Min-max Games with Dependent Strategy Sets

Anonymous Author(s)
Affiliation
Address
email

Abstract

Min-max optimization problems (i.e., games) are attracting a great deal of attention because of their applicability to a wide range of machine learning problems. Although significant progress has been made recently, the literature to date has focused on games with independent strategy sets; little is known about solving games with dependent strategy sets. We introduce a first-order method that solves a large class of convex-concave min-max games with dependent strategy sets, and show that our method converges in polynomial time.

To illustrate the importance of solving min-max games with dependent strategy sets, we observe that the computation of competitive equilibria in Fisher markets is an instance of this model. Further, we demonstrate the efficacy and efficiency of our algorithms in practice by computing competitive equilibria in Fisher markets with varying utility structures. Our experiments suggest potential ways to extend our theoretical results, by demonstrating how different smoothness properties can affect the convergence rate of our algorithms.

1 Introduction

Min-max optimization problems have attracted a great deal of attention recently because of their applicability to a wide range of machine learning problems. Examples of settings in which min-max optimization problems arise include, but are not limited to, reinforcement learning [13], generative adversarial networks [54], fairness in machine learning [12, 20, 40, 56, 69], adversarial learning [60], generative adversarial imitation learning [9, 26], and statistical learning (e.g., learning parameters of exponential families) [12], . These applications often require solving a **constrained min-max optimization problem** (with independent feasible sets): i.e., $\min_{\boldsymbol{x}\in X}\max_{\boldsymbol{y}\in Y} f(\boldsymbol{x},\boldsymbol{y})$, where $f : X \times Y \to \mathbb{R}$ is continuous, and $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ are non-empty and compact.

A **convex-concave** constrained min-max optimization problem is one in which f is convex in $\boldsymbol{x}$ and concave in $\boldsymbol{y}$. In the special case of convex-concave objective functions, the seminal minimax theorem holds: i.e., $\min_{\boldsymbol{x}\in X}\max_{\boldsymbol{y}\in Y} f(\boldsymbol{x},\boldsymbol{y}) = \max_{\boldsymbol{y}\in Y}\min_{\boldsymbol{x}\in X} f(\boldsymbol{x},\boldsymbol{y})$ [46]. This theorem guarantees the existence of a saddle point (i.e., a point that is simultaneously a minimum of f in the $\boldsymbol{x}$-direction and a maximum of f in the $\boldsymbol{y}$-direction), and allows us to interpret the optimization problem as a simultaneous-move, zero-sum game, where $\boldsymbol{y}^*$ (resp. $\boldsymbol{x}^*$) is a best-response of the outer (resp. inner) player to their opponent's action $\boldsymbol{x}^*$ (resp. $\boldsymbol{y}^*$), in which case a saddle point is also called a minimax point or a Nash equilibrium.

In this paper, we show that the computation of competitive equilibria in Fisher markets [8], a central solution concept for one of the most well-studied market models in algorithmic game theory [47], can be understood as a convex-concave min-max optimization problem, albeit with *dependent* feasible sets. We define a **constrained min-max optimization problem with dependent feasible sets** as an optimization problem of the following form: $\min_{\boldsymbol{x}\in X}\max_{\boldsymbol{y}\in Y:\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y})\geq \boldsymbol{0}} f(\boldsymbol{x},\boldsymbol{y})$, where $f : X \times Y \to \mathbb{R}$ is continuous, $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ are non-empty and compact, and $\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y}) =$

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.

$(g_1(\boldsymbol{x}, \boldsymbol{y}), \ldots, g_K(\boldsymbol{x}, \boldsymbol{y}))^T$ with $g_k : X \times Y \to \mathbb{R}$.[1] Unfortunately, certain desirable properties of standard convex-concave min-max optimization problems with independent feasible sets do not carry over to the more general dependent setting. First and foremost, the minimax theorem [46] does not hold, which in turn precludes the existence of Nash equilibria:

Example 1.1. *Consider the following constrained min-max optimization problem with dependent feasible sets:* $\min_{x \in [-1,1]} \max_{y \in [-1,1]: x+y \leq 0} x^2 + y + 1$. *The optimum is* $x^* = 1/2, y^* = -1/2$, *with value* $3/4$. *Now, consider the same constrained problem, with the order of the* min *and the* max *reversed:* $\max_{y \in [-1,1]} \min_{x \in [-1,1]: x+y \leq 0} x^2 + y + 1$. *The optimum is now* $x^* = -1, y^* = 1$, *with value* 3.

Without a minimax theorem, constrained min-max optimization problems with dependent feasible sets are more appropriately viewed as sequential games, where the outer player chooses $\boldsymbol{x} \in X$ before the inner player responds with their choice of $\boldsymbol{y}(\boldsymbol{x}) \in X$ s.t. $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{x})) \geq \boldsymbol{0}$. The relevant equilibrium concept is then a Stackelberg [66] equilibrium,[2] in which the outer player optimizes their choice assuming the inner player will best-respond: i.e., optimize their choice in turn.

In this paper, we develop algorithms that converge to Stackelberg equilibria in constrained convex-concave min-max optimization problems with dependent feasible sets—hereafter, min-max games with dependent strategy sets. In these games, the **value function**[3] $V : X \to \mathbb{R}$ for the outer player is defined as $V(\boldsymbol{x}) = \max_{\boldsymbol{y} \in Y: \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) \geq 0} f(\boldsymbol{x}, \boldsymbol{y})$. This function represents the outer player's loss, assuming the inner player chooses a feasible best-response, so it is the function the outer player seeks to minimize. The inner player's value function, which they seek to maximize, is simply the objective function. The first-order necessary and sufficient conditions for a tuple $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X \times Y$ to be a Stackelberg equilibrium in a convex-concave min-max game with dependent strategy sets are given by the KKT stationarity conditions for the problems $\min_{\boldsymbol{x} \in X} V(\boldsymbol{x})$ for the $\boldsymbol{x}$-player, and $V(\boldsymbol{x})$ for the $\boldsymbol{y}$-player. *The algorithms developed in this paper are first-order methods, with the flavor of gradient descent/ascent, which follow the gradients of the players' value functions.*

In the independent strategy setting, Danskin's theorem [15], states that $\nabla_{\boldsymbol{x}} V(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} f(\boldsymbol{x}, \boldsymbol{y}^*)$, where $\boldsymbol{y}^* \in \arg\max_{\boldsymbol{y} \in Y} f(\boldsymbol{x}, \boldsymbol{y})$. In other words, when there is no dependence among the players' strategy sets, the gradient of the value function coincides with that of the objective function. As the first-order necessary and sufficient conditions for a tuple $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ to be a saddle point are for it to be a stationary point of f (i.e., $\nabla_{\boldsymbol{x}} f(\boldsymbol{x}^*, \boldsymbol{y}^*) = \nabla_{\boldsymbol{y}} f(\boldsymbol{x}^*, \boldsymbol{y}^*) = 0$). It is therefore logical for players in the independent setting to follow the gradient of the objective function. More generally, however, the direction of steepest descent (resp. ascent) for the outer (resp. inner) player is the gradient of their value function.

Example 1.2. *Consider, once again, the problem posed in Example 1.1, and recall Jin, Netrapalli, and Jordan's gradient descent with max-oracle algorithm [34]:* $\boldsymbol{x}^{(t+1)} = \boldsymbol{x}^{(t)} - \eta \nabla_{\boldsymbol{x}} f\left(\boldsymbol{x}^t, \boldsymbol{y}^*(\boldsymbol{x}^{(t)})\right)$, *where* $\boldsymbol{y}^*(\boldsymbol{x}^{(t)}) \in \arg\max_{\boldsymbol{y} \in Y: \boldsymbol{g}(\boldsymbol{x}^{(t)}, \boldsymbol{y}) \geq 0} f(\boldsymbol{x}^{(t)}, \boldsymbol{y})$ *for* $\eta > 0$. *Applied to this sample problem, with* $\eta = 1$, *this algorithm yields the following update rule:* $x^{(t+1)} = x^{(t)} - 2x^{(t)} = -x^{(t)}$. *Thus, letting* $x^{(0)}$ *equal any feasible* x, *the output cycles between* x *and* $-x$, *so that the average of the iterates converges to* $x^* = 0$ *(with* $y^* = 0$*), which is not a Stackelberg equilibrium, as the Stackelberg equilibrium of this game is* $x^* = 1/2, y^* = -1/2$.

