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

Anonymous Authors

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 fixed points or local equilibria. A number of techniques have been proposed including optimistic and extra-gradient algorithms (Daskalakis & Panageas, 2018; Daskalakis et al., 2018; Mertikopoulos et al., 2019), gradient adjustments (Balduzzi et al., 2018; Mescheder et al., 2017), and opponent modeling methods (Foerster et al., 2018; Letcher et al., 2019; Schäfer & Anandkumar, 2019; Zhang & Lesser, 2010). However, only a select number of algorithms can guarantee convergence to stable fixed points satisfying sufficient conditions for a local Nash equilibrium (LNE) (Adolphs et al., 2019; Mazumdar 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, 1966; 1967; 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 (Jin et al., 2019; Metz et al., 2017). In the work of Metz et al. (2017) on unrolled GANs, the follower (discriminator) is allowed to ‘roll out’ parameter updates until reaching an approximate local optimum between each update of the leader (generator). This behavior intuitively approximates a hierarchical order of play. The success of the unrolling method as a training mechanism provides some evidence supporting the efficacy of local Stackelberg equilibrium (LSE) as a solution concept. However, much remains to be discovered on the
role of such equilibria in the game optimization landscape. The methods of Metz et al. (2017) are formalized by Jin et al. (2019), who show in zero-sum games the stable fixed points of simultaneous gradient descent with a timescale separation between players approaching infinity satisfy sufficient conditions for a LSE. It remains an open question to develop implementable algorithms with similar guarantees.

**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 stable fixed points satisfying sufficient conditions for a LSE in zero-sum games.

We provide a characterization of LSE via sufficient conditions on player 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 fixed points of simultaneous gradient descent dynamics that are DSE and not DNE. This insight gives meaning to a broad class of fixed 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 fixed points of the dynamics are DSE and such equilibria must be stable fixed 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.

This work finishes with an empirical study. We demonstrate our proposed dynamics mitigate rotational dynamics when training GANs and provide an example in which simultaneous gradient descent avoids DNE and converges to the neighborhood of a DSE. Given recent evidence that standard training techniques do not converge to DNE but instead to stable points that are locally optimal for the discriminator (Berard et al., 2019), this motivates future investigation to determine if successful training methods often reach DSE.

### 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 $X = X_1 \times X_2$ with $X_1$ and $X_2$ denoting the action spaces of the leader and follower, respectively. Here $X$ is a subset of an $m$-dimensional Euclidean space and similarly, each $X_i$ is a subset of an $m_i$-dimensional Euclidean space. 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 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. This game is called a *Stackelberg game*.

In a Stackelberg game, the leader aims to solve the optimization problem given by $\min_{x_1 \in X_1} \{ f_1(x_1, x_2) \mid x_2 \in \arg \min_{y \in X_2} f_2(x_1, y) \}$ and the follower aims to solve the optimization problem $\min_{x_2 \in X_2} f_2(x_1, x_2)$. We can compare the structure of the Stackelberg game with a simultaneous play game in which each player 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.

#### 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 (LNIE)). 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 \( x^*_i \in U_i \) is a local Stackelberg solution for the leader if, \( \forall x_1 \in U_1 \),

\[
\sup_{x_2 \in R_{U_2}(x^*_1)} f_1(x^*_1, x_2) \leq \sup_{x_2 \in R_{U_2}(x_1)} f_1(x_1, x_2),
\]

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 equilibrium 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. The following gives sufficient conditions for a LNIE 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 \( D^2_{ij} f_i(x^*) > 0 \) for each \( i \in \{1, 2\} \).

Analogous sufficient conditions can be stated to characterize a LSE from Definition 2.

**Definition 4** (Differential Stackelberg (DSE)). The joint strategy \( x^* = (x^*_1, x^*_2) \in X \) is a differential Stackelberg equilibrium if \( D f_1(x^*) = 0 \), \( D^2 f_2(x^*) > 0 \), and \( D^2_{22} f_2(x^*) > 0 \) where \( x^*_2 = r(x^*_1) \) and \( r(\cdot) \) implicitly defined by \( D^2_{22} f_2(x^*) = 0 \).

Game Jacobians play a key role in determining stability of critical points. Let \( \omega(x) = (D_1 f_1(x), D_2 f_2(x)) \) be the vector of individual gradients for the simultaneous play game and \( \omega_S(x) = (D f_1(x), D^2_{22} f_2(x)) \) as the equivalent for the Stackelberg game. Observe that \( D f_1 \) 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 believes the follower will play a best response to \( x_1 \). We note that the reaction curve of the follower to \( x_1 \) may not be unique. However, under sufficient conditions on a local Stackelberg solution \( x \), locally \( D^2_{22} f_2(x) = 0 \) and \( \text{det}(D^2_{22} f_2(x)) \neq 0 \) so that \( D f_1 \) is well defined via the implicit mapping theorem (Lee, 2012).

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

\[
J(x) = \begin{bmatrix}
D^2_{12} f_1(x) & D_{12} f_1(x) \\
D_{21} f_2(x) & D^2_{22} f_2(x)
\end{bmatrix}.
\]

Similarly, the vector field \( \omega_S(x) \) serves as the foundation of the learning dynamics we formulate in Section 2.3 and analyze throughout. The Jacobian of the Stackelberg vector field \( \omega_S(x) \) is given by

\[
J_S(x) = \begin{bmatrix}
D_1 f_1(x) & D_2 f_1(x) \\
D_{21} f_2(x) & D^2_{22} f_2(x)
\end{bmatrix}.
\]

We say a critical point is stable with respect to a vector field \( V(x) \) if the spectrum of the Jacobian of \( V \) is in the open left-half complex plane; e.g., \( x \) such that \( \omega(x) = 0 \) is stable if \( \text{spec}(-J(x)) \subset C_- \). A critical point is non-degenerate if the determinant of the vector field Jacobian is non-zero.

Noting that the Schur complement of \( J_S(x) \) with respect to \( D^2_{22} f_2(x) \) is identically \( D^2 f_2(x, r(x_1)) \), we give alternative but equivalent sufficient conditions as those in Definition 4 in terms of \( J_S(x) \). This will be used to explicitly draw a connection between the Stackelberg learning dynamics and the sufficient conditions for a local Stackelberg. Here, \( S_1(\cdot) \) denotes the Schur complement of \( \cdot \) with respect to the bottom block matrix in \( \cdot \).

**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 \( D^2 f_2(x^*) \neq 0 \) and \( \omega_S(J_S(x^*)) \neq 0 \) if and only if \( x^* \) is a DSE. Moreover, in zero-sum games, \( S_1(J_S(x^*)) = S_1(J(x)) \).

The proof is in Appendix G.1.

**Genericity, Structural Stability, and Bilinear Games.** 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 LNIE and structurally stable in the classes of zero-sum and general-sum continuous games, respectively (Mazumdar & Ratliff, 2019; Ratliff
We obtain analogous results for we study for each player in discrete time is given by det(\(D_f(x_1, x_2)\)) = 0 \forall x \in X. Since such games belong to a degenerate class in the context of the generality 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 B. Finally, for bimatrix games with finite actions it is common to reparameterize the details can be found in Appendix B. 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 D on a parameterized bilinear game.

2.3. Stackelberg Learning Dynamics

Recall that \(\omega_S(x) = (Df_1(x), D_2f_2(x))\) is the vector field for Stackelberg games and it, along with its Jacobian \(J_S(x)\), characterize sufficient conditions for a DSE. Letting \(\omega_{S,i}\) be the \(i\)-th component of \(\omega_S\), the leader total derivative is \(\omega_{S,1}(x) = D_1f_1(x) - D_{21}f_2(x) \cdot (D_2^2f_2(x))^{-1}D_2f_1(x)\) where \(Dr \equiv -(D_2^2f_2(x))^{-1}D_{21}f_2(x)\) with \(r\) defined by the implicit function theorem (Lee, 2012) in a neighborhood of a differential Stackelberg, meaning where \(\omega_{S,2}(x) = 0\) with \(\det(D_2^2f_2(x)) \neq 0\) (which is also holds generically at critical points by Lemma 7). The Stackelberg learning rule we study for each player in discrete time is given by

\[
x_{i,k+1} = x_{i,k} - \gamma_{i,k}h_{S,i}(x_k).
\]

In deterministic learning players have oracle gradient access so \(h_{S,i}(x) = \omega_{S,i}(x)\). In stochastic learning players have unbiased gradient estimates and \(h_{S,i}(x_k) = \omega_{S,i}(x_k) + w_{k+1,i}\) where \(\{w_{i,k}\}\) is player \(i\)'s noise process.

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 E.

Let us first show that for zero-sum games, all stable critical points of \(\dot{x} = -\omega_S(x)\) 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 \(\dot{x} = -\omega_S(x)\) if and only if \(x\) is a DSE. Moreover, if \(f\) is generic, a point \(x\) is a stable critical point of \(\dot{x} = -\omega_S(x)\) if and only if it is a LSE.

The result follows from the structure of the Jacobian of \(\omega_S(x)\), 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 & Ratliff (2018)).

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.

**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 is obtained by examining the first Schur complement of the Jacobian of the simultaneous gradient descent dynamics, and noting that it is exactly the second order sufficient conditions for the leader in the Stackelberg game. The second statement follows from the fact 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.

The preceding result indicates that recent works seeking DNE are also seeking DSE. This leaves the question of the meaning of stable points of \(\dot{x} = -\omega(x)\) which are not DNE. **Finding Meaning in Spurious Stable Fixed 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 \(D_1^2f(x)\) and \(-D_2^2f(x)\) are both not positive definite but \(J(x)\) has positive eigenvalues. 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
reaching DNE, critical points show a number of successful approaches to
result since recent works have proposed schemes to avoid
able fixed points which are not DNE, implying some stable fixed points of
points of simultaneous gradient descent which are DNE, but not DNE.

In Appendix F, we provide necessary and sufficient condi-
tions for attractors at which the follower’s Hessian is positive
definite to be DSE. Taking intuition from the expression
$S_1(J(x)) = D^2_f(x) - D_{21}f(x)^T (D^2_f(x))^{-1} D_{21}f(x)$,
the conditions are derived from relating $\text{spec}(D^2_f)$ to
$\text{spec}(D^2_{2j})$ via $D_{12}f$. To illustrate this fact, consider the fol-
lowing example in which stable points are DSE and not DNE—
meaning points $x \in X$ at which $D^2_f(x) \neq 0$, $-D^2_f(x) > 0$
and $\text{spec}(-J(x^*)) \subset \mathbb{C}^n$ and $S_1(J(x)) > 0$.

Example: Non-Nash Attractors are Stackelberg. Con- 
sider the zero-sum game defined by
$$f(x) = -e^{-0.01(x_1^2 + x_2^2)}((ax_1^2 + x_2)^2 + (bx_2^2 + x_1)^2).$$

Let player 1 be the leader who aims to minimize $f$ with re-
spect 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 $-D^2_f(x) > 0$
are DSE in games on $\mathbb{R}^2$ (see Corollary 1, Appendix F).

This example, along with Propositions 8 and 9 in Ap-
pendix F, implies some stable fixed 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 em-
pirical 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
local optimum (Berard et al., 2019). 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 $D^2_f(x) = 0$.
Under this assumption, we show stable critical points
which are not DNE are DSE given $-D^2_{2j}(x) > 0$.

Proposition 4. Consider a zero-sum GAN satisfying the
realizable assumption. Any stable critical point of $\dot{x} =
-\omega(x)$ at which $-D^2_{2j}(x) > 0$ is a DSE and a stable critical
point of $\dot{x} = -\omega_S(x)$.

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 G and the
stochastic results in Appendix H.

4.1. Deterministic Setting

Consider the deterministic Stackelberg update
$$x_{k+1} = x_k - \gamma_2 \omega_{S_j}(x_k)$$
where $\omega_{S_j}(x_k)$ is the $m$-dimensional vector with entries
$$\tau^{-1}(D_1 f_1(x_k) - D_{12} f_2(x_k) (D^2_{2j} f_2(x_k))^{-1} D_{21} f_1(x_k)) \in \mathbb{R}^{m_1} \text{ and } D_{2j} f_2(x_k) \in \mathbb{R}^{m_2}, \text{ and } \tau = \gamma_2/\gamma_1$$
is the “timescale” separation with $\gamma_2 > \gamma_1$. We refer to these
dynamics as the $\tau$-Stackelberg update.

To get convergence guarantees, we apply well known results
from discrete time dynamical systems. For a dynamical sys-
tem $x_{k+1} = F(x_k)$, 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 (see Prop. 10,
Appendix G). In particular, $\rho(DF(x^*)) \leq c < 1$ implies
that $\|DF\| \leq c + \varepsilon < 1$ for $\varepsilon > 0$ on a neigh-
borhood of $x^*$ (Ortega & Rheinboldt, 1970, 2.2.8). Hence,
Prop. 10 implies that if $\rho(DF(x^*)) = 1 - \alpha < 1$ for
some $\alpha$, 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 $K > 0$,
$\|x_k - x^*\|_2 \leq K (1 - \alpha/2)^k \|x_0 - x^*\|_2$ using $\varepsilon = \alpha/4$.

For a zero-sum setting defined by cost function $f \in C^0(X, \mathbb{R})$ with $q \geq 2$, recall that
$S_1(J(x)) = D^2_f(x) - D_{21}f(x)^T (D^2_f(x))^{-1} D_{21}f(x)$ is the first Schur
complement of the Jacobian $J(x)$. Let $F = (f, -f) : X \to \mathbb{R}^2$ and $DF$ be its Jacobian.

Theorem 1 (Zero-Sum Convergence.). Consider a zero-
sum game defined by $f \in C^q(X, \mathbb{R})$ with $q \geq 2$. For a DSE
$x^*$, with $\alpha = \min \{\lambda_{\min}(-D^2_f(x^*)), \lambda_{\min}\{1/2 S_1(J(x^*)))\}$
and $\beta = \rho(DF(x^*))$, the $\tau$–Stackelberg update converges
locally with a rate of $O(1 - (\alpha/\beta)^2).$
Implicit Learning Dynamics in Stackelberg Games

The proof leverages the structure of the Jacobian $J_{S_*}$, 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 $J_{S_*}(x)$ is the union of the spectrum of $\tau^{-1}S_1(J(x))$ and $-D_2^f(x)$ for zero-sum settings. In Appendix G, we also provide a local convergence rate to an $\varepsilon$LSE given the agents initialize in a radius $p$ ball $B_p(x^*)$.

**Proposition 5.** Consider a zero sum game defined by $f \in C^q(X, \mathbb{R})$, $q \geq 2$. Suppose that $\gamma_2 < 1/L$ where $\max\{\operatorname{spec}(\frac{1}{2}S_1(J(x))) \cup \operatorname{spec}(-D_2^f(x))\} \leq L$. Then, $x$ is a stable critical point of $\tau$-Stackelberg update if and only if $x$ is a DSE.

The proof follows from Theorem 1 and Proposition 2.

**Theorem 2** (Almost Sure Avoidance of Saddles). Consider a general sum game defined by $f_i \in C^q(X, \mathbb{R})$, $q \geq 2$ for $i = 1, 2$ and where, without loss of generality, player 1 is the leader. Suppose that $\omega_{S_*}$ is $L$-Lipschitz with $\tau > 1$ and that $\gamma_2 < 1/L$. The $\tau$-Stackelberg learning dynamics converge to saddle points of $\dot{x} = -\omega_{S_*}(x)$ on a set of measure zero.

In the zero-sum case, $\omega_{S_*}$ being Lipschitz is equivalent to $\max\{\operatorname{spec}(\frac{1}{2}S_1(J(x))) \cup \operatorname{spec}(-D_2^f(x))\} \leq L$. In this case, using the structure of the Jacobian $J_{S_*}$, we know that the eigenvalues are real, and hence the only admissible types of critical points are stable, unstable, or saddle points. The above results shows that $\tau$-Stackelberg learning avoids saddle points almost surely. Moreover, we know that the only stable critical points are DSE and DSE are never saddle points for zero-sum games.

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 DSE since there is potentially stable critical points which are not DSE. This can be seen by examining the Jacobian which is no longer lower block triangular.

Given a stable DSE $x^*$, let $B_p(x^*)$ be the largest ball of radius $p > 0$ contained in the region of attraction on which $\frac{1}{2}(J^2_{S_*} + J_{S_*})$ is positive definite. Define $\alpha = \min_{x \in B_p(x^*)} \lambda^2_{\min}(\frac{1}{2}(J^2_{S_*} + J_{S_*}))$ and $\beta = \max_{x \in B_p(x^*)} \lambda_{\max}(J_{S_*}(x)^T J_{S_*}(x))$.

