

A Pseudo-Games and Generalized Nash Equilibria

Pseudo-games, or **abstract economies** [4], are optimization problems that are closely related to min-max Stackelberg games, but which are technically not games, as noted by Facchinei and Kanzow [26, 27], because each player’s strategy set is not fixed at the outset (i.e., before they have to make a decision), but instead depends on the other players’ choices. [39]. In this appendix, we formally define two-player, zero-sum pseudo-games¹⁰ and discuss how they differ from min-max Stackelberg games. We also define the equilibrium concept *par excellence* of pseudo-games, namely **generalized Nash equilibrium**, and juxtapose its definition with vanilla Nash equilibrium.

A two-player, zero-sum **pseudo-game** comprises two players, with respective payoff functions $-f(\mathbf{x}, \mathbf{y})$ and $f(\mathbf{x}, \mathbf{y})$, and respective strategy spaces given by the correspondences $\mathcal{X} : Y \rightrightarrows X$ and $\mathcal{Y} : X \rightrightarrows Y$, i.e., set valued mappings that depend on the choice the other player takes. More specifically, $\mathcal{X}(\mathbf{y}) = \{\mathbf{x} \in X \mid \forall k \in [K], g_k(\mathbf{x}, \mathbf{y}) \geq 0\}$ and $\mathcal{Y}(\mathbf{x}) = \{\mathbf{y} \in Y \mid \forall k \in [K], g_k(\mathbf{x}, \mathbf{y}) \geq 0\}$. We denote a pseudo-game by (X, Y, f, \mathbf{g}) .

Pseudo-games are closely related to min-max Stackelberg games, as they both comprise agents with the same objectives and the same space of feasible strategy profiles, namely $\{(\mathbf{x}, \mathbf{y}) \in X \times Y \mid \forall k \in [K], g_k(\mathbf{x}, \mathbf{y}) \geq 0\}$. However, whereas play proceeds sequentially in Stackelberg games, so that inner player chooses a strategy $\mathbf{y} \in Y$ s.t. $\mathbf{g}(\hat{\mathbf{x}}, \mathbf{y}) \geq 0$ *after* the outer player chooses $\hat{\mathbf{x}} \in X$, in pseudo-games players choose their strategies simultaneously. In other words, the space of feasible strategy profiles is an emergent feature of a pseudo-game; it depends on the players’ choices, which are made simultaneously, yet must somehow still be feasible, even though one player’s choice could render another player’s choice infeasible.

A **generalized Nash equilibrium (GNE)** of pseudo-game (X, Y, f, \mathbf{g}) is a strategy profile $(\mathbf{x}^*, \mathbf{y}^*) \in X \times Y$ s.t. $\mathbf{g}(\mathbf{x}^*, \mathbf{y}^*) \geq 0$ such that

$$\max_{\mathbf{y} \in Y: \mathbf{g}(\mathbf{x}^*, \mathbf{y}) \geq 0} f(\mathbf{x}^*, \mathbf{y}) \leq f(\mathbf{x}^*, \mathbf{y}^*) \leq \min_{\mathbf{x} \in X: \mathbf{g}(\mathbf{x}, \mathbf{y}^*) \geq 0} f(\mathbf{x}, \mathbf{y}^*) . \quad (4)$$

That is, at a GNE the players choose best responses to the other players’ strategies from within the space of strategies defined by the other players’ choices. Under Assumption 3.1, a GNE is guaranteed to exist for any pseudo-game (X, Y, f, \mathbf{g}) . A Nash equilibrium, however, is not guaranteed to exist.

A **Nash equilibrium (NE)** of a min-max Stackelberg game (X, Y, f, \mathbf{g}) is a tuple $(\mathbf{x}^*, \mathbf{y}^*) \in X \times Y$ such that

$$\max_{\mathbf{y} \in Y} \min_{\mathbf{x} \in X: \mathbf{g}(\mathbf{x}, \mathbf{y}) \geq 0} f(\mathbf{x}, \mathbf{y}) \leq f(\mathbf{x}^*, \mathbf{y}^*) \leq \min_{\mathbf{x} \in X} \max_{\mathbf{y} \in Y: \mathbf{g}(\mathbf{x}, \mathbf{y}) \geq 0} f(\mathbf{x}, \mathbf{y}) . \quad (5)$$

That is, at a NE the players best respond to one another’s choices, under the assumption that the other player is doing the same: i.e., best responding to their choice.

We summarize what is known about min-max Stackelberg games vs. pseudo-games, and Stackelberg vs. generalized Nash (and vanilla Nash) equilibria, as follows. Min-max Stackelberg games are proper games, while pseudo-games are not well-defined games. Stackelberg min-max games are played sequentially, while pseudo-games are played simultaneously. When Nash equilibria exist in min-max Stackelberg games, they coincide with Stackelberg equilibria. Additionally, Nash equilibria coincide with generalized Nash equilibria in min-max games with independent strategy sets. There exist no polynomial-time computation guarantees for generalized Nash equilibria in pseudo-games, even in the two-player, zero-sum setting, while we provide polynomial-time computation guarantees for Stackelberg equilibria in min-max Stackelberg games.

B Omitted Proofs Section 2

Proposition B.1. *Consider a min-max Stackelberg game (X, Y, f, \mathbf{g}) . Then if Assumption 3.1 holds, the outer player’s value function, $V(\mathbf{x}) = \max_{\mathbf{y} \in Y: \mathbf{g}(\mathbf{x}, \mathbf{y}) \geq 0} f(\mathbf{x}, \mathbf{y})$, is continuous and convex.*

Proof. By Berge’s maximum theorem [7], the outer player’s value function is continuous.

¹⁰We refer the reader to Facchinei and Kanzow’s [26, 27] survey on pseudo-games for a more detailed exposition, beyond two-player, zero-sum pseudo-games.

Define $\mathcal{L} : Y \times \mathbb{R}_+^K \times X \rightarrow \mathbb{R}$ s.t. $\mathcal{L}(\mathbf{y}, \boldsymbol{\lambda}; \mathbf{x}) = f(\mathbf{x}, \mathbf{y}) + \sum_{k=1}^K \lambda_k g_k(\mathbf{x}, \mathbf{y})$ and $h : Y \times X \rightarrow \mathbb{R}$ s.t. $h(\mathbf{y}; \mathbf{x}) = \min_{\boldsymbol{\lambda} \in \mathbb{R}_+^K} \mathcal{L}(\mathbf{y}, \boldsymbol{\lambda}; \mathbf{x})$. Since Slater's condition is satisfied under Assumption 3.1, the KKT theorem [44] applies, which means that for all $\mathbf{x} \in X$ and $\mathbf{y} \in Y$, the optimal KKT multipliers $\boldsymbol{\lambda}^*$ exist and are well-defined: i.e., there exists $c \in \mathbb{R}_+$ such that $\forall k \in [K], 0 \leq \lambda_k^* \leq c < \infty$. We can thus re-express the outer player's value function $V(\mathbf{x})$ as the following Lagrangian saddle-point problems:

$$V(\mathbf{x}) = \max_{\mathbf{y} \in Y: g(\mathbf{x}, \mathbf{y}) \geq \mathbf{0}} f(\mathbf{x}, \mathbf{y}) = \max_{\mathbf{y} \in Y} \min_{\boldsymbol{\lambda} \in \mathbb{R}_+^K} \mathcal{L}(\mathbf{y}, \boldsymbol{\lambda}; \mathbf{x}) = \max_{\mathbf{y} \in Y} \min_{\boldsymbol{\lambda} \in [0, c]^K} \mathcal{L}(\mathbf{y}, \boldsymbol{\lambda}; \mathbf{x}) = \max_{\mathbf{y} \in Y} h(\mathbf{y}; \mathbf{x}). \quad (6)$$

Notice that \mathcal{L} is concave in \mathbf{y} , convex in $\boldsymbol{\lambda}$, and convex in \mathbf{x} . Since h is the min-projection of \mathcal{L} w.r.t. $\boldsymbol{\lambda}$ onto the compact set $[0, c]^K$, h must be convex in \mathbf{x} ([65], Proposition 2.22). Further, since V is the maximum of convex functions over a compact set, i.e., $\max_{\mathbf{y} \in Y} h(\mathbf{y}; \mathbf{x})$, V must be convex ([65], Theorem 5.5). \square

Proposition B.2. *Given any min-max Stackelberg game (X, Y, f, g) , a Stackelberg equilibrium always exists and its value is unique.*

Proof of Proposition B.2 By Berge's maximum theorem [7], the outer player's value function $V(\mathbf{x}) = \max_{\mathbf{y} \in Y: g(\mathbf{x}, \mathbf{y}) \geq \mathbf{0}} f(\mathbf{x}, \mathbf{y})$ is continuous, and the inner player's solution correspondence $Y^*(\mathbf{x}) = \arg \max_{\mathbf{y} \in Y: g(\mathbf{x}, \mathbf{y}) \geq \mathbf{0}} f(\mathbf{x}, \mathbf{y})$ is non-empty, for all $\mathbf{x} \in X$. Since V is continuous and X is compact and non-empty, by the extreme value theorem [62], there exists a minimizer $\mathbf{x}^* \in X$ of V . Hence $(\mathbf{x}^*, \mathbf{y}^*(\mathbf{x}^*))$, with $\mathbf{y}^*(\mathbf{x}^*) \in Y^*(\mathbf{x}^*)$, is a Stackelberg equilibrium of (X, Y, f, g) .

