SIAM J. MATH. DATA SCI.
Vol. 2, No. 1, pp. 103–131

On Gradient-Based Learning in Continuous Games[*]

Eric Mazumdar[†], Lillian J. Ratliff[‡], and S. Shankar Sastry[†]

Abstract. We introduce a general framework for competitive gradient-based learning that encompasses a wide breadth of multiagent learning algorithms, and analyze the limiting behavior of competitive gradient-based learning algorithms using dynamical systems theory. For both general-sum and potential games, we characterize a nonnegligible subset of the local Nash equilibria that will be avoided if each agent employs a gradient-based learning algorithm. We also shed light on the issue of convergence to non-Nash strategies in general- and zero-sum games, which may have no relevance to the underlying game, and arise solely due to the choice of algorithm. The existence and frequency of such strategies may explain some of the difficulties encountered when using gradient descent in zero-sum games as, e.g., in the training of generative adversarial networks. To reinforce the theoretical contributions, we provide empirical results that highlight the frequency of linear quadratic dynamic games (a benchmark for multiagent reinforcement learning) that admit global Nash equilibria that are almost surely avoided by policy gradient.

Key words. continuous games, gradient-based algorithms, multiagent learning

AMS subject classifications. 68T05, 37N40

DOI. 10.1137/18M1231298

1. Introduction.

With machine learning algorithms increasingly being deployed in real world settings, it is crucial that we understand how the algorithms can interact, and the dynamics that can arise from their interactions. In recent years, there has been a resurgence in research efforts on multiagent learning, and learning in games. The recent interest in adversarial learning techniques also serves to show how game theoretic tools can be being used to *robustify* and improve the performance of machine learning algorithms. Despite this activity, however, machine learning algorithms are still being treated as black-box approaches and being naïvely deployed in settings where other algorithms are actively changing the environment. In general, outside of highly structured settings, there exists no guarantees on the performance or limiting behaviors of learning algorithms in such settings.

Indeed, previous work on understanding the collective behavior of coupled learning algorithms, either in competitive or cooperative settings, has mainly looked at games where the global structure is well understood like bilinear games [44, 19, 25, 23], convex games [27, 40], or potential games [28], among many others. Such games are more conducive to the statement of global convergence guarantees since the assumed global structure can be exploited.

[*] Received by the editors December 10, 2018; accepted for publication (in revised form) November 21, 2019; published electronically February 18, 2020.
https://doi.org/10.1137/18M1231298

Funding: This work was supported by a National Science Foundation award, CNS:1656873, and by the Defense Advanced Research Projects Agency, FA8750-18-C-0101.

[†] University of California, Berkeley, Berkeley, CA 94720 (mazumdar@berkeley.edu, sastry@eecs.berkeley.edu).

[‡] University of Washington, Seattle, WA 98195 (ratliffl@uw.edu).

In games with fewer assumptions on the players' costs, however, there is still a lack of understanding of the dynamics and limiting behaviors of learning algorithms. Such settings are becoming increasingly prevalent as deep learning is increasingly being used in game theoretic settings [17, 15, 1, 49].

Gradient-based learning algorithms are extremely popular in a variety of these multiagent settings due to their versatility, ease of implementation, and dependence on local information. There are numerous recent papers in multiagent reinforcement learning that employ gradient-based methods (see, e.g., [1, 15, 49]), yet even within this well-studied class of learning algorithms, a thorough understanding of their convergence and limiting behaviors in general continuous games is still lacking.

Generally speaking, in both the game theory and the machine learning communities, two of the central questions when analyzing the dynamics of learning in games are the following:

Q1. *Are all attractors of the learning algorithms employed by agents' equilibria relevant to the underlying game?*

Q2. *Are all equilibria relevant to the game also attractors of the learning algorithms agents employ?*

In this paper, we provide some answers to the above questions for the class of gradient-based learning algorithms by analyzing their limiting behavior in general continuous games. In particular, we leverage the continuous time limit of the more naturally discrete multiagent learning algorithms. This allows us to draw on the extensive theory of dynamical systems and stochastic approximation to make statements about the limiting behaviors of these algorithms in both deterministic and stochastic settings. The latter is particularly relevant since it is common for stochastic gradient methods to be used in multiagent machine learning contexts.

Analyzing gradient-based algorithms through the lens of dynamical systems theory has recently yielded new insights into their behavior in the classical optimization setting [48, 42, 22]. We show that a similar type of analysis can also help to understand the limiting behaviors of gradient-based algorithms in games. We remark, however, that there is a *fundamental difference* between the dynamics that are analyzed in much of the single-agent, gradient-based learning and optimization literature and the ones we analyze in the competitive multiagent case: the combined dynamics of gradient-based learning schemes in games *do not necessarily correspond to a gradient flow*. This may seem to be a subtle point, but it it turns out to be extremely important.

Gradient flows admit desirable convergence guarantees, e.g., almost sure convergence to local minimizers, due to the fact that they preclude flows with the *worst geometries* [39]. In particular, they do not exhibit nonequilibrium limiting behavior such as periodic orbits. Gradient-based learning in games, on the other hand, does not preclude such behavior. Moreover, as we show, asymmetry in the dynamics of gradient-play in games can lead to surprising behaviors such as nonrelevant limiting behaviors being attracting under the flow of the game dynamics and relevant limiting behaviors, such as a subset of the Nash equilibria being almost surely avoided.

1.1. Related work.

The study of continuous games is quite extensive (see, e.g., [2, 30]), though in large part the focus has been on games admitting a fair amount of structure. The behavior of learning algorithms in games is also well studied (see, e.g., [16]). In this section, we comment on the most relevant prior work and defer a more comprehensive discussion of our results in the context of prior work to section 6.

As we noted, previous work on learning in games in both the game theory literature, and more recently from the machine learning community, has largely focused on addressing (Q1) whether all attractors of the learning dynamics are game-relevant equilibria, and (Q2) whether all game-relevant equilibria are also attractors of the learning dynamics. The primary type of game-relevant equilibrium considered in the investigation of these two questions is a Nash equilibrium.

The majority of the existing work has focused on Q1. In fact, a large body of prior work focuses on games with structures that preclude the existence of non-Nash equilibria. Consequently, answering Q1 reduces to analyzing the convergence of various learning algorithms (including gradient-play) to the unique Nash equilibrium or the set of Nash equilibria. This is often shown by exploiting the game structure. Examples of classes of games where gradient-play has been well studied are potential games [28], concave or monotone games [40, 8, 27], and gradient-play over the space of stochastic policies in two-player finite-action bilinear games [44]. In the latter setting, other gradient-like algorithms such as multiplicative weights have also been studied fairly extensively [19], and have been shown to converge to cycling behaviors.

Some works have also attempted to address Q1 in the context of gradient-play in two-player zero-sum games. Concurrently with this paper, for a general class of "sufficiently smooth" two-player, zero-sum games it was shown that there exist stationary points for gradient-play that are non-Nash [12].[1] In such games, it has also been shown that gradient-play can converge to cycles (see, e.g., [25, 47, 19]).

There is also related work in more general games on the analysis of when Nash equilibria are attracting for gradient-based approaches (i.e., Q2). Sufficient conditions for this to occur are the conditions for stable differential Nash equilibria introduced in [34, 35, 36] and the condition for variational stability later analyzed in [27]. We remark that these conditions are equivalent for the classes of games we consider. Neither of these works give conditions under which Nash equilibria are avoided by gradient-play or comment on other attracting behaviors.

Expanding on this rich body of literature (only the most relevant of which is covered in our short review), in this paper we provide answers to Q1 without imposing structure on the game outside regularity conditions on the cost functions by exploiting the observation that gradient-based learning dynamics are not gradient flows. We also provide answers to Q2 by demonstrating that a nontrivial set of games admit Nash equilibria that are almost surely avoided by gradient-play. We give explicit conditions for when this occurs. Using similar analysis tools, we also provide new insights into the behavior of gradient-based learning in structured classes of games such as zero-sum and potential games.

1.2. Contributions and organization. We present a general framework for modeling competitive gradient-based learning that applies to a broad swath of learning algorithms. In section 3, we draw connections between the limiting behavior of this class of algorithms and game theoretic and dynamical systems notions of equilibria. In particular, we construct general-sum and zeros-sum games that admit non-Nash attracting equilibria of the gradient dynamics.

[1] This paper was under review at the time that [12] became publicly available. Our results show the existence of these non-Nash equilibria and attracting cycles in both general-sum and zero-sum games.

Such points are attracting under the learning dynamics, yet at least one player—*and potentially all of them*—has a direction in which they could unilaterally deviate to decrease their cost. Thus, these non-Nash equilibria are of questionable game theoretic relevance and can be seen as artifacts of the players' algorithms.