**Theorem 3.** 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. Suppose that $x^* \in X$ is a stable DSE and $x_0 \in B_p(x^*)$. Further, let $\gamma_2 = \sqrt{\alpha}/\beta$ where $\alpha < \beta$. Given $\varepsilon > 0$, the number of iterations required by $\tau$-Stackelberg learning to obtain an $\varepsilon$-DSE is bounded by $O(2\beta \log(\varepsilon) / \alpha)$.

**4.2. Stochastic Setting**

In the stochastic setting, players use updates of the form

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

where $\gamma_{i,k} = o(\gamma_{2,k})$ and $\{w_{i,k+1}\}$ is a stochastic process for each $i = 1, 2$. The results in this section assume the following. The maps $Df_1 : \mathbb{R}^m \to \mathbb{R}^m$, $D_2^f : \mathbb{R}^m \to \mathbb{R}^{m_2}$ are Lipschitz, and $\|Df_2\| < \infty$. For each $i \in \{1, 2\}$, the learning rates satisfy $\sum_k \gamma_{i,k} = \infty$, $\sum_k \gamma_{i,k}^2 < \infty$.

The 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 now show the dynamics avoid saddle points in the stochastic learning regime.

**Theorem 4** (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 $\mathbb{E}[\|w_{i,t}\|^2 + \|J_{S,i}\|] \geq b_i$ 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.

We also give asymptotic convergence guarantees. In particular, Theorem 7 in Appendix H.3 provides a global convergence guarantee in general-sum games to the set of stable attractors which contains the set of DSE under assumptions on the global asymptotic stability of attractors of the continuous time limit singularly perturbed dynamical system. The set of attractors may contain critical points which are not DSE. However, Corollary 5 in Appendix H.3 gives a global convergence guarantee in zero-sum games to the set of stable attractors of which the only critical points are DSE under identical assumptions using the structure of the Jacobian for the dynamics.

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 limit singularly perturbed system, and thus to stable DSE.

**Proposition 6.** 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 $\gamma_{1,k} = o(\gamma_{2,k})$. Given a DSE $x^*$, let $B_p(x^*) = B_{p_1}(x_1^*) \times B_{p_2}(x_2^*)$ with $p_1, p_2 > 0$ on which $\det(D_2^f(x^*)) \neq 0$. Suppose $x_0 \in B_p(x^*)$. If the dynamics $\dot{x}_2 = -D_2^f(x^*)$ have a locally asymptotically stable attractor $\gamma(x_1)$ uniformly in $x_1$ on $B_{p_2}(x_2^*)$ and the dynamics $\dot{x}_1 = Df_1(x_1^*, r(x_1))$ have a locally asymptotically stable attractor on $B_{p_1}(x_1^*)$, then $x_k \to x^*$ almost surely.

Beyond these asymptotic guarantees, in Appendix H, we leverage results stochastic approximation to provide finite-time, high probability concentration bounds which guarantee that the players’ stochastic updates get ‘locked in’ to an $\varepsilon$-ball around a DSE.
5. Experiments

We present several illustrative 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 (\texttt{simgrad}). We provide implementation details of the Stackelberg leader update, the techniques to compute the relevant eigenvalues, and hyperparameters in Appendix C and D. In Appendix D.3, we apply the Stackelberg learning to a DCGAN on the MNIST dataset, demonstrating that the update can be implemented for large-scale problems.

**Example 1: Learning a Covariance Matrix.** We consider a data generating process of $x \sim \mathcal{N}(0, \Sigma)$, where the covariance $\Sigma$ 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 $G_V(z) = V z$. 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 and it minimizes a regularized cost function defined by $-f(V, W) + \frac{\eta}{2} \text{Tr}(W^\top W)$, where $\eta \geq 0$ is a tunable regularization parameter. The game is formally defined by the costs $(f_1, f_2) = (f(V, W), -f(V, W) + \frac{\eta}{2} \text{Tr}(W^\top W))$, 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 dynamics and \texttt{simgrad}, and analyze the distance from equilibrium as a function of time. We plot $||\Sigma - VV^\top||_2$ for the generator’s performance and $||\frac{1}{2}(W + W^\top)||_2$ for the discriminator’s performance in Fig. 2 for varying dimensions $m$ with learning rates $\gamma_1 = \gamma_2/2 = 0.015$ and fixed regularization terms $\eta = m/5$.

We observe that Stackelberg learning converges to an equilibrium in fewer iterations than \texttt{simgrad}. 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 \texttt{simgrad}, whose Jacobian can admit complex eigenvalues, known to cause rotational forces in the dynamics. While there may be imaginary eigenvalues in the Jacobian of Stackelberg dynamics in general-sum games, this example demonstrates empirically that the Stackelberg dynamics rotate less.
Example 2: Learning a Mixture of Gaussians. We train a GAN to learn a mixture of Gaussian distribution, where the generator is the leader and the discriminator is the follower. 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, and the default ADAM optimizer configuration in PyTorch version 1.2 with decaying learning rates.

We ensure the follower’s Hessian to be well-conditioned by modifying the leader’s update,

\[ x_1^{n+1} = x_1 - \gamma_1(D_1 f_1(x) + D r_\eta(x)^{\top} D_2 f_2(x)) \]

where \( D r_\eta(x)^{\top} = -D_{21} f_2(x)^{\top} (D_2^2 f_2(x) + \eta I)^{-1} \) and \( \eta = 1 \) is the regularization parameter. We provide details on its derivation in Appendix B. We also employ regularization in the follower’s implicit map to compute the eigenvalues of the Schur complement to determine whether a critical point is in a neighborhood of a DNE for the zero-sum game.

Diamond configuration. We demonstrate that both simgrad and Stackelberg learning can converge to a neighborhood of a non-Nash DSE. This experiment uses the saturating GAN objective (Goodfellow et al., 2014). In Fig. 3a–3b and Fig. 3g–3h we show a sample of the generator and the discriminator for simgrad and the Stackelberg dynamics after 40,000 training batches. 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 that simgrad appears in a neighborhood of a DSE that is not a DNE since the individual Hessians for the leader and follower are indefinite and positive definite, respectively, and the Schur complement is positive definite. Moreover, the eigenvalues of the leader’s individual Hessian are nearly zero, which reflects the realizable assumption (cf. Sec. 3). The Stackelberg learning dynamics also converge to DSE of the zero-sum game which is not a DNE, evident from the Schur complement of \( J \) being positive. This example demonstrates that standard GAN training can converge to stable critical points that are DSE and not DNE which produce good generator and discriminator performance. This indicates that it may not be necessary to look only for DNE.

Circle configuration. We demonstrate improved performance when using Stackelberg learning dynamics. 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. 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 DNE and the Stackelberg dynamics converge again to a DSE that is not a DNE. This provides further evidence that DSE may be easier to reach, and can provide suitable performance.

**Figure 4.** MoG: Improved Stability on the Learning Path. Convergence to DNE for simgrad in (b)–(e) and to non-Nash DSE for Stackelberg learning in (f)–(i). To determine definiteness of the Jacobian, its Schur complement and individual Hessians, we show the six smallest and six largest real eigenvalues in (j)–(m) for simgrad and (n)–(q) for Stackelberg. The eigenvalues indicate simgrad converged to a DNE, yet simgrad exhibits unstable learning (b–e), while Stackelberg learning is stable (f–i) and converges to a non-Nash DSE.
6. Discussion

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.

References


Implicit Learning Dynamics in Stackelberg Games


**A. Guide to the Appendix**

**Appendix B.** Derivations for the regularized Stackelberg learning dynamics including necessary and sufficient conditions.

**Appendix C.** Details on numerically computing the Stackelberg update and Schur complement.

**Appendix D.** Additional details on numerical simulations for the GANs and bimatrix game, comparing the Stackelberg update with simgrad.

**Appendix E.** Proofs for the main results in Section 3.

**Appendix F.** Necessary and sufficient conditions for non-Nash attractors such that \(-D_2^2 f(x) > 0\) to be differential Stackelberg equilibria in zero-sum settings.

**Appendix G.** Convergence proofs of the results in Section 4.1, and additional results for the deterministic setting not covered in Section 4.1.

**Appendix H.** Convergence proofs for the results in Section 4.2. In addition, this section includes a number of extensions for the stochastic setting not included in the main body. In particular, we provide convergence guarantees for the leader assuming the follower plays an exact best response, and an extended two timescale analysis for the case where the follower is performing individual gradient play. We also introduce finite-time, high-probability concentration bounds which guarantee that the agents’ updates get locked-in to an \(\varepsilon\)-ball around a differential Stackelberg equilibrium.

**Appendix I.** Proofs of stability and genericity in zero-sum continuous Stackelberg games.

**B. 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

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

in which we regularize the inverse of \(D_2^2 f_2\) term. This update can be derived from the following perspective. Suppose player 1 views player 2 as optimizing a linearized version of its cost with a regularization term which captures the leader’s lack of confidence in the local linearization holding globally:

\[
\arg \min_y (y - x_{2,k})^\top D_2 f_2(x_k) + \frac{\eta}{2} \|y - x_{2,k}\|^2.
\]

The first-order optimality conditions for this problem are

\[
\begin{align*}
 0 &= D_2 f_2(x_k) + (y - x_{2,k})^\top D_2^2 f_2(x_k) + \eta (y - x_{2,k}) \\
 0 &= D_2 f_2(x_k) - (\eta I + D_2^2 f_2(x_k)) x_{2,k} + (D_2^2 f_2(x_k) + \eta I) y.
\end{align*}
\]

Hence, if the leader views the follower as updating along the gradient direction determined by these first order conditions, then the follower’s response map is given by

\[
x_{k+1,2} = x_{k,2} - (D_2^2 f_2(x_k) + \eta I)^{-1} D_2 f_2(x_k).
\]

Ignoring higher order terms in the derivative of the response map, the approximate Stackelberg update is given by

\[
\begin{align*}
 x_{k+1,1} &= x_{k,1} - \gamma_1 (D_1 f_1(x_k) - D_{21} f_2(x_k) + \eta I)^{-1} D_2 f_1(x_k)) \\
 x_{k+1,2} &= x_{k,2} - \gamma_2 D_2 f_2(x_k).
\end{align*}
\]

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.

**Proposition 7** (Regularized Stackelberg: Sufficient Conditions). A point \(x^*\) such that the first order conditions \(D_1 f_1(x) - D_{21} f_2(x)^\top (D_2^2 f_2(x) + \eta I)^{-1} D_2 f_2(x) = 0\) and \(D_2 f_2(x) = 0\) hold, and such that \(D(D_1 f_1(x) - D_{21} f_2(x)^\top (D_2^2 f_2(x) + \eta I)^{-1} D_2 f_2(x)) > 0\) and \(D_2^2 f_2(x) > 0\) is a differential Stackelberg equilibrium with respect to the regularized dynamics.
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.

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

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

C. 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 \( D_{12} f_2(x) \) term. These operations can be done efficiently in Python by utilizing Jacobian-vector products in auto-differentiation libraries combined with the \texttt{sparse.LinearOperator} class in \texttt{scipy}. These objects can also be used to compute their eigenvalues, inverses, or the Schur complement of the game dynamics using the \texttt{scipy.sparse.linalg} package. We found that the conjugate gradient method \texttt{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 \((f_1, f_2)\):

$$J(x) = \begin{bmatrix} D_{11}^2 f_1(x) & D_{12} f_1(x) \\ D_{21} f_2(x) & D_{22}^2 f_2(x) \end{bmatrix}.$$  

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 \texttt{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 \texttt{eigsh} or \texttt{eigs} respectively. Selecting to compute the smallest or largest \( k \) eigenvalues—sorted by either magnitude, real or imaginary values—allows one to examining 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 \( x \in X_1 \times X_2 \), \( D_{11}^2 f_1(x) - D_{12} f_1(x)(D_{22}^2 f_2)\^{-1}(x)D_{21} f_2(x) \). 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: \( u = D_{21} f_2(x)v, w = (D_{22}^2 f_2)\^{-1}(x)u, q = D_{12} f_1(x)w, \) and \( p = D_{11}^2 f_1(x)v \). 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 which can be computationally expensive. It solves the linear equation \( D_{22}^2 f_2(x)w = u \) 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. Particularly, when the Hessian is poorly conditioned, some methods may fail to converge. More investigation is required to determine which method is best suited for specific uses. For example, a fast iteration method with warm start might be appropriate for computing the leader update online, while a residual-based method might be better for computing the the eigenvalues of the Schur complement. Specifically, for our mixture of gaussians and MNIST GANs, we found that computing the leader update using the conjugate gradient method with maximum of 5 iterations and warm-start works well. We compared using the true Hessian for smaller scale problems and found the estimate to be within numerical precision.

D. Additional Numerical Simulations and Details

This section includes complete details on the training process and hyper-parameters selected in the mixture of Gaussian and MNIST experiments. We also present a numerical experiment of a parameterized bilinear game.

D.1. Mixture of Gaussians GAN

The underlying data distribution for the diamond experiment consists of Gaussian distributions with means given by \( \mu = [1.5 \sin(\omega), 1.5 \cos(\omega)] \) for \( \omega \in \{k \pi/2\}_{k=0}^3 \) and each with covariance \( \sigma^2 I \) where \( \sigma^2 = 0.15 \). Each sample of real data given to the discriminator is selected uniformly at random from the set of Gaussian distributions. The underlying data distribution for the circle experiment consists of Gaussian distributions with means given by \( \mu = [\sin(\omega), \cos(\omega)] \) for
Implicit Learning Dynamics in Stackelberg Games

\[ \omega \in \{k\pi/4\}_{k=0}^{\pi} \text{ and each with covariance } \sigma^2 I \text{ where } \sigma^2 = 0.3. \] 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 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 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 are 0.0001 and 0.0004 in the diamond and circle experiments, respectively. The objective for the game in the diamond experiment is the 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 \( \gamma_{t+1} = \gamma_t / \nu_{t+1}. \) 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. We regularize the implicit map of the follower as detailed in Appendix B using the parameter \( \eta = 1. \)

D.2. 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 adapted to handle 28 \( \times \) 28 images. We train a GAN on an MNIST dataset consisting of only the digits 0 and 1 from the training images and on an MNIST dataset containing the entire set of training images. We train using a batch size of 256 and a latent dimension of 100 and use the ADAM optimizer with the default parameters for the DCGAN network. We regularize the implicit map of the follower as detailed in Appendix B using the parameter \( \eta = 5000. \) 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 mixture of Gaussian experiments.

We show the results in Fig. 5 after 2900 batches. For each dataset we show a sample of 16 digits to get a clear view of the generator performance and a sample of 256 digits to get a broader view of the generator output. The Stackelberg dynamics are able to converge to a solution that generates realistic handwritten digits. 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 large scale networks with millions of parameters. We believe the tools we develop for our implementation can be helpful to researchers working on GANs since a number of theoretical works on this topic require second order information to strengthen the convergence guarantees.

The underlying data distribution for the MNIST experiments consists of digits 0 and 1 from the MNIST training dataset or each digit from the MNIST training dataset. We scale each image to the range \([-1, 1]\). Each sample of real data given to the discriminator is selected sequentially from a shuffled version of the dataset. The batch size for each player is 256. We train the generator using latent \( z \in \mathbb{R}^{100} \) sampled from a standard normal distribution in each training batch. The discriminator is trained using input vectorized images \( x \in \mathbb{R}^{28 \times 28} \) sampled from the underlying distribution in each training batch. We use the DCGAN architecture (Radford et al., 2015) for our generator and discriminator. Since DCGAN was built for 64 \( \times \) 64 images, we adapt it to handle 28 \( \times \) 28 images in the final layer. We follow the parameter choices from the DCGAN paper (Radford et al., 2015). This means we initialize the weights 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

\[ \text{Figure 5. We demonstrate Stackelberg learning on the MNIST dataset for digits for 0s and 1s in (a)-(b) and for all digits in (c)-(d).} \]
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 \(\varepsilon\)-neighborhood of a DNE at \((x_1^*, x_2^*) = (-1.4, -4.4)\). (c)-(d): for the Stackelberg learning dynamics, we observe convergence to an \(\varepsilon\)-neighborhood of a DSE at \((x_1^*, x_2^*) = (-1.4, -1.6)\) 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.

set the initial learning rates to be 0.0002. The learning rate for each player is decayed exponentially such that \(\gamma_i = \gamma_i^{0k}\) and \(\nu_i = 1 - 10^{-5}\) and \(\nu_i = 1 - 10^{-7}\). We regularize the implicit map of the follower as detailed in Appendix B using the parameter \(\eta = 5000\). 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 mixture of Gaussian experiments.