Now consider an algorithm that updates based not on gradient of the objective function, but rather of the value function, namely $\boldsymbol{x}^{(t+1)} = \boldsymbol{x}^{(t)} - \eta \nabla_{\boldsymbol{x}} V(\boldsymbol{x}^t)$. *The value function is* $V(x) = \max_{y \in [-1,1]: x+y \leq 0} x^2 + y + 1 = x^2 - x + 1$, *with gradient* $V'(x) = 2x - 1$. *Thus, when* $\eta = 1$, *this algorithm yields the following update rule:* $x^{(t+1)} = x^{(t)} - V'(x^{(t)}) = x^{(t)} - 2x^{(t)} + 1 = -x^{(t)} + 1$. *If we run this algorithm from initial point* $x^{(0)} = 1/8$, *we get* $x^{(1)} = -1/8 + 1 = 7/8$, $x^{(2)} = -7/8 + 1 = 1/8$, *and so on. The average of the iterates* $1/8, 7/8, 1/8, \ldots$ *converges to* $x^* = 1/2\,(1/8 + 7/8) = 1/2$, *and correspondingly* $y^* = -1/2$, *which is indeed the Stackelberg equilibrium.*

[1] Although notationally similar, we note that this model is more general than that of Daskalakis, Skoulakis, and Zampetakis [16]. The authors state that under the assumptions of the minimax theorem [46], a Nash equilibrium is guaranteed to exist in their model. When there is dependence among the players' feasible sets, however, this claim does not hold (see Example 1.1).

[2] In constrained convex-concave min-max optimization problems with independent feasible sets, the notions of Nash and Stackelberg equilibria coincide. They also coincide in the dependent setting, *when Nash equilibria exist*, but they do not coincide in general, since Nash equilibria need not exist.

[3] Note that this use of the term value function in economics is distinct from its use in reinforcement learning.

In this paper, we introduce two first-order (subgradient) methods that solve min-max games with dependent strategy sets—to our knowledge the first such methods. Our approach relies on a new generalization of a series of fundamental results in mathematical economics known as **envelope theorems** [1, 41]. Envelope theorems generalize aspects of Danskin's theorem, by providing explicit formulas for the gradient of the outer player's value function in dependent strategy settings *when a derivative is guaranteed to exist.* To sidestep the differentiability issue, we introduce a generalized envelope theorem that gives an explicit formula for the subdifferential of the outer player's value function in dependent strategy settings.

Our first algorithm follows Jin, Netrapalli, and Jordan [34], assuming access to a max-oracle that returns $\boldsymbol{y}^* \in \arg\max_{\boldsymbol{y} \in Y : \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) \geq 0} f(\boldsymbol{x}, \boldsymbol{y})$, given $\boldsymbol{x} \in X$. Hence, our first algorithm solves only for an optimal $\boldsymbol{x}^*$, while our second algorithm explicitly solves for both $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$. We show that both algorithms converge in polynomial time to a Stackelberg equilibrium of any convex-concave min-max game with nonempty, compact and convex dependent strategy sets. We summarize the iteration complexities of our algorithms (i.e., the number of iterations required to achieve an ε-approximate equilibrium, where ε is the desired precision) in Table 1.

Table 1: Iteration complexities of Algorithms 1 and 2 for min-max games with dependent strategy sets. Here, $\mu_{\boldsymbol{x}}$ and $\mu_{\boldsymbol{y}}$ are strong convexity/concavity parameters.

Properties of f	Iteration Complexity	
	Algorithm 1	Algorithm 2
$\mu_{\boldsymbol{x}}$-Strongly-Convex-$\mu_{\boldsymbol{y}}$-Strongly-Concave	$O\left(\varepsilon^{-1}\right)$	$\tilde{O}(\varepsilon^{-1})$
$\mu_{\boldsymbol{x}}$-Strongly-Convex-Concave		$O(\varepsilon^{-2})$
Convex-$\mu_{\boldsymbol{y}}$-Strongly-Concave	$O\left(\varepsilon^{-2}\right)$	$\tilde{O}(\varepsilon^{-2})$
Convex-Concave		$O(\varepsilon^{-3})$

Finally, we apply our results to the computation of competitive equilibria in Fisher markets. In this context, our method for solving a min-max game with dependent strategy sets reduces to solving the market in a decentralized manner using a natural market dynamic called tâtonnement [67]. We demonstrate the efficacy and efficiency of our algorithms in practice by running a series of experiments in which we compute competitive equilibria in Fisher markets with varying utility structures—specifically, linear, Cobb-Douglas, and Leontief. Although our theoretical results do not apply to all these Fisher markets—Leontief utilities, in particular, are not differentiable—tâtonnement converges in all experiments. That said, the rate of convergence does seem to depend on the precise smoothness characteristics of the utility structures; we observe slower convergence for Leontief utilities, and faster convergence than our theory predicts for Cobb-Douglas utilities, which are not only differentiable, but whose value function is also differentiable.

Related Work To the best of our knowledge, algorithms designed to solve min-max games with dependent strategy sets are few, and are not accompanied by polynomial-time computation guarantees. Shimizu and Aiyoshi [58, 59] proposed the first such algorithm via a relaxation to a constrained optimization problem with infinitely many constraints, which nonetheless seems to perform well in practice. More recently, Segundo, Krohling, and Cosme [57] proposed an evolutionary algorithm designed to solve min-max games with dependent strategy sets, but they provide no guarantees. Extensive-form games in which players' strategy sets can depend on other players' actions have been studied by Davis, Waugh, and Bowling [17] assuming payoffs are bilinear, and by Farina, Kroer, and Sandholm [22] for another specific class of convex-concave payoffs.

The study of the computation of competitive equilibria in Fisher markets was initiated by Devanur et al. [18], who provided a polynomial-time method for the case of linear utilities. Jain, Vazirani, and Ye [33] subsequently showed that a large class of Fisher markets could be solved in polynomial-time using interior point methods. Recently, Gao and Kroer [23] proposed first-order methods for solving Fisher markets (only; not general min-max games with dependent strategy sets), assuming linear, quasilinear, and Leontief utilities, as such methods can be more efficient when markets are large. See appendix F for a detailed discussion of recent progress on solving min-max games assuming independent strategy sets, both in the convex-concave case, and the non-convex-concave case.

2 Preliminaries

Notation We use Roman uppercase letters to denote sets (e.g., X), bold uppercase letters to denote matrices (e.g., $\boldsymbol{X}$), bold lowercase letters to denote vectors (e.g., $\boldsymbol{p}$), and Roman lowercase letters to denote scalar quantities, (e.g., c). We denote the ith row vector of a matrix (e.g., $\boldsymbol{X}$) by the corresponding bold lowercase letter with subscript i (e.g., $\boldsymbol{x}_i$). Similarly, we denote the jth entry of a vector (e.g., $\boldsymbol{p}$ or $\boldsymbol{x}_i$) by the corresponding Roman lowercase letter with subscript j (e.g., p_j or x_{ij}). We denote the set of integers $\{1, \ldots, n\}$ by $[n]$, the set of natural numbers by $\mathbb{N}$, the set of real numbers by $\mathbb{R}$, the set of non-negative real numbers by $\mathbb{R}_+$.

