in arbitrary online min-max Stackelberg games. We proved that in min-max games (with independent strategy sets) OMD dynamics closely track the changing Stackelberg equilibria of a game. As we were not able to extend these theoretical robustness guarantees to min-max Stackelberg games (with dependent strategy sets), we instead ran a series of experiments with online Fisher markets, which are canonical examples of min-max Stackelberg games. Our experiments suggest that OMD dynamics are robust for min-max Stackelberg games so that perhaps the robustness guarantees we have provided for OMD dynamics in min-max games (with independent strategy sets) can be extended to min-max Stackelberg games (with dependent strategy sets).

The theory developed in this paper opens the door to extending the myriad applications of Stackelberg games in AI to incorporating dependent strategy sets. Such models promise to be more expressive, and as a result could provide decision makers with better solutions to problems in security, environmental protection, etc.

ACKNOWLEDGMENTS

We thank several anonymous reviewers for their feedback on an earlier draft of this paper.

This work was partially supported by NSF Grant CMMI-1761546.

REFERENCES

- [1] Jacob Abernethy, Kevin A Lai, and Andre Wibisono. 2019. Last-iterate convergence rates for min-max optimization. *arXiv preprint arXiv:1906.02027* (2019).
- [2] Mohammad Alkousa, Darina Dvinskikh, Fedor Stonyakin, Alexander Gasnikov, and Dmitry Kovalev. 2020. Accelerated methods for composite non-bilinear saddle point problem. *arXiv:1906.03620* [math.OC]
- [3] Kenneth Arrow and Gerard Debreu. 1954. Existence of an equilibrium for a competitive economy. *Econometrica: Journal of the Econometric Society* (1954), 265–290.
- [4] Yahav Behavod, Christopher Jung, and Steven Z Wu. 2020. Metric-free individual fairness in online learning. *Advances in neural information processing systems* 33 (2020), 11214–11225.
- [5] Aharon Ben-Tal and Arkadi Nemirovski. 2000. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical programming* 88, 3 (2000), 411–424.
- [6] Stephen Boyd, Stephen P Boyd, and Lieven Vandenbergh. 2004. *Convex optimization*. Cambridge university press.
- [7] William C Brainard, Herbert E Scarf, et al. 2000. *How to compute equilibrium prices in 1891*. Citeseer.
- [8] Qi Cai, Mingyi Hong, Yongxin Chen, and Zhaoran Wang. 2019. On the Global Convergence of Iteration Learning: A Case for Linear Quadratic Regulator. *arXiv:1901.03674* [cs.LG]
- [9] Yun Kuen Cheung, Martin Hoefer, and Paresh Nakhe. 2019. Tracing Equilibrium in Dynamic Markets via Distributed Adaptation. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems* (Montreal QC, Canada) (AAMAS '19). International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1225–1233.
- [10] Bo Dai, Hanjun Dai, Arthur Gretton, Le Song, Dale Schuurmans, and Niao He. 2019. Kernel Exponential Family Estimation via Doubly Dual Embedding. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics (Proceedings of Machine Learning Research, Vol. 89)*, Kamalika Chaudhuri and Masashi Sugiyama (Eds.). PMLR, 2321–2330. <http://proceedings.mlr.press/v89/dai19a.html>
- [11] Bo Dai, Albert Shaw, Lihong Li, Lin Xiao, Niao He, Zhen Liu, Jianshu Chen, and Le Song. 2018. SBEED: Convergent Reinforcement Learning with Nonlinear Function Approximation. In *Proceedings of the 35th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 80)*, Jennifer Dy and Andreas Krause (Eds.). PMLR, 1125–1134. <http://proceedings.mlr.press/v80/dai18c.html>