### D.3. Parameterized Bilinear Game

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 \(G = (f_1, f_2)\) with costs given by 

\[ f_1(x_1, x_2) = \pi(x_1)^T A \pi(x_2) + \eta \|x_1\|^2 \]

and

\[ f_2(x_1, x_2) = \pi(x_1)^T B \pi(x_2) + \frac{\eta}{2} \|x_2\|^2 \]

where 

\[ A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1/2 & 1 \\ 1 & 1/2 \end{bmatrix} \]

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

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

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 \((\pi_1, \pi_2)\) where \(\pi_1 = (1/2, 1/2)\) and any policy \(\pi_2\) 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 \(\omega\) and \(\omega_2\) 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 \((0, y)\) 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
This appendix contains proofs for results given in Section 3. To be clear, we restate each result before providing the proof.

**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 \(\dot{x} = -\omega_S(x)\) if and only if it is a DSE. Moreover, if \(f\) is generic, a point \(x\) is a stable critical point of \(\dot{x} = -\omega_S(x)\) if and only if it is a LSE.

**Proof.** The Jacobian of the Stackelberg limiting dynamics \(\dot{x} = -\omega_S(x)\) at a stable critical point is

\[
J_S(x) = \begin{bmatrix}
D_1(Df)(x) & 0 \\
-D_{21}f(x) & -D_2^2f(x)
\end{bmatrix} > 0.
\]

The structure of the Jacobian \(J_S(x)\) follows from the fact that

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

The eigenvalues of a lower triangular block matrix are the union of the eigenvalues in each of the block diagonal components. This implies that if \(J_S(x) > 0\), then necessarily \(D_1(Df)(x) > 0\) and \(-D_2^2f(x) > 0\). Consequently, any stable critical point of the Stackelberg limiting dynamics must be a differential Stackelberg equilibrium by definition.

**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.

**Proof.** Consider an arbitrary sufficiently smooth zero-sum game \((f, -f)\) on continuous strategy spaces. Suppose \(x\) is a
stable differential Nash equilibrium so that by definition \( D_1^2f(x) > 0, -D_2^2f(x) > 0 \), and

\[
J(x) = \begin{bmatrix} D_1^2f(x) & D_{12}f(x) \\ -D_{21}f(x) & -D_2^2f(x) \end{bmatrix} > 0.
\]

Then, the Schur complement of \( J(x) \) is also positive definite:

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

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

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

**Proof.** Consider an attractor \( x^* \) of \( \dot{x} = -\omega(x) \) such that \( -D_2^2f(x^*) \geq 0 \). Note that the realizable assumption implies that the Jacobian of \( \omega \) is

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

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

\[
-D_{21}f(x^*)(D_2^2f)^{-1}(x^*)D_{21}f(x^*) > 0.
\]

Since \( x^* \) is an attractor, \( D_1f(x^*) = 0 \) and \( D_2f(x^*) = 0 \) so that

\[
Df(x^*) = D_1f(x^*) + D_2f(x^*)(D_2^2f(x^*))^{-1}D_{21}f(x^*) = 0
\]

Consequently, the necessary conditions for a local Stackelberg equilibrium are satisfied. Moreover, since both \( -D_2^2f(x^*) > 0 \) and the Schur complement \( D_1^2f(x^*) - D_{21}f(x^*)(D_2^2f(x^*))^{-1}D_{21}f(x^*) > 0 \), the Jacobian of \( \omega_S \) has positive real eigenvalues so the point \( x^* \) is stable.

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

As alluded to in the main body, an interesting question is when non-Nash attractors of simultaneous gradient play—i.e., critical points \( x^* \) of \( \dot{x} = -\omega(x) \) at which \( \text{spec}(−J(x^*)) \subset \mathbb{C}_{−} \)—are differential Stackelberg equilibria. Attracting critical points \( x^* \) of the dynamics \( \dot{x} = -\omega(x) \) that are non-Nash equilibria are such that either \( D_1^2f(x^*) \) or \( -D_2^2f(x^*) \) are not positive definite. Without loss of generality, considering player 1 to be the leader, an attractor of the Stackelberg dynamics \( \dot{x} = -\omega_S(x) \) requires both \( -D_2^2f(x^*) \) and \( D_1^2f(x^*) - D_{21}f(x^*)(D_2^2f(x^*))^{-1}D_{21}f(x^*) \) to be positive definite. Hence, if \( -D_2^2f(x^*) \) is not positive definite at a non-Nash attractor of \( \dot{x} = -\omega(x) \), then \( x^* \) will also not be an attractor of \( \dot{x} = -\omega_S(x) \). A central question is when can this occur zero-sum games? Towards answering this, let us consider the non-Nash attractors\(^3\) with \( -D_2^2f(x^*) > 0 \) and determine when the Schur complement above is positive definite, so that \( x^* \) is an attractor of \( \dot{x} = \omega_S(x) \).

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 non-Nash attractor \( x^* \), let \( \text{spec}(D_1^2f(x^*)) = \{\mu_j, j \in \{1, \ldots, m\}\} \) where

\[
\mu_1 \leq \cdots \leq \mu_r < 0 \leq \mu_{r+1} \leq \cdots \leq \mu_m.
\]

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

**Proposition 8** (Necessary conditions). Consider a non-Nash attractor \( x^* \) of the individual gradient dynamics \( \dot{x} = -\omega(x) \) such that \( -D_2^2f(x^*) > 0 \). Given \( \kappa > 0 \) such that \( |D_{21}f(x^*)| \leq \kappa, \) if \( D_1^2f(x^*) - D_{21}f(x^*)(D_2^2f(x^*))^{-1}D_{21}f(x^*) > 0 \), then \( r \leq n \) and \( \kappa^2 \lambda_i + \mu_i > 0 \) for all \( i \in \{1, \ldots, r-p\} \).

For a matrix \( W \), let \( W^\top \) denote the conjugate transpose.

\(^3\)We note that we could study an analogous setup in which we characterize when non-Nash attractors with \( D_1^2f(x^*) > 0 \) are such that \( -D_2^2f(x^*) + D_{12}f(x^*)(D_1^2f(x^*))^{-1}D_{12}f(x^*) > 0 \), thereby switching the roles of leader and follower.
Proposition 9 (Sufficient conditions). Let $x^*$ be a stable non-Nash attractor of the individual gradient dynamics $\dot{x} = -\omega(x)$ 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 $D_{12} f(x^*) = \Sigma$ 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 $i \in \{1, \ldots, r\}$, then $x^*$ is a differential Stackelberg equilibrium and an attractor of $\dot{x} = -\omega_S(x)$.

The proofs of the propositions follow from linear algebra results. Before diving into the proofs, we provide some commentary. Essentially, this says that if $D_1^2 f(x^*) = W_1 \Sigma W_1^\top$ with $W_1 W_1^\top = I_{n \times n}$ and $M$ diagonal, and $-D_2^2 f(x^*) = W_2 \Sigma W_2^\top$ with $W_2 W_2^\top = I_{m \times m}$ and $\Lambda$ diagonal, then $D_{12} f(x^*)$ can be written as $W_1 \Sigma W_1^\top$ for some diagonal matrix $\Sigma \in \mathbb{R}^{n \times m}$ (not necessarily positive). Note that since $\Sigma$ does not necessarily have positive values, $W_1 \Sigma W_1^\top$ 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 $\Sigma$ has non-zero entries, then the remaining assumptions are still sufficient to guarantee that

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

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

The results in this subsection follow from the theory of block operator matrices and indefinite linear algebra (Tretter, 2008).

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 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

$$\lambda_j(UWU^\top) \leq ||U||^2 \lambda_j(W)$$

for $j = 1, \ldots, \min\{k, m, \text{rank}(UWU^\top)\}$.

Let us define $|M| = (MM^\top)^{1/2}$ for a matrix $M$. Recall also that for Propositions 8 and 9, we have defined $\text{spec}(D_1^2 f(x^*)) = \{\mu_j, j \in \{1, \ldots, m\}\}$ where $\mu_1 \leq \cdots \leq \mu_r < 0 < \mu_{r+1} \leq \cdots \leq \mu_m$, and $\text{spec}(-D_2^2 f(x^*)) = \{\lambda_i, i \in \{1, \ldots, n\}\}$ where $\lambda_1 \geq \cdots \geq \lambda_n > 0$, given an attractor $x^*$.

We can now use the above Lemma to prove Proposition 8. The proof follows the main arguments in the proof of Lemma 3.2 in the work by Berger et al. (2018) with some minor changes due to the nature of our problem.

Proof of Proposition 8. Let $x^*$ be a stable attractor of $\dot{x} = -\omega_S(x)$ 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 $x_1 \in \mathbb{R}^n$ and $x_2 \in \mathbb{R}^m$. Suppose that $A - BC^{-1} B^\top > 0$.

Claim: $r \leq n$ is necessary. We argue by contradiction. Suppose not—i.e., assume that $r > n$. Note that if $m < n$, then this is not possible. In this case, we automatically satisfy that $r \leq n$. Otherwise, $r \geq m > n$. Let $S_1 = \ker(B(-C^{-1} + |C^{-1}|)B^\top)$ and consider the subspace $S_2$ of $\mathbb{C}^m$ spanned by the all the eigenvectors of $A$ corresponding to non-positive eigenvalues. Note that $\dim S_1 = m - \text{rank}(B(-C^{-1} + |C^{-1}|)B^\top) \geq m - \text{rank}(-C^{-1} + |C^{-1}|) = m - n$ by assumption, we have that $\dim S_2 = r$ so that, since $r > n$,

$$\dim S_1 + \dim S_2 \geq (m - n) + r = m + (r - n) > m.$$ 

Thus, $S_1 \cap S_2 \neq \{0\}$. Now, $S_1 = \ker(B(-C^{-1} + |C^{-1}|)B^\top)$. Hence, for any non-trivial vector $v \in S_1 \cap S_2$, $(BC^{-1} B^\top - B|C^{-1}|B^\top) v = 0$ so that we have

$$\langle (A - BC^{-1} B^\top) v, v \rangle = \langle Av, v \rangle - \langle B|C^{-1}|B^\top v, v \rangle \leq 0.$$  (7)
Note that the inequality in (7) holds because 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

\[
0 < \lambda_m(A - BC^{-1}B^\top) \leq \lambda_1(A) + \lambda_{m-1+i}(-BC^{-1}B^\top), \quad i \in \{1, \ldots, m\}.
\]

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

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

which gives the desired result.

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

Now, let us prove Proposition 9 which gives sufficient conditions for when a stable non-Nash attractor \( x^* \) of \( \dot{x} = -\omega(x) \) is a differential Stackelberg equilibrium. Then, combining this with Proposition 2, we have a sufficient condition under which stable non-Nash attractors are in fact stable attractors of \( \dot{x} = -\omega_S(x) \).

**Proof of Proposition 9.** Let \( x^* \) be a stable non-Nash attractor of \( \dot{x} = -\omega(x) \) 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 \( D_i^2 f(x^*) = W_i M W_i^\top \) with \( W_i W_i^\top = I_{n \times n} \) and \( M = \text{diag}(\mu_1, \ldots, \mu_m) \), and \( -D_2^2 f(x^*) = W_2 A W_2^\top \) with \( W_2 W_2^\top = I_{m \times m} \) and \( \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n) \).

By assumption, there exists a diagonal matrix \( \Sigma \in \mathbb{R}^{m \times n} \) such that \( D_1^2 f(x^*) = W_1 \Sigma W_2^\top \) 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,

\[
D_2^2 f(x^*) - D_21 f(x^*)^\top (D_2^2 f(x^*))^{-1} D_21 f(x^*) = W_1 M W_1^\top + W_1 \Sigma W_2^\top (W_2 A W_2^\top)^{-1} W_2 \Sigma W_1^\top
\]

\[
= W_1 (M + \Sigma \Lambda^{-1} \Sigma^\top) W_1^\top
\]

Hence, to understand the eigenstructure of the Schur complement, 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.

Hence, by Proposition 2 it is an attractor of \( \dot{x} = -\omega_S(x) \).

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 \( J(x^*) \) has positive eigenvalues, \( -D_2^2 f(x^*) > 0 \), and the Schur complement \( D_1^2 f(x^*) - D_21 f(x^*)^\top (D_2^2 f(x^*))^{-1} D_21 f(x^*) \) is not positive definite. In the scalar case, 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, 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 \( \sup_x \|D_{12} f(x)\| \leq \kappa < \infty \) and \( K = W_1 \Sigma W_2^\top \) with \( \Sigma \) again a (not necessarily positive) diagonal matrix, then the results of Proposition 9 hold. From a control point of view, one can think about the leader’s update as having a feedback 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.

\(^{3}\)Functions such that derivative of \( f \) is Lipschitz will satisfy this condition.
F.1. equivalence in scalar games

**Corollary 1.** Suppose that $x_i \in \mathbb{R}$ for $i = 1, 2$ and that player 1 is the leader and player 2 is the follower in a zero-sum setting defined by $f : \mathbb{R}^2 \to \mathbb{R}$. Then, any non-Nash attractors of $\dot{x} = -\omega(x)$ at which $-D^2 f(x) > 0$ are differential Stackelberg equilibria.

The proof follows directly from examining the necessary and sufficient conditions for when a $2 \times 2$ matrix is positive definite.

G. Deterministic Convergence Results

Consider the deterministic Stackelberg update

$$x_{k+1,1} = x_{k,1} - \gamma_1 (D_1 f_1(x_k) - D_2 f_1(x_k))$$
$$x_{k+1,2} = x_{k,2} - \gamma_2 D_2 f_2(x_k)$$

which is equivalent to

$$x_{k+1,1} = x_{k,1} - \frac{\gamma_2}{\tau} (D_1 f_1(x_k) - D_2 f_1(x_k))$$
$$x_{k+1,2} = x_{k,2} - \gamma_2 D_2 f_2(x_k)$$

where $\tau = \gamma_2 / \gamma_1$ is the “timescale” separation. Then,

$$x_{k+1} = x_k - \gamma_2 \omega_{S_1}(x_k)$$

where

$$\omega_{S_1}(x_k) = \left( \tau^{-1} (D_1 f_1(x_k) - D_2 f_1(x_k))^{-1} D_2 f_2(x_k), D_2 f_2(x_k) \right).$$

In the following subsections, we provide the proof of Proposition 1 which offers alternative sufficient conditions for differential Stackelberg relevant to our dynamics as well as convergence guarantees for the zero-sum and general-sum settings, respectively.

G.1. Proof of Sufficient Conditions for Differential Stackelberg in General Sum Settings

Proposition 1 provides alternative sufficient conditions for differential Stackelberg in terms of the Jacobian of $\omega_{S_1}$ which, in turn, defines the local behavior of the $\tau$-Stackelberg learning dynamics.

Consider now a general-sum, continuous Stackelberg game defined by $(f_1, f_2)$ with $f_i \in C^{q}(\mathbb{R}^{m_i}, \mathbb{R})$, $q \geq 2$. Note that

$$D f_1(x_1, r(x_1)) = D_1 f_1(x_1, r(x_1)) + D r(x_1)^\top D_2 f_1(x_1, r(x_1))$$

and $\omega_S(x_1, x_2) = (D f_1(x_1, x_2), D_2 f_2(x_1, x_2))$ so that the hierarchical play Jacobian of the vector field $\omega_S(x)$ is given by

$$J_S(x) = \begin{bmatrix} D_1(D f_1(x_1, x_2)) & D_2(D f_1(x_1, x_2)) \\ D_2 f_2(x_1, x_2) & D_2 f_2(x_1, x_2) \end{bmatrix}$$

Equilibrium conditions are such that $\omega_S(x_1, x_2) = 0$ and $x_2 = r(x_1)$ where $r(\cdot)$ is implicitly defined by $D_2 f_2(x_1, x_2) = 0$. The following result gives equivalent conditions for a differential Stackelberg equilibrium in terms of $J_S(x)$, which determines the local behavior of the dynamical system defined by the learning dynamics.

We use the notation $S_1(\cdot)$ to denote the Schur complement of $(\cdot)$ with respect to the bottom block diagonal matrix.

**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 $D f_1(x^*) = 0$ and $S_1(J_S(x^*)) > 0$ if and only if $x^*$ is a DSE. Moreover, in zero-sum games, $S_1(J_S(x)) = S_1(J(x))$.