Let $(\mathbf{x}_1, \mathbf{y}_1)$ and $(\mathbf{x}_2, \mathbf{y}_2)$ be two different Stackelberg equilibria. WLOG, suppose $f(\mathbf{x}_1, \mathbf{y}_1) > f(\mathbf{x}_2, \mathbf{y}_2)$, so that $V(\mathbf{x}_1) = f(\mathbf{x}_1, \mathbf{y}_1) > f(\mathbf{x}_2, \mathbf{y}_2) = V(\mathbf{x}_2)$, where the first and last equality follow from the definition of Stackelberg equilibrium. But then $(\mathbf{x}_1, \mathbf{y}_1)$ cannot be a Stackelberg equilibrium, since \mathbf{x}_1 is not a minimizer of the outer player's value function. \square

Proposition B.3. *The set of Stackelberg equilibria of any min-max Stackelberg game (X, Y, f, g) is non-empty, compact, and convex.*

Proof. By Proposition B.2, the set of Stackelberg equilibria of any min-max Stackelberg game is non-empty. Additionally, under Assumption 3.1, we have that for all $\mathbf{x} \in X$, $f(\mathbf{x}, \cdot)$ is concave, and $\{\mathbf{y} \in Y \mid g(\mathbf{x}, \mathbf{y}) \geq \mathbf{0}\}$ is convex. Hence, by Theorem 2.6 of Rockafeller [65], the set of solutions $Y^*(\mathbf{x}) = \arg \max_{\mathbf{y} \in Y: g(\mathbf{x}, \mathbf{y}) \geq \mathbf{0}} f(\mathbf{x}, \mathbf{y})$ is compact- and convex- valued. Similarly, by Proposition B.1, under Assumption 3.1, $\max_{\mathbf{y} \in Y: g(\mathbf{x}, \mathbf{y}) \geq \mathbf{0}} f(\mathbf{x}, \mathbf{y})$ is continuous and convex. Hence, the set of solutions $X^* = \arg \min_{\mathbf{x} \in X} \max_{\mathbf{y} \in Y: g(\mathbf{x}, \mathbf{y}) \geq \mathbf{0}} f(\mathbf{x}, \mathbf{y})$ is compact- and convex-valued. Since the composition of two compact-convex-valued correspondences is again compact-convex-valued (Proposition 5.52 of Rockafeller [65]), we conclude that the set of Stackelberg equilibria, namely $X^*(Y^*)$, is compact and convex. \square

C Envelope Theorem

Danskin's theorem [17] offers insights into optimization problems of the form: $\max_{\mathbf{y} \in Y} f(\mathbf{x}, \mathbf{y})$, where $Y \subset \mathbb{R}^m$ is compact and non-empty. Among other things, Danskin's theorem allows us to compute the gradient of the objective function of this optimization problem with respect to \mathbf{x} .

Theorem C.1 (Danskin's Theorem). *Consider an optimization problem of the form: $\max_{\mathbf{y} \in Y} f(\mathbf{x}, \mathbf{y})$, where $Y \subset \mathbb{R}^m$ is compact and non-empty. Suppose that Y is convex and that f is concave in \mathbf{y} . Let $V(\mathbf{x}) = \max_{\mathbf{y} \in Y} f(\mathbf{x}, \mathbf{y})$ and $Y^*(\mathbf{x}) = \arg \max_{\mathbf{y} \in Y} f(\mathbf{x}, \mathbf{y})$. Then V is differentiable at $\hat{\mathbf{x}}$, if the solution correspondence $Y^*(\hat{\mathbf{x}})$ is a singleton: i.e., $Y^*(\hat{\mathbf{x}}) = \{\mathbf{y}^*(\hat{\mathbf{x}})\}$. Additionally, the gradient at $\hat{\mathbf{x}}$ is given by $V'(\hat{\mathbf{x}}) = \nabla_{\mathbf{x}} f(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}}))$.*

Unfortunately, Danskin's theorem does not hold when the set Y is replaced by a correspondence, which occurs in min-max Stackelberg games: i.e., when the inner problem is $\max_{\mathbf{y} \in Y: g(\mathbf{x}, \mathbf{y}) \geq \mathbf{0}} f(\mathbf{x}, \mathbf{y})$.

Example C.2 (Danskin's theorem does not apply to min-max Stackelberg games). Consider the optimization problem:

$$\max_{y \in \mathbb{R}: y+x \geq 0} -y^2 + y + 2x + 2. \quad (7)$$

The solution to this problem is unique, given any $x \in X$, meaning the solution correspondence $Y^*(x)$ is singleton-valued. We denote this unique solution by $y^*(x)$. After solving, we find that

$$y^*(x) = \begin{cases} 1/2 & \text{if } x \geq -1/2 \\ -x & \text{if } x < -1/2 \end{cases} \quad (8)$$

The value function $V(x) = \max_{y \in \mathbb{R}: y+x \geq 0} -y^2 + y + 2x + 2$ is then given by:

$$V(x) = f(x, y^*(x)) \quad (9)$$

$$= -y^*(x)^2 + y^*(x) + 2x + 2 \quad (10)$$

$$= \begin{cases} -1/4 + 1/2 + 2x + 2 & \text{if } x \geq -1/2 \\ -x^2 - x + 2x + 2 & \text{if } x < -1/2 \end{cases} \quad (11)$$

$$= \begin{cases} 9/4 + 2x & \text{if } x \geq -1/2 \\ -x^2 + x + 2 & \text{if } x < -1/2 \end{cases} \quad (12)$$

The derivative of this value function is:

$$\frac{\partial V}{\partial x} = \begin{cases} 2 & \text{if } x \geq -1/2 \\ 1 - 2x & \text{if } x < -1/2 \end{cases} \quad (13)$$

However, the derivative predicted by Danskin's theorem is 2. Hence, Danskin's theorem does not hold when the constraints are parameterized, i.e., when the problem is of the form $\min_{\mathbf{y} \in Y: \mathbf{g}(\mathbf{x}, \mathbf{y})} f(\mathbf{x}, \mathbf{y})$ rather than $\min_{\mathbf{y} \in Y} f(\mathbf{x}, \mathbf{y})$, where $X \subset \mathbb{R}^n$, $Y \subset \mathbb{R}^m$, and for all $k \in [K]$, $g_k : X \times Y \rightarrow \mathbb{R}$ are continuous.

N.B. For simplicity, we do not assume the constraint set is compact in this example. Compactness of the constraint set can be used to guarantee existence of a solution, but as a solution to this particular problem always exists, we can do away with this assumption.

The following theorem, due to Milgrom and Segal [50], generalizes Danskin's theorem to handle parameterized constraints:

Theorem C.3 (Envelope Theorem [50]). Consider the maximization problem

$$V(\mathbf{x}) = \max_{\mathbf{y} \in Y} f(\mathbf{x}, \mathbf{y}), \text{ subject to } g_k(\mathbf{x}, \mathbf{y}) \geq 0, \text{ for all } k = 1, \dots, K, \quad (14)$$

where $Y \subseteq \mathbb{R}^m$.

Define the solution correspondence $Y^*(\mathbf{x}) = \arg \max_{\mathbf{y} \in X: \mathbf{g}(\mathbf{x}, \mathbf{y}) \geq \mathbf{0}} f(\mathbf{x}, \mathbf{y})$. If Assumption 3.1 holds, then the value function V is absolutely continuous, and at any point $\hat{\mathbf{x}} \in X$ where V is differentiable:

$$\nabla_{\mathbf{x}} V(\hat{\mathbf{x}}) = \nabla_{\mathbf{x}} L(\mathbf{y}^*(\hat{\mathbf{x}}), \boldsymbol{\lambda}^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})), \hat{\mathbf{x}}) = \nabla_{\mathbf{x}} f(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) + \sum_{k=1}^K \lambda_k^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) \nabla_{\mathbf{x}} g_k(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})), \quad (15)$$

where $\boldsymbol{\lambda}^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) = (\lambda_1^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})), \dots, \lambda_K^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})))^T \in \Lambda^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}}))$ are the KKT multipliers associated with $\mathbf{y}^*(\hat{\mathbf{x}}) \in Y^*(\hat{\mathbf{x}})$.