- [12] John M. Danskin. 1966. The Theory of Max-Min, with Applications. *SIAM J. Appl. Math.* 14, 4 (1966), 641–664. <http://www.jstor.org/stable/2946123>
- [13] Constantinos Daskalakis and Ioannis Panageas. 2018. Last-iterate convergence: Zero-sum games and constrained min-max optimization. *arXiv preprint arXiv:1807.04252* (2018).
- [14] Constantinos Daskalakis and Ioannis Panageas. 2018. The limit points of (optimistic) gradient descent in min-max optimization. *arXiv preprint arXiv:1807.03907* (2018).
- [15] Constantinos Daskalakis, Stratis Skoulakis, and Manolis Zampetakis. 2020. The Complexity of Constrained Min-Max Optimization. *arXiv:2009.09623* [cs.CC]
- [16] N. R. Devanur, C. H. Papadimitriou, A. Saberi, and V. V. Vazirani. 2002. Market equilibrium via a primal-dual-type algorithm. In *The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings.* 389–395. <https://doi.org/10.1109/SFCS.2002.1181963>
- [17] Jelena Diakonikolas, Constantinos Daskalakis, and Michael Jordan. 2021. Efficient methods for structured nonconvex-nonconcave min-max optimization. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 2746–2754.
- [18] Steven Diamond and Stephen Boyd. 2016. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research* 17, 83 (2016), 1–5.
- [19] Harrison Edwards and Amos Storkey. 2016. Censoring Representations with an Adversary. *arXiv:1511.05897* [cs.LG]
- [20] Francisco Facchinei and Christian Kanzow. 2007. Generalized Nash equilibrium problems. *4or* 5, 3 (2007), 173–210.
- [21] Francisco Facchinei and Christian Kanzow. 2010. Generalized Nash equilibrium problems. *Annals of Operations Research* 175, 1 (2010), 177–211.
- [22] Fei Fang and Thanh H. Nguyen. 2016. Green Security Games: Apply Game Theory to Addressing Green Security Challenges. *SIGeom Exch.* 15, 1 (Sept. 2016), 78–83. <https://doi.org/10.1145/2994501.2994507>
- [23] Yoav Freund and Robert E Schapire. 1996. Game theory, on-line prediction and boosting. In *Proceedings of the ninth annual conference on Computational learning theory.* 325–332.
- [24] Gauthier Gidel, Hugo Berard, Gaëtan Vignoud, Pascal Vincent, and Simon Lacoste-Julien. 2020. A Variational Inequality Perspective on Generative Adversarial Networks. *arXiv:1802.10551* [cs.LG]
- [25] Denizalp Goktas and Amy Greenwald. 2021. Convex-Concave Min-Max Stackelberg Games. *Advances in Neural Information Processing Systems* 34 (2021).
- [26] Denizalp Goktas, Enrique Areyan Viqueira, and Amy Greenwald. 2021. A Consumer-Theoretic Characterization of Fisher Market Equilibria. In *International Conference on Web and Internet Economics*. Springer, 334–351.
- [27] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in neural information processing systems* 27 (2014).
- [28] Erfan Yazdandoost Hamedani and Necdet Serhat Aybat. 2018. A primal-dual algorithm for general convex-concave saddle point problems. *arXiv preprint arXiv:1803.01401* 2 (2018).
- [29] E. Yazdandoost Hamedani, A. Jalilzadeh, N. S. Aybat, and U. V. Shanbhag. 2018. Iteration Complexity of Randomized Primal-Dual Methods for Convex-Concave Saddle Point Problems. *arXiv:1806.04118* [math.OC]
- [30] Charles R. Harris, K. Jarrod Millman, Stéfan J van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. 2020. Array programming with NumPy. *Nature* 585 (2020), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- [31] J. D. Hunter. 2007. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* 9, 3 (2007), 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- [32] Adam Ibrahim, Waiss Azizian, Gauthier Gidel, and Ioannis Mitliagkas. 2019. Lower bounds and conditioning of differentiable games. *arXiv preprint arXiv:1906.07300* (2019).