**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 $J_S$ is

$$S_1(J_S(x)) = D_1(D f_1(x_1, x_2)) - D_2(D f_1(x_1, x_2))(D_2 f_2(x_1, x_2))^{-1} D_2 f_2(x_1, x_2)$$
Implicit Learning Dynamics in Stackelberg Games

where

\[ D_1(Df_1(x_1, x_2)) = D_1^2 f_1(x_1, x_2) + D_{12} f_1(x_1, x_2) Dr(x_1) + D_2 f_1(x_1, x_2) D^2 r(x_1)^4 \]

and

\[ D_2(Df_1(x_1, x_2)) = D_{12} f_1(x_1, x_2) + Dr(x_1)^T D_2^2 f_1(x_1, x_2). \]

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

\[
D(Df_1(x_1, r(x_1))) = D_1^2 f_1(x_1, r(x_1)) + D_{12} f_1(x_1, r(x_1)) Dr(x_1) + D_2 f_1(x_1, r(x_1))^T D^2 r(x_1) \\
+ (D_{12} f_1(x_1, r(x_1)) + Dr(x_1)^T D_2^2 f_1(x_1, r(x_1)) Dr(x_1)
\]

Note also that by the implicit function theorem, \( Dr(x_1) = - (D^2 f_2(x_1, x_2))^{-1} D_2 f_1(x_1, x_2) \bigg|_{x_2=r(x_1)} \). Hence, we have that

\[
D(Df_1(x_1, r(x_1))) = S_1(J_S(x))|_{x_2=r(x_1)}.
\]

\section*{G.2. Almost Sure Avoidance of Saddles}

\textbf{Theorem 2} (Almost Sure Avoidance of Saddles). Consider a general sum game defined by \( f_i \in C^q(X, \mathbb{R}) \), \( q \geq 2 \) for \( i = 1, 2 \) and where, without loss of generality, player 1 is the leader. Suppose that \( \omega_{S_1} \) is \( L \)-Lipschitz with \( \tau > 1 \) and that \( \gamma_2 < 1/L \). The \( \tau \)-Stackelberg learning dynamics converge to saddle points of \( \dot{x} = -\omega_{S_1}(x) \) on a set of measure zero.

\textit{Proof.} To show this, we follow the arguments in Mazumdar & Ratliff (2018) 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 \( g_{S_1} \) is a diffeomorphism, and then apply the center manifold theorem. First, recall that \( \tau = \gamma_2/\gamma_1 > 1 \).

Let

\[
g_{S_1}(x) = x - \gamma_2 \left( \frac{1}{2} (D_1 f(x) - D_{21} f^T (x) (D_2^2 f(x)))^{-1} D_2 f(x), D_2 f(x) \right)
\]

We claim that \( g_{S_1} : \mathbb{R}^m \rightarrow \mathbb{R}^m \) is a diffeomorphism. If we can show that \( g_{S_1} \) is invertible and a local diffeomorphism, then the claim follows.

Consider \( x \neq y \) and suppose \( g_{S_1}(y) = g_{S_1}(x) \) so that \( y - x = \gamma_2 (\omega_{S_1}(y) - \omega_{S_1}(x)) \). The assumption \( \sup_{x \in \mathbb{R}^m} \|J_{S_1}(x)\|_2 \leq L < \infty \) implies that \( \omega_{S_1} \) satisfies the Lipschitz condition on \( \mathbb{R}^m \). Hence, \( \|\omega_{S_1}(y) - \omega_{S_1}(x)\|_2 \leq L \|y-x\|_2 \). Let \( \Gamma = \gamma_2 \text{diag}(\tau^{-1} I, I) \). Then, \( \|y-x\|_2 \leq L \|\Gamma\|_2 \|y-x\|_2 < \|y-x\|_2 \) since \( \|\Gamma\|_2 \leq \gamma_2 < 1/L \).

Now, observe that \( Dg_{S_1} = I - \gamma_2 J_{S_1}(x) \). If \( Dg_{S_1} \) is invertible, then the implicit function theorem (Lee, 2012, Theorem C.40) implies that \( g_{S_1} \) is a local diffeomorphism. Hence, it suffices to show that \( \gamma_2 J_{S_1}(x) \) 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

\[
\rho(\gamma_2 J_{S_1}(x)) \leq \|\gamma_2 J_{S_1}(x)\|_2 \leq \gamma_2 \sup_{x \in \mathbb{R}^m} \|J_{S_1}(x)\|_2 < \gamma_2 L < 1.
\]

Of course, the spectral radius is the maximum absolute value of the eigenvalues, so that the above implies that all eigenvalues of \( \gamma_2 J_{S_1}(x) \) have absolute value less than 1. Since \( g_{S_1} \) is injective by the preceding argument, its inverse is well-defined and since \( g_{S_1} \) is a local diffeomorphism on \( \mathbb{R}^m \), it follows that \( g_{S_1}^{-1} \) is smooth on \( \mathbb{R}^m \). Thus, \( g_{S_1} \) is a diffeomorphism.

Consider all critical points to the game—i.e. \( \mathcal{A}_1 = \{ x \in X | \omega_{S_1}(x) = 0 \} \). For each \( u \in \mathcal{A}_1 \), let \( B_u \), where \( u \) indexes the point, be the open ball derived from the center manifold theorem (Shub, 1978, Theorem III.7) and let \( B = \cup_u B_u \). Since \( X \subseteq \mathbb{R}^m \), Lindelöf’s lemma (Kelley, 1955)—every open cover has a countable subcover—gives a countable subcover of \( B \). That is, for a countable set of critical points \( \{u_i\}_{i=1}^\infty \) with \( u_i \in \mathcal{A}_1 \), we have that \( B = \cup_{i=1}^\infty B_{u_i} \).

Starting from some point \( x_0 \in X \), if gradient-based learning converges to a strict saddle point, then there exists a \( t_0 \) and index \( i \) such that \( g_{S_1}(x_0) = B_{u_i} \) for all \( t \geq t_0 \). Again, applying the center manifold theorem (Shub, 1978, Theorem III.7) and using that \( g_{S_1}(X) \subseteq X \)—which we note is obviously true if \( X = \mathbb{R}^m \)—we get that \( g_{S_1}(x_0) \in W_{\text{loc}}^s \cap X \).

Using the fact that \( g_{S_1} \) is invertible, we can iteratively construct the sequence of sets defined by \( W_1(u_i) = g_{S_1}^{-1}(W_{\text{loc}}^s \cap X) \) and \( W_{k+1}(u_i) = g_{S_1}^{-1}(W_k(u_i) \cap X) \). Then we have that \( x_0 \in W_k(u_i) \) for all \( t \geq t_0 \). The set \( X_0 = \cup_{i=1}^\infty \cup_{t=0}^\infty W_t(u_i) \)

\footnote{Note that \( D^2 r(x_1) \) is 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 \( (D_2 f_1(x_1)) (D^2 r(x_1))_{i,j,k} \).}
contains all the initial points in X such that gradient-based learning converges to a strict saddle. Since \( u_i \) is a strict saddle, \( I - \gamma_2 D_\omega S_\gamma(u_i) \) has an eigenvalue greater than 1. This implies that the co-dimension of \( E^u \) is strictly less than \( m \). (i.e. \( \dim(W^{\omega \text{loc}}_\gamma) < m \)). Hence, \( W^{\omega \text{loc}}_\gamma \cap X \) has Lebesgue measure zero in \( \mathbb{R}^m \).

Using again that \( g_{S_\gamma} \) is a diffeomorphism, \( g_{S_\gamma}^{-1} \in C^1 \) 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 \( x_0 \) is a measure zero set since it is a countable union of measure zero sets.

There is certainly the problem of spending infinite time at saddles.

G.3. Zero-Sum Setting

For zero-sum settings only, when \( \tau \to \infty \), simultaneous gradient play or \( \tau-\text{GDA} \) with the timescaling, i.e.,

\[
x_{k+1,1} = x_{k,1} - \frac{\gamma_2}{\tau} D_1 f(x_k)
\]

\[
x_{k+1,2} = x_{k,2} + \gamma_2 D_2 f(x_k)
\]

converges to a local minimax equilibrium (Jin et al., 2019). Here, we consider a different update all together (i.e., not \( \tau-\text{GDA} \) but rather \( \tau-\text{Stackelberg} \)) which is defined by the dynamics \( \omega_{S_\gamma} \) given above.

In the zero-sum (deterministic) setting, the analysis is closely related to the \( \varepsilon \)-max oracle in (Jin et al., 2019). However, we provide convergence guarantees using our dynamics as oppose to an exact \( \varepsilon \)-max oracle; that is, the follower is updating by taking a step along the gradient direction as opposed to finding an exact \( \varepsilon \) local minima at each iteration.

Suppose \( \gamma_2 \) is such that \( \rho(I - \gamma_2 J_{S_\gamma}) < 1 \), then standard theory from numerical analysis for dynamical systems provides guarantees on local asymptotic convergence, and we can provide a finite time convergence guarantee to an \( \varepsilon \)-differential Stackelberg equilibrium. Indeed, letting

\[
x_{k+1} = x_k - \gamma_2 \omega_{S_\gamma}(x_k)
\]

be the \( \tau \)-Stackelberg update, then the following well-known result from Ostrowski’s theorem on fixed points gives us exactly such convergence guarantees.

**Proposition 10** (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.

We note that \( \rho(DF(x^*)) < 1 \) implies there exists \( c > 0 \) such that \( \rho(DF(x^*)) \leq c < 1 \). Hence, given any \( \varepsilon > 0 \), there is a norm and a \( c > 0 \) such that \( \|DF\| \leq c + \varepsilon < 1 \) on a neighborhood of \( x^* \) (Ortega & Rheinboldt, 1970, 2.2.8). Thus, the proposition implies that if \( \rho(DF(x^*)) = 1 - \alpha < 1 \) for some \( \alpha > 0 \), 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 \( K > 0 \), \( \|x_k - x^*\| \leq K(1 - \alpha/2)^k \|x_0 - x^*\|_2 \) using \( \varepsilon = \alpha/4 \).

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

\[
S_1(J(x)) = D_1^2 f(x) - D_{21}^2 f(x) \top (D_2^2 f(x))^{-1} D_{21}^2 f(x)
\]

be the first Schur complement of the Jacobian \( J \) of \( \omega(x) = (D_1 f(x), D_2 f(x)) \). Further, let \( F = (f, -f) : X_1 \times X_2 \to \mathbb{R}^2 \) and let \( DF \) denote the Jacobian of the vector-valued function \( F \).

**Theorem 1** (Zero-Sum Convergence.). Consider a zero-sum game defined by \( f \in C^q(X, \mathbb{R}) \) with \( q \geq 2 \). For a DSE \( x^* \), with \( \alpha = \min\{\lambda_{\min}(-D_2^2 f(x^*)), \lambda_{\min}(\frac{1}{\tau} S_1(J(x^*)))\} \) and \( \beta = \rho(DF(x^*)) \), the \( \tau \)-Stackelberg update converges locally with a rate of \( O(1 - (\alpha/\beta)^2) \).

**Proof.** To show convergence, we simply need to show that \( \rho(I - \gamma_2 J_{S_\gamma}) < 1 \) and apply Proposition 10. The structure of the Jacobian \( J_{S_\gamma} \) is lower-block triangular, with symmetric components along the diagonal—i.e., \( \tau^{-1} S_1(J(x)) \) and \( -D_2^2 f(x) \). Due to this structure, we know that \( \text{spec}(J_{S_\gamma}(x)) = \text{spec}(\tau^{-1} S_1(J(x))) \cup \text{spec}(-D_2^2 f(x)) \). Hence, \( \lambda_{\min}(I - \gamma_2 J_{S_\gamma}(x)) = 1 - \gamma_2 \min\{\lambda_{\min}(\tau^{-1} S_1(J(x))), \lambda_{\min}(-D_2^2 f(x))\} = 1 - \gamma_2 \alpha \). Further, with a choice of \( \gamma_2 = \alpha/(2\beta^2) \),

\[
\lambda_{\max}(\tau^{-1} S_1(x)) \leq \|\tau^{-1} D_2^2 f(x^*)\| + \tau^{-1} \|D_{12} f(x^*)\|_2^2 \| (D_2^2 f(x^*))^{-1} \| \leq \tau^{-1} \beta (1 + (1/\alpha)) \leq \beta (1 + \beta/\alpha)
\]
Theorem 5. Consider a zero sum game defined by \( f \in C^q(X, \mathbb{R}) \) 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. Suppose that the largest ball of radius \( p \) contained in the region of attraction on which \( S = \frac{1}{2}(J_{S_1} + JS_2) \) is positive definite. Note that due to \eqref{eq:6}, \(-D^2_x f(x) > 0 \) and \( D^2_x f(x) > 0 \). Define constants 

\[
\alpha = \min_{x \in B_p(x^*)} \lambda^2_{\min}(S(x)) \quad \text{and} \quad \beta = \max_{x \in B_p(x^*)} \lambda_{\max}(J_{S_1}(x)^T J_{S_2}(x)).
\]

**Theorem 5.** Consider a zero sum game \((f, -f)\) defined by \( f \in C^q(X_1 \times X_2, \mathbb{R}) \) 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. Suppose that \( x^* \in X \) is a stable differential Stackelberg equilibrium and \( x_0 \in B_p(x^*) \). Further, let \( \gamma_2 = \sqrt{\alpha/\beta} \) where \( \alpha < \beta \). Given \( \varepsilon > 0 \), the number of iterations required by \( \tau \)-Stackelberg learning to obtain an \( \varepsilon \)-differential Stackelberg equilibrium is bounded by \( O(2\beta \log(p/\varepsilon)/\alpha) \).

The equivalence in \eqref{eq:6} implies that \( x^* \) is a differential Stackelberg equilibrium for which the conditions of Theorem 5 hold if and only if it is a differential Nash equilibrium.

**Corollary 2.** Under the conditions of Theorem 5, for any \( x_0 \in B_p(x^*) \), the number of iterations required by \( \tau \)-Stackelberg learning to obtain an \( \varepsilon \)-differential Nash equilibrium is \( O(2\beta \log(p/\varepsilon)/\alpha) \).

**Proof of Theorem 5.** It suffices to show that for the choice of \( \gamma_2, \rho(I - \gamma_2 J_{S_1}(x)) < 1 \) on \( B_p(x^*) \). Then, an inductive argument can be made with the inductive hypothesis that \( x_k \in B_p(x^*) \).

Since \( \omega_{S_1}(x^*) = 0 \), we have that

\[
\|x_{k+1} - x^*\|_2 = \|x_k - x^* - \gamma_2(\omega_{S_1}(x_k) - \omega_{S_1}(x^*))\|_2 \leq \sup_{x \in B_p(x^*)} \|I - \gamma_2 J_{S_1}(x)\|_2 \|x_k - x^*\|_2.
\]

If \( \sup_{x \in B_p(x^*)} \|I - \gamma_2 J(x)\|_2 \) is less than one, where the norm is the operator 2-norm, then the dynamics are contracting as in Proposition 10. To simplify notation, we drop the explicit dependence on \( x \). Then,

\[
(I - \gamma_2 J_{S_1})^T (I - \gamma_2 J_{S_2}) \leq (1 - 2\gamma_2 \lambda_{\min}(S) + 2\lambda_{\max}(J_{S_1}^T J_{S_2})) I \leq (1 - \alpha/\beta) I.
\]

Using the above to bound \( \sup_{x \in B_p(x^*)} \|I - \gamma_2 J_{S_1}(x)\|_2 \), we have \( \|x_{k+1} - x^*\|_2 \leq (1 - \alpha/\beta)^{1/2} \|x_k - x^*\|_2 \). Since \( \alpha < \beta, (1 - \alpha/\beta) < e^{-\alpha/\beta} \) so that \( \|x_{k+1} - x^*\|_2 \leq e^{-T\alpha/(2\beta)} \|x_0 - x^*\|_2 \). This, in turn, implies that \( x_k \in B_p(x^*) \) for all \( k \geq T = \left[ 2\beta \log(p/\varepsilon) \right] / \alpha \).

**Proposition 5.** Consider a zero sum game defined by \( f \in C^q(X, \mathbb{R}), q \geq 2 \). Suppose that \( \gamma_2 < 1/L \) where \( \max\{\text{spec}(\frac{1}{2} S_1(J(x))) \cup \text{spec}(-D^2_x f(x))\} \leq L \). Then, \( x \) is a stable critical point of \( \tau \)-Stackelberg update if and only if \( x \) is a DSE.

The proof of this proposition follows almost immediately from Theorem 1 and Proposition 2.

**Theorem 6.** Consider a zero sum game defined by \( f \in C^r(X, \mathbb{R}), r \geq 2 \). Suppose that \( \tau > 1 \) and \( \gamma_2 < 1/L \) where \( \max\{\text{spec}(\frac{1}{2} S_1(J(x))) \cup \text{spec}(-D^2_x f(x))\} \leq L \). The \( \tau \)-Stackelberg learning dynamics converge to the set of DSE almost surely.