D Omitted Subdifferential Envelope Theorem Proof (Section 3)

Proof of Theorem 3.2 As usual, let $V(\mathbf{x}) = \max_{\mathbf{y} \in Y: \mathbf{g}(\mathbf{x}, \mathbf{y}) \geq \mathbf{0}} f(\mathbf{x}, \mathbf{y})$. First, note that Proposition B.1 V is subdifferentiable as it is convex [9]. Reformulating the problem as a Lagrangian saddle point problem, for all $\hat{\mathbf{x}} \in X$, it holds that:

$$V(\hat{\mathbf{x}}) = \max_{\mathbf{y} \in Y: \mathbf{g}(\hat{\mathbf{x}}, \mathbf{y}) \geq \mathbf{0}} f(\hat{\mathbf{x}}, \mathbf{y}) \quad (16)$$

$$= \max_{\mathbf{y} \in Y} \min_{\boldsymbol{\lambda} \in \mathbb{R}_{++}^K} \left\{ f(\hat{\mathbf{x}}, \mathbf{y}) + \sum_{k=1}^K \lambda_k g_k(\hat{\mathbf{x}}, \mathbf{y}) \right\} \quad (17)$$

Since f is continuous, Y is compact, and g_1, \dots, g_K are continuous, for all $\hat{\mathbf{x}} \in X$, there exists $\mathbf{y}^*(\hat{\mathbf{x}}) \in \arg \max_{\mathbf{y} \in Y: g(\hat{\mathbf{x}}, \mathbf{y}) \geq 0} f(\hat{\mathbf{x}}, \mathbf{y})$. Furthermore, as Assumption [3.1](#) ensures that an interior solution exists, the Karush-Kuhn-Tucker Theorem [\[44\]](#) applies, so for all $\hat{\mathbf{x}} \in X$ and any associated $\mathbf{y}^*(\hat{\mathbf{x}})$, there exists $\boldsymbol{\lambda}(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) \in \mathbb{R}^K$ that solves Equation [\(17\)](#).

Define the solution correspondence $Y^*(\mathbf{x}) = \arg \max_{\mathbf{y} \in Y: g(\mathbf{x}, \mathbf{y}) \geq 0} f(\mathbf{x}, \mathbf{y})$, and let $\Lambda^*(\mathbf{x}, \mathbf{y}) = \arg \min_{\boldsymbol{\lambda} \in \mathbb{R}_+^K} \left\{ f(\mathbf{x}, \mathbf{y}) + \sum_{k=1}^K \lambda_k g_k(\mathbf{x}, \mathbf{y}) \right\}$. We can then re-express the value function at $\hat{\mathbf{x}}$ as:

$$V(\hat{\mathbf{x}}) = f(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) + \sum_{k=1}^K \lambda_k^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) g_k(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})), \quad \forall \mathbf{y}^*(\hat{\mathbf{x}}) \in Y^*(\hat{\mathbf{x}}), \lambda_k^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) \in \Lambda^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) .$$

Equivalently, we can take the maximum over \mathbf{y}^* 's and $\boldsymbol{\lambda}^*$'s to obtain:

$$V(\hat{\mathbf{x}}) = \max_{\mathbf{y}^*(\hat{\mathbf{x}}) \in Y^*(\hat{\mathbf{x}})} \max_{\lambda_k^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) \in \Lambda^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}}))} \left\{ f(\hat{\mathbf{x}}, \mathbf{y}) + \sum_{k=1}^K \lambda_k^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) g_k(\hat{\mathbf{x}}, \mathbf{y}) \right\} .$$

Note that for fixed $\mathbf{y}^*(\hat{\mathbf{x}}) \in Y^*(\hat{\mathbf{x}})$ and corresponding fixed $\lambda_k^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) \in \Lambda^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}}))$, $f(\hat{\mathbf{x}}, \mathbf{y}) + \sum_{k=1}^K \lambda_k^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) g_k(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}}))$ is differentiable, since f, g_1, \dots, g_K are differentiable.

Additionally, recall the pointwise maximum subdifferential property^{[\[1\]](#)} i.e., if $f(\mathbf{x}) = \max_{\alpha \in \mathcal{A}} f_\alpha(\mathbf{x})$ for a family of functions $\{f_\alpha\}_{\alpha \in \mathcal{A}}$, then $\partial_{\mathbf{x}} f(\mathbf{a}) = \text{conv} \left(\bigcup_{\alpha \in \mathcal{A}} \{ \partial_{\mathbf{x}} f_\alpha(\mathbf{a}) \mid f_\alpha(\mathbf{a}) = f(\mathbf{a}) \} \right)$, which then gives:

$$\partial_{\mathbf{x}} V(\hat{\mathbf{x}}) = \partial_{\mathbf{x}} \left(\max_{\mathbf{y}^*(\hat{\mathbf{x}}) \in Y^*(\hat{\mathbf{x}})} \max_{\lambda_k^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) \in \Lambda^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}}))} \left\{ f(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) + \sum_{k=1}^K \lambda_k^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) g_k(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) \right\} \right) \quad (18)$$

$$= \text{conv} \left(\bigcup_{\mathbf{y}^*(\hat{\mathbf{x}}) \in Y^*(\hat{\mathbf{x}})} \bigcup_{\lambda_k^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) \in \Lambda^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}}))} \partial_{\mathbf{x}} \left\{ f(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) + \sum_{k=1}^K \lambda_k^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) g_k(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) \right\} \right) \quad (19)$$

$$= \text{conv} \left(\bigcup_{\mathbf{y}^*(\hat{\mathbf{x}}) \in Y^*(\hat{\mathbf{x}})} \bigcup_{\lambda_k^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) \in \Lambda^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}}))} \left\{ \nabla_{\mathbf{x}} f(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) + \sum_{k=1}^K \lambda_k^*(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) \nabla_{\mathbf{x}} g_k(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}})) \right\} \right) . \quad (20)$$

□

E Pseudo-Codes and Convergence Results for Section [3](#)

Lemma E.1 (Lipschitz Objective Function implies Lipschitz Value Function). *Let $f : X \times Y$ be a continuous function, where $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$. If $\nabla_{\mathbf{x}} f$ is continuous in (\mathbf{x}, \mathbf{y}) , X, Y are compact and non-empty, and for all $k \in [K]$, g_k is continuous, then $V(\mathbf{x}) = \max_{\mathbf{y} \in Y: g(\mathbf{x}, \mathbf{y}) \geq 0} f(\mathbf{x}, \mathbf{y})$ is ℓ_f -Lipschitz continuous: i.e., $\|V(\mathbf{x}_1) - V(\mathbf{x}_2)\| \leq \ell_f \|\mathbf{x}_1 - \mathbf{x}_2\|$, with $\ell_f = \max_{(\hat{\mathbf{x}}, \hat{\mathbf{y}}) \in X \times Y} \|\nabla_{\mathbf{x}} f(\hat{\mathbf{x}}, \hat{\mathbf{y}})\|$.*

Proof of Lemma [E.1](#) Let $\ell_f = \max_{(\hat{\mathbf{x}}, \hat{\mathbf{y}}) \in X \times Y} \|\nabla_{\mathbf{x}} f(\hat{\mathbf{x}}, \hat{\mathbf{y}})\|$. Because $\nabla_{\mathbf{x}} f$ is continuous, f is ℓ_f -Lipschitz continuous in \mathbf{x} : i.e., $\forall \mathbf{x}_1, \mathbf{x}_2 \in X, \mathbf{y} \in Y, \|f(\mathbf{x}_1, \mathbf{y}) - f(\mathbf{x}_2, \mathbf{y})\| \leq \ell_f \|\mathbf{x}_1 - \mathbf{x}_2\|$.

Fix $\mathbf{x}_1, \mathbf{x}_2 \in X$. Then, for all $\mathbf{y} \in Y$, we have:

$$f(\mathbf{x}_1, \mathbf{y}) \leq f(\mathbf{x}_1, \mathbf{y}) - f(\mathbf{x}_2, \mathbf{y}) + f(\mathbf{x}_2, \mathbf{y}) \quad (21)$$

$$\leq \ell_f \|\mathbf{x}_1 - \mathbf{x}_2\| + f(\mathbf{x}_2, \mathbf{y}) . \quad (22)$$

¹¹See, for example, [\[9\]](#).

