Implicit Learning Dynamics in Stackelberg Games: Equilibria Characterization, Convergence Analysis, and Empirical Study

Tanner Fiez [1] **Benjamin Chasnov** [1] **Lillian Ratliff** [1]

Abstract

Contemporary work on learning in continuous games has commonly overlooked the hierarchical decision-making structure present in machine learning problems formulated as games, instead treating them as simultaneous play games and adopting the Nash equilibrium solution concept. We deviate from this paradigm and provide a comprehensive study of learning in Stackelberg games. This work provides insights into the optimization landscape of zero-sum games by establishing connections between Nash and Stackelberg equilibria along with the limit points of simultaneous gradient descent. We derive novel gradient-based learning dynamics emulating the natural structure of a Stackelberg game using the implicit function theorem and provide convergence analysis for deterministic and stochastic updates for zero-sum and general-sum games. Notably, in zero-sum games using deterministic updates, we show the only critical points the dynamics converge to are Stackelberg equilibria and provide a local convergence rate. Empirically, our learning dynamics mitigate rotational behavior and exhibit benefits for training generative adversarial networks compared to simultaneous gradient descent.

1. Introduction

The emerging coupling between game theory and machine learning can be credited to the formulation of learning problems as interactions between competing objectives and strategic agents. Indeed, generative adversarial networks (GANs) (Goodfellow et al., 2014), robust supervised learning (Madry et al., 2018), reinforcement and multi-agent reinforcement learning (Dai et al., 2018; Zhang et al., 2019), and hyperparameter optimization (Maclaurin et al., 2015) problems can be cast as zero-sum or general-sum continuous action games. To obtain solutions in a tractable manner, gradient-based algorithms have gained attention.

Given the motivating applications, much of the contemporary work on learning in games has focused on zero-sum games with non-convex, non-concave objective functions and seeking stable critical points or local equilibria. A number of techniques have been proposed including optimistic and extra-gradient algorithms (Daskalakis et al., 2018; Daskalakis & Panageas, 2018; Mertikopoulos et al., 2019), gradient adjustments (Balduzzi et al., 2018; Mescheder et al., 2017), and opponent modeling methods (Zhang & Lesser, 2010; Foerster et al., 2018; Letcher et al., 2019; Schäfer & Anandkumar, 2019). However, only a select number of algorithms can guarantee convergence to stable critical points satisfying sufficient conditions for a *local Nash equilibrium* (`LNE`) (Mazumdar et al., 2019; Adolphs et al., 2019).

The dominant perspective in machine learning applications of game theory has been focused on simultaneous play. However, there are many problems exhibiting a hierarchical order of play, and in a game theoretic context, such problems are known as Stackelberg games. The Stackelberg equilibrium (Von Stackelberg, 2010) solution concept generalizes the min-max solution to general-sum games. In the simplest formulation, one player acts as the leader who is endowed with the power to select an action knowing the other player (follower) plays a best-response. This viewpoint has long been researched from a control perspective on games (Basar & Olsder, 1998) and in the bilevel optimization community (Danskin, 1967; 1966; Zaslavski, 2012).

The work from a machine learning perspective on games with a hierarchical decision-making structure is sparse and exclusively focuses on zero-sum games. In the most relevant theoretical work, Jin et al. (2019) show that all stable critical points of simultaneous gradient descent with a timescale separation between players approaching infinity satisfy sufficient conditions for a *local Stackelberg equilibrium* (`LSE`). The closest empirical work we are aware of is on unrolled GANs (Metz et al., 2017), where the leader (generator) optimizes a surrogate cost function that depends on parameters

[1]Department of Electrical and Computer Engineering, University of Washington, Seattle, Washington. Correspondence to: Tanner Fiez <fiezt@uw.edu>.

Proceedings of the 37^{th} International Conference on Machine Learning, Online, PMLR 119, 2020.

of the follower (discriminator) that have been 'rolled out' until an approximate local optimum is reached. This behavior intuitively approximates a hierarchical order of play and consequently the success of the unrolling method as a training mechanism provides some evidence supporting the LSE solution concept. In this paper, we provide a step toward bridging the gap between theory and practice along this perspective by developing implementable learning dynamics with convergence guarantees to critical points satisfying sufficient conditions for a LSE.

Contributions. Motivated by the lack of algorithms focusing on games exhibiting an order of play, we provide a study of learning in Stackelberg games including equilibria characterization, novel learning dynamics and convergence analysis, and an illustrative empirical study. The primary benefits of this work to the community include an enlightened perspective on the consideration of equilibrium concepts reflecting the underlying optimization problems present in machine learning applications formulated as games and an algorithm that provably converges to critical points satisfying sufficient conditions for a LSE in zero-sum games.

We provide a characterization of LSE via sufficient conditions on the players objectives and term points satisfying the conditions *differential Stackelberg equilibria* (DSE). We show DSE are generic amongst LSE in zero-sum games. This means except on a set of measure zero in the class of zero-sum continuous games, DSE and LSE are equivalent. While the placement of *differential Nash equilibria* (DNE) amongst critical points in continuous games is reasonably well understood, an equivalent statement cannot be made regarding DSE. Accordingly, we draw connections between the solution concepts in the class of zero-sum games. We show that DNE are DSE, which indicates the solution concept in hierarchical play games is not as restrictive as the solution concept in simultaneous play games. Furthermore, we reveal that there exist stable critical points of simultaneous gradient descent dynamics that are DSE and not DNE. This insight gives meaning to a broad class of critical points previously thought to lack game-theoretic meaning and may give some explanation for the adequacy of solutions not satisfying sufficient conditions for LNE in GANs. To characterize this phenomenon, we provide necessary and sufficient conditions for when such points exist.

We derive novel gradient-based learning dynamics emulating the natural structure of a Stackelberg game from the sufficient conditions for a LSE and the implicit function theorem. The dynamics can be viewed as an analogue to simultaneous gradient descent incorporating the structure of hierarchical play games. In stark contrast to the simultaneous play counterpart, we show in zero-sum games the only stable critical points of the dynamics are DSE and such equilibria must be stable critical points of the dynamics. Using this fact and saddle avoidance results, we show the only critical points the discrete time algorithm converges to given deterministic gradients are DSE and provide a local convergence rate. In general-sum games, we cannot guarantee the only critical point attractors of the deterministic learning algorithms are DSE. However, we give a local convergence rate to critical points which are DSE. For stochastic gradient updates, we obtain analogous convergence guarantees asymptotically for each game class.

Empirically, we show that our dynamics result in stable learning compared to simultaneous gradient dynamics when training GANs. To gain insights into the placement of DNE and DSE in the optimization landscape, we analyze the eigenvalues of relevant game objects and observe convergence to neighborhoods of equilibria. Finally, we show that our dynamics can scale to computationally intensive problems.

2. Preliminaries

We now formalize the games we study, present equilibrium concepts accompanied by sufficient condition characterizations, and formulate Stackelberg learning dynamics.

2.1. Game Formalisms

Consider a non-cooperative game between two agents where player 1 is deemed the *leader* and player 2 the *follower*. The leader has cost $f_1 : X \to \mathbb{R}$ and the follower has cost $f_2 : X \to \mathbb{R}$, where $X = X_1 \times X_2 \in \mathbb{R}^m$ with $X_1 \in \mathbb{R}^{m_1}$ and $X_2 \in \mathbb{R}^{m_2}$ denoting the action spaces of the leader and follower, respectively.[1] We assume throughout that each f_i is sufficiently smooth: $f_i \in C^q(X, \mathbb{R})$ for some $q \geq 2$. For zero-sum games, the game is defined by costs $(f_1, f_2) = (f, -f)$. In words, we consider the class of two-player smooth games on continuous, unconstrained actions spaces. The designation of 'leader' and 'follower' indicates the order of play between the agents, meaning the leader plays first and the follower second.

In a *Stackelberg game*, the leader and follower aim to solve the following optimization problems, respectively:

$$\min_{x_1 \in X_1} \{ f_1(x_1, x_2) \big| \, x_2 \in \arg\min_{y \in X_2} f_2(x_1, y) \}, \tag{L}$$

$$\min_{x_2 \in X_2} f_2(x_1, x_2). \tag{F}$$

This contrasts with a *simultaneous* play game in which each player i is faced with the optimization problem $\min_{x_i \in X_i} f_i(x_i, x_{-i})$. The learning algorithms we formulate are such that the agents follow myopic update rules which take steps in the direction of steepest descent for the respective optimizations problems.

[1] Our results hold more generally for action spaces that are precompact subsets of the Euclidean space since they are local.

2.2. Equilibria Concepts and Characterizations

Before formalizing learning rules, let us first discuss the equilibrium concept studied for simultaneous play games and contrast it with that which is studied in the hierarchical play counterpart. The typical equilibrium notion in continuous games is the pure strategy Nash equilibrium in simultaneous play games and the Stackelberg equilibrium in hierarchical play games. Each notion of equilibria can be characterized as the intersection points of the reaction curves of the players (Basar & Olsder, 1998). We focus our attention on local notions of the equilibrium concepts as is standard in learning in games since the objective functions we consider need not be convex or concave.

Definition 1 (Local Nash (LNE))**.** *The joint strategy $x^* \in X$ is a local Nash equilibrium on $U_1 \times U_2 \subset X_1 \times X_2$ if for each $i \in \{1,2\}$, $f_i(x^*) \leq f_i(x_i, x^*_{-i})$, $\forall\, x_i \in U_i \subset X_i$.*

Definition 2 (Local Stackelberg (LSE))**.** *Consider $U_i \subset X_i$ for each $i \in \{1,2\}$. The strategy $x_1^* \in U_1$ is a local Stackelberg solution for the leader if, $\forall x_1 \in U_1$,*

$$\sup\nolimits_{x_2 \in R_{U_2}(x_1^*)} f_1(x_1^*, x_2) \leq \sup\nolimits_{x_2 \in R_{U_2}(x_1)} f_1(x_1, x_2),$$

where $R_{U_2}(x_1) = \{y \in U_2 | f_2(x_1, y) \leq f_2(x_1, x_2), \forall x_2 \in U_2\}$. Moreover, (x_1^*, x_2^*) for any $x_2^* \in R_{U_2}(x_1^*)$ is a local Stackelberg equilibrium on $U_1 \times U_2$.

While characterizing existence of equilibria is outside the scope of this work, we remark that Nash equilibria exist for convex costs on compact and convex strategy spaces and Stackelberg equilibria exist on compact strategy spaces (Basar & Olsder, 1998, Thm. 4.3, Thm. 4.8, & §4.9). This means the class of games on which Stackelberg equilibria exist is broader than on which Nash equilibria exist. Existence of local equilibria is guaranteed if the neighborhoods and cost functions restricted to those neighborhoods satisfy the assumptions of the cited results.

Predicated on existence, equilibria can be characterized in terms of sufficient conditions on player costs. We denote $D_i f_i$ as the derivative of f_i with respect to x_i, $D_{ij} f_i$ as the partial derivative of $D_i f_i$ with respect to x_j, and $D(\cdot)$ as the total derivative.[2] The following gives sufficient conditions for a LNE as given in Definition 1.

Definition 3 (Differential Nash (DNE) Ratliff et al. (2016))**.** *The joint strategy $x^* \in X$ is a differential Nash equilibrium if $D_i f_i(x^*) = 0$ and $D_i^2 f_i(x^*) > 0$ for each $i \in \{1,2\}$.*

Analogous sufficient conditions can be stated to characterize a LSE from Definition 2. Towards this end, given a point x^* at which $D_2 f_2(x^*) = 0$ and $\det(D_2^2 f_2(x^*)) \neq 0$, the implicit function theorem (Abraham et al., 1988, Thm. 2.5.7) implies that there exists a neighborhood U_1 and an implicit map $r : x_1 \mapsto x_2$ defined on U_1. Further, $Dr \equiv -(D_2^2 f_2)^{-1} \circ D_{21} f_2$. Note that $\det(D_2^2 f_2(x)) \neq 0$ is a generic condition (cf. Lemma C.3). Let $Df_1(x_1, r(x_1))$ be the total derivative of f_1 and analogously, let $D^2 f_1$ be the second-order total derivative.

[2]Example: given $f(x, r(x))$, $Df = D_1 f + Dr^\top D_2 f$.

Definition 4 (Differential Stackelberg (DSE))**.** *The joint strategy $x^* = (x_1^*, x_2^*) \in X$ is a differential Stackelberg equilibrium if $Df_1(x^*) = 0$, $D_2 f_2(x^*) = 0$, $D^2 f_1(x^*) > 0$, and $D_2^2 f_2(x^*) > 0$.*

Game Jacobians play a key role in determining stability of critical points. For simultaneous play, let

$$\omega(x) = (D_1 f_1(x), D_2 f_2(x))$$

be the vector of individual gradients and

$$\omega_{\mathcal{S}}(x) = (Df_1(x), D_2 f_2(x))$$

as the equivalent for the Stackelberg game. Observe that Df_1 is the total derivative of f_1 with respect to x_1 given x_2 is implicitly a function of x_1, capturing the fact that the leader operates under the assumption that the follower will play a (local) best response to x_1. The reaction curve of the follower may not be unique. However, sufficient conditions on a local Stackelberg solution x—i.e., $D_2 f_2(x) = 0$ and $\det(D_2^2 f_2(x)) \neq 0$—guarantee that Df_1 is well defined (cf. implicit mapping theorem).

The vector field $\omega(x)$ forms the basis of the well-studied simultaneous gradient learning dynamics and the Jacobian of the dynamics is given by

$$J(x) = \begin{bmatrix} D_1^2 f_1(x) & D_{12} f_1(x) \\ D_{21} f_2(x) & D_2^2 f_2(x) \end{bmatrix}.$$

Similarly, the vector field $\omega_{\mathcal{S}}(x)$ serves as the foundation of the learning dynamics we formulate in Section 2.4 and analyze throughout. The Jacobian of the Stackelberg vector field $\omega_{\mathcal{S}}(x)$ is given by

$$J_{\mathcal{S}}(x) = \begin{bmatrix} D_1(Df_1(x)) & D_2(Df_1(x)) \\ D_{21} f_2(x) & D_2^2 f_2(x) \end{bmatrix}. \tag{1}$$

A critical point is called non-degenerate if the determinant of the vector field Jacobian is non-zero. We denote by $\mathbb{C}_-^\circ$ and $\mathbb{C}_+^\circ$ the open left and right half complex planes. Moreover, a critical point x^* of $\dot{x} = -\omega(x)$ is stable if $\operatorname{spec}(-J(x^*)) \subset \mathbb{C}_-^\circ$ or equivalently $\operatorname{spec}(J(x^*)) \subset \mathbb{C}_+^\circ$. Similarly, a critical point x^* of $\dot{x} = -\omega_{\mathcal{S}}(x)$ is stable if $\operatorname{spec}(-J_{\mathcal{S}}(x^*)) \subset \mathbb{C}_-^\circ$ or equivalently $\operatorname{spec}(J_{\mathcal{S}}(x^*)) \subset \mathbb{C}_+^\circ$.

Noting that the Schur complement of $J_{\mathcal{S}}(x)$ with respect to $D_2^2 f_2(x)$ is identically $D^2 f(x_1, r(x_1))$, we give alternative but equivalent sufficient conditions as those in Definition 4 in terms of $J_{\mathcal{S}}(x)$. For a two-by-two block matrix such as $J_{\mathcal{S}}$, we denote by $\mathtt{S}_1(J_{\mathcal{S}})$ the Schur complement of $J_{\mathcal{S}}$ with respect to $D_2^2 f_2$. The proof of the following result is in Appendix B.

Proposition 1. *Consider a game (f_1, f_2) defined by $f_i \in C^q(X, \mathbb{R})$, $i = 1, 2$ with $q \geq 2$ and player 1 (without loss of generality) taken to be the leader. Let x^* satisfy $D_2 f_2(x^*) = 0$ and $D_2^2 f_2(x^*) > 0$. Then $Df_1(x^*) = 0$ and $\mathtt{S}_1(J_{\mathcal{S}}(x^*)) > 0$ if and only if x^* is a* DSE*. Moreover, in zero-sum games, $\mathtt{S}_1(J_{\mathcal{S}}(x)) = \mathtt{S}_1(J(x))$.*

2.3. Genericity and Structural Stability

A natural question is how common is it for local equilibria to satisfy sufficient conditions, meaning in a formal mathematical sense, what is the gap between necessary and sufficient conditions in games. Towards addressing this, it has been shown that DNE are generic amongst LNE and structurally stable in the classes of zero-sum and general-sum continuous games, respectively (Ratliff et al., 2016; Mazumdar & Ratliff, 2019). The results say that except on a set of measure zero in each class of games, DNE = LNE and the equilibria persist under sufficiently smooth perturbations to the costs. We give analogous results for DSE in the class of zero-sum games in this section and provide proofs in Appendix C. The following result allows us to conclude that for a generic zero-sum game, DSE = LSE.

Theorem 1. *For the class of two-player, zero-sum continuous games $(f, -f)$ where $f \in C^q(\mathbb{R}^m, \mathbb{R})$ with $q \geq 2$,* DSE *are generic amongst* LSE*. That is, given a generic $f \in C^q(\mathbb{R}^m, \mathbb{R})$, all* LSE *of the game $(f, -f)$ are* DSE*.*

A critical point x^* of the vector field $\omega_{\mathcal{S}}(x)$ is hyperbolic if there are no eigenvalues of $J_{\mathcal{S}}(x^*)$ with zero real part. We now show that in generic zero-sum games, LSE are hyperbolic critical points of the vector field $\omega_{\mathcal{S}}(x)$, which is desirable property owing to the convergence implications.

Corollary 1. *For the class of two-player, zero-sum continuous games $(f, -f)$ where $f \in C^q(\mathbb{R}^m, \mathbb{R})$ with $q \geq 2$,* LSE *are generically non-degenerate, hyperbolic critical points of the vector field $\omega_{\mathcal{S}}(x)$.*

As a final result in this section, we show that DSE are structural stable in the class of zero-sum games. Structural stability ensures that differential Stackelberg equilibria are robust and persist under smooth perturbations.

Theorem 2. *For the class of two-player, zero-sum continuous games $(f, -f)$ where $f \in C^q(\mathbb{R}^m, \mathbb{R})$ with $q \geq 2$,* DSE *are structurally stable: given $f \in C^q(\mathbb{R}^{m_1} \times \mathbb{R}^{m_2}, \mathbb{R})$, $\zeta \in C^q(\mathbb{R}^{m_1} \times \mathbb{R}^{m_2}, \mathbb{R})$, and a* DSE *$(x_1, x_2) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$, there exists neighborhoods $U \subset \mathbb{R}$ of zero and $V \subset \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$ such that $\forall\, t \in U$ there exists a unique* DSE *$(\tilde{x}_1, \tilde{x}_2) \in V$ for the zero-sum game $(f + t\zeta, -f - t\zeta)$.*

Before moving on, we remark that important classes of non-generic games certainly exist. In games where the cost function of the follower is bilinear, LSE can exist which do not satisfy the sufficient conditions outlined in Definition 4.

Algorithm 1 Deterministic Stackelberg Learning Dynamics

Input: $x_0 \in X$, learning rates $\gamma_1, \gamma_2 > 0$
for $k = 0, 1, \dots$ **do**
  $\omega_{\mathcal{S},1} \leftarrow D_1 f_1(x_k) - D_{21} f_2(x_k)^\top (D_2^2 f_2(x_k))^{-1} D_2 f_1(x_k)$
  $\omega_{\mathcal{S},2} \leftarrow D_2 f_2(x_k)$
  $x_{1,k+1} \leftarrow x_{1,k} - \gamma_1 \omega_{\mathcal{S},1}$
  $x_{2,k+1} \leftarrow x_{2,k} - \gamma_2 \omega_{\mathcal{S},2}$
end for

As a simple example, $x^* = (0, 0)$ is a LSE for the zero-sum game defined by $f(x_1, x_2) = x_1 x_2$ and not a DSE since $D_2^2 f_2(x) = 0\ \forall\ x \in X$. Since such games belong to a degenerate class in the context of the genericity result we provide, they naturally deserve special attention and algorithmic methods. While we do not focus our attention on this class of games, we do propose some remedies to allow our proposed learning algorithm to successfully seek out equilibria in them. In the experiments section, we discuss a regularized version of our dynamics that injects a small perturbation to cure degeneracy problems leveraging the fact that DSE are structurally stable. Further details can be found in Appendix H.1. Finally, for bimatrix games with finite actions it is common to reparameterize the problem using a softmax function to obtain mixed policies on the simplex (Fudenberg et al., 1998). We explore this viewpoint in Appendix H.3 on a parameterized bilinear game.

2.4. Stackelberg Learning Dynamics

Recall that $\omega_{\mathcal{S}}(x) = (Df_1(x), D_2 f_2(x))$ is the vector field for Stackelberg games and it, along with its Jacobian $J_{\mathcal{S}}(x)$, characterize sufficient conditions for a DSE. Letting $\omega_{\mathcal{S},i}$ be the i–th component of $\omega_{\mathcal{S}}$, the leader total derivative is $\omega_{\mathcal{S},1}(x) = D_1 f_1(x) - D_{21} f_2(x)^\top (D_2^2 f_2(x))^{-1} D_2 f_1(x)$. The Stackelberg learning rule we study for each player in discrete time is given by

$$x_{i,k+1} = x_{i,k} - \gamma_{i,k} h_{\mathcal{S},i}(x_k). \tag{2}$$

In deterministic learning players have oracle gradient access so that $h_{\mathcal{S},i}(x) = \omega_{\mathcal{S},i}(x)$. We study convergence for deterministic learning in Section 4.1 and Algorithm 1 provides example pseudocode. In stochastic learning players have unbiased gradient estimates and $h_{\mathcal{S},i}(x_k) = \omega_{\mathcal{S},i}(x_k) + w_{k+1,i}$ where $\{w_{i,k}\}$ is player i's noise process. We provide convergence analysis for stochastic learning in Section 4.2.

3. Implications for Zero-Sum Settings

Before presenting convergence analysis of the update in (2), we draw connections between Nash and Stackelberg equilibria in zero-sum games and discuss the relevance to applications such as adversarial learning. To do so, we evaluate the limiting behavior of the dynamics from a continuous time viewpoint since the discrete time system closely approximates this behavior for suitably selected learning rates. While we provide the intuition behind the results here, the formal proofs of the results are in Appendix D.

Let us first show that for zero-sum games, all stable critical points of $\dot{x} = -\omega_{\mathcal{S}}(x)$ are DSE and vice versa.

Proposition 2. *In zero-sum games $(f, -f)$ with $f \in C^q(X, \mathbb{R})$ for $q \geq 2$, a joint strategy $x \in X$ is a stable critical point of $\dot{x} = -\omega_{\mathcal{S}}(x)$ if and only if x is a* DSE. *Moreover, if f is generic, a point x is a stable critical point of $\dot{x} = -\omega_{\mathcal{S}}(x)$ if and only if it is a* LSE.

The result follows from the structure of the Jacobian of $\omega_{\mathcal{S}}(x)$, which is lower block triangular with player 1 and 2 as the leader and follower, respectively. Proposition 2 implies that with appropriate stepsizes the update rule in (2) will only converge to Stackelberg equilibria and thus, unlike simultaneous gradient descent, will not converge to spurious locally asymptotically stable points that lack game-theoretic meaning (see, e.g., Mazumdar et al. (2020)).

This previous result begs the question of which stable critical points of the dynamics $\dot{x} = -\omega(x)$ are DSE? The following gives a partial answer to the question and also indicates that recent works seeking DNE are also seeking DSE.

Proposition 3. *In zero-sum games $(f, -f)$ with $f \in C^q(X, \mathbb{R})$ for $q \geq 2$,* DNE *are* DSE. *Moreover, if f is generic,* LNE *are* LSE.

This result follows from the facts that the conditions of a DNE imply $\mathtt{S}_1(J(x)) > 0$ and that non-degenerate DNE are generic amongst LNE within the class of zero-sum games (Mazumdar & Ratliff, 2019). In the zero-sum setting, the fact that Nash equilibria are a subset of Stackelberg equilibria for finite games is well-known (Basar & Olsder, 1998). We extend this result locally to continuous action space games. Similar to our work and concurrently, Jin et al. (2019) show that LNE are local min-max solutions.

In Proposition D.1 of Appendix D, we show the previous results imply all DNE are stable critical points of both $\dot{x} = -\omega(x)$ and $\dot{x} = -\omega_{\mathcal{S}}(x)$. This leaves the question of the meaning of stable points of $\dot{x} = -\omega(x)$ which are not DNE.

Finding Meaning in Spurious Stable Critical Points. We focus on the question of when stable fixed points of $\dot{x} = -\omega(x)$ are DSE and not DNE. It was shown by Jin et al. (2019) that not all stable points of $\dot{x} = -\omega(x)$ are local min-max or local max-min equilibria since one can construct a function such that $D_1^2 f(x)$ and $-D_2^2 f(x)$ are *both* not positive definite but the real parts of the eigenvalues of $J(x)$ are positive. It appears to be much harder to characterize when a stable critical point of $\dot{x} = -\omega(x)$ is not a DNE but is a DSE since it requires the follower's individual Hessian to be positive definite. Indeed, it reduces to a fundamental problem in linear algebra in which the relationship between the eigenvalues of the sum of two matrices is largely unknown without assumptions on the structure of the matrices (Knutson & Tao, 2001).