**Proof.** Due to the structure of the Jacobian, we know that the at all critical points, the Jacobian of the \( \tau \)-Stackelberg dynamics satisfies

\[
\text{spec}(J_{S_1}(x)) = \text{spec}(D^2_x f(x) - D_{12} f(x)^\top (D^2_x f(x))^{-1} D_{12} f(x)) \cup \text{spec}(-D^2_x f(x))
\]
Further, both $D_i^2 f(x) - D_{12} f(x)^T (D_2^2 f(x))^{-1} D_{12} f(x)$ and $-D_2^2 f(x)$ are Hermitian. Hence, all critical points are either saddle points (i.e., linearly unstable critical points), unstable points (i.e., $\text{spec}(J_{S_2}(x)) \subset \mathbb{C}^+_\infty$), or stable differential Stackelberg equilibria. Unstable points are not attractors of the dynamics. Hence, what remains to be shown is that the set of initial conditions for which the Stackelberg learning rule converges to a strict saddle point is of measure zero. The remainder of the proof follows the proof of Theorem 2.

Corollary 3. Suppose that $f \in C^q(X_1 \times X_2, \mathbb{R})$, $q \geq 2$. Let $S$ be the set of DSE. Suppose that
\[
\xi < \inf_{x \in S} \{ |\lambda_{\max}(D_1^2 f(x) - D_{12} f(x)^T (D_2^2 f(x))^{-1} D_{12} f(x))|, |\lambda_{\max}(-D_2^2 f(x))| \} < \infty.
\]
Then, if $\tau$–Stackelberg learning converges to a differential Stackelberg equilibrium for all but a measure zero set of initial conditions, then $\gamma_2 < 2/\xi$.

The above corollary provides a necessary condition on the stepsize of the follower for almost sure convergence to differential Stackelberg equilibria.

G.4. General Sum Setting

Analogous to the deterministic zero-sum setting, we can provide convergence guarantees for deterministic general sum settings. However, the convergence guarantee is no longer a global guarantee in that there is potentially other types of attractors in the general sum setting; this can be seen by the structure of the Jacobian which is no longer lower block triangular, nor is it symmetric.

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 $J_{S_2}$ is not lower block triangular and hence, the convergence rate depends more abstractly on the spectral structure of $J_{S_2}$ as opposed to the second-order sufficient conditions for a local Stackelberg equilibrium. It is still an open question as to how the spectrum of $J_{S_2}$ relates to $\text{schur}(J_{S_2})$.

Given a stable differential Stackelberg equilibrium $x^*$, let $B_p(x^*)$ be the largest ball of radius $p > 0$ contained in the region of attraction on which $S = \{ J_{S_2}^T + J_{S_2} \}$ is positive definite.

Define constants $\alpha = \min_{x \in B_p(x^*)} \lambda_{\min}(S(x)^T S(x))$ and $\beta = \max_{x \in B_p(x^*)} \lambda_{\max}(J_{S_2}(x)^T J_{S_2}(x))$.

Theorem 3. 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. Suppose that $x^* \in X$ is a stable DSE and $x_0 \in B_p(x^*)$. Further, let $\gamma_2 = \sqrt{\alpha/\beta}$ where $\alpha < \beta$. Given $\varepsilon > 0$, the number of iterations required by $\tau$–Stackelberg learning to obtain an $\varepsilon$–DSE is bounded by $O(2\beta \log(p/\varepsilon)/\alpha)$.

Proof. It suffices to show that for the choice of $\gamma_2$, $\rho(I - \gamma_2 J_{S_2}(x)) < 1$ on $B_p(x^*)$. Then, an inductive argument can be made with the inductive hypothesis that $x_k \in B_p(x^*)$. Since $\omega_{S_2}(x^*) = 0$, we have that
\[
\|x_{k+1} - x^*\|_2 = \|x_k - x^* - \gamma_2 (\omega_{S_2}(x_k) - \omega_{S_2}(x^*))\|_2 \leq \sup_{x \in B_p(x^*)} \|I - \gamma_2 J_{S_2}(x)\|_2 \|x_k - x^*\|_2.
\]

If $\sup_{x \in B_p(x^*)} \|I - \gamma_2 J_{S_2}(x)\|_2$ is less than one, where the norm is the operator 2–norm, then the dynamics are contracting as in Proposition 10. For notational convenience, we drop the explicit dependence on $x$. Then,
\[
(I - \gamma_2 J_{S_2})^T (I - \gamma_2 J_{S_2}) \leq (1 - 2\gamma_2 \lambda_{\min}(S) + \gamma_2^2 \lambda_{\max}(J_{S_2}^T J_{S_2})) I \leq (1 - \alpha/\beta) I.
\]

Using the above to bound $\sup_{x \in B_p(x^*)} \|I - \gamma_2 J_{S_2}(x)\|_2$, we have $\|x_{k+1} - x^*\|_2 \leq (1 - \alpha/\beta)^{1/2} \|x_k - x^*\|_2$. Since $\alpha < \beta$, $(1 - \alpha/\beta) < e^{-\alpha/\beta}$ so that $\|x_{k+1} - x^*\|_2 \leq e^{-T\alpha/(2\beta)} \|x_0 - x^*\|_2$. This, in turn, implies that $x_k \in B_{\varepsilon}(x^*)$ for all $k \geq T = \lceil 2\varepsilon \log(p/\varepsilon) \rceil$.\qed
H. Stochastic Convergence Results and Extended Analysis:

In this supplementary section, we provide the formal proofs for the stochastic convergence results.

Before diving into the results, let us describe at a higher level the flavor of the results. Beyond the asymptotic convergence guarantees above and the extended analysis in Appendix H.3, we can give concentration bounds which provided guarantees on the probability of getting ’locked in’ to a neighborhood of a differential Stackelberg equilibrium using results in (Borkar & Pattathil, 2018; Thoppe & Borkar, 2019) along with classical dynamical systems tools; however, due to the length of the exposition, we leave this for another paper.

The analysis techniques we employ combine tools from dynamical systems theory with the theory of stochastic approximation. In particular, we leverage the limiting continuous time dynamical systems derived from (2) to characterize concentration bounds for iterates or samples generated by (2). We note that the hierarchical learning update in (2) with timescale separation \( \gamma_{1,k} = o(\gamma_{2,k}) \) has a limiting dynamical system that takes the form of a \textit{singularly perturbed} dynamical system given by

\[
\begin{align*}
\dot{x}_1(t) &= -Df_1(x_1(t), x_2(t)) \\
\dot{x}_2(t) &= -\nu^{-1}Df_2(x_1(t), x_2(t))
\end{align*}
\]

which, in the limit as \( \nu \to 0 \), approximates (4).

The limiting dynamical system has known convergence properties (asymptotic convergence in a region of attraction for a locally asymptotically stable attractor). Such convergence properties can be translated in some sense to the discrete time system by comparing \textit{pseudo-trajectories}—in this case, linear interpolations between sample points of the update process—generated by sample points of (2) and the limiting system flow for initializations containing the set of sample points of (2).

Let us now review some mathematical preliminaries from dynamical systems theory.

H.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 \textit{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 \textit{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 \textit{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.

For the dynamics \( \dot{x} = -\omega(x) \), let \( J(x) \) denote the Jacobian of the vector field \( \omega(x) \). Similarly, for the dynamics \( \dot{x} = -\omega_S(x) \), let \( J_S(x) \) denote the Jacobian of the vector field \( \omega_S(x) \). Then, we say a differential Nash equilibrium of a continuous game with corresponding individual gradient vector field \( \omega \) is stable if \( \text{spec}(J(x)) \subset \mathbb{C}_- \) where \( \text{spec}(\cdot) \) denotes the spectrum of its argument and \( \mathbb{C}_- \) denotes the open left-half complex plane. Similarly, we say differential Stackelberg equilibrium is stable if \( \text{spec}(J_S(x)) \subset \mathbb{C}_- \).

In the stochastic setting, we use chain invariant sets.

**Definition 5.** 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 \((T, \delta)\)-perturbation.

**Lemma 2** (Hirsch Lemma). Given \( \varepsilon > 0, T > 0 \), there exists \( \tilde{\varepsilon} > 0 \) such that for all \( \varepsilon \in (0, \tilde{\varepsilon}) \), every \((T, \delta)\)-perturbation of \( \xi = F(\xi) \) converges to an \( \varepsilon \)-neighborhood of the global attractor set for \( \xi = F(\xi) \).

H.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 \( \text{dim}(X_i) = m_i \) for each \( i \in \{1, 2\} \) and \( m = m_1 + m_2 \).
Assumption 1. The following hold:

A1a. The maps $Df_1 : \mathbb{R}^m \rightarrow \mathbb{R}^{m_1}$, $D_2 f_2 : \mathbb{R}^d \rightarrow \mathbb{R}^{m_2}$ are $L_1$, $L_2$ Lipschitz, and $\|Df_1\| \leq M_1 < \infty$.

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

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 $c_i \geq 0$, $i \in \mathcal{I}$.

Assume 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 $x_{1,k}, x_{2,k+1} \in \arg\min_x f_2(x_{1,k}, x_2)$ for which $D_2 f_2(x_{1,k}, x_2) = 0$ 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 $D_2 f_2(x_1, x_2) = 0$, then the implicit function theorem implies that there exists neighborhoods $U \subset X_1$ and $V \subset X_2$ and a smooth map $r : U \rightarrow V$ such that $r(x_1) = x_2$.

Assumption 2. For every $x_1$, $\dot{x}_2 = -D_2 f_2(x_1, x_2)$ has a globally asymptotically stable equilibrium $r(x_1)$ uniformly in $x_1$ and $r : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^{d_2}$ is $L_r$--Lipschitz.

Consider the leader’s learning rule

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

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

Proposition 11. Suppose that for each $x \in X$, $D_2^2 f_2$ is non-degenerate and Assumption 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 1 holds for $i = 2$ and Assumption 2 holds, then $x_{2,k} \rightarrow x_2^* = r(x_1^*)$ 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 (20) is a stochastic approximation of $x_1 = -Df_1(x, x_2)$ and consequently is expected to track this ODE asymptotically. The main idea behind the analysis is to construct a continuous interpolated trajectory $\tilde{x}(t)$ for $t \geq 0$ and show it asymptotically almost surely approaches the solution set to the ODE. Under Assumptions 1–3, results from (Borkar, 2008, §2.1) imply that the sequence generated from (20) converges almost surely to a compact internally chain transitive set of $\tilde{x}_1 = -Df_1(x_1, x_2)$. 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 $D_2^2 f_1(x_1, r(x_1)) > 0$ and from assumption $D_2^3 f_2(x_1, r(x_1)) > 0$. Finally, from (Borkar, 2008, §2.2), we can conclude that the update from (20) 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 $x_{2,k} \rightarrow r(x_1^*)$ is guaranteed since $r$ is Lipschitz and $x_{1,k} \rightarrow x_1^*$.

The above result can be stated with a relaxed version of Assumption 2.

Corollary 4. Given a differential Stackelberg equilibrium $x^* = (x_1^*, x_2^*)$, let $B_p(x^*) = B_{p_1}(x_1^*) \times B_{p_2}(x_2^*)$ for some $p_1, p_2 > 0$ on which $D_2^2 f_2$ is non-degenerate. Suppose that Assumption 1 holds for $i = 1$ and that $x_{1,0} \in B_{p_1}(x_1^*)$. Then, $x_{1,k}$ converges almost surely to $x_1^*$. Moreover, if Assumption 1 holds for $i = 2$, $r(x_1^*)$ is a locally asymptotically stable equilibrium uniformly in $x_1$ on the ball $B_{p_2}(x_2^*)$, and $x_{2,0} \in B_{p_2}(x_2^*)$, then $x_{2,k} \rightarrow x_2^* = r(x_1^*)$.

The proof follows the same arguments as the proof of Proposition 11.

H.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 $D_2 f_2(x_{1,k}, x_{2,k}) = 0$, but the follower is actually performing the update $x_{2,k+1} = x_{2,k} + g_2(x_{1,k}, x_{2,k})$ where $g_2 \equiv -\gamma_{2,k} E[D_2 f_2]$. The learning dynamics in this setting are then

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

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

Proof.
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 $\nu \to 0$, the above approximates the singularly perturbed system defined by

$$
\dot{x}_1(t) = -Df_1(x_1(t), x_2(t)) \\
\dot{x}_2(t) = -\frac{1}{\nu} D_2f_2(x_1(t), x_2(t))
$$

(23)

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 equilibrated, meaning $D_2f_2(x_1, x_2) = 0$ given $x_1$. From this point of view, the learning dynamics (21)–(22) 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 justification for 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 $D_2f_2(x_{1,k}, x_2) = 0$. Moreover, under sufficient regularity conditions, the implicit mapping theorem (Lee, 2012) gives rise to the implicit map $r : U \to X : x_1 \mapsto x_2$ on a neighborhood $U \subset X_1$ of $x_{1,k}$. We note that when $x$ 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 $U \subset X$ 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,

$$Dr(x_1) = -(D_2^2f_2)^{-1}(x_1, r(x_1))D_2f_2(x_1, r(x_1))$$

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

$$Df_1(x_1, x_2) = D_1f_1(x_1, x_2) + D_2f_1(x_1, x_2)Dr(x_1)$$

(24)

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

H.3.1. Asymptotic Almost Sure Convergence

The following two results are fairly classical results in stochastic approximation. They are leveraged here to make 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 (21)–(22).

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

Under Assumption 1–3, and the assumption that $\gamma_{1,k} = o(\gamma_{2,k})$, classical results imply that the dynamics (21)–(22) converge almost surely to a compact internally chain transitive set $T$ of (23); see, e.g., (Borkar, 2008, §6.1-2), (Bhatnagar et al., 2012, §3.3). Furthermore, it is straightforward to see that stable differential Nash equilibria are internally chain transitive sets since they are stable attractors of the dynamics $\xi_t = F(\xi_t)$ from (23).

Remark 1. There are two important points to remark on at this juncture. First, the flow of the dynamics (23) is not necessarily a gradient flow, meaning that the dynamics may admit non-equilibrium attractors such as periodic orbits. The dynamics correspond to a gradient vector field if and only if $D_2f_2(f_1) = D_2f_2(0)$. This means that the dynamics admit a potential function. Equilibria may also not be isolated unless the Jacobian of $\omega_\Sigma$, say $J_\Sigma$, is non-degenerate at the points. Second, except in the case of zero-sum settings in which $(f_1, f_2) = (f, -f)$, non-Stackelberg locally asymptotically stable equilibria are attractors. That is, convergence does not imply that the players have settled on a Stackelberg equilibrium, and this can occur even if the dynamics admit a potential.
Let $t_k = \sum_{i=0}^{k-1} \gamma_{1,i}$ 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 initialization $x_s$.

**Proposition 12.** Suppose that Assumptions 1 and 2 hold. Then, conditioning on the event $\{\sup_k \sum_i \|x_{i,k}\|^2 < \infty\}$, for any integer $K > 0$, $\lim_{k \to \infty} \sup_{0 \leq k \leq K} \|x_{1,k+h} - \xi_{1,s}(t_{k-h})\|_2 = 0$ 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 $\{\sup_k \sum_i \|x_{i,k}\|^2 < \infty\}$, $(x_{1,k}, x_{2,k}) \to \{(x_1, r(x_1))\} \subset \mathbb{R}^{d_1}$ 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 $(\dot{x}_1 = 0, \dot{x}_2 = -D_2 f_2(x_1, x_2))$. That is, on intervals $[\hat{t}_j, \hat{t}_{j+1}]$ where $\hat{t}_j = \sum_{i=0}^{j-1} \gamma_{2,i}$, the norm difference between interpolated trajectories of the sample paths and the trajectories of $(\dot{x}_1 = 0, \dot{x}_2 = -D_2 f_2(x_1, x_2))$ vanishes a.s. as $k \to \infty$. Since the leader is tracking $\dot{x}_1 = 0$, the follower can be viewed as tracking $\dot{x}_2(t) = -D_2 f_2(x_1, x_2(t))$. Then applying Lemma 2 provided in Appendix H, $\lim_{k \to \infty} \|x_{2,k} - r(x_{1,k})\|_2 = 0$ almost surely.

Now, by Assumption 1, $Df_1$ is Lipschitz and bounded (in fact, independent of $A_{1a}$, since $Df_1 \in C^q$, $q \geq 1$, it is locally Lipschitz 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 Benaim (1999, Prop. 4.1). Moreover, under Assumptions $A_{1b}$ and $A_{1c}$, the assumptions of Benaim (1999, Prop. 4.2) are satisfied, which gives the desired result.

**Theorem 7.** Under Assumption 3 and the assumptions of Proposition 12, $(x_{1,k}, x_{2,k}) \to (x_1^*, x_2^*)$ almost surely conditioned on the event $\{\sup_k \sum_i \|x_{i,k}\|^2 < \infty\}$. That is, the learning dynamics (21)–(22) converge to stable attractors of (23), the set of which includes the stable DSE.