Algorithm 2 Nested Gradient Descent

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

```
1: for  $t = 1, \dots, T_x$  do
2:    $\mathbf{y}^{(t-1)} = \mathbf{y}^{(0)}$ 
3:   for  $s = 1, \dots, T_y$  do
4:      $\mathbf{y}^{(t-1)} = \Pi_{\{\mathbf{y} \in Y: g(\mathbf{x}^{(t-1)}, \mathbf{y}) \geq 0\}} (\mathbf{y}^{(t-1)} + \eta_s^y [\nabla_{\mathbf{y}} f(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t-1)})])$ 
5:   end for
6:   Set  $\boldsymbol{\lambda}^{(t-1)} = \boldsymbol{\lambda}^*(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t-1)})$ 
7:   Set  $\mathbf{x}^{(t)} = \Pi_X \left( \mathbf{x}^{(t-1)} - \eta_t^x \left[ \nabla_{\mathbf{x}} f(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t-1)}) + \sum_{k=1}^K \lambda_k^{(t-1)} \nabla_{\mathbf{x}} g_k(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t-1)}) \right] \right)$ 
8: end for
9:  $\mathbf{y}^{(T)} = \mathbf{y}^{(0)}$ 
10: for  $s = 1, \dots, T_y$  do
11:    $\mathbf{y}^{(T)} = \Pi_{\{\mathbf{y} \in Y: g(\mathbf{x}^{(T)}, \mathbf{y}) \geq 0\}} (\mathbf{y}^{(T)} + \eta_s^y \nabla_{\mathbf{y}} f(\mathbf{x}^{(T)}, \mathbf{y}^{(T)}))$ 
12: end for
13: return  $(\mathbf{x}^{(T)}, \mathbf{y}^{(T)})$ 
```

Taking the max over the \mathbf{y} 's on both sides (which is guaranteed to exist by the continuity of f , and of the g_k 's, and the fact that Y is non-empty and compact) yields:

$$\max_{\mathbf{y} \in Y: g(\mathbf{x}_1, \mathbf{y}) \geq 0} f(\mathbf{x}_1, \mathbf{y}) \leq \ell_f \|\mathbf{x}_1 - \mathbf{x}_2\| + \max_{\mathbf{y} \in Y: g(\mathbf{x}_2, \mathbf{y}) \geq 0} f(\mathbf{x}_2, \mathbf{y}) \quad (23)$$

$$V(\mathbf{x}_1) \leq \ell_f \|\mathbf{x}_1 - \mathbf{x}_2\| + V(\mathbf{x}_2) \quad (24)$$

$$V(\mathbf{x}_1) - V(\mathbf{x}_2) \leq \ell_f \|\mathbf{x}_1 - \mathbf{x}_2\| \quad . \quad (25)$$

Since this inequality holds for arbitrary $\mathbf{x}_1, \mathbf{x}_2 \in X$, we also have:

$$V(\mathbf{x}_2) - V(\mathbf{x}_1) \leq \ell_f \|\mathbf{x}_1 - \mathbf{x}_2\| \quad . \quad (26)$$

Combining these two inequalities, it holds that: $\|V(\mathbf{x}_1) - V(\mathbf{x}_2)\| \leq \ell_f \|\mathbf{x}_1 - \mathbf{x}_2\|$. \square

Proof of Theorem 3.3 By our subdifferential envelope theorem (Theorem 3.2), we have:

$$\begin{aligned} \nabla_{\mathbf{x}} f(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t-1)}) + \sum_{k=1}^K \lambda_k^{(t-1)} \nabla_{\mathbf{x}} g_k(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t-1)}) &\in \partial_{\mathbf{x}} V(\mathbf{x}^{(t-1)}) \quad (27) \\ &= \partial_{\mathbf{x}} \max_{\mathbf{y} \in Y: g(\mathbf{x}^{(t-1)}, \mathbf{y}) \geq 0} f(\mathbf{x}^{(t-1)}, \mathbf{y}) \quad . \quad (28) \end{aligned}$$

For notational clarity, let $\mathbf{h}(t-1) = \nabla_{\mathbf{x}} f(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t-1)}) + \sum_{k=1}^K \lambda_k^{(t-1)} \nabla_{\mathbf{x}} g_k(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t-1)})$. Suppose that $\mathbf{x}^* \in \arg \min_{\mathbf{x} \in X} \max_{\mathbf{y} \in Y: g(\mathbf{x}, \mathbf{y}) \geq 0} f(\mathbf{x}, \mathbf{y})$. Then:

$$\|\mathbf{x}^{(T)} - \mathbf{x}^*\|^2 \quad (29)$$

$$= \left\| \Pi_X \left(\mathbf{x}^{(T-1)} - \eta_T \mathbf{h}(T-1) \right) - \Pi_X(\mathbf{x}^*) \right\|^2 \quad (30)$$

$$\leq \left\| \mathbf{x}^{(T-1)} - \eta_T \mathbf{h}(T-1) - \mathbf{x}^* \right\|^2 \quad (31)$$

$$= \left\| \mathbf{x}^{(T-1)} - \mathbf{x}^* \right\|^2 - 2\eta_T \left\langle \mathbf{h}(T-1), \left(\mathbf{x}^{(T-1)} - \mathbf{x}^* \right) \right\rangle + \eta_T^2 \|\mathbf{h}(T-1)\|^2 \quad (32)$$

$$\leq \left\| \mathbf{x}^{(T-1)} - \mathbf{x}^* \right\|^2 - 2\eta_T \left(f(\mathbf{x}^{(T-1)}, \mathbf{y}^{(T-1)}) - f(\mathbf{x}^*, \mathbf{y}^{(T-1)}) \right) + \eta_T^2 \|\mathbf{h}(T-1)\|^2 \quad , \quad (33)$$

where the first line follows from the subgradient descent rule and the fact that $\mathbf{x}^* \in X$; the second, because the project operator is a non-expansion; the third, by the definition of the norm; and the

fourth, by the definition of the subgradient, i.e., $\langle \mathbf{h}(t-1), (\mathbf{x}^{(t-1)} - \mathbf{x}^*) \rangle \geq f(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t-1)}) - f(\mathbf{x}^*, \mathbf{y}^{(t-1)})$. Applying this inequality recursively, we obtain:

$$\|\mathbf{x}^{(T)} - \mathbf{x}^*\|^2 \leq \|\mathbf{x}^{(0)} - \mathbf{x}^*\|^2 - \sum_{t=1}^T 2\eta_t \left(f(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t-1)}) - f(\mathbf{x}^*, \mathbf{y}^{(t-1)}) \right) + \sum_{t=1}^T \eta_t^2 \|\mathbf{h}(t-1)\|^2. \quad (34)$$

Since $\|\mathbf{x}^{(t)} - \mathbf{x}^*\| \geq 0$, we have:

$$2 \sum_{t=1}^T \eta_t \left(f(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t-1)}) - f(\mathbf{x}^*, \mathbf{y}^{(t-1)}) \right) \leq \|\mathbf{x}^{(0)} - \mathbf{x}^*\|^2 + \sum_{t=1}^T \eta_t^2 \|\mathbf{h}(t-1)\|^2. \quad (35)$$

Let $(\mathbf{x}_{\text{best}}^{(t)}, \mathbf{y}_{\text{best}}^{(t)}) \in \arg \min_{(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}): k \in [t]} f(\mathbf{x}^{(k)}, \mathbf{y}^{(k)})$. Then:

$$\sum_{t=1}^T \eta_t \left(f(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t-1)}) - f(\mathbf{x}^*, \mathbf{y}^{(t-1)}) \right) \quad (36)$$

$$\geq \sum_{t=1}^T \eta_t \left(f(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t-1)}) - \max_{\mathbf{y} \in Y: \mathbf{g}(\mathbf{x}^*, \mathbf{y}) \geq 0} f(\mathbf{x}^*, \mathbf{y}) \right) \quad (37)$$

$$\geq \left(\sum_{t=1}^T \eta_t \right) \min_{t \in [T]} \left(f(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t-1)}) - \max_{\mathbf{y} \in Y: \mathbf{g}(\mathbf{x}^*, \mathbf{y}) \geq 0} f(\mathbf{x}^*, \mathbf{y}) \right) \quad (38)$$

$$= \left(\sum_{t=1}^T \eta_t \right) \left(f(\mathbf{x}_{\text{best}}^{(T)}, \mathbf{y}_{\text{best}}^{(T)}) - \max_{\mathbf{y} \in Y: \mathbf{g}(\mathbf{x}^*, \mathbf{y}) \geq 0} f(\mathbf{x}^*, \mathbf{y}) \right) \quad (39)$$