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

One way to see this game is as a **Stackelberg game**, i.e., a sequential game with two players, where WLOG, we assume that the minimizing player goes first and the maximizing player goes second. The relevant solution concept for Stackelberg games is the **Stackelberg equilibrium**:

Definition 2.1 (Stackelberg Equilibrium)**.** *Consider the min-max game with dependent strategy sets $(X, Y, f, \boldsymbol{g})$. A strategy profile $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X \times Y$ such that $\boldsymbol{g}(\boldsymbol{x}^*, \boldsymbol{y}^*) \geq 0$ is a (ε, δ)-Stackelberg equilibrium if* $\max_{\boldsymbol{y} \in Y : \boldsymbol{g}(\boldsymbol{x}^*, \boldsymbol{y}) \geq 0} f(\boldsymbol{x}^*, \boldsymbol{y}) - \delta \leq f(\boldsymbol{x}^*, \boldsymbol{y}^*) \leq \min_{\boldsymbol{x} \in X} \max_{\boldsymbol{y} \in Y : \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) \geq 0} f(\boldsymbol{x}, \boldsymbol{y}) + \varepsilon$.

Intuitively, a (ε, δ)-Stackelberg equilibrium is a point at which the $\boldsymbol{x}$-player's (resp. $\boldsymbol{y}$-player's) payoff is no more than ε (resp. δ) away from its optimum. By Proposition A.1, a $(0, 0)$-Stackelberg equilibrium is guaranteed to exist in min-max games with dependent strategy sets.

Finally, we define several concepts that are needed in our convergence proofs. Given $A \subset \mathbb{R}^n$, the function $f : A \to \mathbb{R}$ is said to be ℓ_f-**Lipschitz-continuous** iff $\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in X, \|f(\boldsymbol{x}_1) - f(\boldsymbol{x}_2)\| \leq \ell_f \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|$. If the gradient of f, ∇f, is $\ell_{\nabla f}$-Lipschitz-continuous, we then refer to f as $\ell_{\nabla f}$-**Lipschitz-smooth**. A function $f : A \to \mathbb{R}$ is μ-**strongly convex** if $f(\boldsymbol{x}_1) \geq f(\boldsymbol{x}_2) + \langle \nabla_{\boldsymbol{x}} f(\boldsymbol{x}_2), \boldsymbol{x}_1 - \boldsymbol{x}_2 \rangle + {}^{\mu}/_{2} \|\boldsymbol{x}_1 - \boldsymbol{x}_1\|^2$, and μ-**strongly concave** if $-f$ is μ-strongly convex.

3 First-Order Methods via an Envelope Theorem

The envelope theorems, popular tools in mathematical economics, allow for explicit formulas for the gradient of the outer player's value function in min-max games, even when the strategy sets are dependent. Afriat [1] appears to have been the first make use of the Lagrangian to differentiate the value function, though his conclusion was later obtained under weaker assumptions by Milgrom and Segal [41]. Our envelope theorem relies on the following assumptions:

Assumption 3.1. *1. $f, g_1, \ldots, g_K$ are continuous and concave in $\boldsymbol{y}$; 2. $\nabla_{\boldsymbol{x}} f, \nabla_{\boldsymbol{x}} g_1, \ldots, \nabla_{\boldsymbol{x}} g_K$ are continuous in $(\boldsymbol{x}, \boldsymbol{y})$; and 3. $\forall \boldsymbol{x} \in X, \exists \widehat{\boldsymbol{y}} \in Y$ s.t. $g_k(\boldsymbol{x}, \widehat{\boldsymbol{y}}) > 0$, for all $k = 1, \ldots, K$.*

Milgrom and Segal's [41] envelope theorem provides an explicit formula for the gradient of the value function of the outer player. When strategy sets are independent, under mild assumptions, this value function is guaranteed to be differentiable [48]. When strategy sets are dependent, however, this value function is not necessarily differentiable, as seen in Example B.2. As a remedy, we present a subdifferential envelope theorem for non-differentiable value functions:

Theorem 3.2 (Subdifferential Envelope Theorem)**.** *Consider the value function $V(\boldsymbol{x}) = \max_{\boldsymbol{y} \in Y : \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) \geq 0} f(\boldsymbol{x}, \boldsymbol{y})$. Let $Y^*(\boldsymbol{x}) = \arg\max_{\boldsymbol{y} \in Y : \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) \geq 0} f(\boldsymbol{x}, \boldsymbol{y})$ and suppose Assumption 3.1 holds. Then, at any point $\widehat{\boldsymbol{x}} \in X$, $\partial_{\boldsymbol{x}} V(\widehat{\boldsymbol{x}}) =$*

$$\operatorname{conv}\left(\bigcup_{\boldsymbol{y}^*(\widehat{\boldsymbol{x}}) \in Y^*(\widehat{\boldsymbol{x}})} \bigcup_{\lambda_k(\widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\widehat{\boldsymbol{x}})) \in \Lambda(\widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\widehat{\boldsymbol{x}}))} \left\{ \nabla_{\boldsymbol{x}} f\left(\widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\widehat{\boldsymbol{x}})\right) + \sum_{k=1}^{K} \lambda_k(\widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\widehat{\boldsymbol{x}})) \nabla_{\boldsymbol{x}} g_k\left(\widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\widehat{\boldsymbol{x}})\right) \right\} \right), \tag{1}$$

where $\boldsymbol{\lambda}(\widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\widehat{\boldsymbol{x}})) = (\lambda_1(\widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\widehat{\boldsymbol{x}})), \ldots, \lambda_K(\widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\widehat{\boldsymbol{x}})))^T \in \Lambda(\widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\widehat{\boldsymbol{x}}))$ *are the Lagrange multipliers associated with* $\boldsymbol{y}^*(\widehat{\boldsymbol{x}}) \in Y^*(\widehat{\boldsymbol{x}})$ *and* conv *is the convex hull operator.*

This (subdifferential) envelope theorem tell us that the (sub)gradient of the **value function**, $V(\boldsymbol{x}) = \max_{\boldsymbol{y} \in Y : g(\boldsymbol{x}, \boldsymbol{y}) \geq 0} f(\boldsymbol{x}, \boldsymbol{y})$ is a (convex combinations of any possible) gradient of the Lagrangian evaluated at an optimal $(\boldsymbol{y}^*(\boldsymbol{x}), \boldsymbol{\lambda}(\boldsymbol{x}, \boldsymbol{y}^*(\boldsymbol{x}))) \in Y^*(\boldsymbol{x}) \times \Lambda(\boldsymbol{x}, \boldsymbol{y}^*(\boldsymbol{x}))$.

With our envelope theorem in hand, we are now ready to present two gradient-descent/ascent-type algorithms for min-max games with dependent strategy sets, which follow the gradient of the players' value functions.[4] Our first algorithm, **max-oracle gradient-descent**, following Jin, Netrapalli, and Jordan [34], assumes access to a max-oracle, which given $\boldsymbol{x} \in X$, returns a δ-maximum of the inner player's value function. That is, for all $\boldsymbol{x} \in X$, the max-oracle returns $\widehat{\boldsymbol{y}} \in Y$ s.t. $\boldsymbol{g}(\boldsymbol{x}, \widehat{\boldsymbol{y}}) \geq 0$ and $f(\boldsymbol{x}, \widehat{\boldsymbol{y}}) \geq \max_{\boldsymbol{y} \in Y : \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) \geq 0} f(\boldsymbol{x}, \boldsymbol{y}) - \delta$. It then runs (sub)gradient descent on the outer player's value function, using Theorem 3.2 to compute the requisite subgradients. Inspired by the multi-step gradient-descent algorithm of Nouiehed et al. [48], our second algorithm, **nested gradient-descent/ascent**, computes both $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$ explicitly, without oracle access. We simply replace the max-oracle in our max-oracle gradient-descent algorithm by a projected gradient-ascent procedure, which again computes a δ-maximum of the inner player's value function.