**Proof.** Continuing with the conclusion of the proof of Proposition 12, on intervals $[\hat{t}_k, \hat{t}_{k+1}]$ 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 2 gives the result.

As with Corollary 4, we can relax Assumptions 2 and 3 to local asymptotic stability assumptions.

**Proposition 6.** 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 $\gamma_{1,k} = o(\gamma_{2,k})$. Given a DSE $x^*$, let $B_p(x^*) = B_{p_1}(x_1^*) \times B_{p_2}(x_2^*)$ with $p_1, p_2 > 0$ on which $\det(D_2^2 f_2(x)) \neq 0$. Suppose $x_0 \in B_p(x^*)$. If the dynamics $\dot{x}_2 = -D_2 f_2(x)$ have a locally asymptotically stable attractor $r(x_1)$ uniformly in $x_1$ on $B_{p_1}(x_1^*)$ and the dynamics $\dot{x}_1 = Df_1(x_1, r(x_1))$ have a locally asymptotically stable attractor on $B_{p_1}(x_1^*)$, then $x_k \to x^*$ almost surely.

The proof follows the same arguments as the proof of Proposition 12.

In the zero-sum setting, we note that the continuous time limiting system admits only saddle points, and stable and unstable critical points; hence, we conjecture that avoidance of saddles should lead to a global convergence guarantee. In particular, if Stackelberg learning converges, then it converges to the set of DSE.

**H.3.2. 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 & Ratliff, 2018). 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 $X \subset \mathbb{R}^d$ and where $TX$ denotes the tangent space of $X$.

**Theorem 8 (Theorem 1 (Pemantle, 1990)).** Suppose $\gamma_t$ is $F_t$-measurable and $\mathbb{E}[w_t | F_t] = 0$. 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
Implicit Learning Dynamics in Stackelberg Games

point), (ii) $c_1/t^n \leq \gamma_t \leq c_2/t^n$, (iii) $E[(w_i \cdot v)^+|\mathcal{F}_t] \geq c_3/t^n$ for every unit vector $v \in TX$, and (iv) $\|w_i\|_2 \leq c_4/t^n$. Then $P(x_t \to p) = 0$.

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

**Theorem 4 (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 $E[(w_{i,t} \cdot v)^+|\mathcal{F}_{i,t}] \geq b_i$ 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 8, provided the assumptions of the theorem hold. The assumption that $E[(w_{i,t} \cdot v)^+|\mathcal{F}_{i,t}] \geq b_i$ 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.

H.3.3. Zero Sum Convergence

Leveraging the results in Section 3, the convergence guarantees are stronger since in zero-sum settings all attractors are Stackelberg; this contrasts with the Nash equilibrium concept.

**Corollary 5.** Consider a zero-sum setting $(f_1, -f)$ on $X = X_1 \times X_2 = \mathbb{R}^d$. Under the assumptions of Theorem 4 and Assumptions 2–3, conditioning on the event $\{\sup_k \sum_i \|x_{i,k}\|^2 < \infty\}$, the learning dynamics (21)–(22) asymptotically converge to the set of stable attractors almost surely, of which the only critical points are differential Stackelberg equilibria.

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

H.3.4. Finite-Time High-Probability Guarantees

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 (Benaïm, 1999; Bhatnagar et al., 2012; Borkar, 2008; Kushner & Yin, 2003). We also note that it is possible to extend the results to the $N$ follower setting. Due to the length of the current exposition, we have left this out.

While asymptotic guarantees of the proceeding section are useful, high-probability finite-time guarantees can be leveraged more directly in analysis and synthesis, e.g., of mechanisms to coordinate otherwise autonomous agents. In this section, we aim to provide concentration bounds for the purpose of deriving convergence rate and error bounds in support of this objective. The results in this section follow the very recent work by Borkar & Pattathil (2018). We highlight key differences and, in particular, where the analysis may lead to insights relevant for learning in hierarchical decision problems between non-cooperative agents.

Consider a locally asymptotically stable differential Stackelberg equilibrium $x^* = (x_1^*, r(x_1^*)) \in X$ and let $B_{q_0}(x^*)$ be an $q_0 > 0$ radius ball around $x^*$ contained in the region of attraction. Stability implies that the Jacobian $J_x(x_1^*, r(x_1^*))$ is positive definite and by the converse Lyapunov theorem (Sastry, 1999, Chap. 5) there exists local Lyapunov functions for the dynamics $\dot{x}_1(t) = -Df_1(x_1(t), r(x_1(t)))$ and for the dynamics $\dot{x}_2(t) = -\nu^{-1}Df_2(x_1(x_2(t)))$, for each fixed $x_1$. In particular, there exists a local Lyapunov function $V \in C^1(\mathbb{R}^d)$ with $\lim_{\|x_1\| \to \infty} V(x_1) = \infty$, and $(\nabla V(x_1), Df_1(x_1, r(x_1))) < 0$ for $x_1 \neq x_1^*$. For $q > 0$, let $V^q = \{x \in \text{dom}(V) : V(x) \leq q\}$. Then, there is also $q > q_0 > 0$ and $\epsilon_0 > 0$ such that for $\epsilon < \epsilon_0$,

$$\{x_1 \in \mathbb{R}^d_1 \mid \|x_1 - x_1^*\| \leq \epsilon\} \subseteq V^{q_0} \subseteq \mathcal{N}_\epsilon(V^{q_0}) \subseteq V^q \subseteq \text{dom}(V)$$

where

$$\mathcal{N}_\epsilon(V^{q_0}) = \{x \in \mathbb{R}^d_1 \mid \exists x' \in V^{q_0} \text{ s.t. } \|x' - x\| \leq \epsilon_0\}.$$  

An analogously defined $\check{V}$ exists for the dynamics $\check{x}_2$ for each fixed $x_1$.

For now, fix $n_0$ sufficiently large; we specify the values of $n_0$ for which the theory holds before the statement of Theorem 9.

Define the event $\mathcal{E}_n = \{\check{x}_2(t) \in V^q \forall t \in \tilde{t}_{n_0, \tilde{t}_{n_0}}\}$ where

$$\check{x}_2(t) = x_{2,k} + \frac{\epsilon - \epsilon_n}{72 \delta} (x_{2,k+1} - x_{2,k})$$
are linear interpolates—i.e., asymptotic pseudo-trajectories—defined for $t \in (\hat{t}_k, \hat{t}_{k+1})$ with $\hat{t}_{k+1} = \hat{t}_k + \gamma_2 k$ and $\hat{t}_0 = 0$.

The basic idea for constructing a concentration bound is to leverage Alekseev’s formula (Thm. ?) to bound the difference between the asymptotic pseudo-trajectories and the flow of the corresponding limiting differential equation on each continuous time interval between each of the successive iterates $k$ and $k+1$ by sequences of constants that decay asymptotically. Then, a union bound is used over all time intervals after defined for $n \geq n_0$ in order to construct a concentration bound. This is done first for the follower, showing that $x_{2,k}$ tracks the leader’s ‘conjecture’ or belief $r(x_{1,k})$ about the follower’s reaction, and then for the leader.

**Follower’s Sequence Tracks the Leader’s Conjectured Sample Path.** Let us start by showing that $x_{2,k}$ tracks the leader’s ‘conjecture’ or belief $r(x_{1,k})$ about the follower’s reaction.

Following Borkar & Pattathil (2018), we can express the linear interpolates for any $n \geq n_0$ as

$$\bar{x}_2(\hat{t}_{n+1}) = \bar{x}_2(\hat{t}_{n_0}) - \sum_{k=n_0}^{n} \gamma_{2,k} (D_2 f_2(x_k) + w_{2,k+1})$$

where $\gamma_{2,k} D_2 f_2(x_k) = \int_{\hat{t}_k}^{\hat{t}_{k+1}} D_2 f_2(x_k, \bar{x}_2(\hat{t}_k)) \, ds$ and similarly for the $w_{2,k+1}$ term. Adding and subtracting $\int_{\hat{t}_{n_0}}^{\hat{t}_{n+1}} D_2 f_2(x_1(s), \bar{x}_2(s)) \, ds$, Alekseev’s formula can be applied to get

$$\bar{x}_2(t) = x_2(t) + \Phi_2(t, s, x_1(\hat{t}_{n_0}), \bar{x}_2(\hat{t}_{n_0}))(\bar{x}_2(\hat{t}_{n_0}) - x_2(\hat{t}_{n_0})) + \int_{\hat{t}_{n_0}}^{t} \Phi_2(t, s, x_1(s), \bar{x}_2(s)) \zeta_2(s) \, ds$$

where $x_1(t) \equiv x_1$ is constant (since $\dot{x}_1 = 0$), $x_2(t) = r(x_1)$, and

$$\zeta_2(s) = -D_2 f_2(x_1(\hat{t}_k), \bar{x}_2(\hat{t}_k)) + D_2 f_2(x_1(s), \bar{x}_2(s)) + w_{2,k+1}.$$  \hfill (25)

In addition, for $t \geq s$, $\Phi_2(\cdot)$ satisfies linear system

$$\dot{\Phi}_2(t, s, x_0) = J_2(x_1(t), x_2(t)) \Phi_2(t, s, x_0)$$

with $\Phi_2(t, s, x_0) = I$ and $x_0 = (x_{1,0}, x_{2,0})$ and where $J_2$ the Jacobian of $-D_2 f_2(x_1, \cdot)$.

Given that $x^* = (x^*_1, r(x^*_1))$ is a stable differential Stackelberg equilibrium, $J_2(x^*)$ is positive definite. Hence, as in (Thoppe & Borkar, 2019, Lem. 5.3), we can find $M, \kappa_2 > 0$ such that for $t \geq s, x_{2,0} \in V^q$, $\|\Phi_2(t, s, x_{1,0}, x_{2,0})\| \leq M e^{-\kappa_2(t-s)}$; this result follows from standard results on stability of linear systems (see, e.g., Callier & Desoer (1991, §7.2, Thm. 33)) along with a bound on

$$\int_{t-s}^{t} \|D_2 f_2(x_1, x_2(\theta, s, x_0)) - D_2 f_2(x^*)\| d\theta$$

for $x_0 \in V^q$ (see, e.g., Thoppe & Borkar (2019, Lem 5.2)).

Now, an interesting point worth making is that this analysis leads to a very nice result for the leader-follower setting. In particular, through the use of the auxiliary variable $z$, we can show that the follower’s sample path ‘tracks’ the leader’s conjectured sample path. Indeed, consider $z_k = r(x_{1,k}, \bar{z}_k)$, that is, $x_2 = D_2 f_2(x_1, z_k) = 0$. Then, using a Taylor expansion of the implicitly defined conjecture $r$, we get $z_{k+1} = z_k + D_2 f_2(x_1, \bar{z}_k + \bar{z}_k) + \delta_{k+1}$ where $\|\delta_{k+1}\| \leq L_r \|x_{1,k+1} - x_{1,k}\|^2$ is the error from the remainder terms. Plugging in $x_{1,k+1},$

$$z_{k+1} = z_k + \gamma_{2,k} (D_2 f_2(x_1, z_k) + \tau_k D_2 r_2(x_1, w_{1,k+1} - x_{1,k}) + \gamma_{2,k}^{-1} \delta_{k+1})$$

where $\tau_k = \gamma_{1,k}/\gamma_{2,k}$. The terms after $-D_2 f_2$ are $o(1)$, and hence asymptotically negligible, so that this $z$ sequence tracks dynamics as $x_{2,k}$. We show that with high probability, they asymptotically contract, leading to the conclusion that the follower’s dynamics track the leader’s conjecture.

Towards this end, we first bound the normed difference between $x_{2,k}$ and $z_k$. Define constants

$$H_{n_0} = (\|\bar{x}_2(\hat{t}_{n_0}) - x_2(\hat{t}_{n_0})\| + \|\bar{z}(\hat{t}_{n_0}) - x_2(\hat{t}_{n_0})\|),$$

and

$$S_{2,n} = \sum_{k=n_0}^{n-1} (\int_{\hat{t}_k}^{\hat{t}_{k+1}} \Phi_2(\hat{t}_n, s, x_1(\hat{t}_n), \bar{x}_2(\hat{t}_n)) ds) w_{2,k+1}.$$
Lemma 3. For any $n \geq n_0$, there exists $K > 0$ such that conditioned on $\mathcal{E}_n$,
\[
\|x_{2,n} - z_n\| \leq K (\|S_{2,n}\| + e^{-\kappa_2 (t_n - t_0)} H_{n_0} + \sup_{k \leq n+1} \gamma_{2,k} \|w_{2,k+1}\| + \sup_{k \leq n+1} \tau_k \|w_{1,k+1}\|)
\]
where $\tau_k = \gamma_{1,k}/\gamma_{2,k}$.

Using this bound, we can provide an asymptotic guarantee that $x_{2,k}$ tracks $r(x_{1,k})$ and a high-probability guarantee that $x_{2,k}$ gets locked in to a ball around $r(x_1^*)$. Fix $\varepsilon \in (0,1)$ and let $N$ be such that $\gamma_{2,n} \leq \varepsilon/(8K), \tau_n \leq \varepsilon/(8K)$ for all $n \geq N$. Let $n_0 \geq N$ and with $K$ as in Lemma 3, let $T$ be such that $e^{-\kappa_2 (t_n - t_0)} H_{n_0} \leq \varepsilon/(8K)$ for all $n \geq n_0 + T$.

Theorem 9. Suppose that Assumptions 1, 2, and 3 hold and let $\gamma_{1,k} = o(\gamma_{2,k})$. Given a stable differential Stackelberg equilibrium $x^* = (x_1^*, r(x_1^*))$, the follower's sample path generated by (22) with asymptotically track the leader's conjecture $z_k = r(x_{1,k})$ and, given $\varepsilon \in [0,1)$, will get ‘locked in’ to a $\varepsilon$–neighborhood with high probability conditioned on reaching $B_{n_0}(x^*)$ by iteration $n_0$. That is, letting $\bar{n} = n_0 + T + 1$, for some $C_1, C_2, C_3, C_4 > 0$, we have
\[
P(\|x_{2,n} - z_n\| \leq \varepsilon, \forall n \geq \bar{n}|x_{2,n_0}, z_{n_0} \in B_{n_0}) \geq 1 - \sum_{n=n_0}^{\infty} C_1 e^{-C_2 \sqrt{\varepsilon/n_2} + \sum_{n=n_0}^{\infty} C_2 e^{-(C_2/\beta_n)}/\beta_n}.
\]

with $\beta_n = \max_{n_0 \leq k \leq n+1} e^{-\kappa_2 (\sum_{s=k+1}^{n+1} \gamma_{2,s})} \gamma_{2,k}$.

The key technique in proving the above theorem (which is done in detail in Borkar & Pattathil (2018) using results from Thoppe & Borkar (2019)), is taking a union bound of the errors over all the continuous time intervals defined for $n \geq n_0$.