In section 4, we show that policy gradient multiagent reinforcement learning (MARL), generative adversarial networks (GANs), gradient-based multiagent multiarmed bandits, among several other common multiagent learning settings, conform to this framework. The framework is amenable to tools for analysis from dynamical systems theory.

Also in section 4, we show that a subset of the local Nash equilibria in general-sum games and potential games is avoided almost surely when each player employs a gradient-based algorithm. We show that this holds in two broad settings: the full information setting when each player has oracle access to their gradient but randomly initializes their first action, and a partial information setting where each player has access to an unbiased estimate of their gradient.

Thus, we provide a negative answer to both Q1 and Q2 for n-player general-sum games, and highlight the nuances present in zero-sum and potential games. We also show that the dynamics formed from the individual gradients of agents' costs are *not gradient flows*. This in turn implies that competitive gradient-based learning in general-sum games may converge to periodic orbits and other nontrivial limiting behaviors that arise in, e.g., chaotic systems.

To support the theoretical results, we present empirical results in section 5 that show that policy gradient algorithms avoid global Nash equilibria in a large number of linear quadratic (LQ) dynamic games, a benchmark for MARL.

We conclude in section 6 with a discussion of the implications of our results and some links with prior work as well as some comments on future directions.

2. Preliminaries.

Consider n agents indexed by $\mathcal{I} = \{1, \dots, n\}$. Each agent $i \in \mathcal{I}$ has their own decision variable $x_i \in X_i$, where X_i is their finite-dimensional strategy space of dimension m_i. Define $X = X_1 \times \cdots \times X_n$ to be the finite-dimensional joint strategy space with dimension $m = \sum_{i\in\mathcal{I}} m_i$. Each agent is endowed with a cost function $f_i \in C^s(X, \mathbb{R})$ with $s \geq 2$ and such that $f_i : (x_i, x_{-i}) \mapsto f_i(x_i, x_{-i})$, where we use the notation $x = (x_i, x_{-i})$ to make the dependence on the action of the agent x_i, and the actions of all agents excluding agent i, $x_{-i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ explicit. The agents seek to minimize their own cost, but only have control over their own decision variable x_i. In this setup, agents' costs are not necessarily aligned with one another, meaning they are competing.

Given the game $\mathcal{G} = (f_1, \dots, f_n)$, agents are assumed to update their strategies simultaneously according to a gradient-based learning algorithm of the form

$$x_{i,t+1} = x_{i,t} - \gamma_{i,t} h_i(x_{i,t}, x_{-i,t}), \tag{2.1}$$

where $\gamma_{i,t}$ is agent i's step size at iteration t.

We analyze the following two settings:

1. Agents have *oracle access* to the gradient of their cost with respect to their own choice variable, i.e., $h_i(x_{i,t}, x_{-i,t}) = D_i f_i(x_{i,t}, x_{-i,t})$, where $D_i f_i \equiv \partial f_i / \partial x_i$ denotes the derivative of f_i with respect to x_i.

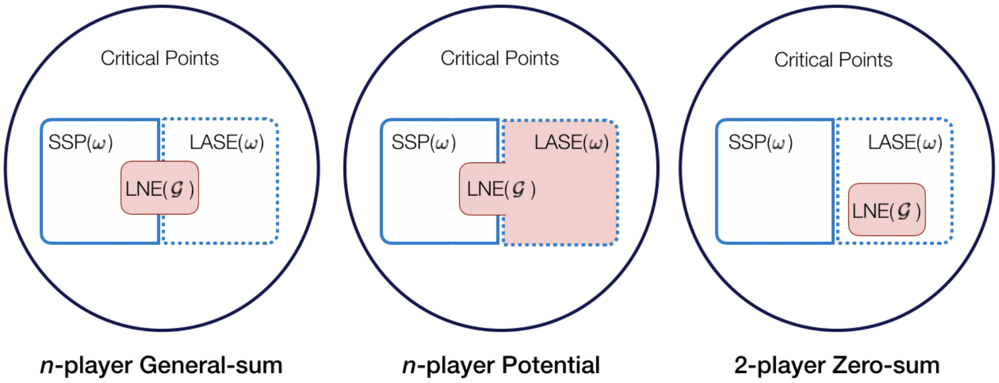

Figure 1. *Links between the equilibria of generic continuous games $\mathcal{G}$ and their properties under the gradient dynamics $\dot{x} = -\omega(x)$.*

2. Agents have an *unbiased estimator* of their gradient, i.e., $h_i(x_{i,t}, x_{-i,t}) = D_i f_i(x_{i,t}, x_{-i,t}) + w_{i,t+1}$, where $\{w_{i,t}\}$ is a zero mean, finite variance stochastic process.

We refer to the former setting as *deterministic* gradient-based learning and the latter setting as *stochastic* gradient-based learning. Assuming that all agents are employing such algorithms, we aim to analyze the limiting behavior of the agents' strategies. To do so, we leverage the following game theoretic notion of a Nash equilibrium.

Definition 2.1. *A strategy $x \in X$ is a local Nash equilibrium for the game $(f_1, \dots, f_n)$ if, for each $i \in \mathcal{I}$, there exists an open set $W_i \subset X_i$ such that $x_i \in W_i$ and $f_i(x_i, x_{-i}) \le f_i(x_i', x_{-i})$ for all $x_i' \in W_i$. If the above inequalities are strict, then we say x is a strict local Nash equilibrium.*

The focus on *local* Nash equilibria is due to our lack of assumptions on the agents' cost functions. If $W_i = X_i$ for each i, then a local Nash equilibrium x is a global Nash equilibrium. This holds in, e.g., the bimatrix games and the LQ games we analyze in section 5. Depending on the agents' costs, a game $(f_1, \dots, f_n)$ may admit anywhere from one to a continuum of local or global Nash equilibria; or none at all.

3. Linking games and dynamical systems.

In this section, we draw links between the limiting behavior of dynamical systems and game theoretic notions of equilibria in three broad classes of continuous games. For brevity, the proofs of the propositions in this section are supplied in Appendix A. A high-level summary of the links we draw is shown in Figure 1.

Define $\omega(x) = (D_1 f_1(x), \dots, D_n f_n(x))$ to be the vector of player derivatives of their own cost functions with respect to their own choice variables. When each player is employing a gradient-based learning algorithm, the joint strategy of the players, (in the limit as the agents' step sizes go to zero) follows the differential equation

$$\dot{x} = -\omega(x).$$

A point $x \in X$ is said to be an equilibrium, critical point, or stationary point of the dynamics if $\omega(x) = 0$. Stationary points of $\dot{x} = -\omega(x)$ are joint strategies from which, under gradient-play, the agents do not move. We note that $\omega(x) = 0$ is a necessary condition for a point $x \in X$ to be a local Nash equilibrium [36]. Hence, all local Nash equilibria are critical points of the joint dynamics $\dot{x} = -\omega(x)$.

Central to dynamical systems theory is the study of limiting behavior and its stability properties. A classical result in dynamical systems theory allows us to characterize the stability properties of an equilibrium x^* by analyzing the Jacobian of the dynamics at x^*. The Jacobian of ω is defined by

$$D\omega(x) = \begin{bmatrix} D_1^2 f_1(x) & \cdots & D_{n1} f_1(x) \\ \vdots & \ddots & \vdots \\ D_{1n} f_n(x) & \cdots & D_n^2 f_n(x) \end{bmatrix}.$$

Since $D\omega$ is a matrix of second derivatives, it is sometimes referred to as the "game Hessian." Similarly to the Hessian matrix of a gradient flow, $D\omega$ allows us to further characterize the critical points of ω by their properties under the flow of $\dot{x} = -\omega(x)$. Let $\lambda_i(x) \in \mathrm{spec}(D\omega(x))$ for $i \in \{1, \ldots, m\}$ denote the eigenvalues of $D\omega$ at x, where $\mathrm{Re}(\lambda_1(x)) \leq \cdots \leq \mathrm{Re}(\lambda_m(x))$, that is, $\lambda_1(x)$ is the eigenvalue with the smallest real part. Of particular interest are asymptotically stable equilibria.

Definition 3.1. *A point $x \in X$ is a locally asymptotically stable equilibrium of the continuous time dynamics $\dot{x} = -\omega(x)$ if $\omega(x) = 0$ and $\mathrm{Re}(\lambda) > 0$ for all $\lambda \in \mathrm{spec}(D\omega(x))$.*

Locally asymptotically stable equilibria have two properties of interest. First, they are isolated, meaning that there exists a neighborhood around them in which no other equilibria exist. Second, they are exponentially attracting under the flow of $\dot{x} = -\omega(x)$, meaning that if agents initialize in a neighborhood of a locally asymptotically stable equilibrium x^* and follow the dynamics described by $\dot{x} = -\omega(x)$, they will converge to x^* exponentially fast [41]. This, in turn, implies that a discretized version of $\dot{x} = -\omega(x)$, namely,

$$x_{t+1} = x_t - \gamma\omega(x_t), \tag{3.1}$$

converges locally for appropriately selected step size γ at a rate of $O(1/t)$. Such results motivate the study of the continuous time dynamical system $\dot{x} = -\omega(x)$ in order to understand convergence properties of gradient-based learning algorithms of the form (2.1).