Hence, we get the following bound:

$$f(\mathbf{x}_{\text{best}}^{(T)}, \mathbf{y}_{\text{best}}^{(T)}) - \max_{\mathbf{y} \in Y: \mathbf{g}(\mathbf{x}^*, \mathbf{y}) \geq 0} f(\mathbf{x}^*, \mathbf{y}) \leq \frac{\|\mathbf{x}^{(0)} - \mathbf{x}^*\|^2 + \sum_{t=1}^T \eta_t^2 \|\mathbf{h}(t-1)\|^2}{2 \left(\sum_{t=1}^T \eta_t \right)} \quad (40)$$

Now since f is ℓ_f -Lipschitz with $\ell_f = \max_{(\hat{\mathbf{x}}, \hat{\mathbf{y}}) \in X \times Y} \|\nabla_{\mathbf{x}} f(\hat{\mathbf{x}}, \hat{\mathbf{y}})\|$, all the subgradients are bounded: i.e., for all $k \in \mathbb{N}$, $\|\mathbf{h}(k-1)\| \leq \ell_f$. So:

$$f(\mathbf{x}_{\text{best}}^{(T)}, \mathbf{y}_{\text{best}}^{(T)}) - \max_{\mathbf{y} \in Y: \mathbf{g}(\mathbf{x}^*, \mathbf{y}) \geq 0} f(\mathbf{x}^*, \mathbf{y}) \leq \frac{\|\mathbf{x}^{(0)} - \mathbf{x}^*\|^2 + \ell_f^2 \sum_{t=1}^T \eta_t^2}{2 \left(\sum_{t=1}^T \eta_t \right)} \quad (41)$$

Equivalently,

$$f(\mathbf{x}_{\text{best}}^{(T)}, \mathbf{y}_{\text{best}}^{(T)}) - \min_{\mathbf{x} \in X} \max_{\mathbf{y} \in Y: \mathbf{g}(\mathbf{x}, \mathbf{y}) \geq 0} f(\mathbf{x}, \mathbf{y}) \leq \frac{\|\mathbf{x}^{(0)} - \mathbf{x}^*\|^2 + \ell_f^2 \sum_{t=1}^T \eta_t^2}{2 \left(\sum_{t=1}^T \eta_t \right)} \quad (42)$$

Recall the assumptions that the step sizes are square-summable but not summable, namely $\sum_{k=1}^T \eta_k^2 \leq \infty$ and $\sum_{k=1}^T \eta_k = \infty$. Now as $T \rightarrow \infty$, Equation (42) becomes:

$$\lim_{k \rightarrow \infty} f(\mathbf{x}_{\text{best}}^{(k)}, \mathbf{y}_{\text{best}}^{(k)}) \leq \min_{\mathbf{x} \in X} \max_{\mathbf{y} \in Y: \mathbf{g}(\mathbf{x}, \mathbf{y}) \geq 0} f(\mathbf{x}, \mathbf{y}). \quad (43)$$

We have thus proven the first inequality of the two that define an $(0, \delta)$ -Stackelberg equilibrium.

The second inequality follows by construction, as for all $k \in \mathbb{N}$, the max oracle returns $\mathbf{y}_{\text{best}}^{(k)}$ that satisfies $f(\mathbf{x}_{\text{best}}^{(k)}, \mathbf{y}_{\text{best}}^{(k)}) \geq \max_{\mathbf{y} \in Y: \mathbf{g}(\mathbf{x}, \mathbf{y}) \geq 0} f(\mathbf{x}_{\text{best}}^{(k)}, \mathbf{y}) - \delta$. Thus, as $k \rightarrow \infty$, the best iterate converges to a $(0, \delta)$ -Stackelberg equilibrium.

Additionally, setting $\eta_t = \frac{\|\mathbf{x}^{(0)} - \mathbf{x}^*\|}{\ell_f \sqrt{T}}$, we find that for all $t \in [T]$,

$$f(\mathbf{x}_{\text{best}}^{(T)}, \mathbf{y}_{\text{best}}^{(T)}) - \min_{\mathbf{x} \in X} \max_{\mathbf{y} \in Y: g(\mathbf{x}, \mathbf{y}) \geq 0} f(\mathbf{x}, \mathbf{y}) \leq \frac{\ell_f \|\mathbf{x}^{(0)} - \mathbf{x}^*\|^2}{\sqrt{T}} \quad (44)$$

Likewise, setting $\varepsilon \leq \frac{\ell_f \|\mathbf{x}^{(0)} - \mathbf{x}^*\|^2}{\sqrt{T}}$, we obtain $T \leq \frac{\ell_f^2 \|\mathbf{x}^{(0)} - \mathbf{x}^*\|^4}{\varepsilon^2}$, and thus $N_T(\varepsilon) \in O(\varepsilon^{-2})$.

Finally, applying the same logic as above to infer the second inequality in the Stackelberg equilibrium definition, the best iterate converges to a (ε, δ) -Stackelberg equilibrium in $O(\varepsilon^{-2})$ iterations. \square

Theorem E.2. Suppose Algorithm [4](#) is run on a convex-concave min-max Stackelberg game given by (X, Y, f, g) where X is convex. Suppose further that Assumption [3.1](#) holds, and that f is μ -strongly convex in \mathbf{x} , i.e., $\forall \mathbf{x}_1, \mathbf{x}_2 \in X, \mathbf{y} \in Y$,

$$f(\mathbf{x}_1, \mathbf{y}) \geq f(\mathbf{x}_2, \mathbf{y}) + \langle \mathbf{h}, (\mathbf{x}_1 - \mathbf{x}_2) \rangle + \frac{\mu}{2} \|\mathbf{x}_1 - \mathbf{x}_2\|^2, \quad (45)$$

where $\mathbf{h} \in \partial_{\mathbf{x}} f(\mathbf{x}_2, \mathbf{y})$. Then, if $(\mathbf{x}_{\text{best}}^{(t)}, \mathbf{y}_{\text{best}}^{(t)}) \in \arg \min_{(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}): k \in [t]} f(\mathbf{x}^{(k)}, \mathbf{y}^{(k)})$, for $\varepsilon \in (0, 1)$, and $\eta_t = \frac{2}{\mu(t+1)}$, if we choose $T \geq N_T(\varepsilon) \in O(\varepsilon^{-1})$, then there exists an iteration $T^* \leq T$ s.t. $(\mathbf{x}_{\text{best}}^{(T^*)}, \mathbf{y}_{\text{best}}^{(T^*)})$ is an (ε, δ) -Stackelberg equilibrium.

Proof of Theorem [E.2](#) By our subdifferential envelope theorem (Theorem [3.2](#)), we have:

$$\nabla_{\mathbf{x}} f(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t-1)}) + \sum_{k=1}^K \lambda_k^{(t-1)} \nabla_{\mathbf{x}} g_k(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t-1)}) \in \partial_{\mathbf{x}} V(\mathbf{x}^{(t-1)}) \quad (46)$$

$$= \partial_{\mathbf{x}} \max_{\mathbf{y} \in Y: g(\mathbf{x}^{(t-1)}, \mathbf{y}) \geq 0} f(\mathbf{x}^{(t-1)}, \mathbf{y}). \quad (47)$$

For notational clarity, let $\mathbf{h}(t-1) = \nabla_{\mathbf{x}} f(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t-1)}) + \sum_{k=1}^K \lambda_k^{(t-1)} \nabla_{\mathbf{x}} g_k(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t-1)})$. Suppose that $\mathbf{x}^* \in \arg \min_{\mathbf{x} \in X} \max_{\mathbf{y} \in Y: g(\mathbf{x}, \mathbf{y}) \geq 0} f(\mathbf{x}, \mathbf{y})$. Then, for all $t \in \mathbb{N}$ s.t. $t \geq 1$, we have:

$$\|\mathbf{x}^{(t)} - \mathbf{x}^*\|^2 \quad (48)$$

$$= \left\| \Pi_X \left(\mathbf{x}^{(t-1)} - \eta_t \mathbf{h}(t-1) \right) - \Pi_X(\mathbf{x}^*) \right\|^2 \quad (49)$$

$$\leq \left\| \mathbf{x}^{(t-1)} - \eta_t \mathbf{h}(t-1) - \mathbf{x}^* \right\|^2 \quad (50)$$

$$= \left\| \mathbf{x}^{(t-1)} - \mathbf{x}^* \right\|^2 - 2\eta_t \langle \mathbf{h}(t-1), (\mathbf{x}^{(t-1)} - \mathbf{x}^*) \rangle + \eta_t^2 \|\mathbf{h}(t-1)\|^2 \quad (51)$$

$$\leq \left\| \mathbf{x}^{(t-1)} - \mathbf{x}^* \right\|^2 - 2\eta_t \left[\frac{\mu}{2} \left\| \mathbf{x}^{(t-1)} - \mathbf{x}^* \right\|^2 + f(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t)}) - f(\mathbf{x}^*, \mathbf{y}^{(t)}) \right] + \eta_t^2 \|\mathbf{h}(t-1)\|^2 \quad (52)$$

$$= \left\| \mathbf{x}^{(t-1)} - \mathbf{x}^* \right\|^2 - \eta_t \mu \left\| \mathbf{x}^{(t-1)} - \mathbf{x}^* \right\|^2 - 2\eta_t \left(f(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t-1)}) - f(\mathbf{x}^*, \mathbf{y}^{(t-1)}) \right) + \eta_t^2 \|\mathbf{h}(t-1)\|^2 \quad (53)$$

$$= (1 - \eta_t \mu) \left\| \mathbf{x}^{(t-1)} - \mathbf{x}^* \right\|^2 - 2\eta_t \left(f(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t-1)}) - f(\mathbf{x}^*, \mathbf{y}^{(t-1)}) \right) + \eta_t^2 \|\mathbf{h}(t-1)\|^2 \quad (54)$$

where the first line follows from the subgradient descent rule and the fact that $\mathbf{x}^* \in X$; the second, because the project operator is a non-expansion; the third, by the definition of the norm; the fourth, by the definition of strong convexity; and the last two via straightforward algebra.