Proofs of Theorem 9 and Lemma 3. Constructing the asymptotic pseudo-trajectories. Recall that the asymptotic pseudo-trajectories for any $n \geq n_0$ are given by
\[
\bar{x}_2(\tilde{t}_{n+1}) = \bar{x}_2(\tilde{t}_{n_0}) - \sum_{k=n_0}^{n} \gamma_{2,k}(D_2 f_2(x_k) + w_{2,k+1}).
\]

Note that
\[
\sum_{k=n_0}^{n} \gamma_{2,k} D_2 f_2(x_k) = \sum_{k=n_0}^{n} \int_{\tilde{t}_k}^{\tilde{t}_{k+1}} D_2 f_2(x_{1,k}, \bar{x}_2(t_k)) \, ds
\]
and similarly for the $w_{2,k+1}$ term, due to the fact that $\tilde{t}_{k+1} - \tilde{t}_k = \gamma_{2,k}$ by construction. Hence, for $s \in [\tilde{t}_k, \tilde{t}_{k+1}]$, the above can be rewritten as
\[
\bar{x}_2(t) = \bar{x}_2(\tilde{t}_{n_0}) + \int_{\tilde{t}_{n_0}}^{t} -D_2 f_2(x_1(s), \bar{x}_2(s)) + \bar{z}_{21}(s) + \bar{z}_{22}(s) \, ds
\]
where $\bar{z}_{21}(s) = -D_2 f_2(x_1(\tilde{t}_k), \bar{x}_2(\tilde{t}_k)) - D_2 f_2(x_1(s), \bar{x}_2(s))$ and $\bar{z}_{22}(s) = -w_{2,k+1}$. Note that $\bar{z}_{2}(s)$ in (25) is defined such that $\bar{z}_2(s) = \bar{z}_{21}(s) + \bar{z}_{22}(s)$. Then, by the nonlinear variation of constants formula (Alekseev's formula), we have
\[
\bar{x}_2(t) = x_2(t) + \Phi_2(t, s, x_1(\tilde{t}_{n_0}), \bar{x}_2(\tilde{t}_{n_0}))(\bar{x}_2(\tilde{t}_{n_0}) - \bar{x}_2(t)) + \int_{\tilde{t}_{n_0}}^{t} \Phi_2(t, s, x_1(s), \bar{x}_2(s)) (\bar{z}_{21}(s) + \bar{z}_{22}(s)) \, ds
\]
where $x_1(t) \equiv x_1$ is constant (since $\dot{x}_1 = 0$) and $x_2(t) = r(x_1)$. Moreover, for $t \geq s$, $\Phi_2(s)$ satisfies linear system
\[
\Phi_2(t, s, x_0) = J_2(x(t), x(t)) \Phi_2(t, s, x_0),
\]
with initial data $\Phi_2(t, s, x_0) = I$ and $x_0 = (x_{1,0}, x_{2,0})$ and where $J_2$ the Jacobian of $-D_2 f_2(x_1, \cdot)$. Given that $x^* = (x_1^*, r(x_1^*))$ is a stable differential Stackelberg equilibrium, $J_2(x^*)$ is positive definite. Hence, as in (Thoppe & Borkar, 2019, Lem. 5.3), we can find $M, \kappa > 0$ such that for $t \geq s, x_{2,0} \in V^r$,
\[
\|\Phi_2(t, s, x_{1,0}, x_{2,0})\| \leq M e^{-\kappa (t-s)}.
\]
This result follows from standard results on stability of linear systems (see, e.g., Callier & Desoer (1991, §7.2, Thm. 33)) along with a bound on $\int_s^t \|D_2^2 f_2(x_1, x_2(\theta, s, \bar{x}_0)) - D_2^2 f_2(x^*)\|d\theta$ for $\bar{x}_0 \in V^q$ (see, e.g., Thoppe & Borkar (2019, Lem. 5.2)).
Recall also that we have the auxiliary sequence representing the leader’s belief (i.e., assumed response model) which we derived using a Taylor expansion of the implicitly defined map \( r \):

\[
z_{k+1} = z_k + Dr(x_{1,k})(x_{1,k+1} - x_{1,k}) + \delta_{k+1}
\]

(27)

where \( \delta_{k+1} \) are the remainder terms which satisfy \( \|\delta_{k+1}\| \leq L_r \|x_{1,k+1} - x_{1,k}\|^2 \) by assumption. Plugging in \( x_{1,k+1} \),

\[
z_{k+1} = z_k + \gamma_{2,k}( - D_2 f_2(x_{1,k}, z_k) + \tau_k Dr(x_{1,k})(w_{1,k+1} - D f_1(x_{1,k}, x_{2,k})) + \gamma_{-1,2,k}(\delta_{k+1}).
\]

The terms after \(-D_2 f_2\) are \( o(1) \), and hence asymptotically negligible, so that this \( z \) sequence tracks dynamics as \( x_{2,k} \). Using similar techniques as above, we can express linear interpolates of the leader’s assumed response model for the follower as

\[
\bar{z}(t) = \bar{z}(\hat{t}_{n_0}) + \int_{\hat{t}_{n_0}}^t - D_2 f_2(x_1(s), \bar{z}(s)) + \sum_{j=1}^d \zeta_j(s) \, ds
\]

where we recall that \( \tau_k = \gamma_{1,k}/\gamma_{2,k} \) and where the \( \zeta_j \)'s are defined as follows:

\[
\begin{align*}
\zeta_{31}(s) &= -D_2 f_2(x_1(\hat{t}_k), \bar{z}(\hat{t}_k)) + D_2 f_2(x_1(s), \bar{z}(s)) \\
\zeta_{32}(s) &= \tau_k Dr(x_{1,k})w_{1,k+1} \\
\zeta_{33}(s) &= -\tau_k D f_1(x_{1,k}, x_{2,k})Dr(x_{1,k}) \\
\zeta_{34}(s) &= \frac{1}{\gamma_{2,k}} \delta_{k+1}
\end{align*}
\]

Once again, Alekseev’s formula can be applied where \( x_2(t) = r(x_1) \) and \( \Phi_2 \) is the same as in the application of Alekseev’s to \( x_{2,k} \). Indeed, this gives us

\[
\bar{z}(\hat{t}_n) = x_2(\hat{t}_n) + \Phi_2(\hat{t}_n, \hat{t}_{n_0}, x_1(\hat{t}_{n_0}), \bar{z}(\hat{t}_{n_0}))(\bar{z}(\hat{t}_{n_0}) - x_2(\hat{t}_{n_0}))
\]

(28)

\[
+ \sum_{k=0}^{n-1} \int_{\hat{t}_k}^{\hat{t}_{k+1}} \Phi_2(\hat{t}_n, s, x_1(s), \bar{z}(s))(-D_2 f_2(x_1(\hat{t}_k), \bar{z}(\hat{t}_k)) + D_2 f_2(x_1(s), \bar{z}(s))) \, ds \\
+ \sum_{k=0}^{n-1} \int_{\hat{t}_k}^{\hat{t}_{k+1}} \Phi_2(\hat{t}_n, s, x_1(s), \bar{z}(s))\tau_k Dr(x_{1,k})w_{1,k+1} \, ds \\
- \sum_{k=0}^{n-1} \int_{\hat{t}_k}^{\hat{t}_{k+1}} \Phi_2(\hat{t}_n, s, x_1(s), \bar{z}(s))\tau_k D f_1(x_{1,k}, x_{2,k})Dr(x_{1,k}) \, ds \\
+ \sum_{k=0}^{n-1} \int_{\hat{t}_k}^{\hat{t}_{k+1}} \Phi_2(\hat{t}_n, s, x_1(s), \bar{z}(s))\frac{1}{\gamma_{2,k}} \delta_{k+1} \, ds
\]

Applying the linear system stability results, we get that

\[
\|\Phi_2(\hat{t}_n, \hat{t}_{n_0}, x_1(\hat{t}_{n_0}), \bar{z}(\hat{t}_{n_0}))(\bar{z}(\hat{t}_{n_0}) - x_2(\hat{t}_{n_0}))\| \leq e^{-\gamma_2(\hat{t}_n - \hat{t}_{n_0})}\|\bar{z}(\hat{t}_{n_0}) - x_2(\hat{t}_{n_0})\|
\]

Each of the terms (a)–(d) can be bound as in Lemma III.1–5 in (Borkar & Patthathil, 2018). The bounds are fairly straightforward using (28).

**Tracking bounds and Proof of Lemma 3.** Now that we have each of these asymptotic pseudo-trajectories, we can show that with high probability, \( x_{2,k} \) and \( z_k \) asymptotically contract to one another, leading to the conclusion that the follower’s dynamics track the leader’s belief about the follower’s reaction.

Moreover, we can bound the difference between \( x_{2,k} \), using \( \bar{x}_2(t_{2,k}) = x_{2,k} \), and the continuous flow \( x_2(t) \) on each interval \([t_{2,k}, t_{2,k+1}]\) where \( t_{2,k} = \hat{t}_k \). The normed-difference bound can then be leveraged to obtain concentration bounds by taking a union bound across all continuous time intervals defined after sufficiently large \( n_0 \) and conditioned on the events \( \mathcal{E}_n = \{\bar{x}_2(t) \in V^q \forall t \in [\hat{t}_{n_0}, \hat{t}_{n}]\} \).

Towards this end, define \( H_{n_0} = (\|\bar{x}_2(\hat{t}_{n_0}) - x_2(\hat{t}_{n_0})\| + \|\bar{z}(\hat{t}_{n_0}) - x_2(\hat{t}_{n_0})\|) \), and

\[
S_{2,n} = \sum_{k=n_0}^{n-1} \left( \int_{\hat{t}_k}^{\hat{t}_{k+1}} \Phi_2(\hat{t}_n, s, x_1(\hat{t}_k), \bar{x}_2(\hat{t}_k))ds \right) w_{2,k+1}.
\]

Applying Lemma 5.8 (Thoppe & Borkar, 2019), conditioned on \( \mathcal{E}_n \), we get there exists some constant \( K > 0 \) such that

\[
\|\bar{x}_2(\hat{t}_n) - x_2(\hat{t}_n)\| \leq \|\Phi_2(\hat{t}_n, \hat{t}_{n_0}, x_1, \bar{x}_2(\hat{t}_{n_0}))(\bar{x}_2(\hat{t}_{n_0}) - x_2(\hat{t}_{n_0}))\| + K \left( \|S_{2,n}\| \\
+ \sup_{n_0 \leq k \leq n-1} \gamma_{2,k} + \sup_{n_0 \leq k \leq n-1} \gamma_{2,k} \|w_{2,k+1}\|^2 \right).
\]
Implicit Learning Dynamics in Stackelberg Games

Using the bound on the linear system $\Phi_2(\cdot)$, this exactly leads to the bound
\[
\|\tilde{x}_2(t_n) - x_2(t_n)\| \leq K \left( e^{-\gamma_2(t_n - t_{n_0})} \|\tilde{x}_2(t_{n_0}) - x_2(t_{n_0})\| \right.
\]
\[
+ \|S_{2,n}\| + \sup_{n_0 \leq k \leq n-1} \gamma_2 k + \sup_{n_0 \leq k \leq n-1} \gamma_2 k \|w_{2,k+1}\|^2 \right)
\]

Thus, leveraging Lemma III.1–5 (Thoppe & Borkar, 2019), we obtain the result of Lemma 3.

**Proving Concentration Bounds in Theorem 9.** Now, to obtain the concentration bounds, fix $\varepsilon \in [0,1)$ and let $N$ be such that $\gamma_2 n \leq \varepsilon/(8K)$, $\tau_n \leq \varepsilon/(8K)$ for all $n \geq N$. Let $n_0 \geq N$ and with $K$ as in Lemma 3, let $T$ be such that $e^{-\gamma_2(t_{n_0} - t_{n_0})}H_{n_0} \leq \varepsilon/(8K)$ for all $n \geq n_0 + T$.

Using Lemma 3 and Lemma 3.1 (Thoppe & Borkar, 2019),
\[
P(\|x_{2,n} - z_n\| \leq \varepsilon, \forall n \geq n_0 | x_{2,n_0}, z_{n_0} \in B_{B_0}) 
\geq 1 - P(\bigcup_{n=n_0}^{\infty} A_{1,n} \cup \bigcup_{n=n_0}^{\infty} A_{2,n} \cup \bigcup_{n=n_0}^{\infty} A_{3,n} | x_{2,n_0}, z_{n_0} \in B_{B_0})
\]
where
\[
A_{1,n} = \{E_n, \|S_{2,n}\| > \frac{\varepsilon}{8K}\},
A_{2,n} = \{E_n, \gamma_2 k \|w_{2,k+1}\|^2 > \frac{\varepsilon}{8K}\},
A_{3,n} = \{E_n, x_n \|w_{n+1}\|^2 > \frac{\varepsilon}{8K}\}.
\]

Taking a union bound gives
\[
P(\|x_{2,n} - z_n\| \leq \varepsilon, \forall n \geq n_0 | x_{2,n_0}, z_{n_0} \in B_{B_0}) \geq 1 - \sum_{n=n_0}^{\infty} P(A_{1,n} | x_{2,n_0}, z_{n_0} \in B_{B_0})
\]
\[
+ \sum_{n=n_0}^{\infty} P(A_{2,n} | x_{2,n_0}, z_{n_0} \in B_{B_0})
\]
\[
+ \sum_{n=n_0}^{\infty} P(A_{3,n} | x_{2,n_0}, z_{n_0} \in B_{B_0}).
\]

Theorem 6.2 (Thoppe & Borkar, 2019), gives bounds
\[
\sum_{n=n_0}^{\infty} P(A_{2,n} | x_{2,n_0}, z_{n_0} \in B_{B_0}) \leq K_1 \sum_{n=n_0}^{\infty} \exp \left( -\frac{K^2 \varepsilon^2}{\sqrt{2\pi}} \right),
\]
and
\[
\sum_{n=n_0}^{\infty} P(A_{3,n} | x_{2,n_0}, z_{n_0} \in B_{B_0}) \leq K_1 \sum_{n=n_0}^{\infty} \exp \left( -\frac{K^2 \varepsilon^2}{\sqrt{2\pi}} \right).
\]

By Theorem 6.3 (Thoppe & Borkar, 2019), we also have that
\[
\sum_{n=n_0}^{\infty} P(A_{1,n} | x_{2,n_0}, z_{n_0} \in B_{B_0}) \leq K_2 \sum_{n=n_0}^{\infty} \exp \left( -\frac{K_3 e^{2\beta_n}}{\beta_n} \right)
\]
with
\[
\beta_n = \max_{n_0 \leq k \leq n-1} e^{-\gamma_2 (\sum_{i=k+1}^{n-1} \gamma_2 k)}
\]
for some $K_1, K_2, K_3 > 0$. This proves the result of Theorem 9 with $C_1 = K_1$, $C_2 = K^2$, $C_3 = K_2$, $C_4 = K_3$.

The above theorem can be restated to give a guarantee on getting locked-in to an $\varepsilon$-neighborhood of a stable differential Stackelberg equilibria $x^*$ if the learning processes are initialized in $B_{B_0}(x^*)$.

**Corollary 6.** Fix $\varepsilon \in [0,1)$ and suppose that $\gamma_2 n \leq \varepsilon/(8K)$ for all $n \geq 0$. With $K$ as in Lemma 3, let $T$ be such that $e^{-\gamma_2(t_{n_0} - t_{n_0})}H_{n_0} \leq \varepsilon/(8K)$ for all $n \geq T$. Under the assumptions of Theorem 9, $x_{2,k}$ will get ‘locked in’ to an $\varepsilon$-neighborhood with high probability conditioned on $x_0 \in B_{B_0}(x^*)$ where the high-probability bound is given in (26) with $n_0 = 0$.

**Leader’s Sample Path Gets Locked in to a Neighborhood of the Stackelberg Equilibrium.** Given that the follower’s action $x_{2,k}$ tracks $r(x_{1,k})$, we can also show that $x_{1,k}$ gets locked into an $\varepsilon$-neighborhood of $x^*_1$ after a finite time with high probability. First, a similar bound as in Lemma 3 can be constructed for $x_{1,k}$.
Theorem 10. Suppose that Assumptions 1–3 hold and that \( \gamma_{1,k} = o(\gamma_{2,k}) \). Given a stable differential Stackelberg equilibrium \( x^* \) and \( \varepsilon \in (0, 1) \), \( x_0 \) will get ‘locked in’ to a \( \varepsilon \)-neighborhood of \( x^* \) with high probability conditioned reaching \( B_{\eta_0}(x^*) \) by iteration \( n_0 \). That is, letting \( \bar{n} = n_0 + T + 1 \), for some constants \( C_j > 0 \), \( j \in \{1, \ldots, 6\} \),

\[
P(\|x_{1,n} - x(\hat{t}_n)\| \leq \varepsilon, \forall n \geq \bar{n}, x_0, x_{\hat{t}_n} \in B_{\eta_0}) \]
\[
\geq 1 + \sum_{n=n_0}^{\infty} \bar{C}_1 e^{-\bar{C}_2 \varepsilon^2 / \sqrt{n}} - \sum_{n=n_0}^{\infty} \bar{C}_1 e^{-\bar{C}_2 \varepsilon^2 / \sqrt{n}}
\]
\[
- \sum_{n=n_0}^{\infty} \bar{C}_3 e^{-\bar{C}_4 \varepsilon^2 / \beta n} - \sum_{n=n_0}^{\infty} \bar{C}_5 e^{-\bar{C}_6 \varepsilon^2 / \eta_n}
\]

(32)

with \( \eta_n = \max_{n_0 \leq k \leq n-1} \left(e^{-\bar{C}_1 (\sum_{n=n_0}^{k+1} \gamma_{1,k})}\right) \gamma_{1,k} \).

The proof of this theorem is exactly analogous to that of Theorem 9.

Proof Sketch of Theorem 10 and Lemma 4. Using the linear interpolates \( \bar{x}_1(t) \) as above, Alekseev’s formula can again be applied to get

\[
\bar{x}_1(t) = x_1(t, \hat{t}_n, y(\hat{t}_n)) + \Phi_1(t, \hat{t}_n, \bar{x}_1(\hat{t}_n))(\bar{x}_1(\hat{t}_n) - x_1(\hat{t}_n)) + \int_{\hat{t}_n}^t \Phi_1(s, \bar{x}_1(s)) d\zeta_1(s) d\zeta_2(s) d\zeta_3(s)
\]

where \( x_1(t) = x_1^* \) again, since \( \bar{x}_1 = 0 \) and the following hold:

\[
\zeta_1(s) = Df_1(x_1, r(\bar{x}_1(s)))
\]
\[
\zeta_2(s) = Df_2(x_1) - Df_1(x_1, r(\bar{x}_1(s)))
\]
\[
\zeta_3(s) = w_1, k+1
\]
Moreover, $\Phi_1$ is the solution to a linear system with dynamics $J_1(x^*_1, r(x^*_1))$, the Jacobian of $-Df_1(\cdot, r(\cdot))$, and with initial data $\Phi_1(s, s, x_{1,0}) = I$. This linear system, as above, has bound

$$\|\Phi_1(t, s, x_{1,0})\| \leq M_1 e^{\kappa_1(t-1)}$$

for some $M_1, \kappa_1 > 0$.