Once $\widehat{\boldsymbol{y}}$ is found at iteration t, one can compute optimal Lagrangian multipliers $\lambda_1(\boldsymbol{x}^{(t)}, \widehat{\boldsymbol{y}}(\boldsymbol{x}^{(t)}))$, $\ldots$, $\lambda_K(\boldsymbol{x}^{(t)}, \widehat{\boldsymbol{y}}(\boldsymbol{x}^{(t)}))$ for the outer player's value function, either via a system of linear equations using the complementary slackness conditions and the value of the objective function at the optimal, namely $(\boldsymbol{x}^{(t)}, \widehat{\boldsymbol{y}}(\boldsymbol{x}^{(t)}))$, or by running gradient descent on the Lagrangian of the dual variables. Additionally, most algorithms solving convex programs will return $\boldsymbol{\lambda}(\boldsymbol{x}^{(t)}, \widehat{\boldsymbol{y}}(\boldsymbol{x}^{(t)})) = (\lambda_1(\boldsymbol{x}^{(t)}, \widehat{\boldsymbol{y}}(\boldsymbol{x}^{(t)})), \ldots, \lambda_K(\boldsymbol{x}^{(t)}, \widehat{\boldsymbol{y}}(\boldsymbol{x}^{(t)})))$ together with the optimal $\widehat{\boldsymbol{y}}(\boldsymbol{x}^{(t)}))$ without incurring any additional computational expense. As a result, we assume that the optimal Lagrangian multipliers $\boldsymbol{\lambda}(\boldsymbol{x}^{(t)}, \widehat{\boldsymbol{y}}(\boldsymbol{x}^{(t)}))$ associated with a solution $\widehat{\boldsymbol{y}}(\boldsymbol{x}^{(t)}))$ can be computed in constant time.

Having explained our two procedures, our next task is to derive their convergence rates. In this vein, an important question that remains to be answered is what assumptions are needed on the objective function and constraints, and how convergence rates vary with these assumptions.

It turns out that under very mild assumptions, i.e., when Assumption 3.1 holds, the outer player's value function is Lipschitz continuous in $\boldsymbol{x}$. More precisely, the objective function f is ℓ_f-Lipschitz in $\boldsymbol{x}$, where $\ell_f = \max_{(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}) \in X \times Y} \|\nabla_{\boldsymbol{x}} f(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}})\|$.[5] As f is ℓ_f-Lipschitz continuous in $\boldsymbol{x}$, the value function is also ℓ_f-Lipschitz.[6] This fact suggests that an $(\varepsilon, \varepsilon)$-Stackelberg equilibrium should be computable in $O(\varepsilon^{-2})$ iterations by our max-oracle gradient descent algorithm (Algorithm 1), since our method is a subgradient method.

Theorem 3.3. *Consider a convex-concave min-max game with dependent strategy sets,* $(X, Y, f, \boldsymbol{g})$*, where* X *is convex, and* $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ *is a* $(0, \delta)$*-Stackelberg equilibrium, and suppose Assumption 3.1 holds. If Algorithm 1 is run for* T *iterations, with step sizes* η_t *s.t.* $\sum_{k=1}^T \eta_k^2 \leq \infty, \sum_{k=1}^T \eta_k = \infty$*, and if* $(\boldsymbol{x}_{\text{best}}^{(t)}, \boldsymbol{y}_{\text{best}}^{(t)}) \in \arg\min_{(\boldsymbol{x}^{(k)}, \boldsymbol{y}^{(k)}) : k \in [t]} f(\boldsymbol{x}^{(k)}, \boldsymbol{y}^{(k)})$*, it holds that* $\lim_{T \to \infty} f(\boldsymbol{x}_{\text{best}}^{(T)}, \boldsymbol{x}_{\text{best}}^{(T)}) = f(\boldsymbol{x}^*, \boldsymbol{y}^*)$*. Furthermore, for* $\varepsilon \in (0, 1)$*, if we choose* T *s.t.* $T \geq N_T(\varepsilon) := O(\varepsilon^{-2})$*, then there exists an iteration* $T^* \leq T$ *such that* $(\boldsymbol{x}_{\text{best}}^{(T^*)}, \boldsymbol{y}_{\text{best}}^{(T^*)})$ *is an* (ε, δ)*-Stackelberg equilibrium.*

As is expected, the $O(\varepsilon^{-2})$ iteration complexity can be improved to $O(\varepsilon^{-1})$, if additionally, f is strongly convex in $\boldsymbol{x}$. (See Appendix D, Theorem D.2).

Combining the convergence results for our max-oracle gradient descent algorithm with convergence results for gradient descent [6], we obtain the following convergence rates for the nested gradient descent-ascent algorithm (Algorithm 2). We include the formal proof and statement for the case when Assumption 3.1 holds and f is Lipschitz-smooth in Appendix D (Theorem D.3). The other results follow similarly.

Theorem 3.4. *Consider a convex-concave min-max game with dependent strategy sets,* $(X, Y, f, \boldsymbol{g})$*, where both* X *and* Y *are convex. Suppose further that Assumption 3.1 holds.*

[4] We include pseudocode for both our algorithms in Appendix D (Algorithm 1 and Algorithm 2).

[5] This max-norm is well-defined since $\nabla_{\boldsymbol{x}} f$ is continuous and the constraint set is non-empty and compact.

[6] A formal statement of this claim can be found in Appendix D (Lemma D.1).

Then, under standard assumptions on the step sizes, the iteration complexities given in Table 2 hold for the computation of a $(\varepsilon, \varepsilon)$-Stackelberg equilibrium:

Table 2: Summary of the iteration complexity of Algorithm 2 under different assumptions on f.

	f is $\ell_{\nabla f}$-smooth	f is $\ell_{\nabla f}$-smooth + f Strongly Concave in $\boldsymbol{y}$
Assumption 3.1	$O(\varepsilon^{-3})$	$O\left(\varepsilon^{-2}\log(\varepsilon)\right)$
Assumption 3.1 + f Strongly Convex in $\boldsymbol{x}$	$O(\varepsilon^{-2})$	$O(\varepsilon^{-1}\log(\varepsilon))$

Since the value function in the convex-concave dependent setting is not guaranteed to be differentiable (see Example B.2), we cannot ensure that the objective function is Lipschitz-smooth in general. Thus, unlike previous results for the independent setting that required this latter assumption to achieve faster convergence (e.g., [48]), in our analysis of Algorithm 1, we assume only that the objective function is continuously differentiable, which leads to a more widely applicable, albeit slower, convergence rate. Note, however, that we assume Lipschitz-smoothness in our analysis of Algorithm 2, as it allows for faster convergence of the optimal strategy of the inner player, but this assumption could also be done away with again, at the cost of a slower convergence rate.

4 An Economic Application: Fisher Markets

The Fisher market model, attributed to Irving Fisher [8], has received a great deal of attention recently, in particular by computer scientists, as its applications to fair division and mechanism design have proven useful for the design of automated markets in many online marketplaces. In this section, we argue that a competitive equilibrium in Fisher markets can be understood a Stackelberg equilibrium of a convex-concave min-max game with dependent strategy sets. We then apply our first-order methods to compute these equilibria in various Fisher markets.

A **Fisher market** consists of n buyers and m divisible goods [8]. Each buyer $i \in [n]$ has a budget $b_i \in \mathbb{R}_{++}$ and a utility function $u_i : \mathbb{R}^m_+ \to \mathbb{R}$. As is standard in the literature, we assume that there is one divisible unit of each good in the market [47]. An instance of a Fisher market is given by a tuple $(n, m, U, \boldsymbol{b})$, where $U = \{u_1, \ldots, u_n\}$ is a set of utility functions, one per buyer, and $\boldsymbol{b} \in \mathbb{R}^n_+$ is the vector of buyer budgets. We abbreviate as $(U, \boldsymbol{b})$ when n and m are clear from context.