Re-organizing Equation [\(54\)](#), yields:

$$f(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t-1)}) - f(\mathbf{x}^*, \mathbf{y}^{(t-1)}) \quad (55)$$

$$\leq \frac{1 - \eta_t \mu}{2\eta_t} \left\| \mathbf{x}^{(t-1)} - \mathbf{x}^* \right\|^2 - \frac{1}{2\eta_t} \left\| \mathbf{x}^{(t)} - \mathbf{x}^* \right\|^2 + \frac{\eta_t}{2} \|\mathbf{h}(t-1)\|^2. \quad (56)$$

Next, setting $\eta_t = \frac{2}{\mu(t+1)}$, we get:

$$f(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t-1)}) - f(\mathbf{x}^*, \mathbf{y}^{(t-1)}) \quad (57)$$

$$\leq \frac{\mu(t-1)}{4} \|\mathbf{x}^{(t-1)} - \mathbf{x}^*\|^2 - \frac{\mu(t+1)}{4} \|\mathbf{x}^{(t)} - \mathbf{x}^*\|^2 + \frac{1}{\mu(t+1)} \|\mathbf{h}(t-1)\|^2. \quad (58)$$

Multiplying both sides by t , we now have:

$$t \left(f(\mathbf{x}^{(t-1)}, \mathbf{y}^{(t-1)}) - f(\mathbf{x}^*, \mathbf{y}^{(t-1)}) \right) \quad (59)$$

$$\leq \frac{\mu t(t-1)}{4} \|\mathbf{x}^{(t-1)} - \mathbf{x}^*\|^2 - \frac{\mu t(t+1)}{4} \|\mathbf{x}^{(t)} - \mathbf{x}^*\|^2 + \frac{1}{\mu} \|\mathbf{h}(t-1)\|^2. \quad (60)$$

Summing up across all iterations on both sides:

$$\sum_{t=0}^T t \left(f(\mathbf{x}^{(t)}, \mathbf{y}^{(t)}) - f(\mathbf{x}^*, \mathbf{y}^{(t)}) \right) \quad (61)$$

$$\leq \sum_{t=0}^T \frac{\mu t(t-1)}{4} \|\mathbf{x}^{(t-1)} - \mathbf{x}^*\|^2 - \sum_{t=0}^T \frac{\mu t(t+1)}{4} \|\mathbf{x}^{(t)} - \mathbf{x}^*\|^2 + \sum_{t=0}^T \frac{1}{\mu} \|\mathbf{h}(t-1)\|^2 \quad (62)$$

$$= \sum_{t=0}^T \frac{\mu t(t-1)}{4} \|\mathbf{x}^{(t-1)} - \mathbf{x}^*\|^2 - \sum_{t=1}^{T+1} \frac{\mu(t-1)t}{4} \|\mathbf{x}^{(t-1)} - \mathbf{x}^*\|^2 + \sum_{t=0}^T \frac{1}{\mu} \|\mathbf{h}(t-1)\|^2 \quad (63)$$

$$= -\frac{\mu T(T+1)}{4} \|\mathbf{x}^{(T)} - \mathbf{x}^*\|^2 + \sum_{t=0}^T \frac{1}{\mu} \|\mathbf{h}(t-1)\|^2 \quad (64)$$

$$\leq \sum_{t=0}^T \frac{1}{\mu} \|\mathbf{h}(t-1)\|^2 \quad (65)$$

$$\leq \frac{T}{\mu} \ell_f, \quad (66)$$

where the last line holds because f is ℓ_f -Lipschitz with $\ell_f = \max_{(\hat{\mathbf{x}}, \hat{\mathbf{y}}) \in X \times Y} \|\nabla_{\mathbf{x}} f(\hat{\mathbf{x}}, \hat{\mathbf{y}})\|$, which implies that all the subgradients are bounded: i.e., for all $k \in \mathbb{N}$, $\|\mathbf{h}(k-1)\| \leq \ell_f$.

Let $(\mathbf{x}_{\text{best}}^{(t)}, \mathbf{y}_{\text{best}}^{(t)}) \in \arg \min_{(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}): k \in [t]} f(\mathbf{x}^{(k)}, \mathbf{y}^{(k)})$. Then:

$$\sum_{t=0}^T t \left(f(\mathbf{x}^{(t)}, \mathbf{y}^{(t)}) - f(\mathbf{x}^*, \mathbf{y}^{(t)}) \right) \leq \frac{T}{\mu} \ell_f \quad (67)$$

$$\sum_{t=0}^T t \left(f(\mathbf{x}^{(t)}, \mathbf{y}^{(t)}) - \max_{\mathbf{y} \in Y: \mathbf{g}(\mathbf{x}^*, \mathbf{y}) \geq \mathbf{0}} f(\mathbf{x}^*, \mathbf{y}) \right) \leq \frac{T}{\mu} \ell_f \quad (68)$$

$$\left(\sum_{t=0}^T t \right) \min_{t \in [T]} \left(f(\mathbf{x}^{(t)}, \mathbf{y}^{(t)}) - \max_{\mathbf{y} \in Y: \mathbf{g}(\mathbf{x}^*, \mathbf{y}) \geq \mathbf{0}} f(\mathbf{x}^*, \mathbf{y}) \right) \leq \frac{T}{\mu} \ell_f \quad (69)$$

$$\left(\sum_{t=0}^T t \right) \left(f(\mathbf{x}_{\text{best}}^{(T)}, \mathbf{y}_{\text{best}}^{(T)}) - \max_{\mathbf{y} \in Y: \mathbf{g}(\mathbf{x}^*, \mathbf{y}) \geq \mathbf{0}} f(\mathbf{x}^*, \mathbf{y}) \right) \leq \frac{T}{\mu} \ell_f \quad (70)$$

$$\left(\frac{(T+1)T}{2} \right) \left(f(\mathbf{x}_{\text{best}}^{(T)}, \mathbf{y}_{\text{best}}^{(T)}) - \max_{\mathbf{y} \in Y: \mathbf{g}(\mathbf{x}^*, \mathbf{y}) \geq \mathbf{0}} f(\mathbf{x}^*, \mathbf{y}) \right) \leq \frac{T}{\mu} \ell_f \quad (71)$$

$$\left(\frac{(T+1)T}{2} \right) \left(f(\mathbf{x}_{\text{best}}^{(T)}, \mathbf{y}_{\text{best}}^{(T)}) - \max_{\mathbf{y} \in Y: \mathbf{g}(\mathbf{x}^*, \mathbf{y}) \geq \mathbf{0}} f(\mathbf{x}^*, \mathbf{y}) \right) \leq \frac{T}{\mu} \ell_f \quad (72)$$

$$f(\mathbf{x}_{\text{best}}^{(T)}, \mathbf{y}_{\text{best}}^{(T)}) - \max_{\mathbf{y} \in Y: \mathbf{g}(\mathbf{x}^*, \mathbf{y}) \geq \mathbf{0}} f(\mathbf{x}^*, \mathbf{y}) \leq \frac{2\ell_f}{\mu(T+1)}. \quad (73)$$

That is, as the number of iterations increases, the best iterate converges to a $(0, \delta)$ -Stackelberg equilibrium. Likewise, by the same logic we applied at the end of the proof as Theorem [3.3](#), the best iterate converges to a (ε, δ) -Stackelberg equilibrium in $O(\varepsilon^{-1})$ iterations. \square

We now present a theorem which covers one of the cases given in Theorem 3.4. The proofs of the theorems that cover the other cases are similar to the proof below. We note that gradient ascent converges in $O(\varepsilon^{-1})$ iterations to a ε -maximum for a Lipschitz-smooth objective, and in $O(\log(\varepsilon))$ iterations to a ε -maximum for a Lipschitz-smooth and strongly-concave objective function [8].