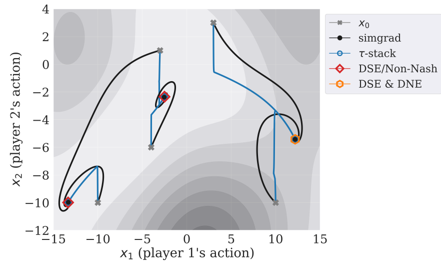


Figure 1. Example demonstrating existence of DSE and DSE that are not DNE: $\mathcal{G} = (f, -f)$ where f is defined in (3) with $a = 0.15, b = 0.25$. There are two stable points of simultaneous gradient descent which are DSE, but not DNE.

In Appendix E, we provide necessary and sufficient conditions for attractors at which the follower's Hessian is positive definite to be DSE. Taking intuition from the expression $\mathtt{S}_1(J(x)) = D_1^2 f(x) - D_{21} f(x)^\top (D_2^2 f(x))^{-1} D_{21} f(x)$, the conditions are derived from relating $\mathrm{spec}(D_1^2 f)$ to $\mathrm{spec}(D_2^2 f)$ via $D_{12} f$. To illustrate this fact, consider the following example in which stable points are DSE and not DNE—meaning points $x \in X$ at which $D_1^2 f(x) \not> 0$, $-D_2^2 f(x) > 0$ and $\mathrm{spec}(-J(x^*)) \subset \mathbb{C}_-^\circ$ and $\mathtt{S}_1(J(x)) > 0$.

Example: Non-Nash Attractors are Stackelberg. Consider the zero-sum game defined by

$$f(x) = -e^{-0.01(x_1^2 + x_2^2)}((ax_1^2 + x_2)^2 + (bx_2^2 + x_1)^2). \quad (3)$$

Let player 1 be the leader who aims to minimize f with respect to x_1 taking into consideration that player 2 (follower) aims to minimize $-f$ with respect to x_2. In Fig. 1, we show the trajectories for different initializations for this game; it can be seen that simultaneous gradient descent can lead to stable critical points which are DSE and not DNE. In fact, it is the case that all stable critical points with $-D_2^2 f(x) > 0$ are DSE in games on $\mathbb{R}^2$ (see Corollary E.1, Appendix E).

This example, along with Propositions E.1 and E.2 in Appendix E, implies some stable critical points of $\dot{x} = -\omega(x)$ which are not DNE are in fact DSE. This is a meaningful result since recent works have proposed schemes to avoid stable critical points which are not DNE as they have been thought to lack game-theoretic meaning (Adolphs et al., 2019; Mazumdar et al., 2019). Moreover, some recent empirical studies show a number of successful approaches to training GANs do not converge to DNE, but rather to stable fixed points of the dynamics at which the follower is at a local optimum (Berard et al., 2020). This may suggest reaching DSE is desirable in GANs.

The 'realizable' assumption in the GAN literature says the discriminator network is zero near an equilibrium parameter configuration (Nagarajan & Kolter, 2017). The assumption implies the Jacobian of $\dot{x} = -\omega(x)$ is such that $D_1^2 f(x) = 0$. Under this assumption, we show stable critical points which are not $\texttt{DNE}$ are $\texttt{DSE}$ given $-D_2^2 f(x) > 0$.

Proposition 4. *Consider a zero-sum GAN satisfying the realizable assumption. Any stable critical point of $\dot{x} = -\omega(x)$ at which $-D_2^2 f(x) > 0$ is a $\texttt{DSE}$ and a stable critical point of $\dot{x} = -\omega_{\mathcal{S}}(x)$.*

4. Convergence Analysis

In this section, we provide convergence guarantees for both the deterministic and stochastic settings. In the former, players have oracle access to their gradients at each step while in the latter, players are assumed to have an unbiased estimator of the gradient appearing in their update rule. Proofs of the deterministic results can be found in Appendix F and the stochastic results in Appendix G.

4.1. Deterministic Setting

Consider the deterministic Stackelberg update

$$x_{k+1} = x_k - \gamma_1 \omega_{\mathcal{S}_\tau}(x_k) \tag{4}$$

where $\omega_{\mathcal{S}_\tau}(x_k)$ is the m-dimensional vector with entries $D_1 f_1(x_k) - D_{21}^\top f_2(x_k)(D_2^2 f_2(x_k))^{-1} D_2 f_1(x_k) \in \mathbb{R}^{m_1}$ and $\tau D_2 f_2(x_k) \in \mathbb{R}^{m_2}$, and $\tau = \gamma_2/\gamma_1$ is the "timescale" separation. We refer to (4) as the τ-Stackelberg update. The Jacobian of $\omega_{\mathcal{S}_\tau}(x)$ is denoted $J_{\mathcal{S}_\tau}(x)$; it is equivalent to $J_{\mathcal{S}}$ with the $m_2 \times m$ block row multiplied by τ.

To get convergence guarantees, we apply well known results from discrete time dynamical systems. For a dynamical system $x_{k+1} = F(x_k)$, when the spectral radius $\rho(DF(x^*))$ of the Jacobian at fixed point is less than one, F is a contraction at x^* so that x^* is locally asymptotically stable (cf. Proposition F.1, Appendix F). In particular, $\rho(DF(x^*)) \le c < 1$ implies that $\|DF\| \le c + \varepsilon < 1$ for $\varepsilon > 0$ on a neighborhood of x^* (Ortega & Rheinboldt, 1970, 2.2.8). Hence, Proposition F.1 implies that if $\rho(DF(x^*)) = 1 - \kappa < 1$ for some κ, then there exists a ball $B_p(x^*)$ of radius $p > 0$ such that for any $x_0 \in B_p(x^*)$, and some constant $K > 0$, $\|x_k - x^*\|_2 \le K(1 - \frac{\kappa}{2})^k \|x_0 - x^*\|_2$ using $\varepsilon = \frac{\kappa}{4}$.

For a zero-sum setting defined by cost function $f \in C^q(X, \mathbb{R})$ with $q \ge 2$, recall that $\texttt{S}_1(J(x)) = D_1^2 f(x) - D_{21} f(x)^\top (D_2^2 f(x))^{-1} D_{21} f(x)$ is the first Schur complement of the Jacobian $J(x)$.

Theorem 3 (Zero-Sum Rate of Convergence.)**.** *Consider a zero-sum game defined by $f \in C^q(X, \mathbb{R})$ with $q \ge 2$. For a $\texttt{DSE}$ x^* with $\alpha = \min\{\lambda_{\min}(\texttt{S}_1(J(x^*))), \lambda_{\min}(-\tau D_2^2 f(x^*))\}$ and $\beta = \max\{\lambda_{\max}(\texttt{S}_1(J(x^*))), \lambda_{\max}(-\tau D_2^2 f(x^*))\}$ and learning rate $\gamma_1 = 1/(2\beta)$, the τ–Stackelberg update converges locally with a rate of $O((1 - \frac{\alpha}{4\beta})^k)$.*

Corollary 2 (Zero-Sum Finite Time Guarantee)**.** *Given $\varepsilon > 0$, under the assumptions of Theorem 3, τ-Stackelberg learning obtains an ε-$\texttt{DSE}$ in $\lceil \frac{4\beta}{\alpha} \log(\|x_0 - x^*\|/\varepsilon) \rceil$ iterations for any $x_0 \in B_\delta(x^*)$ with $\delta = \alpha/(4L\beta)$ where L is the local Lipschitz constant of $I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)$.*

The proofs leverage the structure of the Jacobian $J_{\mathcal{S}_\tau}$, which is lower block diagonal, along with the above noted result from dynamical systems theory. The key insight is that at a given x, the spectrum of $J_{\mathcal{S}_\tau}(x)$ is the union of the spectrum of $\texttt{S}_1(J(x))$ and $-\tau D_2^2 f(x)$ for zero-sum settings.

We now show a discrete-time analogue to Proposition 2.

Proposition 5. *Consider a zero-sum game defined by $f \in C^q(X, \mathbb{R})$, $q \ge 2$. Suppose that $\gamma_1 \le 1/L$ where $\max\{\text{spec}(\texttt{S}_1(J(x))) \cup \text{spec}(-\tau D_2^2 f(x))\} \le L$. Then, x is a stable critical point of τ–Stackelberg update if and only if x is a $\texttt{DSE}$.*

The next result shows that τ-Stackelberg avoids saddle points almost surely in general-sum games. We remark that $\texttt{DSE}$ are never saddle points in zero-sum games.

Theorem 4 (Almost Sure Avoidance of Saddles)**.** *Consider a general sum game defined by $f_i \in C^q(X, \mathbb{R})$, $q \ge 2$ for $i = 1, 2$ and where, without loss of generality, player 1 is the leader. Suppose that $\omega_{\mathcal{S}_\tau}$ is L-Lipschitz and that $\gamma_1 < 1/L$. The τ–Stackelberg learning dynamics converge to saddle points of $\dot{x} = -\omega_{\mathcal{S}_\tau}(x)$ on a set of measure zero.*

In the zero-sum setting, $\omega_{\mathcal{S}_\tau}$ being Lipschitz is equivalent to $\max\{\text{spec}(\texttt{S}_1(J(x))) \cup \text{spec}(-\tau D_2^2 f(x))\} \le L$. The only critical points of τ-Stackelberg learning in the zero-sum case are either saddles, unstable points, or $\texttt{DSE}$ which comprise all the stable critical points due to the structure of the Jacobian $J_{\mathcal{S}_\tau}$. Consequently, the previous pair of results imply that the only critical points τ-Stackelberg learning converges to in zero-sum games are $\texttt{DSE}$ almost surely.

We now provide a convergence guarantee for deterministic general-sum games. However, the convergence guarantee is no longer a global guarantee to the set of attractors of which critical points are $\texttt{DSE}$ since there is potentially stable critical points which are not $\texttt{DSE}$. This can be seen by examining the Jacobian which is no longer lower block triangular.

Given a critical point x^*, let $\alpha = \lambda_{\min}^2(\frac{1}{2}(J_{\mathcal{S}_\tau}^\top(x^*) + J_{\mathcal{S}_\tau}(x^*)))$ and $\beta = \lambda_{\max}(J_{\mathcal{S}_\tau}(x^*)^\top J_{\mathcal{S}_\tau}(x^*))$.

Theorem 5 (General Sum Rate of Convergence)**.** *Consider a general sum game (f_1, f_2) with $f_i \in C^q(X, \mathbb{R})$, $q \ge 2$ for $i = 1, 2$ and where, without loss of generality, player 1 is the leader. For a $\texttt{DSE}$ x^* such that $J_{\mathcal{S}_\tau}^\top(x^*) + J_{\mathcal{S}_\tau}(x^*) > 0$, the τ–Stackelberg update with learning rate $\gamma_1 = \sqrt{\alpha}/\beta$ converges locally with a rate of $O((1 - \frac{\alpha}{2\beta})^{k/2})$.*

Corollary 3 (General Sum Finite Time Guarantee). *Given $\varepsilon > 0$, under the assumptions of Theorem 5, τ–Stackelberg learning obtains an ε–DSE in $\lceil \frac{4\beta}{\alpha} \log(\|x_0 - x^*\|/\varepsilon) \rceil$ iterations for any $x_0 \in B_\delta(x^*)$ with $\delta = \alpha/(2L\beta)$ where L is the local Lipschitz constant of $I - \gamma_1 J_{\mathcal{S}_\tau}(x)$.*

4.2. Stochastic Setting

In the stochastic setting, players use updates of the form

$$x_{i,k+1} = x_{i,k} - \gamma_{i,k}(\omega_{\mathcal{S},i}(x_k) + w_{i,k+1}) \qquad (5)$$

where $\gamma_{1,k} = o(\gamma_{2,k})$ and $\{w_{i,k+1}\}$ is a stochastic process for each $i = 1, 2$. The results in this section assume the following. The maps $Df_1 : \mathbb{R}^m \to \mathbb{R}^{m_1}$, $D_2 f_2 : \mathbb{R}^m \to \mathbb{R}^{m_2}$ are Lipschitz, and $\|Df_1\| < \infty$. For each $i \in \{1, 2\}$, the learning rates satisfy $\sum_k \gamma_{i,k} = \infty$, $\sum_k \gamma_{i,k}^2 < \infty$. The noise processes $\{w_{i,k}\}$ are zero mean, martingale difference sequences: given the filtration $\mathcal{F}_k = \sigma(x_s, w_{1,s}, w_{2,s},\ s \le k)$, $\{w_{i,k}\}_{i\in\mathcal{I}}$ are conditionally independent, $\mathbb{E}[w_{i,k+1}|\ \mathcal{F}_k] = 0$ a.s., and $\mathbb{E}[\|w_{i,k+1}\||\ \mathcal{F}_k] \le c_i(1 + \|x_k\|)$ a.s. for some constants $c_i \ge 0$, $i \in \mathcal{I}$.

The primary technical machinery we use in this section is stochastic approximation theory (Borkar, 2008) and tools from dynamical systems. The convergence guarantees in this section are analogous to that for deterministic learning but asymptotic in nature. We first provide a non-convergence guarantee: the dynamics avoid saddle points in the stochastic learning regime.

Theorem 6 (Almost Sure Avoidance of Saddles.). *Consider a game (f_1, f_2) with $f_i \in C^q(\mathbb{R}^{m_1} \times \mathbb{R}^{m_2}, \mathbb{R})$, $q \ge 2$ for $i = 1, 2$ and where without loss of generality, player 1 is the leader. Suppose that for each $i = 1, 2$, there exists a constant $b_i > 0$ such that $\mathbb{E}[(w_{i,t} \cdot v)^+ | \mathcal{F}_{i,t}] \ge b_i$ for every unit vector $v \in \mathbb{R}^{m_i}$. Then, Stackelberg learning converges to strict saddle points of the game on a set of measure zero.*

We also give asymptotic convergence results. These results, combined with the non-convergence guarantee in Theorem 6, provide a broad convergence analysis for this class of learning dynamics. Theorem G.3 in Appendix G.3 provides a global convergence guarantee in general-sum games to the stable critical point, which may or may not be a DSE, under assumptions on the global asymptotic stability of critical points of the continuous time limiting singularly perturbed dynamical system. In zero-sum games, we know that the only critical points of the continuous time limiting system are DSE. Hence, Corollary G.2 in Appendix G.3 gives a global convergence guarantee in zero-sum games to the DSE under identical assumptions.

Relaxing these assumptions, the following proposition provides a local convergence result which ensures that sample points asymptotically converge to locally asymptotic trajectories of the continuous time limiting singularly perturbed system, and thus to stable DSE.

Theorem 7. *Consider a general sum game (f_1, f_2) with $f_i \in C^q(X, \mathbb{R})$, $q \ge 2$ for $i = 1, 2$ and where, without loss of generality, player 1 is the leader and $\gamma_{1,k} = o(\gamma_{2,k})$. Consider a differential Stackelberg equilibrium $x^* = (x_1^*, x_2^*)$. There exists a neighborhood $U = U_1 \times U_2$ of $x^* = (x_1^*, x_2^*)$ such that for any $x_0 \in U$, x_k converges almost surely to x^*.*

5. Experiments

We now present experiments showing the role of DSE in the optimization landscape of GANs and the empirical benefits of training GANs with Stackelberg learning compared to simultaneous gradient descent (`simgrad`). All detailed experiment information is given in Appendix H.

Example 1: Learning a Covariance Matrix. We consider a data generating process of $x \sim \mathcal{N}(0, \Sigma)$, where the covariance Σ is unknown and the objective is to learn it using a Wasserstein GAN. The discriminator is configured to be the set of quadratic functions defined as $D_W(x) = x^\top W x$ and the generator is a linear function of random input noise $z \sim \mathcal{N}(0, I)$ defined by $G_V(z) = Vz$. The matrices $W \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{m\times m}$ are the parameters of the discriminator and the generator, respectively. The Wasserstein GAN cost for the problem $f(V, W) = \sum_{i=1}^m \sum_{j=1}^m W_{ij}(\Sigma_{ij} - \sum_{k=1}^m V_{ik}V_{jk})$. We consider the generator to be the leader minimizing $f(V, W)$. The discriminator is the follower and it minimizes a regularized cost function defined by $-f(V, W) + \frac{\eta}{2}\operatorname{Tr}(W^\top W)$, where $\eta \ge 0$ is a tunable regularization parameter. The game is formally defined by the costs $(f_1, f_2) = (f(V, W), -f(V, W) + \frac{\eta}{2}\operatorname{Tr}(W^\top W))$, where player 1 is the leader and player 2 is the follower. In equilibrium, the generator picks V^* such that $V^*(V^*)^\top = \Sigma$ and the discriminator selects $W^* = 0$. Further details are given in Appendix C from Daskalakis et al. (2018).

We compare the deterministic gradient update for Stackelberg learning with simultaneous learning, and analyze the distance from equilibrium as a function of time. We plot $\|\Sigma - VV^\top\|_2$ for the generator's performance and $\|W + W^\top\|_2$ for the discriminator's performance in Fig. 2 for varying dimensions m with learning rates $\gamma_1 = \gamma_2/4 = 0.01$ and a fixed regularization of $\eta = 0.5$. The covariance matrix is chosen to be $\Sigma = UU^T + I$ where $U \sim \mathcal{N}(0, 1)$. We observe that Stackelberg learning converges to an equilibrium in fewer iterations. For zero-sum games, our theory provides reasoning for this behavior since at any critical point the eigenvalues of the game Jacobian are purely real. This is in contrast to simultaneous gradient descent, whose Jacobian can admit complex eigenvalues, known to cause rotational forces in the dynamics.

GAN training details. We now train GANs in which each player is parameterized by a neural network. The generator is always taken to be the leader and the discriminator

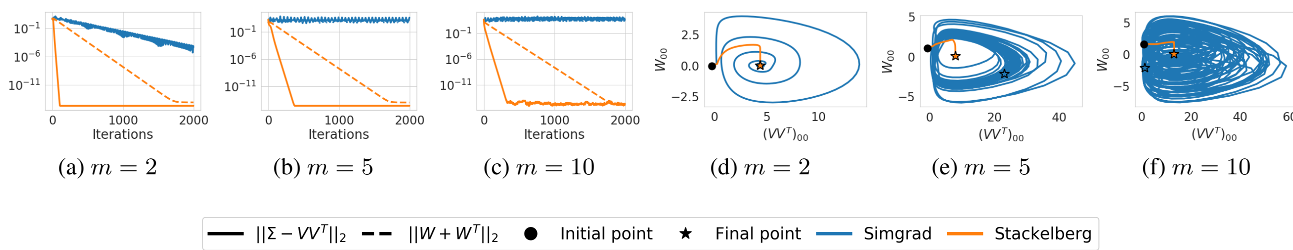


Figure 2. Stackelberg learning more effectively estimates covariance Σ as compared to `simgrad`. Errors given by $\|\Sigma - VV^T\|_2$ and $\|W + W^T\|_2$ are shown in (a)–(c) and trajectory plots of elements of W and VV^T in (d)–(f) showing the cycling of `simgrad`.

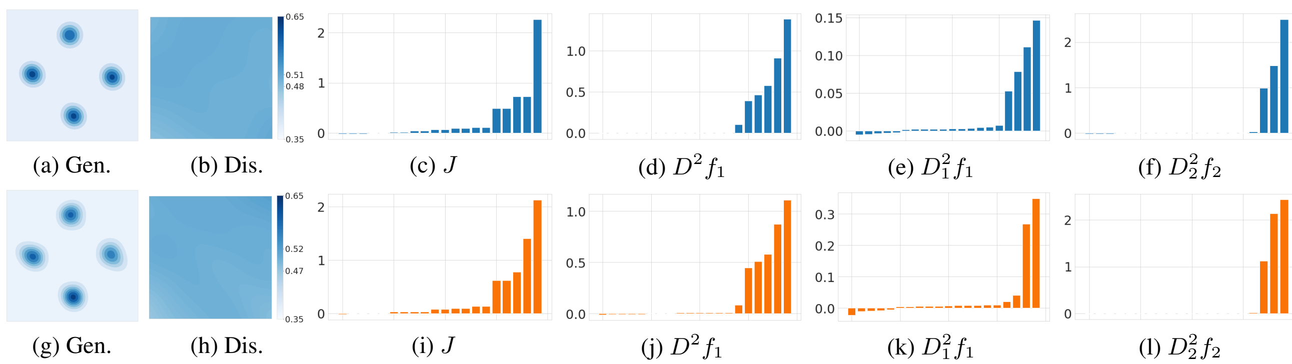


Figure 3. The generator and discriminator performances for `simgrad` and Stackelberg are shown in (a)–(b) and (g)–(h), respectively. We show the 5 smallest and 15 largest real eigenvalues parts of relevant game objects in (c)–(f) for `simgrad` and (i)–(l) for Stackelberg.

the follower in this set of experiments. Moreover, for both Stackelberg learning and simultaneous gradient descent we pass the gradient information of each player into the Adam optimizer (Kingma & Ba, 2015). To ensure the follower's Hessian is well-conditioned in the leader update, we regularize the implicit map of the follower so that the leader gradient is given by $\omega_{\mathcal{S},1} = D_1 f_1(x) + Dr_\eta(x)^\top D_2 f_1(x)$ where $Dr_\eta(x)^\top = -D_{21} f_2(x)^\top \left(D_2^2 f_2(x) + \eta I\right)^{-1}$ and η is the regularization parameter. We also employ regularization in the follower's implicit map when computing eigenvalues of $D^2 f_1(x)$ to determine whether an approximate critical point is in a neighborhood of a DSE. We provide details on the derivation of the regularized leader update along with a notion of a regularized DSE and specifics on the eigenvalue computation in Appendix H.1 and H.2.

Example 2: Learning a Mixture of Gaussians. We train a GAN to learn a mixture of Gaussian distribution. The generator and discriminator networks have two and one hidden layers, respectively; each hidden layer has 32 neurons. We train using a batch size of 256, a latent dimension of 16, with decaying learning rates. For both the diamond and circle configurations, 10 initial seeds were simulated for each set of learning dynamics and behavior was generally consistent across them for both algorithms. The experiments were run for 60,000 batches and the eigenvalues evaluated at that stopping point. We show detailed information for the best run of each algorithm in terms of KL-divergence and in Appendix H.4.1 examine all runs.

Diamond configuration. This experiment uses the saturating GAN objective and Tanh activations. In Fig. 3a–3b and Fig. 3g–3h we show a sample of the generator and the discriminator for `simgrad` and the Stackelberg dynamics at the end of training. Each learning rule converges so that the generator can create a distribution that is close to the ground truth and the discriminator is nearly at the optimal probability throughout the input space. In Fig. 3c–3f and Fig. 3i–3l, we show eigenvalues from the game that present a deeper view of the convergence behavior. We observe from the eigenvalues of J that both sets of dynamics converge to neighborhoods of points that are stable for the simultaneous dynamics and they appear to be in a neighborhood of a DSE since the eigenvalues of $D^2 f_1$ and $D_2^2 f_2$ are nearly all positive. Interestingly, however, since the eigenvalues of $D_1^2 f_1$ are nearly all zero and not all positive and this was consistent across the runs, it appears that the result may reflect the realizable assumption (cf. Sec. 3) as well as convergence to a DSE that is not a DNE. Given the good generator and discriminator performance, it is worth further empirical investigation to determine if DSE that are not DNE are desirable in GANs and if successful methods reach them.

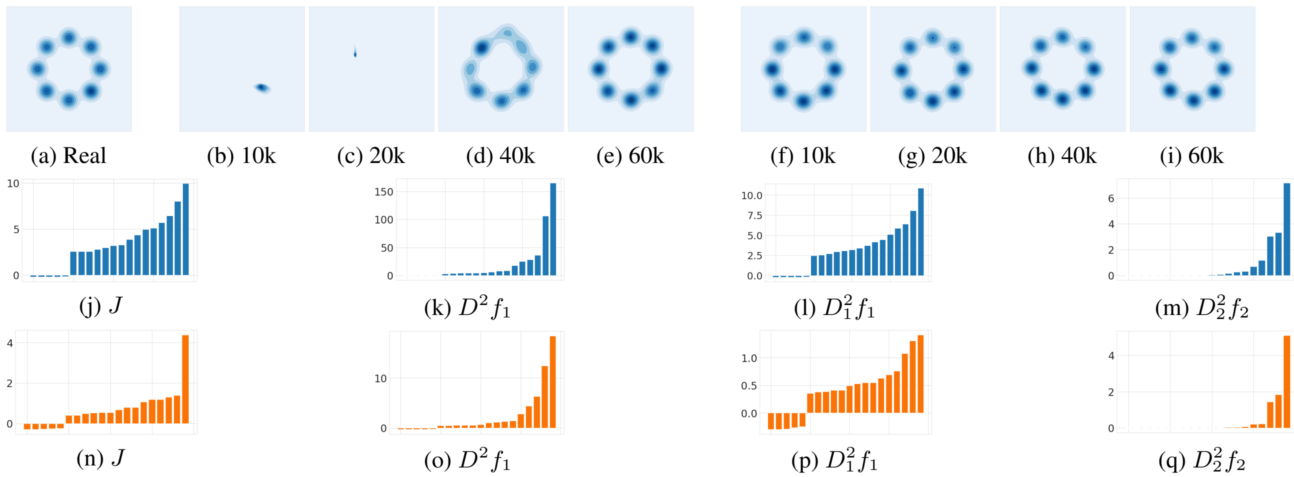
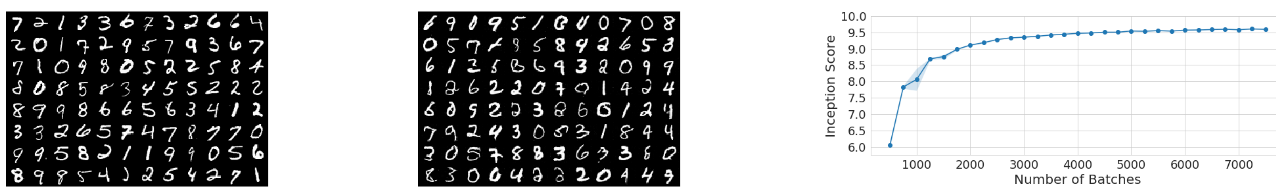


Figure 4. Stackelberg learning improves learning stability: `simgrad` generator in (b)–(e) and Stackelberg learning generator in (f)–(i). We show the 5 smallest and 15 largest real eigenvalue parts of relevant game objects in (j)–(m) for `simgrad` and (n)–(q) for Stackelberg.