An **allocation** $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T \in \mathbb{R}^{n\times m}_+$ is a map from 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** $\boldsymbol{p} = (p_1, \ldots, p_m)^T \in \mathbb{R}^m_+$. A tuple $(\boldsymbol{p}^*, \boldsymbol{X}^*)$ is said to be a **competitive (or Walrasian) equilibrium** of Fisher market $(U, \boldsymbol{b})$ if 1. buyers are utility maximizing, constrained by their budget, i.e., $\forall i \in [n], \boldsymbol{x}^*_i \in \arg\max_{\boldsymbol{x}:\boldsymbol{x}\cdot\boldsymbol{p}^* \leq b_i} u_i(\boldsymbol{x})$; and 2. the market clears, i.e., $\forall j \in [m], p^*_j > 0 \Rightarrow \sum_{i\in[n]} x^*_{ij} = 1$ and $p^*_j = 0 \Rightarrow \sum_{i\in[n]} x^*_{ij} \leq 1$.

We now formulate the problem of computing a competitive equilibrium $(\boldsymbol{p}^*, \boldsymbol{X}^*)$ of a Fisher market $(U, \boldsymbol{b})$, where U is a set of continuous, concave, and homogeneous[7] utility functions, as a convex-concave min-max game with dependent strategy sets, a perspective which to our knowledge has not been taken before.

Fisher markets can by solved via the Eisenberg-Gale convex program [21]. Recently, Cole and Tao [11] derived a new dual formulation of the Eisenberg-Gale program, namely

$$\min_{\boldsymbol{p}\geq\boldsymbol{0}} \sum_{j\in[m]} p_j + \sum_{i\in[n]} b_i \log\left(\max_{\boldsymbol{x}_i\geq\boldsymbol{0}:\boldsymbol{x}_i\cdot\boldsymbol{p}\leq b_i} u_i(\boldsymbol{x}_i)\right) \quad . \tag{2}$$

Rearranging, we obtain the following convex-concave min-max game with dependent strategy sets:

$$\min_{\boldsymbol{p}\geq\boldsymbol{0}} \max_{\boldsymbol{X}\geq\boldsymbol{0}:\boldsymbol{X}\boldsymbol{p}\leq\boldsymbol{b}} \sum_{j\in[m]} p_j + \sum_{i\in[n]} b_i \log\left(u_i(\boldsymbol{x}_i)\right) \quad . \tag{3}$$

[7] A function $f : \mathbb{R}^m \to \mathbb{R}$ is said to be **homogeneous of degree** k if $\forall \boldsymbol{x} \in \mathbb{R}^m, \lambda > 0, f(\lambda\boldsymbol{x}) = \lambda^k f(\boldsymbol{x})$. Unless otherwise indicated, a homogeneous function is assumed to be homogeneous of degree 1.

This min-max game is played by a fictitious (Walrasian) auctioneer and a set of buyers, who effectively play as a team. The objective function in this game is then the sum of the auctioneer's welfare (i.e., the sum of the prices) and the Nash social welfare of buyers (i.e., the second summation). As the buyer's strategy set is dependent on the price vector $\boldsymbol{p}$ selected by the auctioneer, we cannot use existing first-order methods to solve this problem. However, we can use Algorithms 1 and 2.

Starting from Equation (3), define the auctioneer's value function $V(\boldsymbol{p}) = \max_{\boldsymbol{X} \geq \boldsymbol{0}: \boldsymbol{X}\boldsymbol{p} \leq \boldsymbol{b}} \sum_{j \in [m]} p_j + \sum_{i \in [n]} b_i \log(u_i(\boldsymbol{x}_i))$, and the buyer i's demand set $X_i^*(\boldsymbol{p}, \boldsymbol{b}) = \arg\max_{\boldsymbol{x}_i \geq \boldsymbol{0}: \boldsymbol{x}_i \boldsymbol{p} \leq \boldsymbol{b}} u_i(\boldsymbol{x}_i)$. Theorem 3.2 then provides the relevant gradients so that we can run Algorithm 1 and Algorithm 2, namely $\partial_{\boldsymbol{p}} V(\boldsymbol{p}) = \boldsymbol{1}_m - \sum_{i \in [n]} X_i^*(\boldsymbol{p}, \boldsymbol{b})$ and $\nabla_{\boldsymbol{x}_i} \left(\sum_{j \in [m]} p_j + \sum_{i \in [n]} b_i \log(u_i(\boldsymbol{x}_i)) \right) = \frac{b_i}{u_i(\boldsymbol{x}_i)} \nabla_{\boldsymbol{x}_i} u_i(\boldsymbol{x}_i)$, using the Minkowski sum to add set-valued quantities.[8]

Cheung, Cole, and Devanur [10] observed that solving the dual of the Eisenberg-Gale program (Equation 2) via (sub)gradient descent [18] is equivalent to solving for a competitive equilibrium in a Fisher market using an auction-like economic price adjustment process named tâtonnement that was first proposed by Léon Walras in the 1870s [67]. The tâtonnement process increases the prices of goods that are overdemanded and decreases the prices of goods that are underdemanded. Mathematically, the (vanilla) tâtonnement process [4, 67] is defined as $\boldsymbol{p}(t) = \max\left\{ \boldsymbol{p}(t-1) + \eta_t \left(\sum_{i \in [n]} \boldsymbol{x}_i^*(\boldsymbol{p}(t), \boldsymbol{b}) - 1 \right), 0 \right\}$ for $\boldsymbol{p}(0) \in \mathbb{R}_+^m$, where $\boldsymbol{x}_i^*(\boldsymbol{p}(t), b_i) \in X_i^*(\boldsymbol{p}(t), b_i) = \arg\max_{\boldsymbol{x}_i \geq \boldsymbol{0}: \boldsymbol{x}_i \boldsymbol{p}(t) \leq b_i} u_i(\boldsymbol{x}_i)$ is the demand set of buyer i. The max-oracle algorithm applied to Equation (3) is then equivalent to a tâtonnemement process where the buyers report a δ-utility maximizing demand. Further, we have the following corollary of Theorem 3.3.

Corollary 4.1. *Let $(U, \boldsymbol{b})$ be a Fisher market with equilibrium price vector $\boldsymbol{p}^*$, where U is a set of continuous, concave, homogeneous, and continuously differentiable utility functions. Consider the vanilla tâtonnement process. Assume that the step sizes η_t satisfy the usual conditions: $\sum_{k=1}^T \eta_k^2 \leq \infty$ and $\sum_{k=1}^T \eta_k = \infty$. If $\boldsymbol{p}_{\text{best}}^{(t)} \in \arg\min_{\boldsymbol{p}^{(k)}: k \in [t]} V(\boldsymbol{p}^{(k)})$, then $\lim_{k \to \infty} V(\boldsymbol{p}_{\text{best}}^{(k)}) = V(\boldsymbol{p}^*)$. Additionally, tâtonnement converges to an ε-competitive equilibrium in $O(\varepsilon^{-2})$ iterations.*

If we also apply the nested gradient-descent-ascent algorithm to Equation (3), we arrive at an algorithm that is arguably more descriptive of market dynamics than tâtonnement itself, as it also includes the demand-side market dynamics of buyers optimizing their demands, potentially in a decentralized manner. The nested tâtonnement algorithm essentially describes a two-step trial-and-error (i.e., tâtonnement) process, where first the buyers try to discover their optimal demand by increasing their demand for goods in proportion to the marginal utility the goods provide, and then the seller/auctioneer adjusts the market clearing prices by decreasing the prices of goods that are underdemanded and increasing the prices of goods that are overdemanded. As buyers can calculate their demands in a decentralized fashion, the nested tâtonnement algorithm offers a more complete picture of market dynamics then the classic tâtonnement process.

Experiments In order to better understand, the iteration complexity of Algorithms 1 and 2, we ran a series of experiments on Fisher markets with three different classes of utility functions.[9] Each utility structure endows Equation (3) with different smoothness properties, which allows us to compare the efficiency of the algorithms under varying conditions.