Theorem E.3. *Suppose Algorithm 2 is run on a convex-concave min-max Stackelberg game given by (X, Y, f, \mathbf{g}) , where X and Y are convex. Suppose further that Assumption 3.1 holds and that f is $\ell_{\nabla f}$ -smooth, i.e., $\forall (\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2) \in X \times Y$,*

$$\|\nabla f(\mathbf{x}_1, \mathbf{y}_1) - \nabla f(\mathbf{x}_2, \mathbf{y}_2)\| \leq \ell_{\nabla f} \|(\mathbf{x}_1, \mathbf{y}_1) - (\mathbf{x}_2, \mathbf{y}_2)\| . \quad (74)$$

Let $(\mathbf{x}_{\text{best}}^{(t)}, \mathbf{y}_{\text{best}}^{(t)}) \in \arg \min_{(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}) : k \in [t]} f(\mathbf{x}^{(k)}, \mathbf{y}^{(k)})$. For $\varepsilon \in (0, 1)$, if we choose T_x and T_y s.t. $T_x \geq N_{T_x}(\varepsilon) \in O(\varepsilon^{-2})$ and $T_y \geq N_{T_y}(\varepsilon) \in O(\varepsilon^{-1})$, then there exists an iteration $T^ \leq T_x T_y = O(\varepsilon^{-3})$ s.t. $(\mathbf{x}_{\text{best}}^{(T^*)}, \mathbf{y}_{\text{best}}^{(T^*)})$ is an $(\varepsilon, \varepsilon)$ -Stackelberg equilibrium.*

Proof of Theorem 3.3 Since f is ℓ_f -smooth, it is well known that, for each outer iterate $\mathbf{x}^{(t)}$, the inner gradient descent procedure returns an ε -maximum \mathbf{y}_ε^* of $f(\mathbf{x}^{(t)}, \mathbf{y})$ s.t. $\mathbf{y}_\varepsilon^* \in Y$ and $\mathbf{g}(\mathbf{x}^{(t)}, \mathbf{y}_\varepsilon^*) \geq \mathbf{0}$, in $O(\varepsilon^{-2})$ iterations [8]. Combining the iteration complexity of the outer and inner loops using this result and Theorem 3.3 we obtain an iteration complexity of $O(\varepsilon^{-2})O(\varepsilon^{-1}) = O(\varepsilon^{-3})$. \square

F An Economic Application: Details

Our experimental goals were two-fold. First, we sought to understand the empirical convergence rate of our algorithms in different Fisher markets, in which the objective function in Equation (3) satisfies different smoothness properties. Second, we wanted to understand how the behavior of our two algorithms, max-oracle and nested gradient descent, differ in terms of the accuracy of the Stackelberg equilibria they find.

To answer these questions, we ran multiple experiments, each time recording the prices and allocations computed by Algorithm 1 with an *exact* max-oracle, and by Algorithm 2 with nested gradient ascent, during each iteration t of the main (outer) loop. For each run of each algorithm on each market with each set of initial conditions, we then computed the objective function’s value for the iterates, i.e., $f(\mathbf{x}^{(t)}, \mathbf{y}^{(t)})$, which we plot in Figure 1.

Hyperparameters We randomly initialized 500 different linear, Cobb-Douglas, Leontief Fisher markets, each with 5 buyers and 8 goods. Buyer i ’s budget b_i was drawn randomly from a uniform distribution ranging from 100 to 1000 (i.e., $U[100, 1000]$), while each buyer i ’s valuation for good j , v_{ij} , was drawn randomly from $U[5, 15]$. We ran both algorithms for 500, 300, and 700 iterations¹² for linear, Cobb-Douglas, and Leontief Fisher markets, respectively. We started both algorithms from two sets of initial conditions, one with high prices (drawn randomly $U[50, 55]$), and a second, with low prices (drawn randomly from $U[5, 15]$). We opted for a learning rate of 5 for both algorithms, after manual hyper-parameter tuning, and picked a decay rate of $t^{-1/2}$, based on our theory, so that $\eta_1 = 5, \eta_2 = 3.54, \eta_3 = 2.89, \eta_4 = 2.5, \eta_5 = 2.24, \dots$

Programming Languages, Packages, and Licensing We ran our experiments in Python 3.7 [77], using NumPy [34], Pandas [73], and CVXPY [21]. Figure 1 was graphed using Matplotlib [37]. To run the first order James test, we imported the data generated by our Python code into R [63]. Our R script manipulated the data using the Tidyverse package [81], and obtained the desired p -values using the STests package [36].

R as a package is licensed under GPL-2 | GPL-3. Python software and documentation are licensed under the PSF License Agreement. Numpy is distributed under a liberal BSD license. Pandas is distributed under a new BSD license. Matplotlib only uses BSD compatible code, and its license is based on the PSF license. CVXPY is licensed under an APACHE license. Tidyverse is distributed under an MIT license.

¹²In Algorithm 3 $T \in \{500, 300, 700\}$, while in Algorithm 4 $T_p \in \{500, 300, 700\}$.

Implementation Details In our execution of Algorithm 1 for linear, Cobb-Douglas, and Leontief Fisher markets, we used an exact Max-Oracle, since there is a closed-form solution for the demand correspondence in these markets [49].

In our execution of Algorithm 2 in order to project each computed allocation onto the consumers' budget set, i.e., $\{\mathbf{X} \in \mathbb{R}_+^{n \times m} \mid \mathbf{X}\mathbf{p} \leq \mathbf{b}\}$, we used the alternating projection algorithm [8] for convex sets, and alternatively projected onto the sets $\mathbb{R}_+^{n \times m}$ and $\{\mathbf{X} \in \mathbb{R}^{n \times m} \mid \mathbf{X}\mathbf{p} \leq \mathbf{b}\}$.

Computational Resources Our experiments were run on MacOS machine with 8GB RAM and an Apple M1 chip, and took about 2 hours to run. Only CPU resources were used.

Code Repository The data our experiments generated, as well as the code used to produce our visualizations and run the statistical tests, can be found in our [code repository](#).

F.1 Fisher Market Algorithms

Algorithm 3 δ -Approximate Tâtonnement for Fisher Markets

Inputs: $U, \mathbf{b}, \eta, T, \mathbf{p}^{(0)}, \delta$

Output: $(\mathbf{X}^*, \mathbf{p}^*)$

- 1: **for** $t = 1, \dots, T$ **do**
 - 2: For all $i \in [n]$, find $\mathbf{x}_i^{(t)}$ s.t. $u_i(\mathbf{x}_i^{(t)}) \geq \max_{\mathbf{x}_i: \mathbf{x}_i \cdot \mathbf{p}^{(t-1)} \leq b_i} u_i(\mathbf{x}_i) - \delta$ & $\mathbf{x}_i^{(t)} \cdot \mathbf{p}^{(t-1)} \leq b_i$
 - 3: Set $\mathbf{p}^{(t)} = \max \left\{ \mathbf{p}^{(t-1)} - \eta_t (1 - \sum_{i \in [n]} \mathbf{x}_i^{(t)}), 0 \right\}$
 - 4: **end for**
 - 5: **return** $(\mathbf{X}^{(T)}, \mathbf{p}^{(T)})$
-

Algorithm 4 δ -Approximate Nested Tâtonnement for Fisher Markets

Inputs: $U, \mathbf{b}, \eta^p, \eta^X T_p, T_X, \mathbf{p}^{(0)}$

Output: $(\mathbf{X}^*, \mathbf{p}^*)$

- 1: **for** $t = 1, \dots, T_p$ **do**
 - 2: **for** $s = 1, \dots, T_X$ **do**
 - 3: For all $i \in [n]$, $\mathbf{x}_i^{(t)} = \Pi_{\{\mathbf{x}: \mathbf{x} \cdot \mathbf{p}^{(t-1)} \leq b_i\}} \left(\mathbf{x}_i^{(t)} + \eta_s^X \frac{b_i}{u_i(\mathbf{x}_i^{(t)})} \nabla_{\mathbf{x}_i} u_i(\mathbf{x}_i^{(t)}) \right)$
 - 4: **end for**
 - 5: Set $\mathbf{p}^{(t)} = \max \left\{ \mathbf{p}^{(t-1)} - \eta_t^p (1 - \sum_{i \in [n]} \mathbf{x}_i^{(t)}), 0 \right\}$
 - 6: **end for**
 - 7: **return** $(\mathbf{X}^{(T)}, \mathbf{p}^{(T)})$
-

G Additional Related Work

Much progress has been made recently in solving min-max games (with independent strategy sets), both in the convex-concave case and in non-convex-concave case. For the former case, when f is μ_x -strongly-convex in \mathbf{x} and μ_y -strongly-concave in \mathbf{y} , Tseng [76], Yurii Nesterov [83], and Gidel et al. [30] proposed variational inequality methods, and Mokhtari, Ozdaglar, and Pattathil [51], gradient-descent-ascent (GDA)-based methods, all of which compute a solution in $\tilde{O}(\mu_y + \mu_x)$ iterations. These upper bounds were recently complemented by the lower bound of $\tilde{\Omega}(\sqrt{\mu_y \mu_x})$, shown by Ibrahim et al. [38] and Zhang, Hong, and Zhang [84]. Subsequently, Lin, Jin, and Jordan [46] and Alkousa et al. [3] analyzed algorithms that converge in $\tilde{O}(\sqrt{\mu_y \mu_x})$ and $\tilde{O}(\min \{\mu_x \sqrt{\mu_y}, \mu_y \sqrt{\mu_x}\})$ iterations, respectively.