(a) Real images. (b) Fake images. (c) Inception scores.

Figure 5. Stackelberg learning on the MNIST dataset.

Circle configuration. We demonstrate improved performance and stability when using Stackelberg learning dynamics in this example. We use ReLU activation functions and the non-saturating objective and show the performance in Fig. 4 along the learning path for the `simgrad` and Stackelberg learning dynamics. The former cycles and performs poorly until the learning rates have decayed enough to stabilize the training process. The latter converges quickly to a solution that nearly matches the ground truth distribution. We observed this behavior consistently across the runs. In a similar fashion as in the covariance example, the leader update is able to reduce rotations. We show the eigenvalues after training and see that for this configuration, `simgrad` converges to a neighborhood of a `DNE` and the Stackelberg dynamics converge again to the neighborhood of a `DSE` that is not a `DNE`. This provides further evidence that `DSE` may be easier to reach, and can provide suitable performance.

Example 3: MNIST GAN. To demonstrate that the Stackelberg learning dynamics can scale to high dimensional problems, we train a GAN on the MNIST dataset using the DCGAN architecture (Radford et al., 2015) adapted to handle 28×28 images. We simulate 10 random seeds and in Fig. 5c show the mean Inception score along the training process along with the standard error of the mean. The Inception score is calculated using a LeNet classifier following (Berard et al., 2020). We show a real sample in Fig. 5a and a fake sample in Fig. 5a after 7500 batches from the run with the fifth highest inception score. The Stackelberg learning dynamics are able to converge to a solution that generates realistic handwritten digits and get close to the maximum inception score in a stable manner. The primary purpose of this example is to show that the learning dynamics including second order information and an inverse is not an insurmountable problem for training with millions of parameters. We detail how the update can be computed efficiently using Jacobian-vector products and the conjugate gradient algorithm in Appendix H.2.

6. Conclusion

We study learning dynamics in Stackelberg games. This class of games pertains to any application in which there is an order of play. However, the problem has not been extensively analyzed in the way the learning dynamics of simultaneous play games have been. Consequently, we are able to give novel convergence results and draw connections to existing work focused on learning Nash equilibria.

Acknowledgements

This work was supported by NSF CAREER award 1844729, a DoD NDSEG Fellowship and the Computational Neuroscience Training Program (NIH 5T90DA032436-09).

References

Abraham, R., Marsden, J. E., and Ratiu, T. *Manifolds, Tensor Analysis, and Applications*. Springer, 2nd edition, 1988. doi: 10.1007/978-1-4612-1029-0.

Adolphs, L., Daneshmand, H., Lucchi, A., and Hofmann, T. Local saddle point optimization: A curvature exploitation approach. In *International Conference on Artificial Intelligence and Statistics*, pp. 486–495, 2019.

Argyros, I. K. A generalization of ostrowski's theorem on fixed points. *Applied Mathematics Letters*, 12:77–79, 1999.

Balduzzi, D., Racaniere, S., Martens, J., Foerster, J., Tuyls, K., and Graepel, T. The mechanics of n-player differentiable games. In *International Conference on Machine Learning*, pp. 354–363, 2018.

Basar, T. and Olsder, G. J. *Dynamic Noncooperative Game Theory*. Society for Industrial and Applied Mathematics, 2nd edition, 1998.

Benaïm, M. Dynamics of stochastic approximation algorithms. In *Seminaire de Probabilites XXXIII*, pp. 1–68, 1999.

Berard, H., Gidel, G., Almahairi, A., Vincent, P., and Lacoste-Julien, S. A closer look at the optimization landscapes of generative adversarial networks. In *International Conference on Learning Representations*, 2020.

Berger, T., Giribet, J., Pería, F. M., and Trunk, C. On a Class of Non-Hermitian Matrices with Positive Definite Schur Complements. *arXiv preprint arxiv:1807.08591*, 2018.

Bhatnagar, S., Prasad, H., and Prashanth, L. *Stochastic recursive algorithms for optimization: simultaneous perturbation methods*, volume 434. Springer, 2012.

Borkar, V. S. *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press, 2008.

Borkar, V. S. and Pattathil, S. Concentration bounds for two time scale stochastic approximation. In *Allerton Conference on Communication, Control, and Computing*, pp. 504–511. IEEE, 2018.

Broer, H. and Takens, F. Chapter 1 - preliminaries of dynamical systems theory. In Henk Broer, F. T. and Hasselblatt, B. (eds.), *Handbook of Dynamical Systems*, volume 3 of *Handbook of Dynamical Systems*, pp. 1 – 42. Elsevier Science, 2010.

Dai, B., Shaw, A., Li, L., Xiao, L., He, N., Liu, Z., Chen, J., and Song, L. Sbeed: Convergent reinforcement learning with nonlinear function approximation. In *International Conference on Machine Learning*, pp. 1125–1134, 2018.

Daneshmand, H., Kohler, J., Lucchi, A., and Hofmann, T. Escaping saddles with stochastic gradients. In *International Conference on Machine Learning*, pp. 1155–1164, 2018.

Danskin, J. M. The theory of max-min, with applications. *SIAM Journal on Applied Mathematics*, 14(4):641–664, 1966.

Danskin, J. M. *The Theory of Max-Min and its Application to Weapons Allocation Problems*. Springer, 1967.

Daskalakis, C. and Panageas, I. The limit points of (optimistic) gradient descent in min-max optimization. In *Advances in Neural Information Processing Systems*, pp. 9236–9246, 2018.

Daskalakis, C., Ilyas, A., Syrgkanis, V., and Zeng, H. Training gans with optimism. In *International Conference on Learning Representations*, 2018.

Foerster, J., Chen, R. Y., Al-Shedivat, M., Whiteson, S., Abbeel, P., and Mordatch, I. Learning with opponent-learning awareness. In *International Conference on Autonomous Agents and MultiAgent Systems*, pp. 122–130, 2018.

Fudenberg, D., Drew, F., Levine, D. K., and Levine, D. K. *The theory of learning in games*, volume 2. MIT press, 1998.

Gibson, C. G., Wirthmüller, K., du Plessis, A. A., and Looijenga, E. J. N. Topological stability of smooth mappings. In *Lecture Notes in Mathematics*, volume 552. Springer-Verlag, 1976.

Golubitsky, M. and Guillemin, V. *Stable Mappings and Their Singularities*. Springer-Verlag, 1973.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pp. 2672–2680, 2014.

Hirsch, M. W. *Differential topology*. Springer New York, 1976.

Hirsch, M. W. *Differential Topology*, volume 33. Springer Science & Business Media, 2012.

Horn, R. and Johnson, C. *Topics in Matrix Analysis*. Cambridge University Press, 2011.

Jin, C., Netrapalli, P., and Jordan, M. I. What is local optimality in nonconvex-nonconcave minimax optimization? *arXiv preprint arXiv:1902.00618*, 2019.

Kelley, J. *General Topology*. Van Nostrand Reinhold Company, 1955.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

Knutson, A. and Tao, T. Honeycombs and sums of hermitian matrices. *Notices of the American Mathematical Society*, 2001.

Kushner, H. J. and Yin, G. G. *Stochastic approximation and recursive algorithms and applications*, volume 35. Springer Science & Business Media, 2003.

Lee, J. *Introduction to smooth manifolds*. Springer, 2012.

Lee, J. D., Simchowitz, M., Jordan, M. I., and Recht, B. Gradient descent only converges to minimizers. In *Conference on Learning Theory*, pp. 1246–1257, 2016.

Letcher, A., Foerster, J., Balduzzi, D., Rocktäschel, T., and Whiteson, S. Stable opponent shaping in differentiable games. In *International Conference on Learning Representations*, 2019.

Maclaurin, D., Duvenaud, D., and Adams, R. Gradient-based hyperparameter optimization through reversible learning. In *International Conference on Machine Learning*, pp. 2113–2122, 2015.

Madry, A., Makelov, A., Schmidt, L., Tsipras, D., and Vladu, A. Towards deep learning models resistant to adversarial attacks. In *International Conference on Learning Representations*, 2018.

Mazumdar, E. and Ratliff, L. J. Local nash equilibria are isolated, strict local nash equilibria in 'almost all' zero-sum continuous games. In *IEEE Conference on Decision and Control*, 2019.

Mazumdar, E., Jordan, M., and Sastry, S. S. On finding local nash equilibria (and only local nash equilibria) in zero-sum games. *arxiv:1901.00838*, 2019.

Mazumdar, E., Ratliff, L. J., and Sastry, S. S. On gradient-based learning in continuous games. *SIAM Journal on Mathematics of Data Science*, 2(1):103–131, 2020.

Mertikopoulos, P., Lecouat, B., Zenati, H., Foo, C.-S., Chandrasekhar, V., and Piliouras, G. Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile. In *International Conference on Learning Representations*, 2019.

Mescheder, L., Nowozin, S., and Geiger, A. The numerics of gans. In *Advances in Neural Information Processing Systems*, pp. 1825–1835, 2017.

Metz, L., Poole, B., Pfau, D., and Sohl-Dickstein, J. Unrolled generative adversarial networks. In *International Conference on Learning Representations*, 2017.

Nagarajan, V. and Kolter, J. Z. Gradient descent gan optimization is locally stable. In *Advances in Neural Information Processing Systems*, pp. 5585–5595, 2017.

Ortega, J. M. and Rheinboldt, W. C. *Iterative Solutions to Nonlinear Equations in Several Variables*. Academic Press, 1970.

Panageas, I. and Piliouras, G. Gradient Descent Only Converges to Minimizers: Non-Isolated Critical Points and Invariant Regions. In *Innovations in Theoretical Computer Science*, 2017.

Pemantle, R. Nonconvergence to unstable points in urn models and stochastic approximations. *Annuals Probability*, 18(2):698–712, 04 1990.

Radford, A., Metz, L., and Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

Ratliff, L. J., Burden, S. A., and Sastry, S. S. Genericity and structural stability of non-degenerate differential nash equilibria. In *American Control Conference*, pp. 3990–3995. IEEE, 2014.

Ratliff, L. J., Burden, S. A., and Sastry, S. S. On the Characterization of Local Nash Equilibria in Continuous Games. *IEEE Transactions on Automatic Control*, 61(8):2301–2307, 2016.

Sastry, S. S. *Nonlinear Systems Theory*. Springer, 1999.

Schäfer, F. and Anandkumar, A. Competitive gradient descent. In *Advances in Neural Information Processing Systems*, pp. 7623–7633, 2019.

Shub, M. *Global Stability of Dynamical Systems*. Springer-Verlag, 1978.

Thoppe, G. and Borkar, V. A concentration bound for stochastic approximation via alekseev's formula. *Stochastic Systems*, 9(1):1–26, 2019.

Tretter, C. *Spectral Theory of Block Operator Matrices and Applications*. Imperial College Press, 2008.

Von Stackelberg, H. *Market structure and equilibrium*. Springer Science & Business Media, 2010.

Zaslavski, A. J. Necessary optimality conditions for bilevel minimization problems. *Nonlinear Analysis: Theory, Methods & Applications*, 75(3):1655–1678, 2012.

Zhang, C. and Lesser, V. Multi-agent learning with policy prediction. In *AAAI Conference on Artificial Intelligence*, 2010.

Zhang, K., Yang, Z., and Başar, T. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *arXiv preprint arXiv:1911.10635*, 2019.

A. Guide to the Appendix

Appendix B. Proof of the sufficient conditions for differential Stackelberg equilibria in terms of the Schur complement given in Proposition 1.

Appendix C. Proofs of the genericity and structural stability of differential Stackelberg equilibria in zero-sum games from Section 2.3.

Appendix D. Proofs of the results on the connections between the limit points of simultaneous play learning and Stackelberg play learning along with Nash and Stackelberg equilibria in zero-sum games from Section 3.

Appendix E. Necessary and sufficient conditions for stable critical points of the simultaneous play dynamics $\dot{x} = -\omega(x)$ such that $D_1^2 f(x) \not> 0$ and $-D_2^2 f(x) > 0$ to be differential Stackelberg equilibria in zero-sum games.

Appendix F. Proofs of the deterministic convergence results for Stackelberg learning from Section 4.1.

Appendix G. Proofs of the stochastic convergence results from Section 4.2. This appendix includes a number of extensions not included in the paper. In particular, we provide convergence guarantees for the leader assuming the follower plays an exact best response, and extended analysis for the case where the follower is performing individual gradient updates.

Appendix H Further details on the numerical experiments and supplemental experiments beyond that included in the paper.

B. Proof of Sufficient Conditions for Differential Stackelberg in Terms of Schur Complement

Proposition 1. *Consider a game (f_1, f_2) defined by $f_i \in C^q(X, \mathbb{R})$, $i = 1, 2$ with $q \geq 2$ and player 1 (without loss of generality) taken to be the leader. Let x^* satisfy $D_2 f_2(x^*) = 0$ and $D_2^2 f_2(x^*) > 0$. Then $Df_1(x^*) = 0$ and $\mathtt{S}_1(J_{\mathcal{S}}(x^*)) > 0$ if and only if x^* is a differential Stackelberg equilibrium. Moreover, in zero-sum games, $\mathtt{S}_1(J_{\mathcal{S}}(x)) = \mathtt{S}_1(J(x))$.*

Proof. The implicit function theorem implies that there exists neighborhoods U_1 of x_1^* and W of $D_2 f_2(x_1^*, x_2^*)$ and a unique C^q mapping $r : U_1 \times W \to \mathbb{R}^{m_2}$ on which $D_2 f_2(x_1, r(x_1)) = 0$. The first Schur complement of $J_{\mathcal{S}}$ is

$$\mathtt{S}_1(J_{\mathcal{S}}(x)) = D_1(Df_1(x_1, x_2)) - D_2(Df_1(x_1, x_2))(D_2^2 f_2(x_1, x_2))^{-1} D_{21} f_2(x_1, x_2)$$

where

$$D_1(Df_1(x_1, x_2)) = D_1^2 f_1(x_1, x_2) + D_{12} f_1(x_1, x_2) Dr(x_1) + D_2 f_1(x_1, x_2) D^2 r(x_1)^{3}$$

and

$$D_2(Df_1(x_1, x_2)) = D_{12} f_1(x_1, x_2) + Dr(x_1)^\top D_2^2 f_1(x_1, x_2).$$

Now, we also have that the total derivative of $Df_1(x_1, r(x_1))$ is given by

$$\begin{aligned} D(Df_1(x_1, r(x_1))) = D_1^2 f_1(x_1, r(x_1)) + D_{12} f_1(x_1, r(x_1)) Dr(x_1) + D_2 f_1(x_1, r(x_1))^\top D^2 r(x_1) \\ + (D_{12} f_1(x_1, r(x_1)) + Dr(x_1)^\top D_2^2 f_1(x_1, r(x_1))) Dr(x_1) \end{aligned}$$

Note also that by the implicit function theorem, $Dr(x_1) = -(D_2^2 f_2(x_1, x_2))^{-1} D_{21} f_2(x_1, x_2)|_{x_2 = r(x_1)}$. Hence, we have that $D(Df_1(x_1, r(x_1))) = \mathtt{S}_1(J_{\mathcal{S}}(x))|_{x_2 = r(x_1)}$. □

C. Structural Stability and Genericity in Zero-Sum Continuous Stackelberg Games

This section is dedicated to showing that for a generic zero-sum q-differentiable game, all local Stackelberg equilibria of the game are differential Stackelberg equilibria, and further they are structurally stable. Analogous results are known for differential Nash equilibria (Ratliff et al., 2014; Mazumdar & Ratliff, 2019).

We first show that all differential Stackelebrg equilibria are non-degenerate—that is, $\det(J_{\mathcal{S}}(x)) \neq 0$ for any differential Stackelberg equilibrium x.

Proposition C.1. *In zero-sum q-differentiable continuous games, all differential Stackelberg equilibria are non-degenerate, hyperbolic critical points of the vector field $\omega_{\mathcal{S}}(x)$.*

[3] Note that $D^2 r(x_1)$ denotes the appropriately dimensioned tensor so that $D_2 f_1(x_1, x_2) D^2 r(x_1)$ is $m_1 \times m_1$ dimensional. In particular, $D_2 f_1(x_1, x_2)$ is a $1 \times m_2$ dimensional row vector and we take $D^2 r(x_1)$ to be a $m_2 \times m_1 \times m_1$ dimensional tensor so that $D_2 f_1(x_1, x_2) D^2 r(x_1)$ means in Einstein summation notation $(D_2 f_1(x))_i (D^2 r(x_1))_{ijk}$.

Proof. The Jacobian of the vector field $\omega_{\mathcal{S}}(x)$ given in (1) is lower block triangular in zero-sum games at a differential Stackelberg equilibria since $D_2(Df_1(x)) = 0$. Hence, $\det(J_{\mathcal{S}}(x)) \neq 0$ if and only if $\det(D_1(Df(x))) \neq 0$ and $\det(-D_2^2 f(x)) \neq 0$. Since differential Stackelberg are such that $D_1(Df(x)) > 0$ and $-D_2^2 f(x) > 0$, the fact that all differential Stackelberg are non-degenerate follows immediately. Further, the lower block triangular structure implies that $\mathrm{spec}(J_{\mathcal{S}}(x)) = \mathrm{spec}(\mathtt{S}_1(J(x))) \cup \mathrm{spec}(-D_2^2 f(x))$. Hence, all differential Stackelberg equilibria are hyperbolic. $\square$

For a zero-sum q-differentiable game $\mathcal{G} = (f, -f)$, if we let $\mathtt{DSE}(\mathcal{G})$ be the differential Stackelberg equilibria, $\mathtt{NDSE}(\mathcal{G})$ the non-degenerate differential Stackelberg equilibria, and $\mathtt{LSE}(\mathcal{G})$ the local Stackelberg equilibria, then we know that

$$\mathtt{NDSE}(\mathcal{G}) = \mathtt{DSE}(\mathcal{G}) \subseteq \mathtt{LSE}(\mathcal{G})$$

where the first equality is a consequence of Proposition C.1. What we show in the remainder is that for generic $f \in C^q(X, \mathbb{R})$, the game $\mathcal{G} = (f, -f)$ is such that

$$\mathtt{LSE}(\mathcal{G}) = \mathtt{DSE}(\mathcal{G}).$$

In particular, we show that the set of zero-sum q-differentiable games admitting any local Stackelberg equilibria which are degenerate differential Stackelberg equilibria is of measure zero in $C^q(\mathbb{R}^m, \mathbb{R})$.

C.1. Mathematical Preliminaries

In this appendix, we provide some additional mathematical preliminaries; the interested reader should see standard references for a more detailed introduction (Lee, 2012; Abraham et al., 1988).

A *smooth manifold* is a topological manifold with a smooth atlas. Euclidean space, as considered in this paper, is a smooth manifold. For a vector space E, we define the vector space of continuous $(p+s)$–multilinear maps $T_s^p(E) = L^{p+s}(E^*, \dots, E^*, E, \dots, E; \mathbb{R})$ with s copies of E and p copes of E^* and where E^* denotes the dual. Elements of $T_s^p(E)$ are *tensors* on E, and $T_s^p(X)$ denotes the vector bundle of such tensors (Abraham et al., 1988, Definition 5.2.9).

Consider smooth manifolds X and Y of dimension n_x and n_y respectively. An *k–jet* from X to Y is an equivalence class $[x, f, U]_k$ of triples (x, f, U) where $U \subset X$ is an open set, $x \in U$, and $f : U \to Y$ is a C^k map. The equivalence relation satisfies $[x, f, U]_k = [y, g, V]_k$ if $x = y$ and in some pair of charts adapted to f at x, f and g have the same derivatives up to order k. We use the notation $[x, f, U]_k = j^k f(x)$ to denote the k–jet of f at x. The set of all k–jets from X to Y is denoted by $J^k(X, Y)$. The jet bundle $J^k(X, Y)$ is a smooth manifold (see Hirsch, 1976, Chapter 2 for the construction). For each C^k map $f : X \to Y$ we define a map $j^k f : X \to J^k(X, Y)$ by $x \mapsto j^k f(x)$ and refer to it as the *k–jet extension*.

Definition C.1. *Let X, Y be smooth manifolds and $f : X \to Y$ be a smooth mapping. Let Z be a smooth submanifold of Y and u a point in X. Then f* intersects Z transversally at u *(denoted $f \pitchfork Z$ at u) if either $f(u) \notin Z$ or $f(u) \in Z$ and $T_{f(u)}Y = T_{f(u)}Z + (f_*)_u(T_u X)$.*

For $1 \leq k < s \leq \infty$ consider the jet map $j^k : C^s(X, Y) \to C^{s-k}(X, J^k(X, Y))$ and let $Z \subset J^k(X, Y)$ be a submanifold. Define

$$\textstyle\bigcap\!\!\!\!\bigcap^s(X, Y; j^k, Z) = \{h \in C^s(X, Y) |\ j^k h \pitchfork Z\}. \tag{6}$$

A subset of a topological space X is *residual* if it contains the intersection of countably many open–dense sets. We say a property is *generic* if the set of all points of X which possess this property is residual (Broer & Takens, 2010).

Theorem C.1. *(Jet Transversality Theorem, Hirsch, 1976, Chapter 2). Let X, Y be C^∞ manifolds without boundary, and let $Z \subset J^k(X, Y)$ be a C^∞ submanifold. Suppose that $1 \leq k < s \leq \infty$. Then, $\bigcap\!\!\!\!\bigcap^s(X, Y; j^k, Z)$ is residual and thus dense in $C^s(X, Y)$ endowed with the strong topology, and open if Z is closed.*

Proposition C.2. *(Golubitsky & Guillemin, 1973, Chapter II.4, Proposition 4.2). Let X, Y be smooth manifolds and $Z \subset Y$ a submanifold. Suppose that $\dim X < \mathrm{codim} Z$. Let $f : X \to Y$ be smooth and suppose that $f \pitchfork Z$. Then, $f(X) \cap Z = \emptyset$.*

C.2. Genericity

We show that local Stackelberg equilibria of zero-sum games are generically non-degenerate differential Stackelberg equilibria. Towards this end, we utilize the well-known fact that non-degeneracy of critical points is a generic property of sufficiently smooth functions.