Another important class of critical points of a dynamical system are saddle points.

Definition 3.2. *A point $x \in X$ is a saddle point of the dynamics $\dot{x} = -\omega(x)$ if $\omega(x) = 0$ and $\lambda_1(x) \in \mathrm{spec}(D\omega(x))$ is such that $\mathrm{Re}(\lambda_1(x)) \leq 0$. A saddle point such that $\mathrm{Re}(\lambda_i) < 0$ for $i \in \{1, \ldots, \ell\}$ and $\mathrm{Re}(\lambda_j) > 0$ for $j \in \{\ell + 1, \ldots, m\}$ with $0 < \ell < m$ is a strict saddle point of the continuous time dynamics $\dot{x} = -\omega(x)$.*

Strict saddle points are especially relevant to our analysis since their neighborhoods are characterized by stable and unstable manifolds [41]. When the agents evolve according to the dynamics solely on the stable manifold, they converge exponentially fast to the critical point. However, when they evolve solely on the unstable manifold, they diverge from the equilibrium exponentially fast. Agents whose strategies lie on the union of the two manifolds asymptotically avoid the equilibrium. We make use of this general fact in section 4.1.

To better understand the links between the critical points of the gradient dynamics and the Nash equilibria of the game, we make use of an equivalent characterization of strict local

Nash that leverages first and second order conditions on player cost functions. This makes them simpler objects to link to the various dynamical systems notions of equilibria than local Nash equilibria.

Definition 3.3 (see [34, 36]). *A point $x \in X$ is a differential Nash equilibrium for the game defined by $(f_1, \ldots, f_n)$ if $\omega(x) = 0$ and $D_i^2 f_i(x) \succ 0$ for each $i \in \mathcal{I}$.*

In [35], it was shown that local Nash equilibria are generically differential Nash equilibria, where $\det(D\omega(x)) \neq 0$ (i.e., $D\omega$ is nondegenerate). Thus, in the space of games where the agents' costs are at least twice differentiable, the set of games that admit local Nash equilibria that are not nondegenerate differential Nash equilibria is of measure zero [35]. In [35] it was also shown that nondegenerate Nash equilibria are structurally stable, meaning that small perturbations to the agents' cost functions will not change the fundamental nature of the equilibrium. This also implies that gradient-play with slightly biased estimators of the gradient will not have vastly different behaviors in neighborhoods of equilibria.

Given these different equilibrium notions of the learning dynamics and the underlying game, let us define the following sets which will be useful in stating the results in the following sections. For a game $\mathcal{G} = (f_1, \ldots, f_n)$, denote the sets of strict saddle points and locally asymptotically stable equilibria of the gradient dynamics, $\dot{x} = -\omega(x)$, as $\texttt{SSP}(\omega)$ and $\texttt{LASE}(\omega)$, respectively, where we recall that $\omega(x) = (D_1 f_1(x), \ldots, D_n f_n(x))$. Similarly, denote the set of local Nash equilibria, differential Nash equilibria, and nondegenerate differential Nash equilibria of $\mathcal{G}$ as $\texttt{LNE}(\mathcal{G})$, $\texttt{DNE}(\mathcal{G})$, and $\texttt{NDDNE}(\mathcal{G})$, respectively. As previously mentioned, $\texttt{NDDNE}(\mathcal{G}) = \texttt{LNE}(\mathcal{G})$ in almost all continuous games. The key takeaways of this section are summarized in Figure 1.

3.1. General-sum games.

We first analyze the properties of local Nash equilibria under the joint gradient dynamics in n-player general-sum games.

Proposition 3.4. *A nondegenerate differential Nash equilibrium is either a locally asymptotically stable equilibrium or a strict saddle point of $\dot{x} = -\omega(x)$, i.e., $\texttt{NDDNE}(\mathcal{G}) \subset \texttt{SSP}(\omega) \cup \texttt{LASE}(\omega)$.*

Locally asymptotically stable differential Nash equilibria satisfy the notion of variational stability introduced in [27]. In fact, a simple analysis shows that the definitions of variationally stable equilibria and locally asymptotically stable differential Nash equilibria [34] are equivalent in the games we consider, i.e., games where each players' cost is at least twice continuously differentiable. We remark that, from the definition of asymptotic stability, the gradient dynamics have an $O(1/t)$ convergence rate in the neighborhood of such equilibria.

An important point to make is that not every locally asymptotically stable equilibrium of $\dot{x} = -\omega(x)$ is a nondegenerate differential Nash equilibrium. Indeed, the following proposition provides an entire class of games whose corresponding gradient dynamics admit locally asymptotically stable equilibria that are not local Nash equilibria.

Proposition 3.5. *In the class of general-sum continuous games, there exists a continuum of games containing games $\mathcal{G}$ such that $\texttt{LASE}(\omega) \not\subset \texttt{NDDNE}(\mathcal{G})$ and, moreover, $\texttt{LASE}(\omega) \not\subset \texttt{LNE}(\mathcal{G})$.*

Proof. Consider a two-player game $\mathcal{G} = (f_1, f_2)$ on $\mathbb{R}^2$ where

$$f_1(x_1, x_2) = \frac{a}{2}x_1^2 + bx_1x_2, \quad \text{and} \quad f_2(x_1, x_2) = \frac{d}{2}x_2^2 + cx_1x_2$$

for constants $a, b, c, d \in \mathbb{R}$. The Jacobian of ω is given by

$$D\omega(x_1, x_2) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \forall (x_1, x_2) \in \mathbb{R}^2. \tag{3.2}$$

If $a > 0$ and $d < 0$, then the unique stationary point $x = (0, 0)$ is neither a differential Nash nor a local Nash equilibrium since the necessary conditions are violated (i.e., $d < 0$). However, if $a > -d$ and $ad > cb$, the eigenvalues of $D\omega$ have positive real parts and $(0, 0)$ is asymptotically stable. Further, this clearly holds for a continuum of games. Thus, the set of locally asymptotically stable equilibria that are not Nash equilibria may be arbitrarily large. ■

The preceding proposition shows that there exist attracting critical points of the gradient dynamics in general-sum continuous games that are not Nash equilibria and may not be even relevant to the game. Thus, this provides a negative answer to Q2 (whether all attracting equilibria in general games are game relevant for the learning dynamics).

Remark 3.6. We note that, by definition, the non-Nash locally asymptotically stable equilibria (or non-Nash equilibria) do not satisfy the second order conditions for Nash equilibria. Thus, at these joint strategies, at least one player—and maybe all of them—has a direction in which they would unilaterally deviate if they were not using gradient descent. As such, we view convergence to these points to be undesirable.

3.2. Zero-sum games. Let us now restrict our attention to two-player zero-sum games, which often arise when training GANs, in adversarial learning, and in MARL [17, 29, 9]. In such games, one player can be seen as minimizing f with respect to their decision variable and the other as minimizing $-f$ with respect to theirs. The following proposition shows that all differential Nash equilibria in two-player zero-sum games are locally asymptotically stable equilibria under the flow of $\dot{x} = -\omega(x)$.

Proposition 3.7. *For an arbitrary two-player zero-sum game, $(f, -f)$ on $\mathbb{R}^m$, if x is a differential Nash equilibrium, then x is both a nondegenerate differential Nash equilibrium and a locally asymptotically stable equilibrium of $\dot{x} = -\omega(x)$, that is,* $\mathtt{DNE}(\mathcal{G}) \equiv \mathtt{NDDNE}(\mathcal{G}) \subset \mathtt{LASE}(\omega)$.

This result guarantees that the differential Nash equilibria of zero-sum games are isolated and exponentially attracting under the flow of $\dot{x} = -\omega(x)$. This in turn guarantees that simultaneous gradient-play has a local linear rate of convergence to all local Nash equilibria in all zero-sum continuous games. Thus, the answer to Q1 is the context of zero-sum games is “yes,” since all Nash equilibria are attracting for the gradient dynamics.

The converse of the preceding proposition, however, is not true. Not every locally asymptotically stable equilibrium in two-player zero-sum games is a nondegenerate differential Nash equilibrium. Indeed, there may be many locally asymptotically stable equilibria in a zero-sum game that are not local Nash equilibria. The following proposition highlights this fact.