For the special case where f is μ_x -strongly convex in \mathbf{x} and linear in \mathbf{y} , Juditsky, Nemirovski, et al. [42], Hamedani and Aybat [33], and Zhao [86] all present methods that converge to an ε -approximate solution in $O(\sqrt{\mu_x/\varepsilon})$ iterations. When the strong concavity or linearity assumptions of f on \mathbf{y} are dropped, and f is assumed to be μ_x -strongly-convex in \mathbf{x} but only concave in \mathbf{y} , Thekumparampil

et al. [74] provide an algorithm that converges to an ε -approximate solution in $\tilde{O}(\mu_x/\varepsilon)$ iterations, and Ouyang and Xu [60] provide a lower bound of $\tilde{\Omega}(\sqrt{\mu_x/\varepsilon})$ iterations on this same computation. Lin, Jin, and Jordan then went on to develop a faster algorithm, with iteration complexity of $\tilde{O}(\sqrt{\mu_x/\varepsilon})$, under the same conditions.

When f is simply assumed to be convex-concave, Nemirovski [53], Nesterov [54], and Tseng [75] describe algorithms that solve for an ε -approximate solution with $\tilde{O}(\varepsilon^{-1})$ iteration complexity, and Ouyang and Xu [60] prove a corresponding lower bound of $\tilde{\Omega}(\varepsilon^{-1})$.

When f is assumed to be non-convex- μ_y -strongly-concave, and the goal is to compute a first-order Nash, Sanjabi et al. [67] provide an algorithm that converges to ε -an approximate solution in $O(\varepsilon^{-2})$ iterations. Jin, Netrapalli, and Jordan [41], Rafique et al. [64], Lin, Jin, and Jordan [45], and Lu, Tsaknakis, and Hong [47] provide algorithms that converge in $\tilde{O}(\mu_y^2\varepsilon^{-2})$ iterations, while Lin, Jin, and Jordan [46] provide an even faster algorithm, with an iteration complexity of $\tilde{O}(\sqrt{\mu_y}\varepsilon^{-2})$.

When f is non-convex-non-concave and the goal to compute is an approximate first-order Nash equilibrium, Lu, Tsaknakis, and Hong [47] provide an algorithm with iteration complexity $\tilde{O}(\varepsilon^{-4})$, while Nouiehed et al. [58] provide an algorithm with iteration complexity $\tilde{O}(\varepsilon^{-3.5})$. More recently, Ostrovskii, Lowy, and Razaviyayn [59] and Lin, Jin, and Jordan [46] proposed an algorithm with iteration complexity $\tilde{O}(\varepsilon^{-2.5})$.

When f is non-convex-non-concave and the desired solution concept is a “local” Stackelberg equilibrium, Jin, Netrapalli, and Jordan [41], Rafique et al. [64], and Lin, Jin, and Jordan [45] provide algorithms with a $\tilde{O}(\varepsilon^{-6})$ complexity. More recently, Thekumparampil et al. [74], Zhao [85], and Lin, Jin, and Jordan [46] have proposed algorithms that converge to an ε -approximate solution in $\tilde{O}(\varepsilon^{-3})$ iterations.

We summarize the literature pertaining to the convex-concave and the non-convex-concave settings in Tables 3 and 4, respectively.

Table 3: Iteration complexities for min-max games (with independent strategy sets) in convex-concave settings. Note that these results assume that the objective function is Lipschitz-smooth.

Setting	Reference	Iteration Complexity
μ_x -Strongly-Convex- μ_y -Strongly-Concave	[76]	$\tilde{O}(\mu_x + \mu_y)$
	[83]	
	[30]	
	[51]	
	[3]	$\tilde{O}(\min\{\mu_x\sqrt{\mu_y}, \mu_y\sqrt{\mu_x}\})$
	[46]	$\tilde{O}(\sqrt{\mu_x\mu_y})$
	[38]	$\tilde{\Omega}(\sqrt{\mu_x\mu_y})$
	[84]	
μ_x -Strongly-Convex-Linear	[42]	$O(\sqrt{\mu_x/\varepsilon})$
	[33]	
	[86]	
μ_x -Strongly-Convex-Concave	[74]	$\tilde{O}(\mu_x/\sqrt{\varepsilon})$
	[46]	$\tilde{O}(\sqrt{\mu_x/\varepsilon})$
	[60]	$\tilde{\Omega}(\sqrt{\mu_x/\varepsilon})$
Convex-Concave	[53]	$O(\varepsilon^{-1})$
	[54]	
	[75]	
	[46]	$\tilde{O}(\varepsilon^{-1})$
	[60]	$\Omega(\varepsilon^{-1})$

Table 4: Iteration complexities for min-max games (with independent strategy sets) in non-convex-concave settings. Note that although all these results assume that the objective function is Lipschitz-smooth, some authors make additional assumptions: e.g., [58] obtain their result for objective functions that satisfy the Lojasiewicz condition.

Setting	Reference	Iteration Complexity
Nonconvex- μ_y -Strongly-Concave, First Order Nash Equilibrium or Local Stackelberg Equilibrium	[41]	$\tilde{O}(\mu_y^2 \varepsilon^{-2})$
	[64]	
	[45]	
	[47]	
	[46]	$\tilde{O}(\sqrt{\mu_y} \varepsilon^{-2})$
Nonconvex-Concave, First Order Nash Equilibrium	[47]	$\tilde{O}(\varepsilon^{-4})$
	[58]	$\tilde{O}(\varepsilon^{-3.5})$
	[59]	$\tilde{O}(\varepsilon^{-2.5})$
	[46]	
Nonconvex-Concave, Local Stackelberg Equilibrium	[41]	$\tilde{O}(\varepsilon^{-6})$
	[58]	
	[46]	
	[74]	$\tilde{O}(\varepsilon^{-3})$
	[85]	
	[46]	

H Future Directions

Our experiments suggest that the smoothness properties of the value function 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 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 $\mathbf{x} \in X$. This would require not only that the solution function $Y^*(\mathbf{x})$ be a singleton for all $\mathbf{x} \in \mathcal{X}$, but that the optimal KKT multipliers $\Lambda(\hat{\mathbf{x}}, \mathbf{y}^*(\hat{\mathbf{x}}))$ were also unique. The former can be guaranteed when f is strictly concave in \mathbf{y} , while the latter condition is satisfied when the Linear Independence Constraint Qualification condition [35] holds. However, even when both these conditions hold, and when both the objective function and the (parameterized) constraints (i.e., $\mathbf{g}(\mathbf{x}, \mathbf{y})$) are Lipschitz-smooth, the value function is not guaranteed to be Lipschitz-smooth.

Gao and Kroer [29] 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 Stackelberg games can facilitate the extension of their fast convergence results to a larger class of Fisher markets.

A question of interest both for Fisher market dynamics and convex-concave min-max Stackelberg games is whether gradient-descent-ascent (GDA) converges in the dependent strategy set setting as it does in the independent strategy setting [45]. GDA dynamics for Fisher markets correspond to myopic best-response dynamics, which are of general economic interest (see, for example, [52]).

Finally, our results at present concern only *convex-concave* min-max Stackelberg games. It would be of interest to extend these results to the non-convex-concave setting. Doing so could improve our understanding of competitive equilibria in Fisher markets with non-homogeneous utility functions, and our understanding of optimal auctions. Recently, Dai and Zhang [16] defined the concept of a local minimax point, whose analog in Stackelberg games would be a local Stackelberg equilibria. We believe that any algorithm designed for a non-convex-concave setting would 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. For example, 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 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.