Moreover, we can bound the difference between $x_{1,k}$, using $\hat{x}_1(t_{1,k}) = x_{1,k}$, and the continuous flow $x_1(t)$ on each interval $[t_{1,k}, t_{1,k+1})$ where $t_{1,k} = \hat{t}_k$. The normed-difference bound can then be leveraged to obtain concentration bounds by taking a union bound across all continuous time intervals defined after sufficiently large $n_0$ and conditioned on the events $\mathcal{E}_n = \{\hat{x}_2(t) \in V^q \forall t \in [t_{n_0}, t_{n}]\}$ and $\mathcal{E}_n = \{x_1(t) \in V^q \forall t \in [t_{n_0}, t_{n}]\}$. Indeed, following a similar argument as in the proof of Theorem 9 and Lemma 3, we obtain the bound in Lemma 4. From here, the argument follows exactly the concentration bound argument in the remainder of the proof of Theorem 9.

An analogous corollary to Corollary 6 can be stated for $x_{1,k}$ with $n_0 = 0$; we do not present it to save space.

I. Stability and Genericity in Zero-Sum Continuous Stackelberg Games

It is important to reiterate 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.

Necessary conditions for a local Stackelberg solution for the leader follow from necessary conditions in nonlinear optimization.

**Proposition 13.** Suppose that $x^* = (x^*_1, x^*_2)$ is a local Stackelberg equilibrium such that the follower (player 2) is at a strict local minimum—i.e., $D_2 f_2(x^*_1, x^*_2) = 0$ and $D_2^2 f_2(x^*_1, x^*_2) > 0$. Then, $D_1 f_1(x^*_1, x^*_2) = 0$ and $D_2^2 f_1(x^*_1, x^*_2) \geq 0$.

**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 (Lee, 2012), there exists a $C^q$ map $r : x_1 \mapsto x_2$ is defined on a neighborhood of $x^*_1$ such that $r(x^*_1) = x^*_2$ and $Dr \equiv -(D_2^2 f_2)^{-1} \circ D_2 f_2$. Hence, necessary conditions for the leader reduce to necessary conditions on the problem $\min_{x_1} f_1(x_1, r(x_1))$.

**Proposition 14.** In zero-sum $q$-differentiable continuous games, all differential Stackelberg equilibria are non-degenerate, and hence hyperbolic critical points of the vector field $\omega_G(x)$.

**Proof.** Indeed, the Jacobian of the vector field $\omega_G(x)$ is lower block triangular as shown in (12). Hence, $\det(J_G(x)) \neq 0$ if and only if $\det(D_1(D_2 f(x))) \neq 0$ and $\det(-D_2^2 f(x)) \neq 0$. Since differential Stackelberg are such that $D_1(D_2 f(x)) \geq 0$ and $-D_2^2(f(x)) \geq 0$, the fact that all differential Stackelberg are non-degenerate follows immediately. Further, the lower block triangular structure implies that $\text{spec}(J_G(x)) = \text{spec}(S_1(J(x))) \cup \text{spec}(-D_2^2 f(x))$. Hence, all differential Stackelberg equilibria are hyperbolic.

As with the case of simultaneous play and the corresponding equilibrium concept, the remainder of 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. For a zero-sum $q$-differentiable game $G = (f, -f)$, if we let $DSE(G)$ be the differential Stackelberg equilibria, $NDSE(G)$ the non-degenerate differential Stackelberg equilibria, and $LSE(G)$ the local Stackelberg equilibria, then we know that $NDSE(G) = DSE(G) \subseteq LSE(G)$. What we show is that for generic $f \in C^q(X, \mathbb{R})$, the game $G = (f, -f)$ is such that $LSE(G) = DSE(G)$. In particular, we show that the set of zero-sum $q$-differentiable games admitting any local Stackelberg equilibria which are not non-degenerate differential Stackelberg is of measure zero in $C^q(\mathbb{R}^m, \mathbb{R})$.

I.1. Mathematical Preliminaries

In this appendix, we provide some additional mathematical preliminaries; the interested reader should see standard references for a more detailed introduction (Abraham et al., 1988; Lee, 2012).
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 $T^p_s(E) = L^{p+s}(E^*, \ldots, E^*, E, \ldots, E; \mathbb{R})$ with $s$ copies of $E$ and $p$ copies of $E^*$ and where $E^*$ denotes the dual. Elements of $T^p_s(E)$ are tensors on $E$, and $T^p_s(E)$ 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 6.** 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 $f \cap Z$ at $u$) if either $f(u) \notin Z$ or $f(u) \in Z$ and $T_f(u) Y = T_f(u) Z + (j_s)_u(T_u X)$.

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

$$\bigcap^s(X, Y; j^k, Z) = \{ h \in C^s(X, Y) | j^k h \cap Z \}. \quad (33)$$

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

**Proposition 15.** (Chap. II.4, Proposition 4.2 (Golubitsky & Guillemin, 1973)). Let $X, Y$ be smooth manifolds and $Z \subset Y$ a submanifold. Suppose that $\dim X < \text{codim} Z$. Let $f : X \to Y$ be smooth and suppose that $f \cap Z$. Then, $f(X) \cap Z = \emptyset$.

### I.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 5** (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

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

is non-degenerate at critical points—that is, $\det(H(x)) \neq 0$. For a generic function $f$, we know that $\det(D^2 f(x)) \neq 0$ (cf. Lemma 7). For such a function,

$$S_1(J(x)) = D_1^2 f(x) - D_2 1 f(x) (D_2^2 f(x))^{-1} D_2 1 f(x).$$

Furthermore, we denote the Stackelberg game Jacobian by

$$J_S(x) = \begin{bmatrix} D_1 (D f(x)) & 0 \\ -D_2 1 f(x) & -D_2^2 f(x) \end{bmatrix}.$$

where $J_S(x)$ is obtained by taking the Jacobian of the vector field $(D f(x), -D_2 f(x))$. At critical points—i.e., $x$ such that $(D f(x), -D_2 f(x)) = 0$—we note that

$$J_S(x) = \begin{bmatrix} S_1(J(x)) & 0 \\ -D_2 1 f(x) & -D_2^2 f(x) \end{bmatrix}.$$

\footnote{Note that the structure is lower block triangular since $D_2 (D f(x)) = D_2 1 f(x) + D r(x_1)^T D_2^2 f(x) = D_2 1 f(x) - D_2 1 f(x) (D_2^2 f(x))^{-1} D_2^2 f(x) = 0.$}
Lemma 6. 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 $x \in \mathbb{R}^m$ (i.e., $x \in \{x \in \mathbb{R}^m | \omega_0(x) = 0\}$), $\det(H(x)) \neq 0 \iff \det(J(x)) \neq 0$ and if $\det(-D_2^2 f(x)) \neq 0$, then $\det(H(x)) \neq 0 \iff \det(J_S(x)) \neq 0$.

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 $P$ is 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 $J_S(x)$ is well-defined. Then, using the Schur decomposition of $J(x)$, it is easily seen that $\det(J(x)) = \det(D_1(J(x))) \det(-D_2^2 f(x)) = \det(J_S(x))$ where the last equality holds since $J_S(x)$ is a lower block triangular matrix with $S_1(x)$ and $-D_2^2 f(x)$ on the diagonal at critical points. Hence, the result.

This equivalence between the non-degeneracy of the Hessian and the game Jacobian $J$ (and the relationship to the determinant of $J_S$ 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 15 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 $\dim X < \text{codim } Z$, 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 7. Consider $f \in C^q(\mathbb{R}^{m_1+m_2}, \mathbb{R})$, $q \geq 2$ such that $D_2^2 f \in \mathbb{R}^{m_1 \times m_2}$. It is a generic property that $\det(D_2^2 f(x)) \neq 0$ 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 $D_if(x) = 0$, $i = 1, 2$. Furthermore, the $J^2(\mathbb{R}^m, \mathbb{R})$ bundle associated to $f$ is diffeomorphic to $\mathbb{R}^m \times \mathbb{R} \times \mathbb{R}^{m \times (m+1)/2}$ 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

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

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, $D_i$ is the union of sub-manifolds of codimension at least $m + 1$. Thus, it follows from the Jet Transversality theorem 11 (by way of Proposition 15 since $m + 1 > m$) that for a generic function $f$, the image of the 2-jet extension $j^2 f$ is disjoint from $D_i$. Hence, for such an $f$, for each $x$ that is a critical point, the Hessian of $f$ is such that $\det(D_2^2 f(x)) \neq 0$.

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

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

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, $D$ is the the union of sub-manifolds of codimension at least $m + 2$. Applying the Jet Transversality Theorem 11 again, we get that for such an $f$, for each $x$ that is a critical point, the Hessian of $f$ is such that $\det(D_2^2 f(x)) \neq 0$ 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., (? (Thm. 2.4)), which can then be approximated by a function without critical points such that $\det(D_2^2 f(x)) = 0$ (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.

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 (Mazumdar & Ratliff, 2019; Ratliff et al., 2014) due to the hierarchical structure of the game.
Implicit Learning Dynamics in Stackelberg Games

**Theorem 12.** For two-player, *q*-differentiable zero-sum continuous games with *q* ≥ 2, differential Stackelberg equilibria are generic amongst local Stackelberg equilibria. That is, given a generic *f* ∈ C⁰(R⁰, R⁰), all local Stackelberg equilibria of the game (*f*, −*f*) are differential Stackelberg equilibria.

**Proof.** Let *J*²(R⁰, R⁰) denote the second-order jet bundle containing 2-jets *j*²⁰ such that *f* : R⁰ → R⁰. Then, *J*²(R⁰, R⁰) is locally diffeomorphic to R⁰ × R⁰ × R⁰(m+1) and the 2-jet extension of *f* at any point *x* ∈ R⁰ is given by (*f*, *Df*(*x*), *D²f*(*x*)).

By Lemma 7, we know that

\[ D_2 = R⁰ × R⁰ × \{0\} × Z(m_2) × R⁰ m_1+m_2 × S(m_1) \]

has co-dimension at least *m* + 1 in *J*²(R⁰, R⁰) so that there exists an open dense set of functions *F*₂ ⊂ C⁰(R⁰, R⁰) such that det(*D*₂⁰(*f*)) ≠ 0 at critical points (i.e., where (*D₁⁰(*f*), *D₂⁰(*f*)) = (0, 0)).

Now, also know that there is an open dense set of functions *F*₁ in C⁰(R⁰, R⁰) such that det(*H*(*x*)) ≠ 0 at critical points. The intersection of open dense sets is open dense. Let *F* = *F*₁ ∩ *F*₂. Now, for any *f* ∈ *F*, we have that at critical points det(*H*(*x*)) ≠ 0 and det(*D*₂⁰(*f*)) ≠ 0. Hence, by Lemma 6, det(*J*₂⁰(*x*)) ≠ 0 for all *f* ∈ *F*, and in particular, det(*S*₁(*x*)) ≠ 0.

For all functions *f* ∈ *F*, the critical points of *ω*₂⁰(*x*) = (*Df*(*x*), −*D₂f*(*x*)) coincide with the critical points of the function *f*. Indeed,

\[ (D₁⁰(*f*), D₂⁰(*f*)) = (0, 0) \iff (Df(*x*), −D₂f(*x*)) = 0 \]

since for all *f* ∈ *F*, det(*D*₂⁰(*f*)) ≠ 0 and *D₂f(*x*) = 0 so that the C⁰ implicit map at a critical point *D₂f(*x*) = 0 is well-defined.

Thus, we have constructed an open dense set *F* ⊂ C⁰(R⁰, R⁰) such that for all *f* ∈ *F*, if *x* ∈ R⁰ is a local Stackelberg equilibrium for (*f*, −*f*), then *x* is a non-degenerate differential Stackelberg equilibrium. Indeed, suppose *f* ∈ *F* and *x* ∈ R⁰ is a local Stackelberg equilibrium. Then, a necessary condition is that −*D₂f(*x*) = 0 and −*D₂f(x*) ≥ 0. However, since *f* ∈ *F*, we have that det(−*D₂f(*x*)) = (−1)ᵐ₂ det(*D₂f(*x*)) = 0 so that, in fact, −*D₂f(*x*) > 0. Hence, a local Stackelberg equilibrium such that −*D₂f(*x*) = 0 and −*D₂f(x*) > 0 necessarily satisfies *ω*₂⁰(*x*) = 0 and *S*₁(*x*) ≥ 0.

By Lemma 6, *S*₁(*x*) ≠ 0 so that, in fact, *S*₁(*x*) > 0. Furthermore, due to the lower triangular structure of *J*₂⁰(*x*), det(−*D₂f(*x*)) = (−1)ᵐ₂ det(*D₂f(*x*)) ≠ 0 and det(*S*₁(*x*)) ≠ 0 also imply that det(*J*₂⁰(*x*)) ≠ 0, which completes the proof.

**Corollary 7.** For two-player, *q*-differentiable zero-sum continuous games, local Stackelberg equilibria are generically non-degenerate, hyperbolic critical points of the vector field *ω*₂⁰(*x*).

This corollary follows direction from Theorem 12 and Proposition 13. Hyperbolicity implies that there are no eigenvalues of *J*₂⁰(*x*) with zero real part.

### I.3. Structural Stability

Structural stability follows naturally from genericity. We further show that (non-degenerate) differential Stackelberg are structurally stable, meaning that they persist under smooth perturbations within the class of zero-sum games.

**Theorem 13.** For zero-sum games, differential Stackelberg equilibria are structurally stable: given *f* ∈ C⁰(Rⁿ × Rⁿ, R), *ζ* ∈ C⁰(Rⁿ × Rⁿ, R), and a differential Stackelberg equilibrium (*x₁*, *x₂) ∈ Rⁿ × Rⁿ, there exists neighborhoods *U* ⊂ Rⁿ of zero and *V* ⊂ Rⁿ × Rⁿ such that for all *t* ∈ *U* there exists a unique differential Stackelberg equilibrium (*x̂₁*, *x̂₂) ∈ *V* for the zero-sum game (*f* + *tζ*, −*f* − *tζ*).

**Proof.** Let *R*ⁿ = Rⁿ × Rⁿ. Define the smoothly perturbed cost function *f̂* : Rⁿ × R → R by *f̂*(*x*, *y*, *t*) = *f*(*x*, *y*) + *tζ*(*x*, *y*), and *ω*₂ : Rⁿ × R → T*(Rⁿ) by *ω*₂(*x*, *y*, *t*) = (*Df̂*(*x*, *y*), −*D₂f̂*(*x*, *y*)), for all *t* ∈ R and (*x*, *y*) ∈ Rⁿ.

Since (*x₁*, *x₂) is a differential Stackelberg equilibrium, *Dω*₂(*x*, *y*, 0) is necessarily non-degenerate. Invoking the implicit function theorem (Lee, 2012), there exists neighborhoods *V* ⊂ Rⁿ of zero and *W* ⊂ Rⁿ and a smooth function *σ* ∈ C⁰(*V*, *W*) such that for all *t* ∈ *V* and (*x₁*, *x₂) ∈ *V*, *ω*₂(*x*, *y*, *t*) = (*σ*(*x*), *σ*(*y*)) for all *t* ∈ R and (*x*, *y*) ∈ Rⁿ.

Since (*x₁*, *x₂) is a differential Stackelberg equilibrium, *Dω*₂(*x*, *y*, 0) is necessarily non-degenerate. Invoking the implicit function theorem (Lee, 2012), there exists neighborhoods *V* ⊂ Rⁿ of zero and *W* ⊂ Rⁿ and a smooth function *σ* ∈ C⁰(*V*, *W*) such that for all *t* ∈ *V* and (*x₁*, *x₂) ∈ *V*, *ω*₂(*x*, *y*, *t*) = (*σ*(*x*), *σ*(*y*)) for all *t* ∈ R and (*x*, *y*) ∈ Rⁿ.

Thus, for all *t* ∈ *V*, *σ*(*t*) must be the unique local Stackelberg equilibrium of (*f* + *tζ|*ₘ, *f* − *tζ|ₘ).