- [33] Chi Jin, Praneeth Netrapalli, and Michael I. Jordan. 2020. What is Local Optimality in Nonconvex-Nonconcave Minimax Optimization? *arXiv:1902.00618* [cs.LG]
- [34] Anatoli Juditsky, Arkadi Nemirovski, et al. 2011. First order methods for non-smooth convex large-scale optimization, ii: utilizing problems structure. *Optimization for Machine Learning* 30, 9 (2011), 149–183.
- [35] Sham M Kakade, Shai Shalev-Shwartz, and Ambuj Tewari. 2012. Regularization techniques for learning with matrices. *The Journal of Machine Learning Research* 13, 1 (2012), 1865–1890.
- [36] Debarun Kar, Thanh H Nguyen, Fei Fang, Matthew Brown, Arunesh Sinha, Milind Tambe, and Albert Xin Jiang. 2017. Trends and applications in Stackelberg security games. *Handbook of dynamic game theory* (2017), 1–47.
- [37] Tianyi Lin, Chi Jin, and Michael Jordan. 2020. On gradient descent ascent for nonconvex-concave minimax problems. In *International Conference on Machine Learning*. PMLR, 6083–6093.
- [38] Tianyi Lin, Chi Jin, and Michael I Jordan. 2020. Near-optimal algorithms for minimax optimization. In *Conference on Learning Theory*. PMLR, 2738–2779.
- [39] Songtao Lu, Ioannis Tsaknakis, and Mingyi Hong. 2019. Block alternating optimization for non-convex min-max problems: algorithms and applications in signal processing and communications. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 4754–4758.
- [40] David Madras, Elliot Creager, Toniann Pitassi, and Richard Zemel. 2018. Learning Adversarially Fair and Transferable Representations. *arXiv:1802.06309* [cs.LG]
- [41] Paul Milgrom and Ilya Segal. 2002. Envelope theorems for arbitrary choice sets. *Econometrica* 70, 2 (2002), 583–601.
- [42] Aryan Mokhtari, Asuman Ozdaglar, and Sarath Pattathil. 2020. Convergence Rate of $O(1/k)$ for Optimistic Gradient and Extra-gradient Methods in Smooth Convex-Concave Saddle Point Problems. *arXiv:1906.01115* [math.OC]
- [43] Dov Monderer and Lloyd S Shapley. 1996. Potential games. *Games and economic behavior* 14, 1 (1996), 124–143.
- [44] Angelia Nedic and Asuman Ozdaglar. 2009. Subgradient methods for saddle-point problems. *Journal of optimization theory and applications* 142, 1 (2009), 205–228.
- [45] Arkadi Nemirovski. 2004. 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 (2004), 229–251.
- [46] Arkadij Semenovic Nemirovskij and David Borisovich Yudin. 1983. Problem complexity and method efficiency in optimization. (1983).
- [47] Yurii Nesterov. 2007. Dual extrapolation and its applications to solving variational inequalities and related problems. *Mathematical Programming* 109, 2 (2007), 319–344.
- [48] J v Neumann. 1928. Zur theorie der gesellschaftsspiele. *Mathematische annalen* 100, 1 (1928), 295–320.
- [49] Thanh Hong Nguyen, Debarun Kar, Matthew Brown, Arunesh Sinha, Albert Xin Jiang, and Milind Tambe. 2016. Towards a Science of Security Games. In *Mathematical Sciences with Multidisciplinary Applications*, Bourama Toni (Ed.). Springer International Publishing, Cham, 347–381.
- [50] Maher Nouiehed, Maziar Sanjabi, Tianjian Huang, Jason D Lee, and Meisam Razaviyayn. 2019. Solving a class of non-convex min-max games using iterative first order methods. *arXiv preprint arXiv:1902.08297* (2019).
- [51] Dmitrii M Ostrovskii, Andrew Lowy, and Meisam Razaviyayn. 2020. Efficient search offi rst-order nash equilibria in nonconvex-concave smooth min-max problems. *arXiv preprint arXiv:2002.07919* (2020).
- [52] Yuyuan Ouyang and Yangyang Xu. 2018. Lower complexity bounds offi rst-order methods for convex-concave bilinear saddle-point problems. *arXiv:1808.02901* [math.OC]