Let $\boldsymbol{v}_i \in \mathbb{R}^m$ be a vector of parameters for the utility function of buyer $i \in [n]$. We have the following utility function classes: Linear: $u_i(\boldsymbol{x}_i) = \sum_{j \in [m]} v_{ij} x_{ij}$, Cobb-Douglas: $u_i(\boldsymbol{x}_i) = \prod_{j \in [m]} x_{ij}^{v_{ij}}$, Leontief: $u_i(\boldsymbol{x}_i) = \min_{j \in [m]} \left\{ \frac{x_{ij}}{v_{ij}} \right\}$. Equation (3) satisfies the smoothness properties listed in Table 3 when U is one of these three classes.

Our goals are two-fold. First, we want to understand how the empirical convergence rates of Algorithms 1 and 2 compare to their theoretical guarantees under different utility structures. Second, we want to understand the extent to which the convergence rates of these two algorithms differ in practice. We include a more detailed description of our experimental setup in Appendix E.

[8] We include detailed descriptions of the algorithms applied to Fisher markets in Appendix E.

[9] Our code can be found at https://anonymous.4open.science/r/min-max-fisher-CEFA/

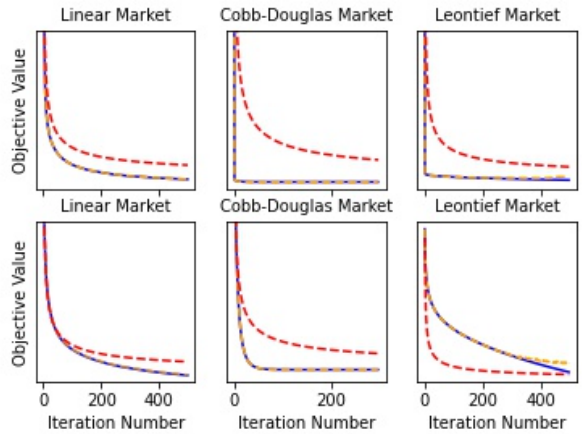

Figure 1: The first row describes the average trajectory of the value of the objective function for a randomly initialized market on each iteration of both Algorithm 3 (in blue) and Algorithm 4 (in orange) when the starting prices are high, while the second row describes the average trajectory of the objective function when starting prices are low for linear, Cobb-Douglas, and Leontief Fisher markets respectively. The dashed red line represents a convergence rate of $O(1/\sqrt{T})$ which corresponds to an iteration complexity of $O(1/\epsilon^2)$.

Table 3: Smoothness properties satisfied by Equation (3) assuming different utility functions. Note that Assumption 3.1 does not hold for Leontief utilities, because they are not differentiable.

	V is differentiable	Assumption 3.1 holds
Linear	$\times$	✓
Cobb-Douglas	✓	✓
Leontief	✓	$\times$

Figure 1 describes the empirical convergence rates of Algorithms 1 and 2 for linear, Cobb-Douglas, and Leontief utilities. We observe that convergence is fastest in Fisher markets with Cobb-Douglas utilities, followed by linear, and then Leontief. We seem to obtain a tight convergence rate of $O(1/\sqrt{T})$ for linear utilities, which seems plausible, as the value function is not differentiable assuming linear utilities, and hence we are unlikely to achieve a better convergence rate. On the other hand, for Cobb-Douglas utilities, both the value and the objective function are differentiable; in fact, they are both twice continuously differentiable, making them both Lipschitz-smooth. These factors combined seem to provide a much faster convergence rate than $O(1/\sqrt{T})$.

Fisher markets with Leontief utilities, in which the objective function is not differentiable, are the hardest markets of the three for our algorithms to solve. Indeed, our theory does not even predict convergence. Still, convergence is not entirely surprising, as Cheung, Cole, and Devanur [10] have shown that buyer demand throughout tâtonnement is bounded for Leontief utilities, which means that the objective function of Equation (3) is locally Lipschitz: i.e., any subgradient computed by the algorithm will be bounded. Overall, our theory suggests that differentiability of the value function is not essential to guarantee convergence of first-order methods in convex-concave games, while our experiments seem to suggest that differentiability of the objective function is more important than differentiability of the value function in regards to the convergence rate.

In order to investigate whether the outputs of Algorithm 1, which uses an exact max-oracle (i.e., 0-max-oracle) are more precise than those of Algorithm 2, we solved 500 randomly initialized markets with both algorithms. We then ran a James' first-order test [2] on the mean output of both algorithms to see if the difference in the average prices were statistically significant [30]. Our calculations produced p-values of 0.69, 0, and 1.06×10^{-18}, for Fisher markets with linear, Cobb-Douglas, and Leontief utilities, respectively. At a significance level of 0.05, these results are *not* statistically significant for linear utilities only. This result can be attributed to the fact that the value function is not differentiable in the linear case, which makes the nested gradient descent/ascent algorithm less precise.

5 Conclusion

In this paper, we proposed two first-order methods to solve a class of constrained convex-concave min-max optimization problems, which we call min-max games with dependent strategy sets. As such games do not afford Nash equilibria in general, we focused on the computation of their Stackelberg equilibria. To solve for Stackelberg equilibria, we introduced a novel subdifferential envelope theorem, which formed the heart of two subgradient methods with polynomial-time iteration complexity that converge to Stackelberg equilibria. Finally, we applied our theory to the computation of competitive equilibria in Fisher markets. This application yielded a new variant of the classic tâtonnement process, where, in addition to the auctioneer iteratively adjusting prices, the buyers iteratively compute their demands.

Our experiments suggest that the smoothness properties of both players' value functions determine the convergence speed to a Stackelberg equilibrium. Additionally, our experiments with Leontief Fisher markets suggest that our convergence results could perhaps be generalized to convex-concave objective functions which are *not* necessarily continuously differentiable; such a result would require a generalization of the subdifferential envelope theorem we have introduced.

Another worthwhile direction would be to try to derive stronger conditions under which the outer player's value function V is Lipschitz-smooth; under stronger assumptions we might be able to achieve faster convergence results. In order for V to be differentiable, the subdifferential given by Theorem 3.2 would have to be a singleton, for all $\boldsymbol{x} \in X$. This would require not only that the solution function $Y^*(\boldsymbol{x})$ be a singleton for all $\boldsymbol{x} \in \mathcal{X}$, but that the optimal Lagrangian multipliers $\Lambda(\widehat{\boldsymbol{x}}, \boldsymbol{y}^*(\widehat{\boldsymbol{x}})))$ were also unique. The former can be guaranteed when f is strictly concave in $\boldsymbol{y}$, while the latter condition is satisfied when the Linear Independence Constraint Qualification condition [29] holds. However, even when both these conditions hold, and when both the inner player's value function and their (parameterized) constraints (i.e., $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y})$) are Lipschitz-smooth, the outer player's value function is not guaranteed to be Lipschitz-smooth.

Gao and Kroer [23] have proposed first-order methods for the efficient computation of competitive equilibria in large Fisher markets. It would be worth exploring whether our view of competitive equilibria in Fisher markets as min-max optimization problems can facilitate the extension of their fast convergence results to a larger class of Fisher markets.

A question of interest both for market dynamics in Fisher markets and convex-concave min-max games with dependent strategy sets is whether two-time scale gradient-descent-ascent (GDA) converges in the dependent strategy set setting as it does in the independent setting [37]. GDA dynamics for Fisher markets correspond to myopic best-response dynamics, which are of general economic interest (see, for example, [43]).

Finally, our results at present concern only *convex-concave* min-max games with dependent strategy sets. It would be of interest to extend these results to the non-convex-concave setting; among other reasons, doing so could improve our understanding of competitive equilibria in Fisher markets with non-homogeneous utility functions. Recently, Dai and Zhang [14] defined the concept of a local minimax point, whose analog in games with dependent strategy sets would be a local Stackelberg equilibria. We believe that any algorithm designed for a non-convex-concave setting will have to aim for a local solution concept, as it is unlikely that (global) Stackelberg equilibria are computable in polynomial time in non-convex-concave settings, as non-convex optimization is hard.