Lemma C.1 (Broer & Takens, 2010, Chapter 1)**.** *For $C^q(\mathbb{R}^m, \mathbb{R})$ functions with $q \geq 2$ it is a generic property that all the critical points are non-degenerate.*

The above lemma implies that for a generic function $f \in C^q(\mathbb{R}^m, \mathbb{R})$, the Hessian

$$H(x) = \begin{bmatrix} D_1^2 f(x) & \cdots & D_{1m} f(x) \\ \vdots & \ddots & \vdots \\ D_{m1} f(x) & \cdots & D_m^2 f(x) \end{bmatrix}$$

is non-degenerate at critical points—that is, $\det(H(x)) \neq 0$. For a generic function f, we know that $\det(D_i^2 f(x)) \neq 0$ (cf. Lemma C.3). The simultaneous learning dynamics game Jacobian in zero-sum games is given by

$$J(x) = \begin{bmatrix} D_1^2 f(x) & D_{12} f(x) \\ -D_{21} f(x) & -D_2^2 f(x) \end{bmatrix}$$

where $J(x)$ is obtained by taking the Jacobian of the vector field $(D_1 f(x), -D_2 f(x))$. Moreover the Schur complement of $J(x)$ is given by

$$\mathtt{S}_1(J(x)) = D_1^2 f(x) - D_{21}^\top f(x)(D_2^2 f(x))^{-1} D_{21} f(x).$$

Furthermore, the Stackelberg game Jacobian in zero-sum games is given by

$$J_{\mathcal{S}}(x) = \begin{bmatrix} D_1(Df(x)) & D_2(Df(x)) \\ -D_{21} f(x) & -D_2^2 f(x) \end{bmatrix}.$$

where $J_{\mathcal{S}}(x)$ is obtained by taking the Jacobian of the vector field $(Df(x), -D_2 f(x))$. At critical points defined by x such that $\omega_{\mathcal{S}}(x) = (Df(x), -D_2 f(x)) = 0$ we note that

$$J_{\mathcal{S}}(x) = \begin{bmatrix} \mathrm{S}_1(J(x)) & 0 \\ -D_{21} f(x) & -D_2^2 f(x) \end{bmatrix}.$$

The structure is lower block triangular since

$$D_2(Df(x)) = D_{12} f(x) + Dr(x_1)^\top D_2^2 f(x) = D_{12} f(x) - D_{12} f(x)(D_2^2 f(x))^{-1} D_2^2 f(x) = 0.$$

Lemma C.2. *Consider $f \in C^q(\mathbb{R}^m, \mathbb{R})$, $q \geq 2$ and the corresponding zero-sum game $\mathcal{G} = (f, -f)$. For any critical point defined by $x \in \mathbb{R}^m$ such that $\omega_{\mathcal{S}}(x) = 0$, $\det(H(x)) \neq 0 \iff \det(J(x)) \neq 0$ and, if $\det(-D_2^2 f(x)) \neq 0$, then $\det(H(x)) \neq 0 \iff \det(J_{\mathcal{S}}(x)) \neq 0$.*

Proof. Consider a fixed $x = (x_1, x_2) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$. Note that $H(x)$ is equal to $J(x)$ with the last m_2 rows scaled each by -1. Hence, $J(x) = PH(x)$ where $P = \text{blockdiag}(I_{m_1}, -I_{m_2})$ with each I_{m_i} the $m_i \times m_i$ identity matrix, so that $\det(H(x)) = (-1)^{m_2} \det(J(x))$, which in turn proves the first equivalence. For the second equivalence, suppose that $\det(-D_2^2 f(x)) \neq 0$ so that $J_{\mathcal{S}}(x)$ is well-defined. Then, using the Schur decomposition of $J(x)$ it is easily seen that $\det(J(x)) = \det(\mathtt{S}_1(J(x))) \det(-D_2^2 f(x)) = \det(J_{\mathcal{S}}(x))$ where the last equality holds since $J_{\mathcal{S}}(x)$ is a lower block triangular matrix with $\mathtt{S}_1(J(x))$ and $-D_2^2 f(x)$ on the diagonal at critical points. Hence, the result. $\square$

This equivalence between the non-degeneracy of the Hessian and the game Jacobian J (and the relationship to the determinant of $J_{\mathcal{S}}$ via the Schur decomposition) allows us to lift the generic property of non-degeneracy of critical points of functions to critical points of the Stackelberg learning dynamics.

The Jet Transversality Theorem and Proposition C.2 can be used to show a subset of a jet bundle having a particular set of desired properties is generic. Indeed, consider the jet bundle $J^k(X, Y)$ and recall that it is a manifold that contains jets $j^k f : X \to J^k(X, Y)$ as its elements where $f \in C^k(X, Y)$. Let $Z \subset J^k(X, Y)$ be the submanifold of the jet bundle that *does not* possess the desired properties. If $\dim X < \operatorname{codim} Z$, then for a generic function $f \in C^k(X, Y)$ the image of the k–jet extension is disjoint from Z implying that there is an open–dense set of functions having the desired properties.

Without loss of generality, we let player 1 be the leader.

Lemma C.3. *Consider $f \in C^q(\mathbb{R}^{m_1+m_2}, \mathbb{R})$, $q \geq 2$ such that $D_i^2 f \in \mathbb{R}^{m_i \times m_i}$. It is a generic property that $\det(D_i^2 f(x)) \neq 0$ for any $i = 1, 2$.*

Proof. Let us start with $f \in C^q$ with $q \geq 3$. First, critical points of a function f are such that $D_i f(x) = 0$, $i = 1, 2$. Furthermore, the $J^2(\mathbb{R}^m, \mathbb{R})$ bundle associated to f is diffeomorphic to $\mathbb{R}^m \times \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}^{m(m+1)/2}$ and the 2-jet extension of f at any point $x \in \mathbb{R}^m$ is given by $(x, f(x), Df(x), D^2 f(x))$.

Now, let us denote by $S(k)$ the space of $k \times k$ symmetric matrices, and consider the subset of $J^2(\mathbb{R}^m, \mathbb{R})$ defined by

$$\mathcal{D}_i = \mathbb{R}^m \times \mathbb{R} \times \{0\} \times Z_i(m_i) \times \mathbb{R}^{m_1 \times m_2} \times S(m - m_i)$$

where $Z(m_i) = \{A \in S(m_i) | \det(A) = 0\}$. The set $Z(m_i)$ is algebraic; hence, we can use the Whitney stratification theorem (Gibson et al., 1976, Ch. 1, Thm. 2.7) to get that each $Z(m_i)$ is the union of submanifolds of co-dimension at least 1. Hence, it is the union of sub-manifolds of codimension at least one and, in turn, $\mathcal{D}_i$ is the union of sub-manifolds of codimension at least $m + 1$. Thus, it follows from the Jet Transversality Theorem C.1 (by way of Proposition C.2 since $m + 1 > m$) that for a generic function f, the image of the 2-jet extension $j^2 f$ is disjoint from $\mathcal{D}_i$. Hence, for such an f, for each x that is a critical point, the Hessian of f is such that $\det(D_i^2 f(x)) \neq 0$.

Furthermore, if we consider the subset $\mathcal{D} \subset J^2(\mathbb{R}^m, \mathbb{R})$ defined by

$$\mathcal{D} = \mathbb{R}^m \times \mathbb{R} \times \{0\} \times Z(m_i) \times \mathbb{R}^{m_1 \times m_2} \times Z(m - m_i),$$

then both $Z(m_i)$ and $Z(m - m_i)$ are algebraic, and so they each are of co-dimension at least one. In turn, $\mathcal{D}$ is the the union of sub-manifolds of codimension at least $m + 2$. Applying the Jet Transversality Theorem C.1 again, we get that for such an f, for each x that is a critical point, the Hessian of f is such that $\det(D_i^2 f(x)) \neq 0$ for $i \in \{1, 2\}$.

The extension to the $q \geq 2$ setting follows directly from the fact that non-degeneracy is an open condition in the C^2 topology, and any function can be C^2 approximated by a C^3 function (see, e.g., Hirsch, 2012, Thm. 2.4), which can then be approximated by a function without critical points such that $\det(D_i^2 f(x)) = 0$ (by the above argument), which we will call coordinate degenerate. This, in turn, implies that functions without critical points that are coordinate-degenerate are dense in the C^2 space of functions. $\square$

While the theorems we leverage from differential geometry and dynamical systems theory are similar, the architecture of the proof of the following theorem deviates quite a bit from (Ratliff et al., 2014; Mazumdar & Ratliff, 2019) due to the heirarchical structure of the game.

Theorem 1. *For the class of two-player, zero-sum continuous games $(f, -f)$ where $f \in C^q(\mathbb{R}^m, \mathbb{R})$ with $q \geq 2$, differential Stackelberg are generic amongst local Stackelberg. That is, given a generic $f \in C^q(\mathbb{R}^m, \mathbb{R})$, all local Stackelberg of the game $(f, -f)$ are differential Stackelberg.*

Proof. Let $J^2(\mathbb{R}^m, \mathbb{R})$ denote the second-order jet bundle containing 2–jets $j^2 f$ such that $f : \mathbb{R}^m \to \mathbb{R}$. Then, $J^2(\mathbb{R}^m, \mathbb{R})$ is locally diffeomorphic to $\mathbb{R}^m \times \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}^{\frac{m(m+1)}{2}}$ and the 2–jet extension of f at any point $x \in \mathbb{R}^m$ is given by $(x, f(x), Df(x), D^2 f(x))$.

By Lemma C.3, we know that

$$\mathcal{D}_2 = \mathbb{R}^m \times \mathbb{R} \times \{0\} \times Z(m_2) \times \mathbb{R}^{m_1 + m_2} \times S(m_1)$$

has co-dimension at least $m + 1$ in $J^2(\mathbb{R}^m, \mathbb{R})$ so that there exists an open dense set of functions $\mathcal{F}_2 \subset C^q(\mathbb{R}^m, \mathbb{R})$ such that $\det(D_2^2 f(x)) \neq 0$ at critical points (i.e., where $(D_1 f(x), D_2 f(x)) = (0, 0)$).

Now, we also know that there is an open dense set of functions $\mathcal{F}_1$ in $C^q(\mathbb{R}^m, \mathbb{R})$ such that $\det(H(x)) \neq 0$ at critical points. The intersection of open dense sets is open dense. Let $\mathcal{F} = \mathcal{F}_1 \cap \mathcal{F}_2$. Now, for any $f \in \mathcal{F}$, we have that at critical points $\det(H(x)) \neq 0$ and $\det(D_2^2 f(x)) \neq 0$. Hence, by Lemma C.2, $\det(J_{\mathcal{S}}(x)) \neq 0$ for all $f \in \mathcal{F}$, and in particular, $\det(S_1(x)) \neq 0$.

For all functions $f \in \mathcal{F}$, the critical points of $\omega_{\mathcal{S}}(x) = (Df(x), -D_2 f(x))$ coincide with the critical points of the function f. Indeed,

$$(D_1 f(x), D_2 f(x)) = (0, 0) \iff (Df(x), -D_2 f(x)) = (0, 0)$$

since for all $f \in \mathcal{F}$, $\det(D_2^2 f(x)) \neq 0$ and $D_2 f(x) = 0$ so that the C^q implicit map at a critical point $D_2 f(x) = 0$ is well-defined.

Thus, we have constructed an open dense set $\mathcal{F} \subset C^q(\mathbb{R}^m, \mathbb{R})$ such that for all $f \in \mathcal{F}$, if $x \in \mathbb{R}^m$ is a local Stackelberg equilibrium for $(f, -f)$, then x is a non-degenerate differential Stackelberg equilibrium. Indeed, suppose $f \in \mathcal{F}$ and $x \in \mathbb{R}^m$ is a local Stackelberg equilibrium. Then, a necessary condition is that $-D_2 f(x) = 0$ and $-D_2^2 f(x) \geq 0$. However, since $f \in \mathcal{F}$, we have that $\det(-D_2^2 f(x)) = (-1)^{m_2} \det(D_2^2 f(x)) \neq 0$ so that, in fact, $-D_2^2 f(x) > 0$. Hence, a local Stackelberg equilibrium such that $-D_2 f(x) = 0$ and $-D_2^2 f(x) > 0$ necessarily satisfies $\omega_{\mathcal{S}}(x) = 0$ and $S_1(x) \geq 0$. However, again since $f \in \mathcal{F}$, and $\det(S_1(x)) \neq 0$ so that, in fact, $S_1(x) > 0$. Furthermore, due to the lower block triangular structure of $J_{\mathcal{S}}(x)$, $\det(-D_2^2 f(x)) = (-1)^{m_2} \det(D_2^2 f(x)) \neq 0$ and $\det(S_1(x)) \neq 0$ also imply that $\det(J_{\mathcal{S}}(x)) \neq 0$, which completes the proof. $\square$

Necessary conditions[4] for a local Stackelberg solution for the leader follow from necessary conditions in nonlinear optimization and can be used in conjunction with Theorem 1 to show that local Stackelberg equilibria are generically hyperbolic—that is, for any local Stackelberg equilibrium x, there are no eigenvalues of $J_{\mathcal{S}}(x)$ with zero real part.

Lemma C.4 (Necessary Conditions for Local Stackelberg)**.** *Consider a game (f_1, f_2) defined by $f_i \in C^q(X, \mathbb{R})$, $i = 1, 2$ with $q \geq 2$ and player 1 (without loss of generality) taken to be the leader. Suppose that $x^* = (x_1^*, x_2^*)$ is a local Stackelberg equilibrium such that the follower (player 2) is at a strict local minimum so that $D_2 f_2(x_1^*, x_2^*) = 0$ and $D_2^2 f_2(x_1^*, x_2^*) > 0$. Then, $Df_1(x_1^*, x_2^*) = 0$ and $D^2 f_1(x_1^*, x_2^*) \geq 0$.*

Proof. The proof is straightforward. Indeed, suppose that given x_1^*, x_2^* is a strict local minimum for the follower. By the implicit function theorem (Abraham et al., 1988, Thm. 2.5.7), there exists a C^q map $r : x_1 \mapsto x_2$ defined on a neighborhood of x_1^* such that $r(x_1^*) = x_2^*$ and $Dr \equiv -(D_2^2 f_2)^{-1} \circ D_{21} f_2$. Hence, necessary conditions for the leader reduce to necessary conditions on the problem $\min_{x_1} f_1(x_1, r(x_1))$. $\square$

The following corollary is a direct consequence of Theorem 1 and Lemma C.4.

Corollary 1. *For the class of two-player, zero-sum continuous games $(f, -f)$ where $f \in C^q(\mathbb{R}^m, \mathbb{R})$ with $q \geq 2$, local Stackelberg equilibria are generically non-degenerate, hyperbolic critical points of the vector field $\omega_{\mathcal{S}}(x)$.*

C.3. Structural Stability

We now show that (non-degenerate) differential Stackelberg are structurally stable, meaning that they persist under smooth perturbations within the class of zero-sum games.

Theorem C.2. *For zero-sum games, differential Stackelberg equilibria are structurally stable: given $f \in C^r(\mathbb{R}^{m_1} \times \mathbb{R}^{m_2}, \mathbb{R})$, $\zeta \in C^r(\mathbb{R}^{m_1} \times \mathbb{R}^{m_2}, \mathbb{R})$, and a differential Stackelberg equilibrium $(x_1, x_2) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$, there exists neighborhoods $U \subset \mathbb{R}$ of zero and $V \subset \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$ such that for all $t \in U$ there exists a unique differential Stackelberg equilibrium $(\tilde{x}_1, \tilde{x}_2) \in V$ for the zero-sum game $(f + t\zeta, -f - t\zeta)$.*

Proof. Let $\mathbb{R}^m = \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$. Define the smoothly perturbed cost function $\tilde{f} : \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}$ by $\tilde{f}(x, y, t) = f(x, y) + t\zeta(x, y)$, and $\tilde{\omega}_{\mathcal{S}} : \mathbb{R}^m \times \mathbb{R} \to T^*(\mathbb{R}^m)$ by $\tilde{\omega}_{\mathcal{S}}(x, y, t) = (D\tilde{f}(x, y)), -D_2\tilde{f}(x, y))$, for all $t \in \mathbb{R}$ and $(x, y) \in \mathbb{R}^m$.

Since (x_1, x_2) is a differential Stackelberg equilibrium, $D\tilde{\omega}_{\mathcal{S}}(x, y, 0)$ is necessarily non-degenerate. Invoking the implicit function theorem (Abraham et al., 1988, Thm. 2.5.7), there exists neighborhoods $V \subset \mathbb{R}$ of zero and $W \subset \mathbb{R}^m$ and a smooth function $\sigma \in C^r(V, W)$ such that for all $t \in V$ and $(x_1, x_2) \in W$, $\tilde{\omega}_{\mathcal{S}}(x_1, x_2, s) = 0 \iff (x_1, x_2) = \sigma(t)$. Since $\tilde{\omega}_{\mathcal{S}}$ is continuously differentiable, there exists a neighborhood $U \subset W$ of zero such that $D\tilde{\omega}_{\mathcal{S}}(\sigma(t), t)$ is invertible for all $t \in U$. Thus, for all $t \in U$, $\sigma(t)$ must be the unique local Stackelberg equilibrium of $(f + t\zeta|_W, -f - t\zeta|_W)$. $\square$

D. Proofs for Results on Connections and Implications for Zero-Sum Settings

This appendix contains proofs for results given in Section 3. To be clear, we restate each result before providing the proof.

Lemma D.1 (Equivalence Between Sets of Critical Points)**.** *Consider a zero-sum game $(f, -f)$ defined by a function $f \in C^q(X, \mathbb{R})$, $q \geq 2$. The critical points of $\dot{x} = -\omega(x)$ and $\dot{x} = -\omega_{\mathcal{S}}(x)$ coincide.*

[4]Recall that the conditions that define or characterize a differential Stackelberg equilibrium are sufficient conditions for a local Stackelberg equilibrium; indeed, if a point x^* is a differential Stackelberg equilibrium, then it is a local Stackelberg equilibrium.

Proof. The result holds since for any $x \in X$, $D_1 f(x) = 0$ and $D_2 f(x) = 0$ if and only if $Df(x) = D_1 f(x) - D_{21} f(x)^\top (D_2^2 f(x))^{-1} D_2 f(x) = 0$ and $D_2 f(x) = 0$. □

The above lemma implies that in zero-sum games the first-order necessary and sufficient conditions for differential Nash equilibria and differential Stackelberg equilibria coincide.

The following is Proposition 2 from the main body.

Proposition 2. *In zero-sum games $(f, -f)$ with $f \in C^q(X, \mathbb{R})$ for $q \geq 2$, a joint strategy $x \in X$ is a stable critical point of $\dot{x} = -\omega_{\mathcal{S}}(x)$ if and only if x is a differential Stackelberg equilibrium. Moreover, if f is generic, a point x is a stable critical point of $\dot{x} = -\omega_{\mathcal{S}}(x)$ if and only if it is a local Stackelberg equilibrium.*

Proof. For a zero-sum game $(f, -f)$, the Jacobian of the Stackelberg limiting dynamics $\dot{x} = -\omega_{\mathcal{S}}(x)$ at a stable critical point is

$$J_{\mathcal{S}}(x) = \begin{bmatrix} D_1(Df)(x) & 0 \\ -D_{21} f(x) & -D_2^2 f(x) \end{bmatrix} > 0. \tag{7}$$

The structure of the Jacobian $J_{\mathcal{S}}(x)$ follows from the fact that

$$D_2(Df(x)) = D_{12} f(x) - D_{12} f(x)(D_2^2 f(x))^{-1} D_2^2 f(x) = 0.$$

The eigenvalues of a lower triangular block matrix are the union of the eigenvalues in each of the block diagonal components. This implies that $J_{\mathcal{S}}(x) > 0$ if and only if $D_1(Df)(x) > 0$ and $-D_2^2 f(x) > 0$. Consequently, invoking Lemma D.1, a point x is a stable critical point of the Stackelberg limiting dynamics if and only if x is a differential Stackelberg equilibrium by the definition.

Now, suppose that $f \in C^q(X, \mathbb{R})$ is a generic function. Then, by Theorem 1 (genericity of differential Stackelberg equilibria in zero sum games), all differential Stackelberg equilibria of $(f, -f)$ are local Stackelberg equilibria so that the final statement of the theorem holds. □

The following is Proposition 3 from the main body.

Proposition 3. *In zero-sum games $(f, -f)$ with $f \in C^q(X, \mathbb{R})$ for $q \geq 2$, differential Nash equilibria are differential Stackelberg equilibria. Moreover, if f is generic, local Nash equilibria are local Stackelberg equilibria.*

Proof. Suppose x is a differential Nash equilibria so that by definition $D_1^2 f(x) > 0$, $-D_2^2 f(x) > 0$. This directly implies that the Schur complement of $J(x)$ is strictly positive definite:

$$D_1^2 f(x) - D_{21} f(x)^\top (D_2^2 f(x))^{-1} D_{21} f(x) > 0.$$

Hence, x is a differential Stackelberg equilibrium since the Schur complement of J is exactly the derivative $D^2 f$ at critical points and $-D_2^2 f(x) > 0$ since x is a differential Nash equilibrium.

Now, suppose that $f \in C^q(X, \mathbb{R})$ is a generic function. Then, by genericity of differential Nash equilibria in zero sum games (Mazumdar & Ratliff, 2019, Thm. 2), all local Nash equilibria of $(f, -f)$ are differential Nash equilibria. Similarly, by Theorem 1 (genericity of differential Stackelberg equilibria in zero sum games), all local Stackelberg equilibria of $(f, -f)$ are differential Stackelberg equilibria so that the final statement of the result holds. □

Moreover, the following lemma has been shown previously (Mazumdar et al., 2020, Prop. 3.7); we provide the proof here for posterity.

Lemma D.2. *Consider a two-player, zero-sum continuous game $(f, -f)$ defined for $f \in C^q(X, \mathbb{R})$ with $q \geq 2$. If x is a differential Nash equilibrium, then it is a stable critical point of $\dot{x} = -\omega(x)$.*

Proof. Consider a zero-sum game $(f, -f)$ and a differential Nash equilibrium x so that $\omega(x) = 0$, $D_1^2 f(x) > 0$, and $-D_2^2 f(x) > 0$. The Jacobian of ω is given by

$$J(x) = \begin{bmatrix} D_1^2 f(x) & D_{12} f(x) \\ -D_{21} f(x) & -D_2^2 f(x) \end{bmatrix} = \begin{bmatrix} D_1^2 f(x) & D_{12} f(x) \\ -D_{12}^\top f(x) & -D_2^2 f(x) \end{bmatrix}.$$

Let (λ, v) be an eigenpair of $J(x)$. The real part of λ, denoted $\mathrm{Re}(\lambda)$, is given by

$$\mathrm{Re}(\lambda) = \tfrac{1}{2}(\bar{\lambda}+\lambda) = \tfrac{1}{2}(v^* J^\top(x) v + v^* J(x) v) = \tfrac{1}{2} v^*(J^\top(x) + J(x))v = \tfrac{1}{2} v^* \begin{bmatrix} D_1^2 f(x) & 0 \\ 0 & -D_2^2 f(x) \end{bmatrix} v > 0$$

where the last equality follows from the positive definiteness of $\mathrm{diag}(D_1^2 f(x), -D_2^2 f(x))$. □

As a consequence of Lemma D.2 and Propositions 2 and 3, any differential Nash equilibrium of a zero-sum game is both a stable critical point of the continuous time simultaneous play dynamics $\dot{x} = -\omega(x)$ and the Stackelberg dynamics $\dot{x} = -\omega_{\mathcal{S}}(x)$, and hence a differential Stackelberg equilibrium. We summarize this result in the following proposition.

Proposition D.1. *Consider a zero-sum game $(f, -f)$ defined by $f \in C^q(X, \mathbb{R})$ with $q \geq 2$. Any differential Nash equilibrium is a stable critical point of $\dot{x} = -\omega(x)$ and $\dot{x} = -\omega_{\mathcal{S}}(x)$, and hence, is a differential Stackelberg equilibrium. For a generic zero-sum game $(f, -f)$, any local Nash equilibrium is a stable critical point of $\dot{x} = -\omega(x)$ and $\dot{x} = -\omega_{\mathcal{S}}(x)$, and hence, is a local Stackelberg equilibrium.*

The following is Proposition 4 from the main body.

Proposition 4. *Consider a zero-sum GAN satisfying the realizable assumption. Any stable critical point of $\dot{x} = -\omega(x)$ at which $-D_2^2 f(x) > 0$ is a differential Stackelberg equilibrium and a stable critical point of $\dot{x} = -\omega_{\mathcal{S}}(x)$.*

Proof. Consider a stable critical point x of $\dot{x} = -\omega(x)$ such that $-D_2^2 f(x) > 0$. Note that the realizable assumption implies that the Jacobian of ω is

$$J(x) = \begin{bmatrix} 0 & D_{12} f(x) \\ -D_{21} f(x) & -D_2^2 f(x) \end{bmatrix}$$

(see, e.g., Nagarajan & Kolter, 2017). Hence, since $-D_2^2 f(x) > 0$,

$$-D_{21}^\top f(x)(D_2^2 f)^{-1}(x) D_{21} f(x) > 0.$$

Moreover, since both $-D_2^2 f(x) > 0$ and the Schur complement $D_1^2 f(x) - D_{21}^\top f(x)(D_2^2 f(x))^{-1} D_{21} f(x) > 0$, we can determine that x is a differential Stackelberg equilibrium. Given that x is a differential Stackelberg equilibrium, it follows that the Jacobian of $\omega_{\mathcal{S}}$ has positive real eigenvalues from the block triangular structure so the point x is stable with respect to $\dot{x} = -\omega_{\mathcal{S}}(x)$. □

E. When are Non-Nash Attractors of Simultaneous Gradient Play Stackelberg Equilibria?

As alluded to in the main body, an interesting question is when stable critical points of simultaneous gradient descent are differential Stackelberg equilibria but not differential Nash equilibria. Attracting critical points x^* of the dynamics $\dot{x} = -\omega(x)$ that are not differential Nash equilibria are such that either $D_1^2 f(x^*)$ or $-D_2^2 f(x^*)$ are not positive definite. Without loss of generality, considering player 1 to be the leader, a stable critical point of the Stackelberg dynamics $\dot{x} = -\omega_{\mathcal{S}}(x)$ requires both $-D_2^2 f(x^*)$ and $D_1^2 f(x^*) - D_{21} f(x^*)^\top (D_2^2 f(x^*))^{-1} D_{21} f(x^*)$ to be positive definite. Furthermore, recall from Proposition 2 that the set of stable critical points for the dynamics $\dot{x} = -\omega_{\mathcal{S}}(x)$ is equivalent to the set of differential Stackelberg equilibria since the conditions for stability are the conditions for a differential Stackelberg equilibria. Hence, if $-D_2^2 f(x^*)$ is not positive definite at a stable critical point of $\dot{x} = -\omega(x)$ that is not a differential Nash equilibria, then x^* will also not be a differential Stackelberg equilibria nor a stable critical point of $\dot{x} = -\omega_{\mathcal{S}}(x)$. Consequently, in this section, we consider stable critical points x^* of $\dot{x} = -\omega(x)$ at which $-D_2^2 f(x^*) > 0$ that are not differential Nash equilibria and determine when $D_1^2 f(x^*) - D_{21} f(x^*)^\top (D_2^2 f(x^*))^{-1} D_{21} f(x^*)$ is positive definite so that x^* is a differential Stackelberg equilibria and also a stable critical point of $\dot{x} = -\omega_{\mathcal{S}}(x)$. [5]

In the following two propositions, we need some addition notion that is common across the two results. Let $x_1 \in \mathbb{R}^m$ and $x_2 \in \mathbb{R}^n$. For a stable critical point x^* of $\dot{x} = -\omega(x)$ that is not a differential Nash equilibria, let $\mathrm{spec}(D_1^2 f(x^*)) = \{\mu_j,\ j \in \{1, \ldots, m\}\}$ where

$$\mu_1 \leq \cdots \leq \mu_r < 0 \leq \mu_{r+1} \leq \cdots \leq \mu_m,$$

and let $\mathrm{spec}(-D_2^2 f(x^*)) = \{\lambda_i,\ i \in \{1, \ldots, n\}\}$ where $\lambda_1 \geq \cdots \geq \lambda_n > 0$, and define $p = \dim(\ker(D_1^2 f(x^*)))$.