Proposition 3.8. *In the class of zero-sum continuous games, there exists a continuum of games such that for each game $\mathcal{G}$, $\mathtt{LASE}(\omega) \not\subset \mathtt{DNE}(\mathcal{G}) \subset \mathtt{LNE}(\mathcal{G})$.*

Proof. Consider the two-player zero-sum game $(f, -f)$ on $\mathbb{R}^2$, where

$$f(x_1, x_2) = \frac{a}{2}x_1^2 + bx_1x_2 + \frac{c}{2}x_2^2,$$

and $a, b, c \in \mathbb{R}$. The Jacobian of ω is given by

$$D\omega(x_1, x_2) = \begin{bmatrix} a & b \\ -b & -c \end{bmatrix} \quad \forall\ (x_1, x_2) \in \mathbb{R}^2.$$

If $a > c > 0$ and $b^2 > ac$, then $D\omega(x_1, x_2)$ has eigenvalues with strictly positive real part, but the unique stationary point is not a differential Nash equilibrium, since $-c < 0$, and, in fact, is not even a Nash equilibrium. Indeed,

$$-f(0, 0) > -f(0, x_2) = -\frac{c}{2}x_2^2 \quad \forall\ x_2 \neq 0.$$

Thus, there exists a continuum of zero-sum games with a large set of locally asymptotically stable equilibria of the corresponding dynamics $\dot{x} = -\omega(x)$ that are not differential Nash. ■

The preceding proposition again shows that there exist non-Nash equilibria of the gradient dynamics in zero-sum continuous games. Thus, this proposition also provides a negative answer to Q2 in the context of zero-sum games.

3.3. Potential games. One last set of games with interesting connections between the Nash equilibria and the critical points of the gradient dynamics is the class known as *potential games*. This particularly nice class of games are ones for which ω corresponds to a gradient flow under a coordinate transformation, that is, there exists a function ϕ (commonly referred to as the potential function) such that for each $i \in \mathcal{I}$, $D_i f_i \equiv D_i \phi$. We remark that due to the equivalence, this class of games is sometimes referred to as an *exact* potential game. Note that a necessary and sufficient condition for $(f_1, \dots, f_n)$ to be a potential game is that $D\omega$ is *symmetric* [28], that is, $D_{ij} f_j \equiv D_{ji} f_i$. This gives potential games the desirable property that the only locally asymptotically stable equilibria of the gradient dynamics are local Nash equilibria.

Proposition 3.9. *For an arbitrary potential game, $\mathcal{G} = (f_1, \dots, f_n)$ on $\mathbb{R}^m$, if x is a locally asymptotically stable equilibrium of $\dot{x} = -\omega(x)$ (i.e., $x \in \mathtt{LASE}(\omega)$), then x is a nondegenerate differential Nash equilibrium (i.e., $x \in \mathtt{NDDNE}(\mathcal{G})$).*

The full proof of Proposition 3.9 is supplied in Appendix A. The preceding proposition rules out non-Nash locally asymptotically stable equilibria of the gradient dynamics in potential games, and implies that every local minimum of a potential game must be a local Nash equilibrium. Thus, in potential games, unlike in general-sum and zero-sum games, the answer to Q2 is positive. However, the following proposition shows that the existence of a potential function is not enough to rule out local Nash equilibria that are saddle points of the dynamics.

Proposition 3.10. *In the class of continuous games, there exist a continuum of potential games containing games $\mathcal{G}$ that admit Nash equilibria that are saddle points of the dynamics $\dot{x} = -\omega(x)$, i.e., $\exists\ \mathcal{G}$ such that for some $x \in \mathtt{LNE}(\mathcal{G})$, $x \in \mathtt{SSP}(\omega)$.*

Proof. Consider the game (f, f) on $X = \mathbb{R}^2$ described by

$$f(x_1, x_2) = \frac{a}{2}x_1^2 + bx_1x_2 + \frac{c}{2}x_2^2,$$

where $a, b, d \in \mathbb{R}$. The Jacobian of ω is given by

$$D\omega(x_1, x_2) = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \quad \forall\ (x_1, x_2) \in \mathbb{R}^2.$$

If $a, c > 0$, then $x = (0, 0)$ is a local Nash equilibrium. However, if $ac < b^2$, $D\omega(x)$ has one positive and one negative eigenvalue and $(0, 0)$ is a saddle point of the gradient dynamics. Thus, there exists a continuum of potential games where a large set of differential Nash equilibria are strict saddle points of $\dot{x} = -\omega(x)$. ■

Proposition 3.10 demonstrates a surprising fact about potential games. Even though all minimizers of the potential function must be local Nash equilibria, *not all local Nash equilibria are minimizers of the potential function.*

3.4. Main takeaways.

The main takeaways of this section are summarized in Figure 1. We note that for zero-sum games, Proposition 3.8 shows that $\texttt{LNE}(\mathcal{G}) \subset \texttt{LASE}(\omega)$. Since the inclusion is strict, the answer to Q2 in such games is "no." For general-sum games, Proposition 3.5 allows us to to conclude that there do exist attracting, non-Nash equilibria. Thus, the answer to Q2 is also no. In potential games, since $\texttt{LASE}(\omega) \subset \texttt{LNE}(\mathcal{G})$ the answer is yes.

In the following sections, we provide answers to Q1 by showing that all local Nash equilibria in $\texttt{LNE}(\mathcal{G}) \cap \texttt{SSP}(\omega)$ are avoided almost surely by gradient-based algorithms in both the deterministic and stochastic settings. In particular, since $\texttt{LNE}(\mathcal{G}) \cap \texttt{SSP}(\omega) \neq \emptyset$ in potential and general-sum games, one cannot give a positive answer to Q1 in either of these classes of games.

4. Convergence of gradient-based learning.

In this section, we provide convergence and nonconvergence results for gradient-based algorithms. We also include a high-level overview of well-known algorithms that fit into the class of learning algorithms we consider; more details can be found in Appendix C.

4.1. Deterministic setting.

We first address convergence to equilibria in the deterministic setting in which agents have oracle access to their gradients at each time step. This includes the case where agents know their own cost functions f_i and observe their own actions as well as their competitors' actions and, hence, can compute the gradient of their cost with respect to their own choice variable.

Since we have assumed that each agent $i \in \mathcal{I}$ has their own *learning rate* (i.e., step sizes γ_i), the joint dynamics of all the players are given by

$$x_{t+1} = g(x_t), \tag{4.1}$$

where $g : x \mapsto x - \gamma \odot \omega(x)$ with $\gamma = (\gamma_i)_{i\in\mathcal{I}}$ and $\gamma > 0$ elementwise. By a slight abuse of notation, $\gamma \odot \omega(x_t)$ is defined to be elementwise multiplication of γ and $\omega(\cdot)$, where γ_1 is multiplied by the first m_1 components of $\omega(\cdot)$, γ_2 is multiplied by the next m_2 components, and so on.

We remark that this update rule immediately distinguishes gradient-based learning in games from gradient descent. By definition, the dynamics of gradient descent in single-agent settings always correspond to gradient flows, i.e., x evolves according to an ordinary differential equation of the form $\dot{x} = -\nabla\phi(x)$ for some function $\phi : \mathbb{R}^d \to \mathbb{R}$. Outside of the class of exact potential games we defined in section 3, the dynamics of players' actions in games are not afforded this luxury; indeed, $D\omega$ is not in general symmetric (which is a necessary condition for a gradient flow). This makes the potential limiting behaviors of $\dot{x} = -\omega(x)$ highly non-trivial to characterize in general-sum games.

The structure present in a gradient flow implies strong properties on the limiting behaviors of x. In particular, it precludes the existence of limit cycles or periodic orbits (limiting behaviors of dynamical systems where the state of system cycles infinitely through a set of states with a finite period) and chaos (an attribute of nonlinear dynamical systems, where the system's behavior can vary extremely due to slight changes in initial position) [41]. We note that both of these behaviors can occur in the dynamics of gradient-based learning algorithms in games.[2]

Despite the wide breadth of behaviors that gradient dynamics can exhibit in competitive settings, we can still make statements about convergence (and nonconvergence) to certain types of equilibria. To do so, we first make the following standard assumptions on the smoothness of the cost functions f_i and the magnitude of the agents' learning rates γ_i.

Assumption 1. For each $i \in \mathcal{I}$, $f_i \in C^s(X, \mathbb{R})$ with $s \geq 2$, $\sup_{x\in X} \|D\omega(x)\|_2 \leq L < \infty$, and $0 < \gamma_i < 1/L$, where $\|\cdot\|_2$ is the induced 2-norm.

Given these assumptions, the following result rules out converging to strict saddle points.