Broader Impact

Our work can be used to expand the scope of many machine learning techniques as discussed in our introduction. Indeed, min-max optimization has been key to the development of fairer classifiers in recent years. On the other hand, our methods could also be used to train GANs, which have been used to create deepfakes with malicious goals. Our work can also be used to improve economic outcomes for companies running online marketplaces, as Fisher markets have been increasingly applied to resource allocation problems—specifically, fair division problems—in recent years. Moreover, with sufficient oversight, progress can be easily monitored, by following the trend of the the objective function across iterations, so that inappropriate use can be detected.

References

[1] S. N. Afriat. "Theory of Maxima and the Method of Lagrange". In: *SIAM Journal on Applied Mathematics* 20.3 (1971), pp. 343–357. ISSN: 00361399. URL: http://www.jstor.org/stable/2099955.

[2] James Algina, T. C. Oshima, and Wen-Ying Lin. "Type I Error Rates for Welch's Test and James's Second-Order Test under Nonnormality and Inequality of Variance When There are Two Groups". In: *Journal of Educational and Behavioral Statistics* 19.3 (1994), pp. 275–291. ISSN: 10769986, 19351054. URL: http://www.jstor.org/stable/1165297.

[3] Mohammad Alkousa et al. *Accelerated methods for composite non-bilinear saddle point problem*. 2020. arXiv: 1906.03620 [math.OC].

[4] Kenneth J. Arrow and Leonid Hurwicz. "On the Stability of the Competitive Equilibrium, I". In: *Econometrica* 26.4 (1958), pp. 522–552. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/1907515.

[5] Claude Berge. *Topological Spaces: including a treatment of multi-valued functions, vector spaces, and convexity*. Courier Corporation, 1997.

[6] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[7] Stephen Boyd and Lieven Vandenberghe. *Subgradients*. Apr. 2018.

[8] William C Brainard, Herbert E Scarf, et al. *How to compute equilibrium prices in 1891*. Citeseer, 2000.

[9] Qi Cai et al. *On the Global Convergence of Imitation Learning: A Case for Linear Quadratic Regulator*. 2019. arXiv: 1901.03674 [cs.LG].

[10] Yun Kuen Cheung, Richard Cole, and Nikhil Devanur. "Tatonnement beyond Gross Substitutes? Gradient Descent to the Rescue". In: *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*. STOC '13. Palo Alto, California, USA: Association for Computing Machinery, 2013, pp. 191–200. ISBN: 9781450320290. DOI: 10.1145/2488608.2488633. URL: https://doi.org/10.1145/2488608.2488633.

[11] Richard Cole and Yixin Tao. *Balancing the Robustness and Convergence of Tatonnement*. 2019. arXiv: 1908.00844 [cs.GT].

[12] Bo Dai et al. "Kernel Exponential Family Estimation via Doubly Dual Embedding". In: *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*. Ed. by Kamalika Chaudhuri and Masashi Sugiyama. Vol. 89. Proceedings of Machine Learning Research. PMLR, 16–18 Apr 2019, pp. 2321–2330. URL: http://proceedings.mlr.press/v89/dai19a.html.

[13] Bo Dai et al. "SBEED: Convergent Reinforcement Learning with Nonlinear Function Approximation". In: *Proceedings of the 35th International Conference on Machine Learning*. Ed. by Jennifer Dy and Andreas Krause. Vol. 80. Proceedings of Machine Learning Research. PMLR, Oct. 2018, pp. 1125–1134. URL: http://proceedings.mlr.press/v80/dai18c.html.

[14] Yu-Hong Dai and Liwei Zhang. *Optimality Conditions for Constrained Minimax Optimization*. 2020. arXiv: 2004.09730 [math.OC].

[15] John M. Danskin. "The Theory of Max-Min, with Applications". In: *SIAM Journal on Applied Mathematics* 14.4 (1966), pp. 641–664. ISSN: 00361399. URL: http://www.jstor.org/stable/2946123.

[16] Constantinos Daskalakis, Stratis Skoulakis, and Manolis Zampetakis. *The Complexity of Constrained Min-Max Optimization*. 2020. arXiv: 2009.09623 [cs.CC].

[17] Trevor Davis, Kevin Waugh, and Michael Bowling. *Solving Large Extensive-Form Games with Strategy Constraints*. 2019. arXiv: 1809.07893 [cs.GT].

[18] N. R. Devanur et al. "Market equilibrium via a primal-dual-type algorithm". In: *The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings*. 2002, pp. 389–395. DOI: 10.1109/SFCS.2002.1181963.

[19] Steven Diamond and Stephen Boyd. "CVXPY: A Python-embedded modeling language for convex optimization". In: *Journal of Machine Learning Research* 17.83 (2016), pp. 1–5.

[20] Harrison Edwards and Amos Storkey. *Censoring Representations with an Adversary*. 2016. arXiv: 1511.05897 [cs.LG].

[21] Edmund Eisenberg and David Gale. "Consensus of subjective probabilities: The pari-mutuel method". In: *The Annals of Mathematical Statistics* 30.1 (1959), pp. 165–168.

[22] Gabriele Farina, Christian Kroer, and Tuomas Sandholm. *Regret Circuits: Composability of Regret Minimizers*. 2019. arXiv: 1811.02540 [cs.LG].

[23] Yuan Gao and Christian Kroer. "First-Order Methods for Large-Scale Market Equilibrium Computation". In: *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*. Ed. by Hugo Larochelle et al. 2020. URL: https://proceedings.neurips.cc/paper/2020/hash/f75526659f31040afeb61cb7133e4e6d-Abstract.html.

[24] Gauthier Gidel et al. *A Variational Inequality Perspective on Generative Adversarial Networks*. 2020. arXiv: 1802.10551 [cs.LG].

[25] Denizalp Goktas. *Introduction to Markets for Computer Scientists*. Apr. 2021.

[26] E. Yazdandoost Hamedani et al. *Iteration Complexity of Randomized Primal-Dual Methods for Convex-Concave Saddle Point Problems*. 2018. arXiv: 1806.04118 [math.OC].

[27] Erfan Yazdandoost Hamedani and Necdet Serhat Aybat. "A primal-dual algorithm for general convex-concave saddle point problems". In: *arXiv preprint arXiv:1803.01401* 2 (2018).

[28] Charles R. Harris et al. "Array programming with NumPy". In: *Nature* 585 (2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2.

[29] R Henrion. "On constraint qualifications". In: *Journal of optimization theory and applications* 72.1 (1992), pp. 187–197.

[30] Freddy Hernandez et al. *stests: Package with useful statistical tests*. R package version 0.1.0. 2021. URL: https://fhernanb.github.io/stests.

[31] J. D. Hunter. "Matplotlib: A 2D graphics environment". In: *Computing in Science & Engineering* 9.3 (2007), pp. 90–95. DOI: 10.1109/MCSE.2007.55.

[32] Adam Ibrahim et al. "Lower bounds and conditioning of differentiable games". In: *arXiv preprint arXiv:1906.07300* (2019).

[33] Kamal Jain, Vijay V Vazirani, and Yinyu Ye. "Market equilibria for homothetic, quasi-concave utilities and economies of scale in production". In: *SODA*. Vol. 5. 2005, pp. 63–71.

[34] Chi Jin, Praneeth Netrapalli, and Michael I. Jordan. *What is Local Optimality in Nonconvex-Nonconcave Minimax Optimization?* 2020. arXiv: 1902.00618 [cs.LG].

[35] Anatoli Juditsky, Arkadi Nemirovski, et al. "First order methods for nonsmooth convex large-scale optimization, ii: utilizing problems structure". In: *Optimization for Machine Learning* 30.9 (2011), pp. 149–183.