[5] We note that we could study an analogous setup in which we characterize when the stable critical points that are not differential Nash equilibria with $D_1^2 f(x^*) > 0$ are such that $-D_2^2 f(x^*) + D_{12} f(x^*)^\top (D_1^2 f(x^*))^{-1} D_{12} f(x^*) > 0$, thereby switching the roles of leader and follower.

Proposition E.1 (Necessary conditions)**.** *Consider a stable critical point x^* of the simultaneous gradient dynamics $\dot{x} = -\omega(x)$ that is not a differential Nash equilibrium and is such that $-D_2^2 f(x^*) > 0$. Given $\kappa > 0$ such that $\|D_{21} f(x^*)\| \leq \kappa$, if x^* is a differential Stackelberg equilibria and a stable critical point of $\dot{x} = -\omega_{\mathcal{S}}(x)$, then $r \leq n$ and $\kappa^2 \lambda_i + \mu_i > 0$ for all $i \in \{1, \dots, r-p\}$.*

For a matrix W, let $W^\dagger$ denote the conjugate transpose.

Proposition E.2 (Sufficient conditions)**.** *Let x^* be a stable critical point of the simultaneous gradient dynamics $\dot{x} = -\omega(x)$ that is not a differential Nash equilibria and is such that $D_1^2 f(x^*)$ and $-D_2^2 f(x^*)$ are Hermitian and $-D_2^2 f(x^*) > 0$. Suppose that there exists a diagonal matrix (not necessarily positive) $\Sigma \in \mathbb{C}^{m \times n}$ with non-zero entries such that $D_{12} f(x^*) = W_1 \Sigma W_2^\dagger$ where W_1 are the orthonormal eigenvectors of $D_1^2 f(x^*)$ and W_2 are orthonormal eigenvectors of $-D_2^2 f(x^*)$. Given $\kappa > 0$ such that $\|D_{21} f(x^*)\| \leq \kappa$, if $r \leq n$ and $\kappa^2 \lambda_i + \mu_i > 0$ for each $i \in \{1, \dots, r-p\}$, then x^* is a differential Stackelberg equilibrium and a stable critical point of $\dot{x} = -\omega_{\mathcal{S}}(x)$.*

The proofs of the propositions follow from linear algebra results. Before diving into the proofs, we provide some commentary. Essentially, this says that if $D_1^2 f(x^*) = W_1 M W_1^\dagger$ with $W_1 W_1^\dagger = I_{m \times m}$ and M diagonal, and $-D_2^2 f(x^*) = W_2 \Lambda W_2^\dagger$ with $W_2 W_2^\dagger = I_{n \times n}$ and Λ diagonal, then $D_{12} f(x^*)$ can be written as $W_1 \Sigma W_2^\dagger$ for some diagonal matrix $\Sigma \in \mathbb{R}^{m \times n}$ (not necessarily positive). Note that since Σ does not necessarily have positive values, $W_1 \Sigma W_2^\dagger$ is not the singular value decomposition of $D_{12} f(x^*)$. In turn, this means that each eigenvector of $D_1^2 f(x^*)$ gets mapped onto a single eigenvector of $-D_2^2 f(x^*)$ through the transformation $D_{12} f(x^*)$ which describes how player 1's variation $D_1 f(x)$ changes as a function of player 2's choice. With this structure for $D_{12} f(x^*)$, we can show that $D_1^2 f(x^*) - D_{21} f(x^*)^\top (D_2^2 f(x^*))^{-1} D_{21} f(x^*) > 0$.

Note that if we remove the assumption that Σ has non-zero entries, then the remaining assumptions are still sufficient to guarantee that

$$D_1^2 f(x^*) - D_{21} f(x^*)^\top (D_2^2 f)^{-1}(x^*) D_{21} f(x^*) \geq 0.$$

This means that x^* does not satisfy the conditions for a differential Stackelberg equilibrium, however, the point does satisfy necessary conditions for a local Stackelberg equilibrium and the point is a marginally stable attractor of the dynamics.

The following lemma is a very well-known result in linear algebra and can be found in nearly any advanced linear algebra text such as (Horn & Johnson, 2011).

Lemma E.1. *Let $W \in \mathbb{C}^{n \times n}$ be Hermitian with k positive eigenvalues (counted with multiplicities) and let $U \in \mathbb{C}^{m \times n}$. Then $\lambda_j(UWU^\dagger) \leq \|U\|^2 \lambda_j(W)$ for each $j \in \{1, \dots, \min\{k, m, \mathrm{rank}(UWU^\dagger)\}\}$.*

Let $|M| = (MM^\top)^{1/2}$ for a matrix M. Recall also that for Propositions E.1 and E.2, we have defined $\mathrm{spec}(D_1^2 f(x^*)) = \{\mu_j,\ j \in \{1, \dots, m\}\}$ where $\mu_1 \leq \cdots \leq \mu_r < 0 \leq \mu_{r+1} \leq \cdots \leq \mu_m$, and $\mathrm{spec}(-D_2^2 f(x^*)) = \{\lambda_i,\ i \in \{1, \dots, n\}\}$ where $\lambda_1 \geq \cdots \geq \lambda_n > 0$, given a stable critical point x^* of $\dot{x} = -\omega(x)$.

We now use the Lemma E.1 to prove Proposition E.1. The proof follows the main arguments in Berger et al. (2018, Lemma 3.2) with some minor changes due to the nature of our problem.

Proof of Proposition E.1. Let x^* be a differential Stackelberg equilibria but not a differential Nash equilibrium and a stable critical point of both $\dot{x} = -\omega(x)$ and $\dot{x} = -\omega_{\mathcal{S}}(x)$ such that $-D_2^2 f(x^*) > 0$. For the sake of presentation, define $A = D_1^2 f(x^*)$, $B = D_{12} f(x^*)$, and $C = D_2^2 f(x^*)$. Recall that $x_1 \in \mathbb{R}^m$ and $x_2 \in \mathbb{R}^n$ and suppose that $A - BC^{-1}B^\top > 0$.

Claim: $r \leq n$ **is necessary.** Assume for the sake of contradiction that $r > n$. Note that if $m \leq n$, then this is not possible. In this case, we automatically satisfy that $r \leq n$. Otherwise, $r \geq m > n$. Let $\mathcal{S}_1 = \ker(B(-C^{-1} + |C^{-1}|)B^\top)$ and consider the subspace $\mathcal{S}_2$ of $\mathbb{C}^m$ spanned by the all the eigenvectors of A corresponding to non-positive eigenvalues. Note that

$$\dim \mathcal{S}_1 = m - \mathrm{rank}(B(-C^{-1} + |C^{-1}|)B^\top) \geq m - \mathrm{rank}(-C^{-1} + |C^{-1}|) = m - n$$

By assumption, we have that $\dim \mathcal{S}_2 = r > n$ so that

$$\dim \mathcal{S}_1 + \dim \mathcal{S}_2 \geq (m-n) + r = m + (r-n) > m.$$

Thus, $\mathcal{S}_1 \cap \mathcal{S}_2 \neq \{0\}$. Now, $\mathcal{S}_1 = \ker(B(-C^{-1} + |C^{-1}|)B^\top)$. Hence, for any non-trivial vector $v \in \mathcal{S}_1 \cap \mathcal{S}_2$, $(BC^{-1}B^\top - B|C^{-1}|B^\top)v = 0$ so that we have

$$\langle (A - BC^{-1}B^\top)v, v\rangle = \langle Av, v\rangle - \langle B|C^{-1}|B^\top v, v\rangle \leq 0. \tag{8}$$

Note that the inequality in (8) holds since the vector v is in the non-positive eigenspace of A and the second term is clearly non-positive. Thus, $A - BC^{-1}B^\top$ cannot be positive definite, which gives a contradiction so that $r \leq n$.

Claim: $\kappa^2\lambda_i + \mu_i > 0$ **is necessary**. Let the maps $\lambda_i(\cdot)$ denote the eigenvalues of its argument arranged in non-increasing order. Then, by the Weyl theorem for Hermitian matrices (Horn & Johnson, 2011), we have that

$$0 < \lambda_m(A - BC^{-1}B^\top) \leq \lambda_i(A) + \lambda_{m-i+1}(-BC^{-1}B^\top),\ i \in \{1, \ldots, m\}.$$

We can now combine this inequality with Lemma E.1. Indeed, we have that

$$0 < \lambda_i(A) + \|B\|^2\lambda_{m-i+1}(-C^{-1}) < \mu_{m-i+1} + \kappa^2\lambda_{m-i+1},\ \ \forall\, i \in \{m - r + p + 1, \ldots, m\}$$

which gives the desired result.

Since we have shown both the necessary conditions, this concludes the proof. $\square$

Now, let us prove Proposition E.2 which gives sufficient conditions for when a stable critical point x^* of $\dot{x} = -\omega(x)$ at which $-D_2^2 f > 0$ is such that it is not a differential Nash equilibrium but it is a differential Stackelberg equilibrium and a stable critical point of $\dot{x} = -\omega_{\mathcal{S}}(x)$.

Proof of Proposition E.2. Let x^* be a stable critical point of $\dot{x} = -\omega(x)$ that is not a differential Nash equilibrium such that $D_1^2 f(x^*)$ and $-D_2^2 f(x^*) > 0$ are Hermitian. Since $D_i^2 f(x^*)$, $i = 1, 2$ are both Hermitian, let $D_1^2 f(x^*) = W_1 M W_1^\dagger$ with $W_1 W_1^\dagger = I_{m\times m}$ and $M = \mathrm{diag}(\mu_1, \ldots, \mu_m)$, and $-D_2^2 f(x^*) = W_2 \Lambda W_2^\dagger$ with $W_2 W_2^\dagger = I_{n\times n}$ and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$.

By assumption, there exists a diagonal matrix $\Sigma \in \mathbb{R}^{m\times n}$ such that $D_{12} f(x^*) = W_1 \Sigma W_2^\dagger$ where W_1 are the orthonormal eigenvectors of $D_1^2 f(x^*)$ and W_2 are orthonormal eigenvectors of $-D_2^2 f(x^*)$. Then,

$$\begin{aligned} D_1^2 f(x^*) - D_{21} f(x^*)^\top (D_2^2 f(x^*))^{-1} D_{21} f(x^*) &= W_1 M W_1^\dagger + W_1 \Sigma W_2^\dagger (W_2 \Lambda W_2^\dagger)^{-1} W_2 \Sigma^\dagger W_1^\dagger \\ &= W_1 (M + \Sigma \Lambda^{-1} \Sigma^\dagger) W_1^\dagger \end{aligned}$$

Hence, to understand the eigenstructure, we simply need to compare the all negative eigenvalues of $D_1^2 f(x^*)$ in increasing order with the most positive eigenvalues of $-D_2^2 f(x^*)$ in decreasing order. Indeed, by assumption, $r \leq n$ and $\kappa^2\lambda_i + \mu_i > 0$ for each $i \in \{1, \ldots, r - p\}$. Thus,

$$D_1^2 f(x^*) - D_{21} f(x^*)^\top (D_2^2 f(x^*))^{-1} D_{21} f(x^*) > 0$$

since it is a symmetric matrix. Combining this with the fact that $-D_2^2 f(x^*) > 0$, x^* is a differential Stackelberg equilibrium and by Proposition 2 it is also a stable critical point of $\dot{x} = -\omega_{\mathcal{S}}(x)$. $\square$

It is also worth noting that the fact that the eigenvalues of $-J(x^*)$ are in the open-left-half complex plane is not used in proving this result. We believe that further investigation could lead to a less restrictive sufficient condition. Empirically, by randomly generating the different block matrices, it is quite difficult to find examples such that the real parts of the eigenvalues of $J(x^*)$ are positive, $-D_2^2 f(x^*) > 0$, and the Schur complement $D_1^2 f(x^*) - D_{21} f(x^*)^\top (D_2^2 f)^{-1}(x^*) D_{21} f(x^*)$ is not positive definite. In the scalar case stated in Corollary E.1, the proof is straightforward; we suspect that using the notion of quadratic numerical range (Tretter, 2008)—a super set of the spectrum of a block operator matrix—along with the fact that the Jacobian of the simultaneous gradient play dynamics, $-J$, has its spectrum in the open left-half complex plane, we may be able to extend the scalar case to arbitrary dimensions.

We also note that the condition depends on conditions that are difficult to check *a priori* without knowledge of x^*. Certain classes of games for which these conditions hold everywhere and not just at the equilibrium can be constructed. For instance, alternative conditions can be given: if the function f which defines the zero-sum game is such that it is concave in x_2 and there exists a K such that

$$D_{12} f(x) = K D_2^2 f(x)$$

where $\sup_x \|D_{12} f(x)\| \leq \kappa < \infty$[6] and $K = W_1 \Sigma W_2^\dagger$ with Σ again a (not necessarily positive) diagonal matrix, then the results of Proposition E.2 hold. From a control point of view, one can think about the leader's update as having a feedback

[6]Functions such that derivative of f is Lipschitz will satisfy this condition.

term with the follower's input. On the other hand, the results are useful for the synthesis of games, such as in reward shaping or incentive design, where the goal is to drive agents to particular desirable behavior.

Corollary E.1. *Consider a zero-sum game $(f, -f)$ defined by $f \in C^q(\mathbb{R} \times \mathbb{R}, \mathbb{R})$ with $q \geq 2$. All stable critical points of $\dot{x} = -\omega(x)$ at which $-D_2^2 f(x) > 0$ that are not differential Nash equilibria are differential Stackelberg equilibria.*

Proof. Consider a zero-sum game $(f, -f)$ defined by $f \in C^q(\mathbb{R} \times \mathbb{R}, \mathbb{R})$ with $q \geq 2$. Suppose x is a stable critical point of $\dot{x} = -\omega(x)$ at which $-D_2^2 f(x) > 0$ and $D_1^2 f(x) \leq 0$ so that it is not a Nash equilibria. The Jacobian $J(x)$ of the vector field $\omega(x)$ is given by

$$J(x) = \begin{bmatrix} D_1^2 f(x) & D_{12} f(x) \\ -D_{21} f(x) & -D_2^2 f(x) \end{bmatrix}.$$

Since x is a stable critical point, the eigenvalues of $J(x)$ have positive real parts. This fact guarantees that the determinant and trace of $J(x)$ must be positive since the eigenvalues are either complex conjugates or both real. As a result, $D_{12} f(x) D_{21} f(x) > D_1^2 f(x) D_2^2 f(x)$ and $D_1^2 f(x^*) > D_2^2 f(x^*)$. It directly follows that the Schur complement of $J(x)$ is positive definite, meaning

$$D_1^2 f(x^*) - D_{12} f(x) (D_2^2 f(x))^{-1} D_{21} f(x) > 0.$$

We conclude that x is a differential Stackelberg equilibrium since $\omega(x) = 0$ if and only if $\omega_{\mathcal{S}}(x) = 0$ by Lemma D.1, the Schur complement of $J(x)$ is the derivative $D^2 f(x)$, and $-D_2^2 f(x) > 0$ was given. □

F. Deterministic Convergence Results

Consider the deterministic Stackelberg update

$$x_{k+1,1} = x_{k,1} - \gamma_1 (D_1 f_1(x_k) - D_{21}^\top f_2(x_k)(D_2^2 f_2(x_k))^{-1} D_2 f_1(x_k)) \tag{9}$$
$$x_{k+1,2} = x_{k,2} - \gamma_2 D_2 f_2(x_k) \tag{10}$$

which is equivalent to the dynamics

$$x_{k+1,1} = x_{k,1} - \gamma_1 (D_1 f_1(x_k) - D_{21}^\top f_2(x_k)(D_2^2 f_2(x_k))^{-1} D_2 f_1(x_k)) \tag{11}$$
$$x_{k+1,2} = x_{k,2} - \gamma_1 \tau D_2 f_2(x_k) \tag{12}$$

where $\tau = \gamma_2/\gamma_1$ is the "timescale" separation. Then, we write the τ–Stackelberg update in "vector"[7] form as

$$x_{k+1} = x_k - \gamma_1 \omega_{\mathcal{S}_\tau}(x_k). \tag{13}$$

where

$$\omega_{\mathcal{S}_\tau}(x_k) = \left(D_1 f_1(x_k) - D_{21}^\top f_2(x_k)(D_2^2 f_2(x_k))^{-1} D_2 f_1(x_k), \tau D_2 f_2(x_k) \right).$$

In the following subsections, we provide the proofs of the convergence guarantees in zero-sum games, almost sure avoidance of saddles, and then the convergence guarantees in general-sum games, respectively.

F.1. Zero-Sum Convergence

Under appropriate choices on the step-size so that the local linearization of the update is a contraction, standard arguments from numerical analysis for dynamical systems give rise to a guarantee on local asymptotic convergence (including a rate of convergence), and a finite-time convergence guarantee to an ε–differential Stackelberg equilibrium. Indeed, recall the following well-known result on on fixed points gives us exactly such convergence guarantees.

Proposition F.1 (Ostrowski's Theorem Argyros, 1999)**.** *Let x^* be a fixed point for the discrete dynamical system $x_{k+1} = F(x_k)$. If the spectral radius of the Jacobian satisfies $\rho(DF(x^*)) < 1$, then F is a contraction at x^* and hence, x^* is asymptotically stable.*

[7] We are in some places treating derivatives as co-vectors and in other places as vectors. The reader should pay attention to context. Here, e.g., $\omega_{\mathcal{S}_\tau}(x) \in \mathbb{R}^{m_1+m_2}$.

We note that $\rho(DF(x^*)) < 1$ implies there exists $c > 0$ such that $\rho(DF(x^*)) \le c < 1$. Hence, given any $\varepsilon > 0$, there is a norm and a $c > 0$ such that $\|DF\| \le c + \varepsilon < 1$ on a neighborhood of x^* (Ortega & Rheinboldt, 1970, 2.2.8). Thus, the proposition implies that if $\rho(DF(x^*)) = 1 - \kappa < 1$ for some κ, then there exists a ball $B_p(x^*)$ of radius $p > 0$ such that for any $x_0 \in B_p(x^*)$, and some constant $K > 0$, $\|x_k - x^*\|_2 \le K(1 - \frac{\kappa}{2})^k \|x_0 - x^*\|_2$ using $\varepsilon = \frac{\kappa}{4}$.

For a zero-sum setting defined by cost function $f \in C^q(X_1 \times X_2, \mathbb{R})$ with $q \ge 2$, let

$$\mathtt{S}_1(J(x)) = D_1^2 f(x) - D_{21} f(x)^\top (D_2^2 f(x))^{-1} D_{21} f(x)$$

be the first Schur complement of the Jacobian $J(x)$ of $\omega(x) = (D_1 f(x), D_2 f(x))$. The game Jacobian for the τ-Stackelberg update at critical points is given by

$$J_{\mathcal{S}_\tau}(x) = \begin{bmatrix} \mathtt{S}_1(J(x)) & 0 \\ -\tau D_{21} f(x) & -\tau D_2^2 f(x) \end{bmatrix} \tag{14}$$

Recall that for function f and g defined on on some subset of real numbers, $f(k) = O(g(k))$ if and only if there exists constants K and C such that $|f(k)| \le C|g(k)|$ for all $k > K$.

Using Proposition F.1, we prove the following two results from the main body.

Theorem 3 (Zero-Sum Rate of Convergence.)**.** *Consider a zero-sum game defined by $f \in C^q(X, \mathbb{R})$ with $q \ge 2$. For a differential Stackelberg equilibrium x^* with $\alpha = \min\{\lambda_{\min}(\mathtt{S}_1(J(x^*))), \lambda_{\min}(-\tau D_2^2 f(x^*))\}$ and $\beta = \max\{\lambda_{\max}(\mathtt{S}_1(J(x^*))), \lambda_{\max}(-\tau D_2^2 f(x^*))\}$ and learning rate $\gamma_1 = 1/(2\beta)$, the τ–Stackelberg update converges locally with a rate of $O((1 - \frac{\alpha}{4\beta})^k)$.*

Before proving the result, we comment that we can prove this result with $\gamma_1 = 1/\beta$ and obtain a rate of $O((1 - \alpha/(2\beta))^k)$; we choose $\gamma_1 = 1/(2\beta)$ in order to ensure that the choice of learning rate satisfies the avoidance of saddles result in Theorem 4.

Proof of Theorem 3. To show convergence, we simply need to show that $\rho(I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)) < 1$ and apply Proposition F.1.

Fix $\gamma_1 = 1/(2\beta)$. The structure of the Jacobian $J_{\mathcal{S}_\tau}(x^*)$ is lower-block triangular, with symmetric components along the diagonal given by $\mathtt{S}_1(J(x))$ and $-\tau D_2^2 f(x)$. From this structure, we know that $\mathrm{spec}(J_{\mathcal{S}_\tau}(x)) = \mathrm{spec}(\mathtt{S}_1(J(x))) \cup \mathrm{spec}(-\tau D_2^2 f(x))$. Then, by the spectral mapping theorem and the fact that the eigenvalues of $J_{\mathcal{S}_\tau}(x^*)$ are real,

$$\max_{\lambda \in \mathrm{spec}(I - \gamma_1 J_{\mathcal{S}_\tau}(x^*))} |\lambda| = |1 - \gamma_1 \min\{\lambda_{\min}(\mathtt{S}_1(J(x^*))), \lambda_{\min}(-\tau D_2^2 f(x^*))\}| = |1 - \gamma_1 \alpha|.$$

so that $\rho(I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)) < 1$ since $\alpha \le \beta$. We also note that $\lambda_{\min}(I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)) = 1 - \gamma_1 \max\{\lambda_{\max}(\mathtt{S}_1(J(x^*))), \lambda_{\max}(-\tau D_2^2 f(x^*))\} \ge 0$ for the choice of γ_1 so that, in fact, $\mathrm{spec}(I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)) \subset [0, 1 - \frac{\alpha}{2\beta}]$.

Applying Proposition F.1 gives the convergence guarantee for τ–Stackelberg. Indeed, since $\rho(I - \gamma_1 J_{\mathcal{S}_\tau}) \le 1 - \frac{\alpha}{2\beta}$, given $\varepsilon > 0$, there exists a matrix norm $\|\cdot\|$ such that $\|I - \gamma_1 J_{\mathcal{S}_\tau}\| \le 1 - \frac{\alpha}{2\beta} + \varepsilon$ (Horn & Johnson, 2011, Lemma 5.6.10).[8] Consider $\varepsilon = \frac{\alpha}{8\beta}$ so that there exists $\|\cdot\|$ such that $\|I - \gamma_1 J_{\mathcal{S}_\tau}\| \le 1 - \frac{3\alpha}{8\beta}$. Taking the Taylor expansion of $I - \gamma_1 \omega_{\mathcal{S}_\tau}(x)$ in a neighborhood of x^*, we have

$$I - \gamma_1 \omega_{\mathcal{S}_\tau}(x) = (I - \gamma_1 \omega_{\mathcal{S}_\tau}(x^*)) + (I - \gamma_1 J_{\mathcal{S}_\tau}(x^*))(x - x^*) + R_1(x - x^*)$$

where $R_1(x - x^*)$ is the remainder term satisfying $R_1(x - x^*) = o(\|x - x^*\|)$.[9] This, in turn, implies that there is a $\delta > 0$ such that $\|R_1(x - x^*)\| \le \frac{\alpha}{8\beta}\|x - x^*\|$ whenever $\|x - x^*\| < \delta$. Hence,

$$\|I - \gamma_1 \omega_{\mathcal{S}_\tau}(x) - (I - \gamma_1 \omega_{\mathcal{S}_\tau}(x^*))\| \le \left(\|I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)\| + \frac{\alpha}{8\beta}\right)\|x - x^*\| \le \left(1 - \frac{\alpha}{4\beta}\right)\|x - x^*\|$$

Thus, for any $x_0 \in \{x | \ \|x - x^*\| < \delta\}$,

$$\|x_k - x^*\| \le \left(1 - \frac{\alpha}{4\beta}\right)^k \|x_0 - x^*\|. \tag{15}$$

[8]In fact, one can show that there is both a weighted induced 1-norm and weighted induced ∞-norm works for any ε so that the construction is not unreasonable.