Theorem 4.1. *Let $f_i : X \to \mathbb{R}$ and γ satisfy Assumption 1. Suppose that $X = X_1 \times \cdots \times X_n \subseteq \mathbb{R}^m$ is open and convex. If $g(X) \subset X$, the set of initial conditions $x \in X$ from which competitive gradient-based learning converges to strict saddle points is of measure zero.*

We remark that the above theorem holds for $X = X_1 \times \cdots \times X_n = \mathbb{R}^m$ in particular, since $g(X) \subset X$ holds trivially in this case. It is also important to note that, as we point out in section 3, local Nash equilibria can be strict saddle points. Thus, all local Nash equilibria that are strict saddle points for $\dot{x} = -\omega(x)$ are avoided almost surely by gradient-play even with oracle gradient access and random initializations. This holds even when players randomly initialize uniformly in an arbitrarily small ball around such Nash equilibria. In section 5, we show that many LQ dynamic games have a strict saddle point as their global Nash equilibrium. For brevity, we provide the proof of Theorem 4.1 in Appendix A, and provide a proof sketch below.

Proof sketch of Theorem 4.1. The core of the proof is the celebrated stable manifold theorem from dynamical systems theory, presented in Theorem A.1. We construct the set of initial positions from which gradient-play will converge to strict saddle points and then use

[2] The Van der Pol oscillator and Lorenz system (see, e.g., [41]) can be seen as the resulting gradient dynamics in a two-player and three-player general-sum game, respectively. The first is a classic example of a system where players converge to cycles and the second is an example of a chaotic system.

the stable manifold theorem to show that the set must have measure zero in the players' joint strategy space. Therefore, with a random initialization players will never evolve solely on the stable manifold of strict saddles and they will consequently diverge from such equilibria.

To be able to invoke the stable manifold theorem, we first show that the mapping $g : \mathbb{R}^m \to \mathbb{R}^m$ is a diffeomorphism, which is nontrivial due to the fact that we have allowed each agent to have their own learning rate γ_i and $D\omega$ is not symmetric. We then iteratively construct the set of initializations that will converge to strict saddle points under the game dynamics. By the stable manifold theorem, and the fact that g is a diffeomorphism, the stable manifold of a strict saddle point must be measure zero. Then, by induction we show that the set of all initial points that converge to a strict saddle point must also be measure zero. ■

In potential games we can strengthen the above nonconvergence result and give convergence guarantees.

Corollary 4.2. *Consider a potential game $(f_1, \dots, f_n)$ on open, convex $X = X_1 \times \cdots \times X_n \subseteq \mathbb{R}^m$ and where each $f_i \in C^s(X, \mathbb{R})$ for $s \geq 3$. Let ν be a prior measure with support X which is absolutely continuous with respect to the Lebesgue measure and assume $\lim_{t\to\infty} g^t(x)$ exists. Then, under Assumption 1, competitive gradient-based learning converges to nondegenerate differential Nash equilibria almost surely. Moreover, the nondegenerate differential Nash to which it converges is generically a local Nash equilibrium.*

Corollary 4.2 guarantees that in potential games, gradient-play will converge to a differential Nash equilibrium. Combining this with Theorem 4.1 guarantees that the differential Nash equilibrium it converges to is a local minimizer of the potential function. A simple implication of this result is that gradient-based learning in potential games cannot exhibit limit cycles or chaos.

Of note is the fact that the agents *do not* need to be performing gradient-based learning on ϕ to converge to Nash almost surely. That is, they do not need to know the function ϕ; they simply need to follow the derivative of their own cost with respect to their own choice variable, and they are guaranteed to converge to a local Nash equilibrium that is a local minimizer of the potential function.

We note that convergence to Nash equilibria is a known characteristic of gradient-play in potential games. However, our analysis also highlights that gradient-play will avoid a subset of the Nash equilibria of the game. This is surprising given the particularly strong structural properties of such games. The proof for Corollary 4.2 is provided in Appendix A and follows from Proposition 3.9, Theorem 4.1, and the fact that $D\omega$ is symmetric in potential games.

4.1.1. Implications and interpretation of convergence analysis.

Both Theorem 4.1 and Corollary 4.2 show that gradient-play in multiagent settings avoids strict saddles almost surely even in the deterministic setting. Combined with the analysis in section 3 which shows that (local) Nash equilibria can be strict saddles of the dynamics for general-sum games, this implies that a subset of the Nash equilibria are almost surely avoided by individual gradient-play, a potentially undesirable outcome in view of Q1 (whether all Nash equilibria are attracting for the learning dynamics). In section 5, we show that the global Nash equilibrium is a saddle point of the gradient dynamics in a large number of randomly sampled LQ dynamic games. This suggests that policy gradient algorithms may fail to converge in such games, which is highly undesired. This is in stark contrast to the single agent setting where policy gradient has been shown to converge to the unique solution of LQ regulator problems [13].

In section 3, we also showed that local Nash equilibria of potential games can be strict saddle points of the potential function. Nonconvergence to such points in potential games is not necessarily a bad result since this in turn implies convergence to a local minimizer of the potential function (as shown in [22, 32]) which is guaranteed to be a local Nash equilibrium of the game. However, these results do imply that *one cannot answer "yes" to* Q1 *in potential games* since some of the Nash equilibria are not attracting under gradient-play.

In zero-sum games, where local Nash equilibria cannot be strict saddle points of the gradient dynamics, our result suggests that *eventually* gradient-based learning algorithms will escape saddle points of the dynamics.

The almost sure avoidance of all equilibria that are saddle points of the dynamics further implies that if (3) converges to a critical point x, then $x \in \texttt{LASE}(\omega)$, i.e., x is locally asymptotically stable for $\dot{x} = -\omega(x)$. This may not be a desired property, however, since we showed in section 3 that zero-sum and general-sum games both admit non-Nash LASE.

Since gradient-play in games generally does not result in a gradient flow, other types of limiting behavior such as limit cycles can occur in gradient-based learning dynamics. Theorem 4.1 says nothing about convergence to other limiting behaviors. In the following sections we prove that the results described in this section extend to the stochastic gradient setting. We also formally define periodic orbits in the context of dynamical systems and state stronger results on avoidance of some more complex limiting behaviors like linearly unstable limit cycles.

4.2. Stochastic setting.

We now analyze the stochastic case in which agents are assumed to have an unbiased estimator for their gradient. The results in this section allow us to extend the results from the deterministic setting to a setting where each agent builds an estimate of the gradient of their loss at the current set of strategies from potentially noisy observations of the environment. Thus, we are able to analyze the limiting behavior of a class of commonly used machine learning algorithms for competitive, multiagent settings. In particular, we show that agents will almost surely not converge to strict saddle points. In Appendix B.1, we show that the gradient dynamics will actually avoid more general limiting behaviors called linearly unstable cycles which we define formally.

To perform our analysis, we make use of tools and ideas from the literature on stochastic approximations (see. e.g., [6]). We note that the convergence of stochastic gradient schemes in the single-agent setting has been extensively studied [37, 33, 7, 26]. We extend this analysis to the behavior of stochastic gradient algorithms in games.

We assume that each agent updates their strategy using the update rule

$$x_{i,t+1} = x_{i,t} - \gamma_{i,t}(D_i f_i(x_{i,t}, x_{-i,t}) + w_{i,t+1}) \tag{4.2}$$

for some zero-mean, finite-variance stochastic process $\{w_{i,t}\}$. Before presenting the results for the stochastic case, let us comment on the different learning algorithms that fit into this framework.

4.2.1. Examples of stochastic gradient-based learning.

The stochastic gradient-based learning setting we study is general enough to include a variety of commonly used multiagent learning algorithms. The classes of algorithms we include is hardly an exhaustive list,

Table 1

Example problem classes that fit into competitive gradient-based learning rules. Details on the derivation of these update rules as gradient-based learning schemes is provided in Appendix C.

Class	Gradient learning rule
Gradient-play	$x_i^+ = x_i - \gamma_i D_i f_i(x_i, x_{-i})$
GANs	$\theta^+ \; = \theta - \gamma\mathbb{E}[D_\theta L(\theta, w)]$ $w^+ = w + \gamma\mathbb{E}[D_w L(\theta, w)]$
MA policy gradient	$x_i^+ = x_i - \gamma_i \mathbb{E}[D_i J_i(x_i, x_{-i})]$
Individual Q-learning	$q_i^+(u_i) = q_i(u_i) + \gamma_i(r_i(u_i, \pi_{-i}(q_i, q_{-i})) - q_i(u_i))$
MA gradient bandits	$x_{i,\ell}^+ = x_{i,\ell} + \gamma_i \mathbb{E}[\beta_i R_i(u_i, u_{-i}) \vert u_i = \ell]$, $\ell = 1, \ldots, m_i$
MA experts	$x_{i,\ell}^+ = x_{i,\ell} + \gamma_i \mathbb{E}[R_i(u_i, u_{-i}) \vert u_i = \ell]$, $\ell = 1, \ldots, m_i$

and indeed many extensions and altogether different algorithms exist that can be considered members of this class. In Table 1, we provide the gradient-based update rule for six different example classes of learning problems: (i) gradient-play in noncooperative continuous games, (ii) GANs, (iii) multiagent policy gradient, (iv) individual Q-learning, (v) multiagent gradient bandits, and (vi) multiagent experts. We provide a detailed analysis of these different algorithms including the derivation of the gradient-based update rules along with some interesting numerical examples in Appendix C. In each of these cases, one can view an agent employing the given algorithm as building an unbiased estimate of their gradient from their observation of the environment.