- [53] The pandas development team. 2020. *pandas-dev/pandas: Pandas*. <https://doi.org/10.5281/zenodo.3509134>
- [54] Hassan Rafique, Mingrui Liu, Qihang Lin, and Tianbao Yang. 2019. Non-Convex Min-Max Optimization: Provable Algorithms and Applications in Machine Learning. arXiv:1810.02060 [math.OC]
- [55] Maziar Sanjabi, Jimmy Ba, Meisam Razaviyayn, and Jason D. Lee. 2018. On the Convergence and Robustness of Training GANs with Regularized Optimal Transport. arXiv:1802.08249 [cs.LG]
- [56] Maziar Sanjabi, Jimmy Ba, Meisam Razaviyayn, and Jason D. Lee. 2018. On the Convergence and Robustness of Training GANs with Regularized Optimal Transport. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems (Montréal, Canada) (NIPS'18)*. Curran Associates Inc., Red Hook, NY, USA, 7091–7101.
- [57] Prasanna Sattigeri, Samuel C. Hoffman, Vijil Chenthamarakshan, and Kush R. Varshney. 2018. Fairness GAN. arXiv:1805.09910 [stat.ML]
- [58] Shai Shalev-Shwartz et al. 2011. Online learning and online convex optimization. *Foundations and trends in Machine Learning* 4, 2 (2011), 107–194.
- [59] Arunesh Sinha, Fei Fang, Bo An, Christopher Kiekintveld, and Milind Tambe. 2018. Stackelberg security games: Looking beyond a decade of success. IJCAI.
- [60] Aman Sinha, Hongseok Namkoong, Riccardo Volpi, and John Duchi. 2020. Certifying Some Distributional Robustness with Principled Adversarial Training. arXiv:1710.10571 [stat.ML]
- [61] Morton Slater. 1959. *Lagrange Multipliers Revisited*. Cowles Foundation Discussion Papers 80. Cowles Foundation for Research in Economics, Yale University. <https://EconPapers.repec.org/RePEc:cwl:cwldpp:80>
- [62] Morton Slater. 2014. *Lagrange Multipliers Revisited*. Springer Basel, Basel, 293–306. https://doi.org/10.1007/978-3-0348-0439-4_14
- [63] Kiran Koshy Thekumparampil, Prateek Jain, Praneeth Netrapalli, and Sewoong Oh. 2019. Efficient Algorithms for Smooth Minimax Optimization. arXiv:1907.01543 [math.OC]
- [64] Paul Tseng. 1995. On linear convergence of iterative methods for the variational inequality problem. *J. Comput. Appl. Math.* 60, 1 (1995), 237–252. [https://doi.org/10.1016/0377-0427\(94\)00094-H](https://doi.org/10.1016/0377-0427(94)00094-H) Proceedings of the International Meeting on Linear/Nonlinear Iterative Methods and Verification of Solution.
- [65] Paul Tseng. 2008. On accelerated proximal gradient methods for convex-concave optimization. *submitted to SIAM Journal on Optimization* 1 (2008).
- [66] Guido Van Rossum and Fred L Drake Jr. 1995. *Python tutorial*. Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands.
- [67] Heinrich Von Stackelberg. 1934. *Marktform und gleichgewicht*. J. springer.
- [68] Léon Walras. 1969. *Elements of Pure Economics; or, The Theory of Social Wealth*. Translated by William Jaffé. Vol. 2. Orion Editions. 457–458 pages.
- [69] Laura Scrimali Yurii Nesterov. 2011. Solving strongly monotone variational and quasi-variational inequalities. *Discrete & Continuous Dynamical Systems* 31, 4 (2011), 1383–1396.
- [70] Junyu Zhang, Mingyi Hong, and Shuzhong Zhang. 2020. On Lower Iteration Complexity Bounds for the Saddle Point Problems. arXiv:1912.07481 [math.OC]
- [71] Renbo Zhao. 2019. Optimal algorithms for stochastic three-composite convex-concave saddle point problems. *arXiv preprint arXiv:1903.01687* (2019).
- [72] Renbo Zhao. 2020. A Primal Dual Smoothing Framework for Max-Structured Nonconvex Optimization. arXiv:2003.04375 [math.OC]