[9]The notation $R_1(x - x^*) = o(\|x - x^*\|)$ as $x \to x^*$ means $\lim_{x \to x^*} \|R_1(x - x^*)\|/\|x - x^*\| = 0$.

Hence, the iteration complexity (or rate of convergence) is $O((1-\alpha/(4\beta))^k)$, since all finite dimensional norms are equivalent. □

Corollary 2 (Zero-Sum Finite Time Guarantee). *Given $\varepsilon > 0$, under the assumptions of Theorem 3, τ-Stackelberg learning obtains an ε-differential Stackelberg equilibrium in $\lceil \frac{4\beta}{\alpha}\log(\|x_0 - x^*\|/\varepsilon)\rceil$ iterations for any $x_0 \in B_\delta(x^*)$ with $\delta = \alpha/(4L\beta)$ where L is the local Lipschitz constant of $I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)$.*

The proof of Corollary 2 follows directly from the conclusion of Theorem 3.

Proof of Corollary 2. Following standard arguments, (15) in the proceeding proof implies a finite time convergence guarantee. Indeed, let $\varepsilon > 0$ be given. Since $0 < \frac{\alpha}{4\beta} < 1$ we have that $(1-\frac{\alpha}{4\beta})^k < \exp(-\frac{k\alpha}{4\beta})$. Hence,

$$\|x_k - x^*\| \le \exp(-k\alpha/(4\beta))\|x_0 - x^*\|$$

This, in turn implies that $x_k \in B_\varepsilon(x^*)$, meaning x_k is a ε-differential Stackelberg equilibrium for all $k \ge \lceil \frac{4\beta}{\alpha}\log(\|x_0 - x^*\|/\varepsilon)\rceil$ whenever $\|x_0 - x^*\| < \delta$.

Now, given that $f_i \in C^q(X,\mathbb{R})$ for $q \ge 2$, $I - \gamma_1 J_{\mathcal{S}_\tau}(x)$ is locally Lipschitz with constant L so that we can find an explicit expression for δ in terms of L. Indeed, recall that $R_1(x-x^*) = o(\|x-x^*\|)$ as $x \to x^*$ which means $\lim_{x\to x^*}\|R_1(x-x^*)\|/\|x-x^*\| = 0$ so that

$$\|R_1(x-x^*)\| \le \int_0^1 \|I - \gamma_1 J_{\mathcal{S}_\tau}(x^* + \eta(x-x^*)) - (I - \gamma_1 J_{\mathcal{S}_\tau}(x^*))\|\|x-x^*\|\, d\eta \le \frac{L}{2}\|x-x^*\|^2$$

Observing that

$$\|R_1(x-x^*)\| \le \frac{L}{2}\|x-x^*\|^2 = \frac{L}{2}\|x-x^*\|\|x-x^*\|,$$

we have that the $\delta > 0$ such that $\|R_1(x-x^*)\| \le \alpha/(8\beta)\|x-x^*\|$ is $\delta = \alpha/(4L\beta)$. □

F.2. Almost Sure Avoidance of Saddles

In this subsection, we prove almost sure avoidance of saddles for the τ-Stackelberg update. To do so, we rely upon the stable manifold theorem.

Theorem F.1. *(Stable Manifold Theorem Shub, 1978, Theorem III.7). Let $x_0 \in X$ be a fixed point for the C^q local diffeomorphism $\phi : U \to \mathbb{R}^m$ where U is an open neighborhood of $x_0 \in \mathbb{R}^m$ and $q \ge 1$. Let $E^s \otimes E^c \otimes E^u$ be the invariant splitting of $\mathbb{R}^m$ into generalized eigenspaces of $D\phi(x_0)$ corresponding to the eigenvalues of absolute value less than one, equal to one, and greater than one. To the $D\phi(x_0)$ invariant subspace $E^s \otimes E^c$ there is an associated local ϕ-invariant C^q embedded disc W^{cs}_{loc} called the local stable center manifold of dimension $\dim(E^s \otimes E^c)$ and ball B (in an adapted norm) around x_0 such that $\phi(W^{cs}_{loc}) \cap B \subset W^{cs}_{loc}$, and if $\phi^t(x) \in B$ for all $t \ge 0$ then $x \in W^{sc}_{loc}$ where $\phi^t = \phi \circ \cdots \circ \phi$ is the t-times composition of the map ϕ.*

We now restate and prove the result of interest for the τ-Stackelberg update.

Theorem 4 (Almost Sure Avoidance of Saddles). *Consider a general sum game defined by $f_i \in C^q(X,\mathbb{R})$, $q \ge 2$ for $i = 1,2$ and where, without loss of generality, player 1 is the leader. Suppose that $\omega_{\mathcal{S}_\tau}$ is L-Lipschitz and that $\gamma_1 < 1/L$. The τ–Stackelberg learning dynamics converge to saddle points of $\dot{x} = -\omega_{\mathcal{S}_\tau}(x)$ on a set of measure zero.*

Proof. To show this, we follow the arguments in Mazumdar et al. (2020) with slight modifications, an argument which also builds on similar results for single player optimization problems (Lee et al., 2016; Panageas & Piliouras, 2017). In particular, we show that $g_{\mathcal{S}_\tau}$ is a diffeomorphism, and then apply the center manifold theorem. Let

$$g_{\mathcal{S}_\tau}(x) = x - \gamma_1\left(D_1 f_1(x) - D_{21}f_2^\top(x)(D_2^2 f_2(x))^{-1}D_2 f_1(x), \tau D_2 f_2(x)\right)$$

We claim that $g_{\mathcal{S}_\tau} : \mathbb{R}^m \to \mathbb{R}^m$ is a diffeomorphism. If we can show that $g_{\mathcal{S}_\tau}$ is invertible and a local diffeomorphism, then the claim follows.

The game Jacobian for the τ-Stackelberg update is given by

$$J_{\mathcal{S}_\tau}(x) = \begin{bmatrix} D_1(Df_1(x_1,x_2)) & D_2(Df_1(x_1,x_2)) \\ \tau D_{21}f_2(x_1,x_2) & \tau D_2^2 f_2(x_1,x_2) \end{bmatrix} \tag{16}$$

We first argue by contradiction that $g_{\mathcal{S}_\tau}$ is invertible. Consider $x \neq y$ and suppose $g_{\mathcal{S}_\tau}(y) = g_{\mathcal{S}_\tau}(x)$ so that $y - x = \gamma_1(\omega_{\mathcal{S}_\tau}(y) - \omega_{\mathcal{S}_\tau}(x))$. The assumption $\sup_{x\in\mathbb{R}^m} \|J_{\mathcal{S}_\tau}(x)\|_2 \leq L < \infty$ implies that $\omega_{\mathcal{S}_\tau}$ satisfies the Lipschitz condition on $\mathbb{R}^m$. Hence, $\|\omega_{\mathcal{S}_\tau}(y) - \omega_{\mathcal{S}_\tau}(x)\|_2 \leq L\|y-x\|_2$. Then, $\|y-x\|_2 \leq L\gamma_1\|y-x\|_2 < \|y-x\|_2$ since $\gamma_1 < 1/L$ which gives rise to a contradiction.

Now, observe that $Dg_{\mathcal{S}_\tau} = I - \gamma_1 J_{\mathcal{S}_\tau}(x)$. If $Dg_{\mathcal{S}_\tau}$ is invertible, then the implicit function theorem (Abraham et al., 1988, Thm. 2.5.7) implies that $g_{\mathcal{S}_\tau}$ is a local diffeomorphism. Hence, it suffices to show that $\gamma_1 J_{\mathcal{S}_\tau}(x)$ does not have an eigenvalue of 1. Indeed, letting $\rho(A)$ be the spectral radius of a matrix A, we know in general that $\rho(A) \leq \|A\|$ for any square matrix A and induced operator norm $\|\cdot\|$ so that

$$\rho(\gamma_1 J_{\mathcal{S}_\tau}(x)) \leq \|\gamma_1 J_{\mathcal{S}_\tau}(x)\|_2 \leq \gamma_1 \sup_{x\in\mathbb{R}^m} \|J_{\mathcal{S}_\tau}(x)\|_2 < \gamma_1 L < 1.$$

Of course, the spectral radius is the maximum absolute value of the eigenvalues, so that the above implies that all eigenvalues of $\gamma_1 J_{\mathcal{S}_\tau}(x)$ have absolute value less than 1. Since $g_{\mathcal{S}_\tau}$ is injective by the preceding argument, its inverse is well-defined and since $g_{\mathcal{S}_\tau}$ is a local diffeomorphism on $\mathbb{R}^m$, it follows that $g_{\mathcal{S}_\tau}^{-1}$ is smooth on $\mathbb{R}^m$. Thus, $g_{\mathcal{S}_\tau}$ is a diffeomorphism.

Consider all critical points to the game, given by $\mathcal{X}_c = \{x \in X \mid \omega_{\mathcal{S}_\tau}(x) = 0\}$. For each $u \in \mathcal{X}_c$, let B_u, where u indexes the point, be the open ball derived from the center manifold theorem stated in Theorem F.1 and let $\mathcal{B} = \cup_u B_u$. Since $X \subseteq \mathbb{R}^m$, Lindelõf's lemma (Kelley, 1955)—every open cover has a countable subcover—gives a countable subcover of $\mathcal{B}$. That is, for a countable set of critical points $\{u_i\}_{i=1}^\infty$ with $u_i \in \mathcal{X}_c$, we have that $\mathcal{B} = \cup_{i=1}^\infty B_{u_i}$.

Starting from some point $x_0 \in X$, if τ-Stackelberg converges to a strict saddle point, then there exists a t_0 and index i such that $g_{\mathcal{S}_\tau}^t(x_0) \in B_{u_i}$ for all $t \geq t_0$. Again, applying the center manifold theorem from Theorem F.1 and using that $g_{\mathcal{S}_\tau}(X) \subset X$, which indeed holds if $X = \mathbb{R}^m$, we get that $g_{\mathcal{S}_\tau}^t(x_0) \in W_{\text{loc}}^{cs} \cap X$ where W_{loc}^{cs} is the local stable center manifold.

Using the fact that $g_{\mathcal{S}_\tau}$ is invertible, we can iteratively construct the sequence of sets defined by $W_1(u_i) = g_{\mathcal{S}_\tau}^{-1}(W_{\text{loc}}^{cs} \cap X)$ and $W_{k+1}(u_i) = g_{\mathcal{S}_\tau}^{-1}(W_k(u_i) \cap X)$. Then we have that $x_0 \in W_t(u_i)$ for all $t \geq t_0$. The set $\mathcal{X}_0 = \cup_{i=1}^\infty \cup_{t=0}^\infty W_t(u_i)$ contains all the initial points in X such that τ-Stackelberg converges to a strict saddle. Since u_i is a strict saddle, $I - \gamma_1 J_{\mathcal{S}_\tau}(u_i)$ has an eigenvalue greater than 1. This implies that the co-dimension of the unstable manifold is strictly less than m—i.e., $\dim(W_{\text{loc}}^{cs}) < m$. Hence, $W_{\text{loc}}^{cs} \cap X$ has Lebesgue measure zero in $\mathbb{R}^m$.

Using again that $g_{\mathcal{S}_\tau}$ is a diffeomorphism, $g_{\mathcal{S}_\tau}^{-1} \in C^1$ so that it is locally Lipschitz and locally Lipschitz maps are null set preserving. Hence, $W_k(u_i)$ has measure zero for all k by induction so that $\mathcal{X}_0$ is a measure zero set since it is a countable union of measure zero sets. □

Proposition 5. *Consider a zero-sum game defined by $f \in C^q(X,\mathbb{R})$, $q \geq 2$. Suppose that $\gamma_1 \leq 1/L$ where $\max\{\text{spec}(\mathtt{S}_1(J(x))) \cup \text{spec}(-\tau D_2^2 f(x))\} \leq L$. Then, x is a stable critical point of τ–Stackelberg update if and only if x is a differential Stackelberg equilibrium.*

The proof of this proposition follows almost immediately from Theorem 3 and Proposition 2. It implies that the only stable attractors[10] which are critical points are differential Stackelberg equilibria. Of course, as noted earlier escaping saddles can be onerous. However, results on escaping saddle points efficiently apply to the τ-Stackelberg learning rule since $\text{spec}(J_{\mathcal{S}_\tau}(x^*)) \subset \mathbb{R}$.

F.3. General-Sum Convergence

Consider a general sum setting defined by $f_i \in C^q(X,\mathbb{R})$ with $q \geq 2$ for $i = 1,2$ and where, without loss of generality, player 1 is the leader and player 2 is the follower. Unlike the zero-sum case, the structure of the Jacobian $J_{\mathcal{S}_\tau}$ is not lower block triangular and hence, the convergence rate depends more abstractly on the spectral structure of $J_{\mathcal{S}_\tau}$ as opposed to the second-order sufficient conditions for a local Stackelberg equilibria. It is still an open question as to how the spectrum of $J_{\mathcal{S}_\tau}$ relates to $\text{schur}(J_{\mathcal{S}_\tau})$.

[10]Note that the set of attractors include non-trivial periodic orbits.

Let $S(x^*) = \frac{1}{2}(J_{\mathcal{S}_\tau}(x^*)^\top + J_{\mathcal{S}_\tau}(x^*))$. Define constants $\alpha = \lambda^2_{\min}\big(S(x^*)\big)$ and $\beta = \lambda_{\max}(J_{\mathcal{S}_\tau}(x^*)^\top J_{\mathcal{S}_\tau}(x^*))$.

Theorem 5 (General Sum Rate of Convergence)**.** *Consider a general sum game (f_1, f_2) with $f_i \in C^q(X, \mathbb{R})$, $q \geq 2$ for $i = 1, 2$ and where, without loss of generality, player 1 is the leader. For a differential Stackelberg equilibrium x^* such that $J^\top_{\mathcal{S}_\tau}(x^*) + J_{\mathcal{S}_\tau}(x^*) > 0$, the τ–Stackelberg update with learning rate $\gamma_1 = \sqrt{\alpha}/\beta$ converges locally with a rate of $O((1 - \frac{\alpha}{2\beta})^{k/2})$.*

Proof of Theorem 5. Let $\gamma_1 = \sqrt{\alpha}/\beta$. Then to bound $\|I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)\|_2^2$ consider the following:

$$(I - \gamma_1 J_{\mathcal{S}_\tau}(x^*))^\top (I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)) \leq (1 - 2\gamma_1 \lambda_{\min}(S(x^*)) + \gamma_1^2 \lambda_{\max}(J^\top_{\mathcal{S}_\tau}(x^*) J_{\mathcal{S}_\tau}(x^*)))I \leq (1 - \alpha/\beta)I. \quad (17)$$

Moreover, $\rho(I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)) \leq \|I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)\|$ for any matrix norm (Horn & Johnson, 2011) so that $\rho(I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)) \leq (1 - \alpha/\beta)^{1/2}$. Taking the Taylor expansion of $I - \gamma_1 \omega_{\mathcal{S}_\tau}(x)$ around x^*, we have

$$I - \gamma_2 \omega_{\mathcal{S}_\tau}(x) = (I - \gamma_1 \omega_{\mathcal{S}_\tau}(x^*)) + (I - \gamma_1 J_{\mathcal{S}_\tau}(x^*))(x - x^*) + R_1(x - x^*)$$

where $R_1(x - x^*)$ is the remainder term satisfying $R_1(x - x^*) = o(\|x - x^*\|_2)$ as $x \to x^*$. This implies that there is a $\delta > 0$ such that $\|R_1(x - x^*)\|_2 \leq \frac{\alpha}{4\beta}\|x - x^*\|_2$ whenever $\|x - x^*\|_2 < \delta$. Hence,

$$\|I - \gamma_1 \omega_{\mathcal{S}_\tau}(x) - (I - \gamma_1 \omega_{\mathcal{S}_\tau}(x^*))\|_2 \leq \left(\|I - \gamma_1 J_{\mathcal{S}_\tau}(x^*)\|_2 + \frac{\alpha}{4\beta}\right)\|x - x^*\|_2 \leq \left(\left(1 - \frac{\alpha}{\beta}\right)^{1/2} + \frac{\alpha}{4\beta}\right)\|x - x^*\|_2$$

We claim that $c(z) = (1 - z)^{1/2} + \frac{z}{4} - \left(1 - \frac{z}{2}\right)^{1/2} \leq 0$ for any $z \in [0, 1]$. Since $c(0) = 0$ and $c(1) = \frac{1}{4} - \frac{1}{\sqrt{2}} \leq 0$, we simply need to show that $c'(z) \leq 0$ on $(0, 1)$ to get that $c(z)$ is a decreasing function on $[0, 1]$, and hence negative on $[0, 1]$. Indeed, $c'(z) = \frac{1}{4} + \frac{1}{2\sqrt{4-2z}} - \frac{1}{2\sqrt{1-z}} \leq 0$ since $(1 - z)^{-1/2} - (4 - 2z)^{-1/2} \geq 1/2$ for all $z \in (0, 1)$.

Note that $\alpha/\beta \in [0, 1]$ since $\alpha \leq \beta$; indeed,

$$\alpha = \lambda^2_{\min}(S(x^*)) \leq \lambda^2_{\max}(S(x^*)) \leq \lambda_{\max}(J^\top_{\mathcal{S}_\tau}(x^*) J_{\mathcal{S}_\tau}(x^*)) = \beta.$$

Further, $\alpha > 0$ and $\beta > 0$ by assumption. Hence,

$$\|I - \gamma_1 \omega_{\mathcal{S}_\tau}(x) - (I - \gamma_1 \omega_{\mathcal{S}_\tau}(x^*))\|_2 \leq \left(1 - \frac{\alpha}{2\beta}\right)^{1/2}\|x - x^*\|_2$$

Thus, for any $x_0 \in \{x | \ \|x - x^*\| < \delta\}$,

$$\|x_k - x^*\|_2 \leq \left(1 - \frac{\alpha}{2\beta}\right)^{k/2}\|x_0 - x^*\|_2 \quad (18)$$

so that the (local) rate of convergence is $O((1 - \alpha/(2\beta))^{k/2})$, since finite dimensional norms are equivalent. □

Analogous to the zero-sum setting, the proof of the finite time convergence guarantee follows directly from the arguments in the proof of Theorem 5.

Corollary 3 (General Sum Finite Time Guarantee)**.** *Given $\varepsilon > 0$, under the assumptions of Theorem 5, τ–Stackelberg learning obtains an ε–differential Stackelberg equilibrium in $\lceil \frac{4\beta}{\alpha} \log(\|x_0 - x^*\|/\varepsilon) \rceil$ iterations for any $x_0 \in B_\delta(x^*)$ with $\delta = \alpha/(2L\beta)$ where L is the local Lipschitz constant of $I - \gamma_1 J_{\mathcal{S}_\tau}(x)$.*

Proof of Corollary 3. Following standard arguments, (18) in the proceeding proof implies a finite time convergence guarantee. Indeed, let $\varepsilon > 0$ be given. Since $(1 - \alpha/(2\beta))^{1/2} \leq \exp(-\alpha/(4\beta))$, we have that

$$\|x_k - x^*\|_2 \leq \exp(-k\alpha/(4\beta))\|x_0 - x^*\|_2.$$

This, in turn, implies that $x_k \in B_\varepsilon(x^*)$ (i.e., x_k is an ε-differential Stackelberg equilibrium) for all $k \geq \lceil \frac{4\beta}{\alpha} \log(\|x_0 - x^*\|_2/\varepsilon) \rceil$ whenever $\|x_0 - x^*\| < \delta$.

Given that $f_i \in C^2(X, \mathbb{R})$ so that $I - \gamma_1 J_{\mathcal{S}_\tau}(x)$ is locally Lipschitz with constant L, we can find an explicit expression for δ in terms of L. Indeed, using similar arguments as in the proof of Corollary 2, $\delta = \alpha/(2L\beta)$.

□

G. Stochastic Convergence Results and Extended Analysis:

In this supplementary section, we provide the formal proofs for the asymptotic stochastic convergence results. Let us now review some mathematical preliminaries from dynamical systems theory.

G.1. Dynamical Systems Theory Primer

Let us first recall some results on stability. Given a sufficiently smooth function $f \in C^q(X, \mathbb{R})$, a critical point x^* of f is said to be *stable* if for all $t_0 \geq 0$ and $\varepsilon > 0$, there exists $\delta(t_0, \varepsilon)$ such that $x_0 \in B_\delta(x^*)$ implies $x(t) \in B_\varepsilon(x^*)$, $\forall t \geq t_0$. Further, x^* is said to be *asymptotically stable* if x^* is additionally attractive—that is, for all $t_0 \geq 0$, there exists $\delta(t_0)$ such that $x_0 \in B_\delta(x^*)$ implies $\lim_{t\to\infty} \|x(t) - x^*\| = 0$. A critical point is said to be *non-degenerate* if the determinant of the Jacobian of the dynamics at the critical point is non-zero. For a non-degenerate critical point, the Hartman-Grobman theorem (Sastry, 1999) enables us to check the eigenvalues of the Jacobian to determine asymptotic stability. In particular, at a non-degenerate critical point, if the eigenvalues of the Jacobian are in the *open left-half* complex plane, then the critical point is asymptotically stable.

In the stochastic setting, we use chain invariant sets.

Definition G.1. *Given $T > 0$, $\delta > 0$, if there exists an increasing sequence of times t_j with $t_0 = 0$ and $t_{j+1} - t_j \geq T$ for each j and solutions $\xi^j(t)$, $t \in [t_j, t_{j+1}]$ of $\dot{\xi} = F(\xi)$ with initialization $\xi(0) = \xi_0$ such that $\sup_{t\in[t_j,t_{j+1}]} \|\xi^j(t) - z(t)\| < \delta$ for some bounded, measurable $z(\cdot)$, the we call z a (T, δ)–perturbation.*

Lemma G.1 (Hirsch Lemma)**.** *Given $\varepsilon > 0$, $T > 0$, there exists $\bar{\delta} > 0$ such that for all $\delta \in (0, \bar{\delta})$, every (T, δ)–perturbation of $\dot{\xi} = F(\xi)$ converges to an ε–neighborhood of the global attractor set for $\dot{\xi} = F(\xi)$.*

G.2. Learning Stackelberg Solutions for the Leader: A Best Response Analysis

In this supplementary section, we provide convergence results for the leader given that the follower is playing a local best response strategy at each iteration. We consider the stochastic setting in which the leader does not have oracle access to their gradients, but do have an unbiased estimator. As an example, players could be performing policy gradient reinforcement learning or alternative gradient-based learning schemes. Let $\dim(X_i) = m_i$ for each $i \in \{1, 2\}$ and $m = m_1 + m_2$.

Assumption G.1. *The following hold:*

A1a. *The maps $Df_1 : \mathbb{R}^m \to \mathbb{R}^{m_1}$, $D_2f_2 : \mathbb{R}^m \to \mathbb{R}^{m_2}$ are L_1, L_2 Lipschitz, and $\|Df_1\| \leq M_1 < \infty$.*

A1b. *For each $i \in \mathcal{I}$, the learning rates satisfy $\sum_k \gamma_{i,k} = \infty$, $\sum_k \gamma_{i,k}^2 < \infty$.*

A1c. *The noise processes $\{w_{i,k}\}$ are zero mean, martingale difference sequences. That is, given the filtration $\mathcal{F}_k = \sigma(x_s, w_{1,s}, w_{2,s},\ s \leq k)$, $\{w_{i,k}\}_{i\in\mathcal{I}}$ are conditionally independent, $\mathbb{E}[w_{i,k+1}|\ \mathcal{F}_k] = 0$ a.s., and $\mathbb{E}[\|w_{i,k+1}\||\ \mathcal{F}_k] \leq c_i(1 + \|x_k\|)$ a.s. for some constants $c_i \geq 0$, $i \in \mathcal{I}$.*

Suppose that the leader (player 1) operates under the assumption that the follower (player 2) is playing a local optimum in each round. That is, given $x_{1,k}$, $x_{2,k+1} \in \arg\min_{x_2} f_2(x_{1,k}, x_2)$ for which $D_2f_2(x_{1,k}, x_2) = 0$ is a first-order local optimality condition. If, for a given $(x_1, x_2) \in X_1 \times X_2$, $D_2^2 f_2(x_1, x_2)$ is invertible and $D_2f_2(x_1, x_2) = 0$, then the implicit function theorem implies that there exists neighborhoods $U \subset X_1$ and $V \subset X_2$ and a smooth map $r : U \to V$ such that $r(x_1) = x_2$.