For example, in multiagent policy gradient (see, e.g., [46, Chapter 13]), agents' costs are defined as functions of a parameter vector x_i that parameterize their policies $\pi_i(x_i)$. The parameters x_i are agent i's choice variable. By following the gradient of their loss function, they aim to tune the parameters in order to converge to an *optimal* policy π_i. Perhaps surprisingly, it is not necessary for agent i to have access to $\pi_{-i}(x_{-i})$ or even x_{-i} in order for them to construct an unbiased estimate of the gradient of their loss with respect to their own choice variable x_i as long as they observe the sequence of actions, say $u_{-i,t}$, of all other agents generated. These actions are implicitly determined by the other agents' policies $\pi_{-i}(x_{-i})(\cdot)$. Hence, in this case if agent i observes $\{(r_{j,t}, u_{j,t}, s_{j,t}) \;\forall\, j \in \mathcal{I}\}$, where (r_j, u_j, s_j) is the reward, action, and state of agent j, then this is enough to construct an unbiased estimate of their gradient. We provide further details on multiagent policy gradient in Appendix C.

4.2.2. Stochastic gradient results.

Returning to the analysis of (4.2), we make the following standard assumptions on the noise processes [37, 38].

Assumption 2. The stochastic process $\{w_{i,t+1}\}$ satisfies the assumptions $\mathbb{E}[w_{i,t+1} \vert\, \mathcal{F}_i^t] = 0$, $t \geq 0$, and $\mathbb{E}[\|w_{i,t+1}\|^2 \vert\, \mathcal{F}_i^t] \leq \sigma^2 < \infty$ a.s. for $t \geq 0$, where $\mathcal{F}_{i,t}$ is an increasing family of σ_i-fields, i.e., filtration, or history generated by the sequence of random variables, given by $\mathcal{F}_{i,t} = \sigma_i(x_{i,k}, w_{i,k}, k \leq t),\ t \geq 0$.

We also make new assumptions on the players' step sizes. These are standard assumptions in the stochastic approximation literature and are needed to ensure that the noise processes are asymptotically controlled.

Assumption 3. For each $i \in \mathcal{I}$, $f_i \in C^s(X, \mathbb{R})$ with $s \geq 2$, $D_i f_i$ is L_i-Lipschitz with $0 < L_i < \infty$, the step sizes satisfy $\gamma_{i,t} \equiv \gamma_t$ for all $i \in \mathcal{I}$ and $\sum_t \gamma_t = \infty$, and $\sum_t (\gamma_t)^2 < \infty$, and $\sup_t \|x_t\| < \infty$ a.s.

Let $(a)^+ = \max\{a, 0\}$ and $a \cdot b$ denotes the inner product. The following theorem extends the results of Theorem 4.1 to the stochastic gradient dynamics in games.

Theorem 4.3. *Consider a game $(f_1, \ldots, f_n)$ on $X = X_1 \times \cdots \times X_n = \mathbb{R}^m$. Suppose each agent $i \in \mathcal{I}$ adopts a stochastic gradient algorithm that satisfies Assumptions 2 and 3. Further, suppose that for each $i \in \mathcal{I}$, there exists a constant $b_i > 0$ such that $\mathbb{E}[(w_{i,t} \cdot v)^+ | \mathcal{F}_{i,t}] \geq b_i$ for every unit vector $v \in \mathbb{R}^{m_i}$. Then, competitive stochastic gradient-based learning converges to strict saddle points of the game on a set of measure zero.*

The proof follows directly from showing that (4.2) satisfies Theorem A.2, provided the assumptions of the theorem hold. The assumption that $\mathbb{E}[(w_{i,t} \cdot v)^+ | \mathcal{F}_{i,t}] \geq b_i$ rules out degenerate cases where the noise forces the stochastic dynamics onto the stable manifold of strict saddle points.

Theorem 4.3 implies that the dynamics of stochastic gradient-based learning defined in (4.2), have the same limiting properties as the deterministic dynamics vis-à-vis saddle points. Thus, the implications described in section 4.1.1 extend to the stochastic gradient setting. In particular, stochastic gradient-based algorithms will avoid a nonnegligible subset of the Nash equilibria in general-sum and potential games. Further, in zero-sum and general-sum games, if the players converge to a critical point, that point may be a non-Nash equilibrium.

4.2.3. Further convergence results for stochastic gradient-play in games. As we demonstrated in section 4.1, outside of potential games, the dynamics of gradient-based learning algorithms in games are not gradient flows. As such, the players' actions can converge to more complex sets than simple equilibria. A particularly prominent class of limiting behaviors for dynamical systems are known as limit cycles (see, e.g., [41]). Limit cycles (or periodic orbits) are sets of states $\mathcal{S}$ such that each state $x \in \mathcal{S}$ is visited at periodic intervals ad infinitum under the dynamics. Thus, if the gradient-based algorithms converge to a limit cycle they will cycle infinitely through the same sequence of actions. Like equilibria, limit cycles can be stable or unstable under the dynamics $\dot{x} = -\omega(x)$, meaning that the dynamics can either converge to or diverge from them depending on their initializations.

We remark that the existence of oscillatory behaviors and limit cycles has been observed in the dynamics of of gradient-based learning in various settings like the training of GANs [11], and multiplicative weights in finite action games [25]. We simply emphasize that the existence of such limiting behaviors is due to the fact that the dynamics are no longer gradient flows. This fact also allows for other complex limiting behaviors like chaos[3] to exist in the dynamics of gradient-based learning in games. We also show in Appendix B.1 that gradient-based learning avoids some limit cycles.

In Appendix B.1, we formalize the notion of a limit cycle and its stability in the stochastic setting. Using these concepts, we then provide an analogous theorem to Theorem 4.3 which states that competitive stochastic gradient-based learning converges to linearly unstable limit

[3] A general term used to characterize dynamical systems where arbitrarily small perturbations in the initial conditions lead to drastically different solutions to the differential equations.

cycles—a parallel notion to strict saddle points but pertaining to more general limit sets—on a set of measure zero, provided that analogous assumptions to those in the statement of Theorem 4.3 hold. Providing such guarantees requires a bit more mathematical formalism, and as such we leave the details of these results to Appendix B.

5. Saddle point local Nash equilibria in LQ dynamic games.

In this section, we present empirical results that show that a nonnegligible subset of two-player LQ games have local Nash equilibria that are strict saddle points of the gradient dynamics. LQ games serve as good benchmarks for analyzing the limiting behavior of gradient-play in a nontrivial setting since they are known to admit global Nash equilibria that can be found by solving a coupled set of Riccati equations [2]. LQ games can also be cast as MARL problems where each agent has a policy that is a linear function of the state and a quadratic reward function. Gradient-play in LQ games can therefore be seen as a form of policy gradient.

The empirical results we now present imply that, even in the relatively straightforward case of linear dynamics, linear feedback policies, and quadratic costs, policy gradient MARL would be unable to find the local Nash equilibrium in a non-negligible subset of problems.

LQ game setup. For simplicity, we consider two-player LQ games in $\mathbb{R}^2$. Consider a discrete time dynamical system defined by

$$z(t+1) = Az(t) + B_1u_1(t) + B_2u_2(t), \tag{5.1}$$

where $z(t) \in \mathbb{R}^2$ is the state at time t, $u_1(t)$ and $u_2(t)$ are the control inputs of players 1 and 2, respectively, and A, B_1, and B_2 are the system matrices. We assume that player i searches for a linear feedback policy of the form $u_i(t) = -K_iz(t)$ that minimizes their loss which is given by

$$f_i(z_0, u_1, u_2) = \textstyle\sum_{t=0}^{\infty} z(t)^TQ_iz(t) + u_i(t)^TR_iu_i(t),$$

where $Q_i \succ 0$ and $R_i \succ 0$ are the cost matrices on the state and input, respectively. We note that the two players are coupled through the dynamics since $z(t)$ is constrained to obey the update equation (5.1). The vector of player derivatives is given by $\omega(K_1, K_2) = (D_1f_1(K_1, K_2), D_2f_2(K_1, K_2))$, where

$$D_if_i(K_1, K_2) = (R_{ii}K_i + B_i^TP_i(B_1K_1 + B_2K_2) - B_i^TP_iA)\textstyle\sum_{t=0}^{\infty} z(t)z(t)^T, \;\; i \in \{1, 2\}.$$

Note that there is a slight abuse of notation here as we are treating D_if_i as a matrix and as the vectorization of a matrix. The matrices P_1 and P_2 can be found by solving the Riccati equations

$$P_i = (A - B_1K_1 - B_2K_2)^TP_i(A - B_1K_1 - B_2K_2) + K_i^TR_iK_i + Q_i, \quad i \in \{1, 2\},$$

for a given (K_1, K_2). As shown in [2], global Nash equilibria of LQ games can be found by solving coupled Ricatti equations. Under the following assumption, this can be done using an analogous method to the method of Lyapunov iterations outlined in [24] for continuous time LQ games.