[36] HW Kuhn and AW Tucker. *Proceedings of 2nd berkeley symposium*. 1951.

[37] Tianyi Lin, Chi Jin, and Michael Jordan. "On gradient descent ascent for nonconvex-concave minimax problems". In: *International Conference on Machine Learning*. PMLR. 2020, pp. 6083–6093.

[38] Tianyi Lin, Chi Jin, and Michael I Jordan. "Near-optimal algorithms for minimax optimization". In: *Conference on Learning Theory*. PMLR. 2020, pp. 2738–2779.

[39] Songtao Lu, Ioannis Tsaknakis, and Mingyi Hong. "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. 2019, pp. 4754–4758.

[40] David Madras et al. *Learning Adversarially Fair and Transferable Representations*. 2018. arXiv: 1802.06309 [cs.LG].

[41] Paul Milgrom and Ilya Segal. "Envelope theorems for arbitrary choice sets". In: *Econometrica* 70.2 (2002), pp. 583–601.

[42] Aryan Mokhtari, Asuman Ozdaglar, and Sarath Pattathil. *Convergence Rate of $\mathcal{O}(1/k)$ for Optimistic Gradient and Extra-gradient Methods in Smooth Convex-Concave Saddle Point Problems*. 2020. arXiv: 1906.01115 [math.OC].

[43] Dov Monderer and Lloyd S Shapley. "Potential games". In: *Games and economic behavior* 14.1 (1996), pp. 124–143.

[44] 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". In: *SIAM Journal on Optimization* 15.1 (2004), pp. 229–251.

[45] Yurii Nesterov. "Dual extrapolation and its applications to solving variational inequalities and related problems". In: *Mathematical Programming* 109.2 (2007), pp. 319–344.

[46] J v Neumann. "Zur theorie der gesellschaftsspiele". In: *Mathematische annalen* 100.1 (1928), pp. 295–320.

[47] Noam Nissan and Tim Roughgarden. *Algorithmic Game Theory*. Cambridge University Press, 2007. DOI: 10.1017/CBO9780511800481.

[48] Maher Nouiehed et al. "Solving a class of non-convex min-max games using iterative first order methods". In: *arXiv preprint arXiv:1902.08297* (2019).

[49] Dmitrii M Ostrovskii, Andrew Lowy, and Meisam Razaviyayn. "Efficient search of first-order nash equilibria in nonconvex-concave smooth min-max problems". In: *arXiv preprint arXiv:2002.07919* (2020).

[50] Yuyuan Ouyang and Yangyang Xu. *Lower complexity bounds of first-order methods for convex-concave bilinear saddle-point problems*. 2018. arXiv: 1808.02901 [math.OC].

[51] Murray H Protter, B Charles Jr, et al. *A first course in real analysis*. Springer Science & Business Media, 2012.

[52] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2013. URL: http://www.R-project.org/.

[53] Hassan Rafique et al. *Non-Convex Min-Max Optimization: Provable Algorithms and Applications in Machine Learning*. 2019. arXiv: 1810.02060 [math.OC].

[54] Maziar Sanjabi et al. *On the Convergence and Robustness of Training GANs with Regularized Optimal Transport*. 2018. arXiv: 1802.08249 [cs.LG].

[55] Maziar Sanjabi et al. "On the Convergence and Robustness of Training GANs with Regularized Optimal Transport". In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. NIPS'18. Montréal, Canada: Curran Associates Inc., 2018, pp. 7091–7101.

[56] Prasanna Sattigeri et al. *Fairness GAN*. 2018. arXiv: 1805.09910 [stat.ML].

[57] Gilberto A.S. Segundo, Renato A. Krohling, and Rodrigo C. Cosme. "A differential evolution approach for solving constrained min–max optimization problems". In: *Expert Systems with Applications* 39.18 (2012), pp. 13440–13450. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2012.05.059. URL: https://www.sciencedirect.com/science/article/pii/S0957417412007750.

[58] K. Shimizu and E. Aiyoshi. "A new computational method for Stackelberg and min-max problems by use of a penalty method". In: *IEEE Transactions on Automatic Control* 26.2 (1981), pp. 460–466. DOI: 10.1109/TAC.1981.1102607.

[59] K. Shimizu and E. Aiyoshi. "Necessary conditions for min-max problems and algorithms by a relaxation procedure". In: *IEEE Transactions on Automatic Control* 25.1 (1980), pp. 62–66. DOI: 10.1109/TAC.1980.1102226.

[60] Aman Sinha et al. *Certifying Some Distributional Robustness with Principled Adversarial Training*. 2020. arXiv: 1710.10571 [stat.ML].

[61] The pandas development team. *pandas-dev/pandas: Pandas*. Version latest. Feb. 2020. DOI: 10.5281/zenodo.3509134. URL: https://doi.org/10.5281/zenodo.3509134.

[62] Kiran Koshy Thekumparampil et al. *Efficient Algorithms for Smooth Minimax Optimization*. 2019. arXiv: 1907.01543 [math.OC].

[63] Paul Tseng. "On accelerated proximal gradient methods for convex-concave optimization". In: *submitted to SIAM Journal on Optimization* 1 (2008).

[64] Paul Tseng. "On linear convergence of iterative methods for the variational inequality problem". In: *Journal of Computational and Applied Mathematics* 60.1 (1995). Proceedings of the International Meeting on Linear/Nonlinear Iterative Methods and Verification of Solution, pp. 237–252. ISSN: 0377-0427. DOI: https://doi.org/10.1016/0377-0427(94)00094-H. URL: https://www.sciencedirect.com/science/article/pii/037704279400094H.

[65] Guido Van Rossum and Fred L Drake Jr. *Python tutorial*. Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands, 1995.

[66] Heinrich Von Stackelberg. *Marktform und gleichgewicht*. J. springer, 1934.

[67] Léon Walras. *Elements of Pure Economics; or, The Theory of Social Wealth. Translated by William Jaffé*. Vol. 2. Orion Editions, 1969, pp. 457–458.

[68] Hadley Wickham et al. "Welcome to the tidyverse". In: *Journal of Open Source Software* 4.43 (2019), p. 1686. DOI: 10.21105/joss.01686.

[69] Depeng Xu et al. *FairGAN: Fairness-aware Generative Adversarial Networks*. 2018. arXiv: 1805.11202 [cs.LG].

[70] Laura Scrimali Yurii Nesterov. "Solving strongly monotone variational and quasi-variational inequalities". In: *Discrete & Continuous Dynamical Systems* 31.4 (2011), pp. 1383–1396.

[71] Junyu Zhang, Mingyi Hong, and Shuzhong Zhang. *On Lower Iteration Complexity Bounds for the Saddle Point Problems*. 2020. arXiv: 1912.07481 [math.OC].

[72] Renbo Zhao. *A Primal Dual Smoothing Framework for Max-Structured Nonconvex Optimization*. 2020. arXiv: 2003.04375 [math.OC].

[73] Renbo Zhao. "Optimal algorithms for stochastic three-composite convex-concave saddle point problems". In: *arXiv preprint arXiv:1903.01687* (2019).

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] Discussion in conclusion
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] Broader impacts section included.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes] Main assumption is given in assumption 3.1 and other auxiliary assumptions are given in theorem statements
   (b) Did you include complete proofs of all theoretical results? [Yes] Yes, all results' proofs can be found in the appendix
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] Information for replication are found on appendix E. Instructions to run our code can be found on the readme.md of `https://anonymous.4open.science/r/min-max-fisher-CEFA/`
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] Details can be found in appendix E
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] Details can be found in appendix E
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] Can be found in appendix E
   (b) Did you mention the license of the assets? [Yes] Can be found in appendix E
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] Code and assets can be found on `https://anonymous.4open.science/r/min-max-fisher-CEFA/`
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]