Assumption G.2. *For every x_1, $\dot{x}_2 = -D_2f_2(x_1, x_2)$ has a globally asymptotically stable equilibrium $r(x_1)$ uniformly in x_1 and $r : \mathbb{R}^{m_1} \to \mathbb{R}^{m_2}$ is L_r–Lipschitz.*

Consider the leader's learning rule

$$x_{1,k+1} = x_{1,k} - \gamma_{1,k}(Df_1(x_{1,k}, x_{2,k}) + w_{1,k+1}) \tag{19}$$

where $x_{2,k}$ is defined via the map r_2 defined implicitly in a neighborhood of $(x_{1,k}, x_{2,k})$.

Theorem G.1. *Suppose that for each $x \in X$, $D_2^2 f_2$ is non-degenerate and Assumption G.1 holds for $i = 1$. Then, $x_{1,k}$ converges almost surely to an (possibly sample path dependent) equilibrium point x_1^* which is a local Stackelberg solution for the leader. Moreover, if Assumption G.1 holds for $i = 2$ and Assumption G.2 holds, then $x_{2,k} \to x_2^* = r(x_1^*)$ so that (x_1^*, x_2^*) is a differential Stackelberg equilibrium.*

Proof. This proof follows primarily from using known stochastic approximation results. The update rule in (19) is a stochastic approximation of $\dot{x}_1 = -Df_1(x_1, x_2)$ and consequently is expected to track this ODE asymptotically. The main idea behind the analysis is to construct a continuous interpolated trajectory $\bar{x}(t)$ for $t \geq 0$ and show it asymptotically almost surely approaches the solution set to the ODE. Under the given assumptions, results from (Borkar, 2008, §2.1) imply that the sequence generated from (19) converges almost surely to a compact internally chain transitive set of $\dot{x}_1 = -Df_1(x_1, x_2)$. Furthermore, it can be observed that the only internally chain transitive invariant sets of the dynamics are differential Stackelberg equilibria since at any stable attractor of the dynamics $D^2f_1(x_1, r(x_1)) > 0$ and from assumption $D_2^2 f_2(x_1, r(x_1)) > 0$. Finally, from (Borkar, 2008, §2.2), we can conclude that the update from (19) almost surely converges to a possibly sample path dependent equilibrium point since the only internally chain transitive invariant sets for $\dot{x}_1 = -Df_1(x_1, x_2)$ are equilibria. The final claim that $x_{2,k} \to r(x_1^*)$ is guaranteed since r is Lipschitz and $x_{1,k} \to x_1^*$. □

The above result can be stated with a relaxed version of Assumption G.2. In particular, if x^* is a differential Stackelberg equilibrium, then there is a neighborhood $U_1 \times U_2$ of x^* on which x_2 is implicitly defined in terms of x_1 and $\dot{x}_2 = -D_2f_2(x_1, x_2)$ has x_2^* as a locally asymptotically stable equilibrium for any $x_1 \in U_1$.

Corollary G.1. *Consider a differential Stackelberg equilibrium $x^* = (x_1^*, x_2^*)$. Suppose that Assumption G.1 holds for $i = 1, 2$. There exists a neighborhood $U = U_1 \times U_2$ of $x^* = (x_1^*, x_2^*)$ such that for any $x_0 \in U$, x_k converges almost surely to x^*.*

Proof. Consider a differential Stackelberg equilibrium x^* so that $Df_1(x^*) = 0$, $D_2f_2(x^*) = 0$, $\mathrm{spec}(D^2f_1(x^*)) \subset \mathbb{R}_+^\circ$ and $\mathrm{spec}(D_2^2f_2(x^*)) \subset \mathbb{R}_+^\circ$. Since $\det(D_2^2f_2(x^*)) \neq 0$ and $D_2f_2(x_1^*, x_2^*) = 0$, the implicit function theorem (Abraham et al., 1988, Theorem 2.5.7) states that there exists a neighborhood W_1 of x_1^* and a unique function $r : W_1 \to \mathbb{R}^{m_2}$ such that $D_2f_2(x_1, r(x_1)) = 0$ for all $x_1 \in W_1$ and $r(x_1^*) = x_2^*$. Due to the fact that eigenvalues vary continuously, there exists a neighborhood W_2 of x_1^* on which $D^2f_1(x_1, r(x_1)) > 0$ and $D_2^2f_2(x_1, r(x_1)) > 0$. Let U_1 be the non-empty, open set whose closure is contained in $W_1 \cap W_2$. Furthermore, since $D_2^2f_2(x_1^*, x_2^*) > 0$ there exists an open neighborhood U_2 of x_2^* such that $D_2^2f_2(x_1, x_2) > 0$ for all $(x_1, x_2) \in U_1 \times U_2$.

From here, the proof follows the same arguments as the proof of Theorem G.1. □

G.3. Learning Stackelberg Equilibria: Two-Timescale Analysis

Now, let us consider the case where the leader again operates under the assumption that the follower is playing (locally) optimally at each round so that the belief is $D_2f_2(x_{1,k}, x_{2,k}) = 0$, but the follower is actually performing the update $x_{2,k+1} = x_{2,k} + g_2(x_{1,k}, x_{2,k})$ where $g_2 \equiv -\gamma_{2,k}\mathbb{E}[D_2f_2]$. The learning dynamics in this setting are then

$$x_{1,k+1} = x_{1,k} - \gamma_{1,k}(Df_1(x_k) + w_{1,k+1}) \quad (20)$$

$$x_{2,k+1} = x_{2,k} - \gamma_{2,k}(D_2f_2(x_k) + w_{2,k+1}) \quad (21)$$

where $Df_1(x) = D_1f_1(x) + D_2f_1(x)Dr(x_1)$. Suppose that $\gamma_{1,k} \to 0$ faster than $\gamma_{2,k}$ so that in the limit $\tau \to \infty$, the above approximates the singularly perturbed system defined by

$$\begin{aligned} \dot{x}_1(t) &= -Df_1(x_1(t), x_2(t)) \\ \dot{x}_2(t) &= -\tau D_2f_2(x_1(t), x_2(t)) \end{aligned} \quad (22)$$

The learning rates can be seen as stepsizes in a discretization scheme for solving the above dynamics. The condition that $\gamma_{1,k} = o(\gamma_{2,k})$ induces a *timescale separation* in which x_2 evolves on a faster timescale than x_1. That is, the fast transient player is the *follower* and the slow component is the *leader* since $\lim_{k\to\infty} \gamma_{1,k}/\gamma_{2,k} = 0$ implies that from the perspective of the follower, x_1 appears quasi-static and from the perspective of the leader, x_2 appears to have equilibriated, meaning $D_2f_2(x_1, x_2) = 0$ given x_1. From this point of view, the learning dynamics (20)–(21) approximate the dynamics in the preceding section. Moreover, stable attractors of the dynamics are such that the leader is at a local optima for f_1, not just along its coordinate axis but in both coordinates (x_1, x_2) constrained to the manifold $r(x_1)$; this is to make a distinction between differential Nash equilibria in which agents are at local optima aligned with their individual coordinate axes.

A note on the use of timescale separation. The reason for this timescale separation is that the leader's update is formulated using the reaction curve of the follower. In the gradient-based learning setting considered, the reaction curve can be characterized by the set of critical points of $f_2(x_{1,k},\cdot)$ that have a local positive definite structure in the direction of x_2, which is

$$\{x_2|\ D_2f_2(x_{1,k},x_2)=0,\ D_2^2f_2(x_{1,k},x_2)>0\}.$$

This set can be characterized in terms of an implicit map r, defined by the leader's belief that the follower is playing a best response to its choice at each iteration, which would imply $D_2f_2(x_{1,k},x_{2,k})=0$. Moreover, under sufficient regularity conditions, the implicit function theorem (Abraham et al., 1988, Theorem 2.5.7) gives rise to the implicit map $r:U\to X_2:x_1\mapsto x_2$ on a neighborhood $U\subset X_1$ of $x_{1,k}$. We note that when r is defined uniformly in x_1 on the domain for which convergence is being assessed, the update in (2) is well-defined in the sense that the component of the derivative Df_1 corresponding to the implicit dependence of the follower's action on x_1 via r is well-defined and locally consistent. In particular, for a given point $x=(x_1,x_2)$ such that $D_2f_2(x_1,x_2)=0$ with $D_2^2f_2(x)$ an isomorphism, the implicit function theorem implies there exists an open set $U\subset X_1$ such that there exists a unique continuously differentiable function $r:U\to X_2$ such that $r(x_1)=x_2$ and $D_2f_2(x_1,r(x_1))=0$ for all $x_1\in U$. Moreover,

$$Dr(x_1)=-(D_2^2f_2)^{-1}(x_1,r(x_1))D_{21}f_2(x_1,r(x_1))$$

on U. Thus, in the limit of the two-timescale setting, the leader sees the follower as having equilibriated (meaning $D_2f_2\equiv 0$) so that

$$Df_1(x_1,x_2)=D_1f_1(x_1,x_2)+D_2f_1(x_1,x_2)Dr(x_1) \tag{23}$$

The map r is an implicit representation of the follower's reaction curve.

We note that without the timescale separation, it is possible to still prove almost sure asymptotic convergence if the overall continuous time deterministic differential equation $\dot{x}=-\omega_{\mathcal{S}}(x)$ is stable. This is a stronger requirement than assuming that x^* is a differential Stackelberg equilibrium such that the system $\dot{x}_1=-Df_1(x_1,r(x_1))$ is locally asymptotically stable and the system $\dot{x}_2=-D_2f_2(x_1,x_2)$ is locally asymptotically stable uniformly in x_1 on the neighborhood obtained via the implicit mapping theorem.

G.3.1. Stochastic Avoidance of Saddles

It is known that stochastic gradient descent in the single player setting with isotropic noise avoids saddles almost surely (Daneshmand et al., 2018). It is also known that gradient play in non-convex games with stochastic gradients and isotropic noise avoids saddle points of the game dynamics (Mazumdar et al., 2020). These results specialize to the case of Stackelberg learning dynamics with stochastic gradients (i.e., unbiased estimators of the true gradient) and sufficiently rich noise.

The follow results from (Pemantle, 1990) implies saddle avoidance in Stackelberg learning. Consider a general stochastic approximation framework $x_{t+1}=x_t+\gamma_t(h(x_t))+\epsilon_t$ for $h:X\to TX$ with $h\in C^2$ and where $X\subset\mathbb{R}^m$ and where TX denotes the tangent space of X.

Theorem G.2 (Theorem 1 (Pemantle, 1990))**.** *Suppose γ_t is $\mathcal{F}_t$–measurable and $\mathbb{E}[w_t|\mathcal{F}_t]=0$. Let the stochastic process $\{x_t\}_{t\geq 0}$ be defined as above for some sequence of random variables $\{\epsilon_t\}$ and $\{\gamma_t\}$. Let $p\in X$ with $h(p)=0$ and let W be a neighborhood of p. Assume that there are constants $\eta\in(1/2,1]$ and $c_1,c_2,c_3,c_4>0$ for which the following conditions are satisfied whenever $x_t\in W$ and t sufficiently large: (i) p is a linear unstable critical point (i.e., a saddle point), (ii) $c_1/t^\eta\leq\gamma_t\leq c_2/t^\eta$, (iii) $\mathbb{E}[(w_t\cdot v)^+|\mathcal{F}_t]\geq c_3/t^\eta$ for every unit vector $v\in TX$, and (iv) $\|w_t\|_2\leq c_4/t^\eta$. Then $P(x_t\to p)=0$.*

The above classical result directly implies avoidance of saddles in Stackelberg learning.

Theorem 6 (Almost Sure Avoidance of Saddles.)**.** *Consider a game (f_1,f_2) with $f_i\in C^q(\mathbb{R}^{m_1}\times\mathbb{R}^{m_2},\mathbb{R})$, $q\geq 2$ for $i=1,2$ and where without loss of generality, player 1 is the leader. Suppose that for each $i=1,2$, there exists a constant $b_i>0$ such that $\mathbb{E}[(w_{i,t}\cdot v)^+|\mathcal{F}_{i,t}]\geq b_i$ for every unit vector $v\in\mathbb{R}^{m_i}$. Then, Stackelberg learning converges to strict saddle points of the game on a set of measure zero.*

The proof follows directly from showing that the Stackelberg learning update satisfies Theorem G.2, provided the assumptions of the theorem hold. The assumption that $\mathbb{E}[(w_{i,t}\cdot v)^+|\mathcal{F}_{i,t}]\geq b_i$ essentially requires the covariance of the noise to be full-rank, and is made to rule out degenerate cases where the noise forces the dynamics to stay on the stable manifold of strict saddle points. Indeed, this is exactly the goal of isotropic noise in stochastic gradient descent.

G.3.2. Asymptotic Almost Sure Convergence

The following two results are fairly classical results in stochastic approximation. They are leveraged here to making conclusions about convergence to Stackelberg equilibria in hierarchical learning settings.

While we do not need the following assumption for all the results in this section, it is required for asymptotic convergence of the two-timescale process in (20)–(21).

Assumption G.3. *The dynamics $\dot{x}_1 = -Df_1(x_1, r(x_1))$ have a globally asymptotically stable equilibrium.*

Under Assumption G.1–G.3, and the assumption that $\gamma_{1,k} = o(\gamma_{2,k})$, classical results imply that the dynamics (20)–(21) converge almost surely to a compact internally chain transitive set $\mathcal{T}$ of (22); see, e.g., (Borkar, 2008, §6.1-2), (Bhatnagar et al., 2012, §3.3). Furthermore, it is straightforward to see that stable differential Stackelberg equilibria are internally chain transitive sets since they are stable attractors of the dynamics from (22).

Let $t_k = \sum_{l=0}^{k-1} \gamma_{1,l}$ be the (continuous) time accumulated after k samples of the slow component x_1. Define $\xi_{1,s}(t)$ to be the flow of $\dot{x}_1 = -Df_1(x_1(t), r(x_1(t)))$ starting at time s from intialization x_s.

Lemma G.2. *Suppose that Assumptions G.1 and G.2 hold. Then, conditioning on the event $\{\sup_k \sum_i \|x_{i,k}\|^2 < \infty\}$, for any integer $K > 0$, $\lim_{k\to\infty} \sup_{0\leq h\leq K} \|x_{1,k+h} - \xi_{1,t_k}(t_{k+h})\|_2 = 0$ almost surely.*

Proof. The proof follows standard arguments in stochastic approximation. We simply provide a sketch here to give some intuition. First, we show that conditioned on the event $\{\sup_k \sum_i \|x_{1,k}\|^2 < \infty\}$, $(x_{1,k}, x_{2,k}) \to \{(x_1, r(x_1))|\ x_1 \in \mathbb{R}^{d_1}\}$ almost surely. Let $\zeta_k = \frac{\gamma_{1,k}}{\gamma_{2,k}}(Df_1(x_k) + w_{1,k+1})$. Hence the leader's sample path is generated by $x_{1,k+1} = x_{1,k} - \gamma_{2,k}\zeta_k$ which tracks $\dot{x}_1 = 0$ since $\zeta_k = o(1)$ so that it is asymptotically negligible. In particular, $(x_{1,k}, x_{2,k})$ tracks $(\dot{x}_1 = 0, \dot{x}_2 = -D_2f_2(x_1, x_2))$. That is, on intervals $[\hat{t}_j, \hat{t}_{j+1}]$ where $\hat{t}_j = \sum_{l=0}^{j-1} \gamma_{2,l}$, the norm difference between interpolated trajectories of the sample paths and the trajectories of $(\dot{x}_1 = 0, \dot{x}_2 = -D_2f_2(x_1, x_2))$ vanishes a.s. as $k \to \infty$. Since the leader is tracking $\dot{x}_1 = 0$, the follower can be viewed as tracking $\dot{x}_2(t) = -D_2f_2(x_1, x_2(t))$. Then applying Lemma G.1 provided in Appendix G, $\lim_{k\to 0} \|x_{2,k} - r(x_{1,k})\| \to 0$ almost surely.

Now, by Assumption G.1, Df_1 is Lipschitz and bounded (in fact, independent of **A1a.**, since $Df_1 \in C^q$, $q \geq 1$, it is locally Lipschtiz and, on the event $\{\sup_k \sum_i \|x_{i,k}\|_2 < \infty\}$, it is bounded). In turn, it induces a continuous globally integrable vector field, and therefore satisfies the assumptions of Benaïm (1999, Prop. 4.1). Moreover, under Assumptions **A1b.** and **A1c.**, the assumptions of Benaïm (1999, Prop. 4.2) are satisfied, which gives the desired result. ☐

Theorem G.3. *Under Assumptions G.2–G.3 and the assumptions of Theorem G.2, $(x_{1,k}, x_{2,k}) \to (x_1^*, r(x_1^*))$ almost surely conditioned on the event $\{\sup_k \sum_i \|x_{i,k}\|^2 < \infty\}$. That is, the learning dynamics (20)–(21) converge to stable critical points of (22), the set of which includes the stable differential Stackelberg equilibria.*

Proof. Continuing with the conclusion of the proof of Lemma G.2, on intervals $[t_k, t_{k+1}]$ the norm difference between interpolates of the sample path and the trajectories of $\dot{x}_1 = -Df_1(x_1, r(x_1))$ vanish asymptotically; applying Lemma G.1 gives the result. ☐

As with Corollary G.1, we can relax Assumption G.2 and G.3 to local asymptotic stability assumptions.

Theorem 7. *Consider a general sum game (f_1, f_2) with $f_i \in C^q(X, \mathbb{R})$, $q \geq 2$ for $i = 1, 2$ and where, without loss of generality, player 1 is the leader and $\gamma_{1,k} = o(\gamma_{2,k})$. Consider a differential Stackelberg equilibrium $x^* = (x_1^*, x_2^*)$. There exists a neighborhood $U = U_1 \times U_2$ of $x^* = (x_1^*, x_2^*)$ such that for any $x_0 \in U$, x_k converges almost surely to x^*.*

Proof. Consider a differential Stackelberg equilibrium x^* so that $Df_1(x^*) = 0$, $D_2f_2(x^*) = 0$, $\mathrm{spec}(D^2f_1(x^*)) \subset \mathbb{R}_+^\circ$ and $\mathrm{spec}(D_2^2f_2(x^*)) \subset \mathbb{R}_+^\circ$. Since $\det(D_2^2f_2(x^*)) \neq 0$ and $D_2f_2(x_1^*, x_2^*) = 0$, the implicit function theorem (Abraham et al., 1988, Theorem 2.5.7) states that there exists a neighborhood W_1 of x_1^* and a unique function $r: W_1 \to \mathbb{R}^{m_2}$ such that $D_2f_2(x_1, r(x_1)) = 0$ for all $x_1 \in W_1$ and $r(x_1^*) = x_2^*$. Due to the fact that eigenvalues vary continuously, there exists a neighborhood W_2 of x_1^* on which $D^2f_1(x_1, r(x_1)) > 0$ and $D_2^2f_2(x_1, r(x_1)) > 0$. Let U_1 be the non-empty, open set whose closure is contained in $W_1 \cap W_2$. Furthermore, since $D_2^2f_2(x_1^*, x_2^*) > 0$ there exists an open neighborhood U_2 of x_2^* such that $D_2^2f_2(x_1, x_2) > 0$ for all $(x_1, x_2) \in U_1 \times U_2$. The remainder of the proof follows the same arguments as the proof of Lemma G.2. ☐

G.3.3. Zero Sum Convergence

Leveraging the results in Section 3, the convergence guarantees are stronger since in zero-sum settings all critical points of the limiting continuous time dynamical system are Stackelberg; this contrasts with the Nash equilibrium concept.

Corollary G.2. *Consider a zero-sum setting $(f, -f)$ on $X = X_1 \times X_2 = \mathbb{R}^m$. Under the assumptions of Theorem G.3, conditioning on the event $\{\sup_k \sum_i \|x_{i,k}\|^2 < \infty\}$, the learning dynamics (20)–(21) asymptotically converge to the stable differential Stackelberg equilibrium almost surely.*

The proof of this theorem follows the above analysis and invokes Proposition 2.

As a final note, which was remarked on previously, it is possible to convert the asymptotic results above, both global and local convergence guarantees, to high-probability concentration bounds using the recent results in (Borkar & Pattathil, 2018; Thoppe & Borkar, 2019). Generally, the results (and their proofs) in this subsection leverage classical results from stochastic approximation (Bhatnagar et al., 2012; Borkar, 2008; Kushner & Yin, 2003; Benaïm, 1999).

H. Additional Numerical Simulations and Details

We now present additional numerical simulations and details. All of our code is available at the following github link `github.com/fiezt/ICML-2020-Implicit-Stackelberg-Learning`.

H.1. Regularizing the Follower's Implicit Map

The derivative of the implicit function used in the leader's update requires the follower's Hessian to be an isomorphism. In practice, this may not always be true along the learning path. Consider the modified update

$$\begin{aligned} x_{k+1,1} &= x_{k,1} - \gamma_1(D_1 f_1(x_k) - D_{21} f_2(x_k)^\top (D_2^2 f_2(x_k) + \eta I)^{-1} D_2 f_1(x_k)) \\ x_{k+1,2} &= x_{k,2} - \gamma_2 D_2 f_2(x_k), \end{aligned}$$

in which we regularize the inverse of $D_2^2 f_2$ term. This update can be derived from the following perspective. This result can be seen by examining first and second order sufficient conditions for the leader's optimization problem given the regularized conjecture about the follower's update, i.e.

$$\arg\min_{x_1} \left\{ f_1(x_1, x_2) |\; x_2 \in \arg\min_y f_2(x_1, y) + \frac{\eta}{2}\|y\|^2 \right\},$$

and for the problem follower is actually solving with its update $\arg\min_{x_2} f_2(x_1, x_2)$.

The leader then views the follower as updating via

$$x_{k+1,2} = x_{k,2} - \gamma_2(D_2 f_2(x_k) + \eta x_{k,2})$$

so that the derivative of the implicit map is given by $(D_2^2 f_2(x_k) + \eta I)^{-1} D_{21} f_2(x_k)$.

Then, the approximate Stackelberg update is given by

$$\begin{aligned} x_{k+1,1} &= x_{k,1} - \gamma_1(D_1 f_1(x_k) - D_{21} f_2(x_k)^\top \left(D_2^2 f_2(x_k) + \eta I\right)^{-1} D_2 f_1(x_k)) \\ x_{k+1,2} &= x_{k,2} - \gamma_2 D_2 f_2(x_k). \end{aligned}$$

In our GAN experiments, we use the regularized update since it is quite common for the discriminator's Hessian to be ill-conditioned if not degenerate. We define sufficient conditions for an equilibrium with respect to the regularized dynamics.

Proposition H.1 (Regularized Stackelberg: Sufficient Conditions)**.** *A point x^* such that the first order conditions $D_1 f_1(x) - D_{21} f_2(x)^\top (D_2^2 f_2(x) + \eta I)^{-1} D_2 f_1(x) = 0$ and $D_2 f_2(x) = 0$ hold, and such that $D(D_1 f_1(x) - D_{21} f_2(x)^\top (D_2^2 f_2(x) + \eta I)^{-1} D_2 f_1(x)) > 0$ and $D_2^2 f_2(x) > 0$ is a differential Stackelberg equilibrium with respect to the regularized dynamics.*

H.2. Computing the Stackelberg Update and Schur Complement

The learning rule for the leader involves computing an inverse-Hessian-vector product for the $D_2^2 f_2(x)$ inverse term and Jacobian-vector product for the $D_{21} f_2(x)$ term. These operations can be done efficiently in Python by utilizing

Jacobian-vector products in auto-differentiation libraries combined with the `sparse.LinearOperator` class in `scipy`. These objects can also be used to compute their eigenvalues, inverses, or the Schur complement of the game dynamics using the `scipy.sparse.linalg` package. We found that the conjugate gradient method `cg` can compute the regularized inverse-Hessian-vector products for the leader update accurately with 5 iterations and a warm start.

The operators required for the leader update can be obtained by the following. Consider the Jacobian of the simultaneous gradient descent learning dynamics $\dot{x} = -\omega(x)$ at a critical point for the general sum game (f_1, f_2):

$$J(x) = \begin{bmatrix} D_1^2 f_1(x) & D_{12} f_1(x) \\ D_{21} f_2(x) & D_2^2 f_2(x) \end{bmatrix}.$$

Its block components consist of four operators $D_{ij} f_i(x) : X_j \to X_i,\ i, j \in \{1, 2\}$ that can be computed using forward-mode or reverse-mode Jacobian-vector products. Instantiating these operators as a linear operator in `scipy` allows us to compute the eigenvalues of the two player's individual Hessians. Properties such as the real eigenvalues of a Hermitian matrix or complex eigenvalues of a square matrix can be computed using `eigsh` or `eigs` respectively. Selecting to compute the smallest or largest k eigenvalues—sorted by either magnitude, real or imaginary values—allows one to examine the positive-definiteness of the operators.