Assumption 4. Either $(A, B_1, \sqrt{Q_1})$ or $(A, B_2, \sqrt{Q_2})$ is stabilizable-detectable.

Further information on the Nash equilibria in LQ games and the method of Lyapunov iterations can be found in [2] and [24], respectively.

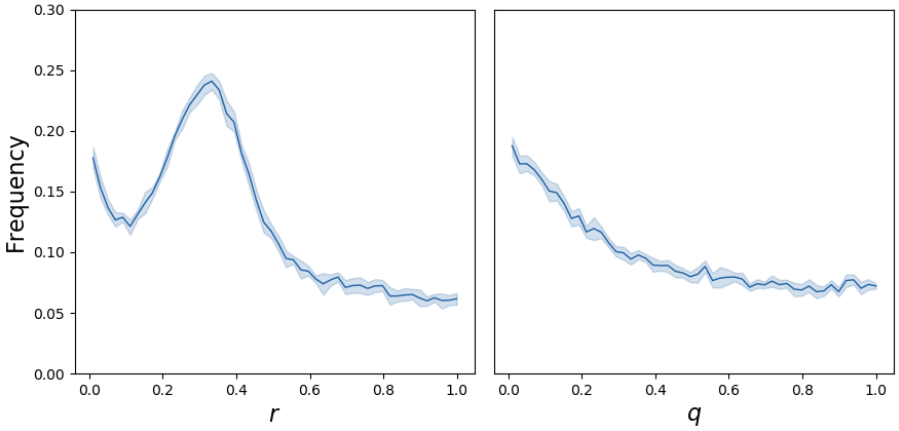

Figure 2. *Frequency (out of* 1000*) of randomly sampled LQ games with global Nash equilibria that are avoided by policy gradient. The experiment was run* 10 *times and the average frequency is shown by the solid line. The shaded region demarcates the* 95% *confidence interval of the experiment. (left)* r *is varied in* $(0,1)$, $q=0.01$*. (right)* q *is varied in* $(0,1)$, $r=0.1$.

Generating LQ games with strict saddle point Nash equilibria. Without loss of generality, we assume $(A, B_1, \sqrt{Q_1})$ is stabilizable-detectable. Given that we have a method of finding the global Nash equilibrium of the LQ game, we now present our experimental setup.

We fix B_1, B_2, Q_1, and R_1 and parameterize Q_2, and R_2 by q and r, respectively. The shared dynamics matrix A has entries that are sampled from the uniform distribution supported on $(0,1)$. For each value of the parameters b, q, and r, we randomly sample 1000 different A matrices. Then, for each LQ game defined in terms of each of the sets of parameters, we find the optimal feedback matrices (K_1^*, K_2^*) using the method of Lyapunov iterations, and we numerically approximate $D\omega(K_1^*, K_2^*)$ using autodifferentiation tools and check its eigenvalues.

The exact values of the matrices are defined as follows: $A \in \mathbb{R}^{2\times 2}$ with each of the entries a_{ij} sampled from the uniform distribution on $(0,1)$,

$$B_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix},\ B_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix},\ Q_1 = \begin{bmatrix} 0.01 & 0 \\ 0 & 1 \end{bmatrix},\ Q_2 = \begin{bmatrix} 1 & 0 \\ 0 & q \end{bmatrix},\ R_1 = 0.01,\ R_2 = r.$$

The results for various combinations of the parameters q and r are shown in Figure 2. For all of the different parameter configurations considered, we found that in anywhere from 0%–25% of the randomly sampled LQ games, there was a global Nash equilibrium that was a strict saddle point of the gradient dynamics. Of particular interest is the fact that for all values of q and r we tested, at least 5% of the LQ games had a global Nash equilibrium with the strict saddle property. In the worst case, around 25% of the LQ games for the given values of q and r admitted such Nash equilibria.

Remark 5.1. These empirical observations imply that multiagent policy gradient, even in the relatively straightforward setting of linear dynamics, linear policies, and quadratic costs, has no guarantees of convergence to the global Nash equilibria in a nonnegligible number of games. Further investigation is warranted to validate this fact theoretically. This in turn supports the idea that for more complicated cost functions, policy classes, and dynamics, local Nash equilibria with the strict saddle property are likely to be very common.

6. Discussion and future directions.

In this paper we provided answers to the following two questions for classes of gradient-based learning algorithms:

Q1. *Are all attractors of the learning algorithms employed by agents' equilibria relevant to the underlying game?*

Q2. *Are all equilibria relevant to the game also attractors of the learning algorithms agents employ?*

We answered these questions in general-sum, zero-sum, and potential games without imposing structure on the game outside regularity conditions on the cost functions by exploiting the observation that gradient-based learning dynamics are not gradient flows. Our analysis is shown in section C to apply to a number of commonly used methods in multiagent learning.

6.1. Links with prior work.

As we noted, previous work on learning in games in both the game theory literature, and more recently from the machine learning community, has largely focused on Q1, though some recent work has analyzed Q2 in the setting of zero-sum games.

In the seminal work by Rosen [40], n-player concave or monotone games are shown to either admit a unique Nash equilibrium or a continuum of Nash equilibria, all of which are attracting under gradient-play. The structure present in these games rules out the existence of non-Nash equilibria.

Two-player, finite-action bilinear games have also been extensively studied. In [44], the authors investigate the convergence of the gradient dynamics in such games. Additionally, the dynamics of other (non-gradient-based) algorithms like multiplicative weights have been studied in [19] among many others. In such settings, the structure guarantees that there exists a unique global Nash equilibrium and no other critical points of the gradient dynamics. As such, non-Nash equilibria, cannot exist.

In the study of learning dynamics in the class of zero-sum games, it has been shown that cycles can be attractors of the dynamics (see, e.g., [25, 47, 19]). Concurrently with our results, [12] also showed the existence of non-Nash attracting equilibria in this setting.

In more general settings, there has been some analysis of the limiting behavior of gradient-play though the focus has been for the most part, on giving sufficient conditions under which Nash equilibria are attracting under gradient-play. For example, [34, 35, 36] introduced the notion of a differential Nash equilibrium which is characterized by first and second order conditions on the players' individual cost functions and which we made extensive use of. Following this body of work, [27] also investigated the local convergence of gradient-play in continuous games. They showed that if a Nash equilibrium satisfies a property known as *variational stability*, the equilibrium is attracting under gradient play. In twice continuously differentiable games, this condition coincides exactly with the definition of stable differential Nash equilibria. Though these works analyze a general class of games, the focus of the analysis is solely on the local characterization and computation (via gradient play) of local Nash equilibria. As such, the issues of nonconvergence that we show in this paper were not discussed.

6.2. Open questions. Our results suggest that gradient-play in multiagent settings has fundamental problems. Depending on the players' costs, in general games and even potential games, which have a particularly *nice* structure, a subset of the Nash equilibria will be almost surely avoided by gradient-based learning when the agents randomly initialize their first action. In zero-sum and general-sum games, even if the algorithms do converge, they may have converged to a point that has no game theoretic relevance, namely, a non-Nash locally asymptotically stable equilibrium.

Last, these results show that limit cycles persist even under a stochastic update scheme. This explains the empirical observations of limit cycles in gradient dynamics presented in [11, 23, 19]. It also implies that gradient-based learning in MARL, multi-armed bandits, GANs, and online optimization all admit limit cycles under certain loss functions. Our empirical results show that these problems are not merely of theoretical interest, but also have great relevance in practice.

Which classes of games have all Nash being attracting for gradient-play and which classes preclude the existence of non-Nash equilibria is an open and particularly interesting question. Further, the question of whether gradient-based algorithms can be constructed for which only game theoretically relevant equilibria are attracting is of particular importance as gradient-based learning is increasingly implemented in game theoretic settings. Indeed, more generally, as learning algorithms are increasingly deployed in markets and other competitive environments, understanding and dealing with such theoretical issues will become increasingly important.