Operators can be combined to compute other operators relatively efficiently for large scale problems without requiring to compute their full matrix representation. For an example, take the Schur complement of the Jacobian above at fixed network parameters $x \in X_1 \times X_2$, $D_1^2(x) - D_{12} f_1(x)(D_2^2 f_2)^{-1}(x) D_{21} f_2(x)$. We create an operator $S_1(x) : X_1 \to X_1$ that maps a vector v to $p - q$ by performing the following four operations: $u = D_{21} f_2(x) v$, $w = (D_2^2 f_2)^{-1}(x) u$, $q = D_{12} f_1(x) w$, and $p = D_1^2(x) v$. Each of the operations can be computed using a single backward pass through the network except for computing w, since the inverse-Hessian requires an iterative method. It solves the linear equation $D_2^2 f_2(x) w = u$ and there are various available methods: we tested (bi)conjugate gradient methods, residual-based methods, or least-squares methods, and each of them provide varying amounts of error when compared with the exact solution. For our mixture of Gaussians and MNIST GANs, we found that computing the leader update using the conjugate gradient method with maximum of five iterations and warm-start works well. We compared using the real Hessian for smaller scale problems and found the estimate to be within numerical precision.

A similar procedure is used to compute a variety higher-order derivatives. For instance, the regularized total derivative of the leader's update is the total derivative of $Df_1(x_1, r(x_1))$. To compute the spectrum of such an operator, we create a function $v \mapsto D(Df_1(x_1, r(x_1)))v$ that takes a vector $v \in \mathbb{R}^{m_1}$ and returns

$$\begin{aligned} D(Df_1(x_1, r(x_1)))v = {} & D_1^2 f_1(x_1, r(x_1))v + D_{12} f_1(x_1, r(x_1)) Dr_\eta(x_1) v \\ & + (D_{12} f_1(x_1, r(x_1)) + Dr_\eta(x_1)^\top D_2^2 f_1(x_1, r(x_1)) Dr_\eta(x_1) v \\ & + D_2 f_1(x_1, r(x_1))^\top D^2 r_\eta(x_1) v \end{aligned} \tag{24}$$

where the last higher order term is assumed to be zero, the regularized variation $Dr_\eta(x_1) = (D_2^2 f_2(x) + \eta I)^{-1} D_{21} f_2(x)$, and regularization term $\eta > 0$. The above derivative can be written as a composition of Jacobian-vector product operators and least squares problems, thus can be computed efficiently with auto-differentiation tools.

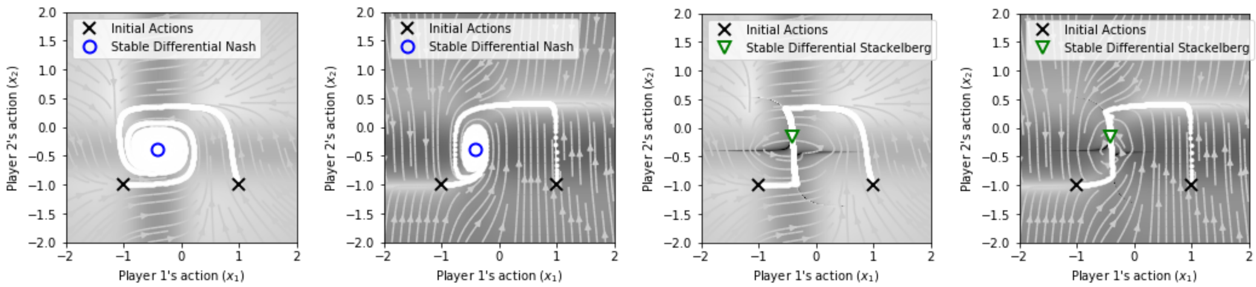
H.3. Parameterized Bilinear Game Experiments

In the continuous game framework, player's actions are continuous. To represent strategies with discrete actions, continuous probability distributions can be employed as mixed strategies over the discrete actions. The gradient-based methods developed in this paper thus can be used to solve for equilibria in the parameterized strategy space. Consider the following game $\mathcal{G} = (f_1, f_2)$ with costs given by $f_1(x_1, x_2) = \pi(x_1)^\top A \pi(x_2) + \frac{\eta_1}{2}\|x_1\|_2^2$ and $f_2(x_1, x_2) = \pi(x_1)^\top B \pi(x_2) + \frac{\eta_2}{2}\|x_2\|_2^2$ where

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},\ B = \begin{bmatrix} 1/2 & 1 \\ 1 & 1/2 \end{bmatrix} \tag{25}$$

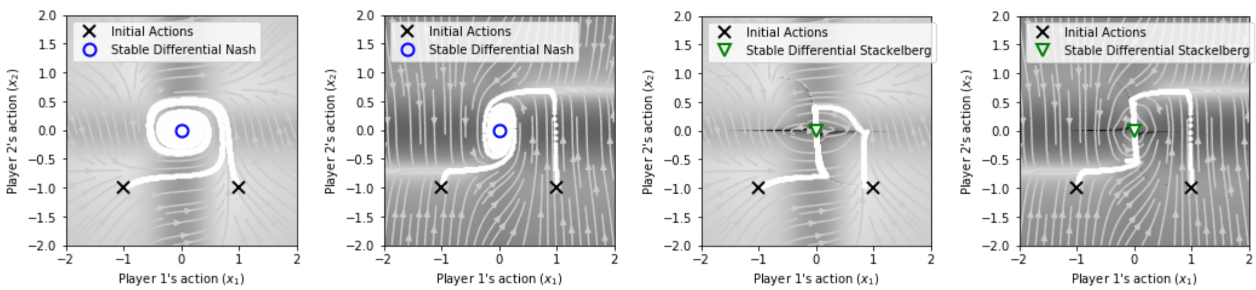
are the matrices representing the bimatrix game with player 1 as the row player and player 2 as the column player. We represent the mixed policy of two discrete actions with a sigmoid-based probability distribution on the simplex, $\pi : \mathbb{R} \to \Delta^1$, given by

$$\pi(x) = (e^{a_1 x + b_1},\ e^{a_2 x + b_2})/(e^{a_1 x + b_1} + e^{a_2 x + b_2})$$



(a) simgrad ($\gamma_1 = \gamma_2$) (b) simgrad ($\gamma_1 = \gamma_2/10$) (c) Stackelberg ($\gamma_1 = \gamma_2$) (d) Stackelberg ($\gamma_1 = \gamma_2/10$)

Figure 6. **Parameterized bilinear game.** Parameters: $(a_1, a_2) = (2.5, -2.5)$ and $(b_1, b_2) = (1, -1)$. (a)–(b): for simgrad, we observe convergence to an ε-neighborhood of a differential Nash equilibria at $(x_1^*, x_2^*) = (-.4, -.4)$. (c)–(d): for the Stackelberg learning dynamics, we observe convergence to an ε-neighborhood of a differential Stackelberg equilibria at $(x_1^*, x_2^*) = (-.4, -.16)$ which corresponds to player 1 choosing the action associated with the top row with probability 0.5 and player 2 choosing the action associated with the first column with probability 0.77. The effects of time-scale separation is visualized as the light colored horizontal path, showing a low gradient norm along player 2's reaction curve.



(a) simgrad ($\gamma_1 = \gamma_2$) (b) simgrad ($\gamma_1 = \gamma_2/10$) (c) Stackelberg ($\gamma_1 = \gamma_2$) (d) Stackelberg ($\gamma_1 = \gamma_2/10$)

Figure 7. **Parameterized bilinear game.** Parameters: $(a_1, a_2) = (2.5, -2.5)$ and $(b_1, b_2) = (0, 0)$. (a)–(b): for simgrad, we observe convergence to an ε-neighborhood of a differential Nash equilibria at $(x_1^*, x_2^*) = (0, 0)$. (c)–(d): for the Stackelberg learning dynamics, we observe convergence to an ε-neighborhood of a differential Stackelberg equilibria at $(x_1^*, x_2^*) = (0, 0)$ which corresponds to player 1 choosing the action associated with the top row with probability 0.5 and player 2 choosing the action associated with the first column with probability 0.5. The effects of time-scale separation is visualized as the light colored horizontal path, showing a low gradient norm along player 2's reaction curve. Note that due to the choice of parameters (b_1, b_2), the regularization is penalizing for any deviation from the equilibrium at $(1/2, 1/2)$ in the policy space.

where the parameters $a_i, b_i, i = 1, 2$ are constants that scale and shift the parameterization. This parameterization scheme can be extended to $d + 1$ actions using d variables. For two actions, we require that a_1 and a_2 have opposite signs. We employ a 2-norm regularization of each player's individual action to regularize each agent towards the interior of the simplex.

The bimatrix game admits a unique mixed Nash equilibrium of $(1/2, 1/2)$ for player 1 and $(1/2, 1/2)$ for player 2. If the game is played sequentially with the leader being player 1, the mixed Stackelberg equilibrium of the game is (π_1, π_2) with $\pi_1 = (1/2, 1/2)$ and any policy π_2 in the simplex for the follower. At this strategy, the cost the leader incurs is independent of the follower's strategy. We refer to Basar & Olsder (1998, §3.6) for discussion on the mixed Stackelberg equilibrium of this bimatrix game. For the softmax parameterized policy class we consider, using $(a_1, a_2) = (2.5, -2.5)$, $(b_1, b_2) = (1, -1)$, $\pi(-0.4) = (1/2, 1/2)$. That is, the parameter $x = -0.4$ corresponds to the policy $(1/2, 1/2)$. On the other hand, if $(a_1, a_2) = (2.5, -2.5)$, $(b_1, b_2) = (0, 0)$, then $\pi(0) = (1/2, 1/2)$.

We plot the the vector field ω and $\omega_{\mathcal{S}}$ and its norm, along with simulations of their discrete time dynamics in Figures 6 and 7. We use parameters $(a_1, a_2) = (2.5, -2.5)$, and regularization $\eta = 0.1$. For the parameters (b_1, b_2), we explore two different pairs: $(b_1, b_2) = (1, -1)$ and $(b_1, b_2) = (0, 0)$. The latter is such that the regularization term is penalizing for any deviation from the equilibrium parameter values, while the former is such that the regularization is penalizing for any deviation from

$(0, 0)$ while the equilibrium is at $(0.4, y)$ for any $y \in [0, 1]$.

The shading of the action space indicates the norm of the dynamics: darker has a larger norm. Different parameterization constants or regularization weights will affect the outcome of the gradient-based learning.

The timescale separation improves the convergence properties of the Stackelberg learning dynamics as it encourages the dynamics to converge to the follower's best-response curve. We observe the distinctly lighter path the shaded plots of Figure 6 (b) and (d), where the follower's response curve runs horizontal to the plot. Comparing plots Stackelberg learning in Figures 6 (c) and (d), we observe that with timescale separation, the paths of the learning dynamics converges first to the manifold on the ridge, then towards the stationary point along the manifold. Doing so prevents the trajectory from being perturbed by the 'cliffs', visualized by the dark cusps with large gradient norm corresponding to area where the follower's Hessian is poorly conditioned. The role of timescale separation is emphasized in this numerical simulation.

H.4. Details on GAN Experiments

This section includes complete details on the training process and hyper-parameters selected in the mixture of Gaussian and MNIST experiments. We also include further experiments for the mixture of Gaussian examples.

H.4.1. Mixture of Gaussians GAN

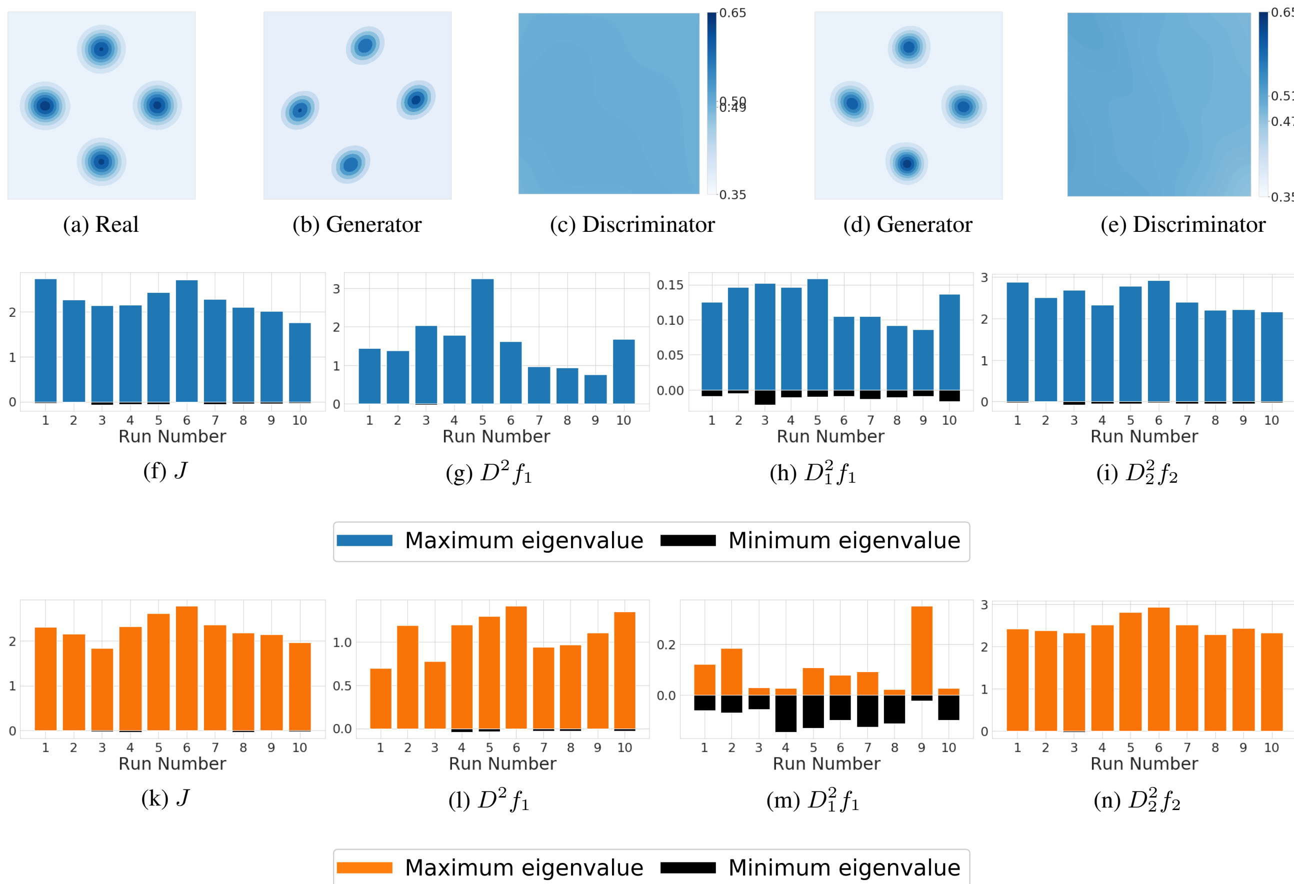
The underlying data distribution for the diamond experiment consists of Gaussian distributions with means given by $\mu = [1.2\sin(\omega), 1.2\cos(\omega)]$ for $\omega \in \{k\pi/2\}_{k=0}^{3}$ and each with covariance $\sigma^2 I$ where $\sigma^2 = 0.15$. The underlying data distribution for the circle experiment consists of Gaussian distributions with means given by $\mu = [\sin(\omega), \cos(\omega)]$ for $\omega \in \{k\pi/4\}_{k=0}^{7}$ and each with covariance $\sigma^2 I$ where $\sigma^2 = 0.05$. Each sample of real data given to the discriminator is selected uniformly at random from the set of Gaussian distributions.

We train the generator using latent vectors $z \in \mathbb{R}^{16}$ sampled from a standard normal distribution in each training batch. The discriminator is trained using input vectors $x \in \mathbb{R}^2$ sampled from the underlying distribution in each training batch. The batch size for each player in the game is 256. The network for the generator contains two hidden layers, each of which contain 32 neurons. The discriminator network consists of a single hidden layer with 32 neurons and it has a sigmoid activation following the output layer. We let the activation function following the hidden layers in the generator network be the Tanh function and the ReLU function in the diamond and circle experiments, respectively. The initial learning rates for each player and for each learning rule is 0.0001 and 0.0004 in the diamond and circle experiments, respectively. The objective for the game in the diamond experiment is the saturating GAN objective and in the circle experiment it is the non-saturating GAN objective. We update the parameters for each player and in each experiment using the Adam optimizer with the default parameters of $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. The learning rate for each player is decayed exponentially such that $\gamma_{i,k} = \gamma_i \nu_i^k$. We let $\nu_1 = \nu_2 = 1 - 10^{-7}$ for simultaneous gradient descent and $\nu_1 = 1 - 10^{-5}$ and $\nu_1 = 1 - 10^{-7}$ for the Stackelberg update. Finally, we regularize the implicit map of the follower as detailed in Appendix H.1 using the parameter $\eta = 1$ and similarly in computing the eigenvalues of $D^2 f_1$ as detailed in (24).

Previously we showed the best runs of the 10 initial seeds we ran for each algorithm for each Gaussian configuration. We now explore further the results over the runs. In general, we found that the conclusions that could be drawn from the experiments were consistent across the runs. To demonstrate this, we provide additional simulation results for the diamond configuration in Figure 8 and the circle configuration in Figure 9. The generator and discriminator outputs we show for taken to be the 5th best of the 10 runs. In the eigenvalue plots in Figure 8 and Figure 9, we show for each of the 10 runs the minimum and maximum real eigenvalue parts as this determines stability and if the dynamics reached an equilibrium. In particular, the black bars show the minimum real parts of the eigenvalues and for several of the plots they are not visible since they are near zero. In Figure 8, we again see reasonable performance for both simultaneous gradient descent and Stackelberg learning in terms of the generator and the discriminator. Moreover, the eigenvalues of the follower Hessian are consistently near zero and include negative values on the scale of the positive values. In Figure 9, we again see that simultaneous gradient cycles along the learning path and Stackelberg learning consistently is stable and generates realistic output. Note that the eigenvalues are more variable for simultaneous gradient descent since for some of the runs the generator was still cycling.

H.4.2. MNIST GAN Details

The GAN trained with Stackelberg learning on MNIST had a version of the DCGAN architecture (Radford et al., 2015) adapted to handle 28×28 images. In Tables 1 and 2 we provide the specific architectures for both the generator and



(a) Real (b) Generator (c) Discriminator (d) Generator (e) Discriminator

(f) J (g) $D^2 f_1$ (h) $D_1^2 f_1$ (i) $D_2^2 f_2$

(k) J (l) $D^2 f_1$ (m) $D_1^2 f_1$ (n) $D_2^2 f_2$

Figure 8. Convergence to differential Stackelberg equilibria that are not differential Nash equilibria for `simgrad` (top) and Stackelberg learning dynamics (bottom). The real distribution is (a) and a sample of a generator and discriminator for each is plotted in (b)–(e) where from (b)–(c) is simultaneous gradient descent and (d)–(e) is Stackelberg learning. The minimum and maximum real parts of the eigenvalues of the game objects (f)–(m) are shown for ten random initial seeds where from (f)–(i) is simultaneous gradient descent and (k)–(n) is Stackelberg learning.

the discriminator in the experiment. Our implementation is in PyTorch and we describe the networks by the parameters passed into nn.Sequential class. Any omitted parameters are the defaults. For our training process, the MNIST images were normalized to the range $[-1, 1]$. Each sample of real data given to the discriminator is selected sequentially from a shuffled version of the dataset. We train using a batch size of 256 and a latent dimension of 100 sampled from a standard normal distribution in each training batch. We initialize the weights of the networks using a zero-centered centered Normal distribution with standard deviation 0.02, optimize using Adam with parameters $\beta_1 = 0.5$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$, and set the initial learning rates to be 2×10^{-4}. The learning rate for each player is decayed exponentially such that $\gamma_{i,k} = \gamma_i \nu_i^k$ and $\nu_1 = 1 - 10^{-5}$ and $\nu_2 = 1 - 10^{-7}$. We regularize the implicit map of the follower as detailed in Appendix H.1 using the parameter $\eta = 10000$. If we view the regularization as a linear function of the number of parameters in the discriminator, then this selection of regularization is nearly equal to that from the Gaussian experiments. The Inception score was calculated using a LeNet classifier following (Berard et al., 2020). Each time we calculated a score we used $N = 5000$ samples and $k = 1$ split. This was simply done to speed up the computation since we observed that for the common choice when using the Inception network of $N = 50000$ samples and $k = 10$ splits the scores were nearly identical to that from using $N = 5000$ samples and $k = 1$ splits.

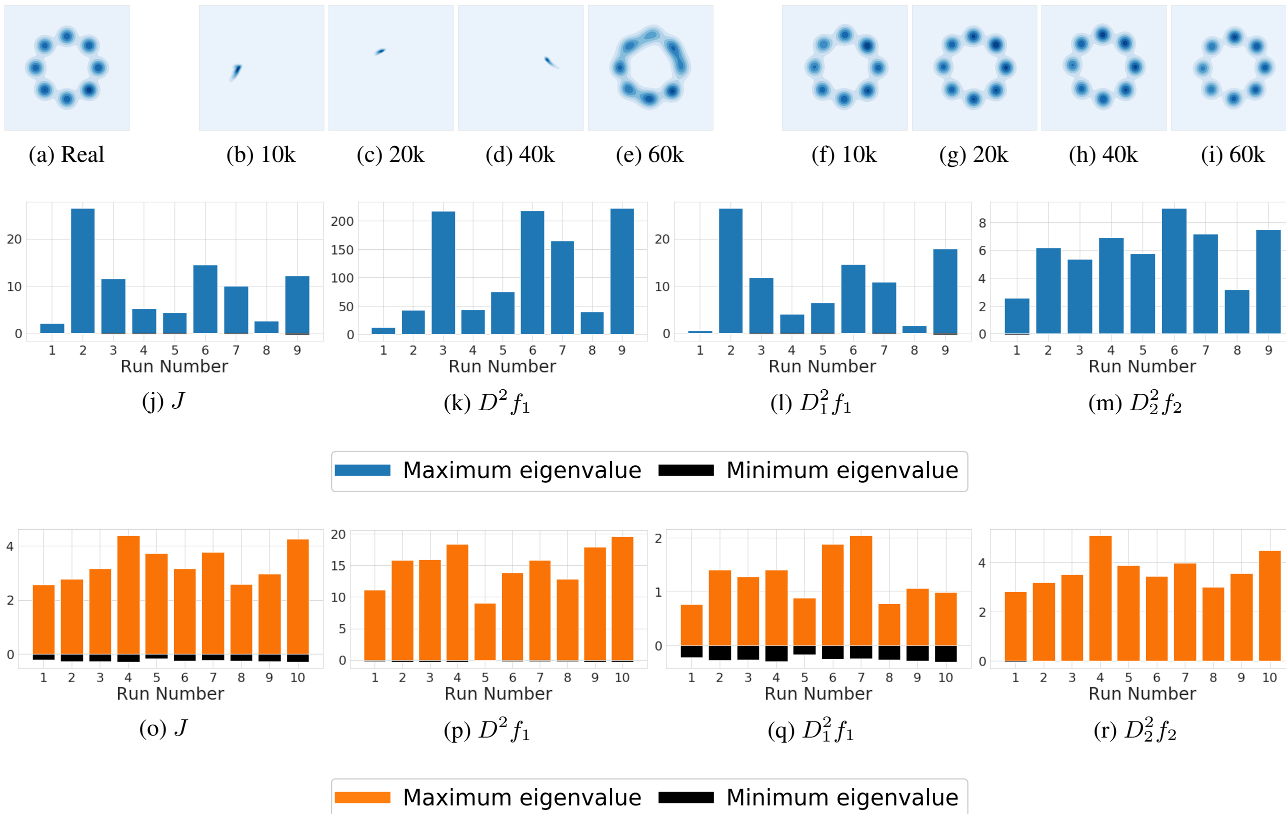


Figure 9. The generator performance along the learning trajectory for simultaneous gradient descent in (b)–(e) and for Stackelberg learning in (f)–(i). The eigenvalues of game objects in (j)–(m) are for simultaneous gradient descent and indicate `simgrad` converges to differential Nash equilibria, while the eigenvalues of game objects (o)–(r) indicate Stackelberg learning converges to a differential Stackelberg equilibria that is not a differential Nash equilibria.

Module	In Channels	Out Channels	Kernel Size	Stride	Padding	Bias
ConvTranspose2d, BatchNorm2d, ReLU	100	512	4	1	0	False
ConvTranspose2d, BatchNorm2d, ReLU	512	256	4	2	1	False
ConvTranspose2d, BatchNorm2d, ReLU	256	128	4	2	1	False
ConvTranspose2d, BatchNorm2d, ReLU	128	512	4	1	0	False
ConvTranspose2d, BatchNorm2d, Tanh	64	1	1	1	2	False

Table 1. Generator Network PyTorch Parameters for the nn.Sequential class in the MNIST experiment.

Module	In Channels	Out Channels	Kernel Size	Stride	Padding	Bias
Conv2d, LeakyReLU(0.2)	1	64	4	2	1	False
Conv2d, BatchNorm2d LeakyReLU(0.2)	64	128	4	2	1	False
Conv2d, BatchNorm2d LeakyReLU(0.2)	128	256	4	2	1	False
Conv2d, Sigmoid	256	1	4	2	1	False

Table 2. Discriminator Network PyTorch Parameters for the nn.Sequential class in the MNIST experiment.