Appendix A. Proofs of the main results.

This appendix contains the full proofs of the results in the paper.

A.1. Proofs on links between dynamical systems and games.

We begin with a proof of Proposition 3.4 that all differential Nash equilibria are either strict saddle points or asymptotically stable equilibria of the gradient dynamics. This relies mainly on the definitions of strict saddle points, locally asymptotically stable equilibria, and nondegenerate differential Nash equilibria and simple linear algebra.

Proof of Proposition 3.4. Suppose that $x \in X$ is a nondegenerate differential Nash equilibrium. We claim that $\mathrm{tr}(D\omega(x)) > 0$. Since x is a differential Nash equilibrium, $D_i^2 f_i(x) \succ 0$ for each $i \in \mathcal{I}$; these are the diagonal blocks of $D\omega(x)$. Further $D_i^2 f_i(x) \succ 0$ implies that $\mathrm{tr}(D_i^2 f_i(x)) > 0$. Since $\mathrm{tr}(D\omega) = \sum_{i=1}^n \mathrm{tr}(D_i^2 f_i(x))$, $\mathrm{tr}(D\omega(x)) > 0$. Thus, it is not possible for all the eigenvalues to have negative real part. Since x is nondegenerate, $\det(D\omega(x)) \neq 0$ so that none of the eigenvalues can have zero real part. Hence, at least one eigenvalue has strictly positive real part.

To complete the proof, we show that the conditions for nondegenerate differential Nash equilibrium are not sufficient to guarantee that x is locally asymptotically stable for the gradient dynamics, that is, not all eigenvalues of $D\omega(x)$ have strictly positive real part. We do this by constructing a class of games with the strict saddle point property. Consider a class of two player games $\mathcal{G} = (f_1, f_2)$ on $\mathbb{R} \times \mathbb{R}$ defined as follows:

$$(f_1(x_1, x_2), f_2(x_1, x_2)) = \left(\frac{a}{2}x_1^2 + bx_1x_2, \frac{d}{2}x_2^2 + cx_1x_2\right).$$

In this game, the Jacobian of the gradient dynamics is given by

$$D\omega(x) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{A.1}$$

with $a, b, c, d \in \mathbb{R}$. If x is a nondegenerate differential Nash equilibria, $a, d > 0$ and $\det(D\omega(x)) \neq 0$ which implies that $ad \neq cb$. Choosing c, d such that $ad < cb$ will guarantee that one of the eigenvalues of $D\omega(x)$ is negative and the other is positive, making x a strict saddle point. This shows that nondegenerate differential Nash equilibria can be strict saddle points of the combined gradient dynamics.

Hence, for any game $(f_1, \dots, f_n)$, a nondegenerate differential Nash equilibrium is either a locally asymptotically stable equilibrium or a strict saddle point, but it not strictly unstable or strictly marginally stable (i.e., having eigenvalues all on the imaginary axis). ■

The proof of Proposition 3.7, which claims that all differential Nash equilibria in zero-sum games are locally asymptotically stable, again just relies on basic linear algebra and the definition of a differential Nash equilibrium.

Proof of Proposition 3.7. Consider a two-player game $(f, -f)$ on $X_1 \times X_2 = \mathbb{R}^m$ with $X_i = \mathbb{R}^{m_i}$. For such a game,

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

Note that $D_{21} f(x) = (D_{12} f(x))^T$. Suppose that $x = (x_1, x_2)$ is a differential Nash equilibrium and let $v = [v_1, v_2] \in \mathbb{R}^m$ with $v_1 \in \mathbb{R}^{m_1}$ and $v_2 \in \mathbb{R}^{m_2}$. Then, $v^T D\omega(x) v = v_1^T D_1^2 f(x) v_1 - v_2^T D_2^2 f(x) v_2 > 0$ since $D_1^2 f(x) \succ 0$ and $-D_2^2 f(x) \succ 0$ for x, a differential Nash equilibrium. Since v is arbitrary, this implies that $D\omega(x)$ is positive definite and, hence, clearly nondegenerate. Thus, for two-player zero-sum games, all differential Nash equilibria are both nondegenerate differential Nash equilibria and locally asymptotically stable equilibria of $\dot{x} = -\omega(x)$. ■

The proof that all locally asymptotically stable equilibria in potential games are differential Nash equilibria relies on the symmetry of $D\omega$ in potential games.

Proof of Proposition 3.9. The proof follows from the definition of a potential game. Since $(f_1, \dots, f_n)$ is a potential game, it admits a potential function ϕ such that $D_i f_i(x) = D_i \phi(x)$ for all x. This, in turn, implies that at a locally asymptotically stable equilibrium of $\dot{x} = -\omega(x)$, $D\omega(x) = D^2\phi(x)$, where $D^2\phi$ is the Hessian matrix of the function ϕ. Further $D^2\phi(x)$ must have strictly positive eigenvalues for x to be a locally asymptotically stable equilibrium of $\dot{x} = -\omega(x)$. Since the Hessian matrix of a function must be symmetric, $D^2\phi(x)$, must be positive definite, which through Sylvester's criterion ensures that each of the diagonal blocks of $D^2\phi(x)$ is positive definite. Thus, we have that the existence of a potential function guarantees that the only locally asymptotically stable equilibria of $\dot{x} = -\omega(x)$, are differential Nash equilibria. ■

A.2. Proofs for deterministic setting. We now present the proof of Theorem 4.1 and its corollaries. The proof of relies on the celebrated stable manifold theorem [43, Theorem III.7],

[45]. Given a map ϕ, we use the notation $\phi^t = \phi \circ \cdots \circ \phi$ to denote the t–times composition of ϕ.

Theorem A.1 (Center and stable manifolds [43, Theorem III.7], [45]). *Let x_0 be a fixed point for the C^r local diffeomorphism $f : U \to \mathbb{R}^d$, where $U \subset \mathbb{R}^d$ is an open neighborhood of x_0 in $\mathbb{R}^d$ and $r \geq 1$. Let $E^s \oplus E^c \oplus E^u$ be the invariant splitting of $\mathbb{R}^d$ into generalized eigenspaces of $D\phi(x_0)$ corresponding to eigenvalues of absolute value less than one, equal to one, and greater than one. To the $D\phi(x_0)$ invariant subspace $E^s \oplus E^c$ there is an associated local ϕ-invariant C^r embedded disk W^{cs}_{loc} called the local stable center manifold of dimension $\dim(E^s \oplus E^c)$ and ball B around x_0 such that $\phi(W^{cs}_{loc}) \cap B \subset W^{cs}_{loc}$, and if $\phi^t(x) \in B$ for all $t \geq 0$, then $x \in W^{sc}_{loc}$.*

Some parts of the proof follow similar arguments to the proofs of results in [22, 32] which apply to (single-agent) gradient-based optimization. Due to the different learning rates employed by the agents and the introduction of the differential game form ω, the proof differs.

Proof of Theorem 4.1. The proof is composed of two parts: (a) the map g is a diffeomorphism, and (b) application of the stable manifold theorem to conclude that the set of initial conditions is measure zero.

(a) *g is a diffeomorphism.* We claim the mapping $g : \mathbb{R}^m \to \mathbb{R}^m$ is a diffeomorphism. If we can show that g is invertible and a local diffeomorphism, then the claim follows. Consider $x \neq y$ and suppose $g(y) = g(x)$ so that $y - x = \gamma \cdot (\omega(y) - \omega(x))$. The assumption $\sup_{x \in \mathbb{R}^m} \|D\omega(x)\|_2 \leq L < \infty$ implies that ω satisfies the Lipschitz condition on $\mathbb{R}^m$. Hence, $\|\omega(y) - \omega(x)\|_2 \leq L \, \|y - x\|_2$. Let $\Gamma = \text{diag}(\Gamma_1, \ldots, \Gamma_n)$, where $\Gamma_i = \text{diag}((\gamma_i)_{j=1}^{m_i})$, that is, Γ_i is an $m_i \times m_i$ diagonal matrix with γ_i repeated on the diagonal m_i times. Then, $\|x - y\|_2 \leq L \|\Gamma\|_2 \|y - x\|_2 < \|y - x\|_2$ since $\|\Gamma\|_2 = \max_i |\gamma_i| < 1/L$.

Now, observe that $Dg = I - \Gamma D\omega(x)$. If Dg is invertible, then the implicit function theorem [21, Theorem C.40] implies that g is a local diffeomorphism. Hence, it suffices to show that $\Gamma D\omega(x)$ 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 $\rho(\Gamma D\omega(x)) \leq \|\Gamma D\omega(x)\|_2 \leq \|\Gamma\|_2 \sup_{x \in \mathbb{R}^m} \|D\omega(x)\|_2 < \max_i |\gamma_i| L < 1$ Of course, the spectral radius is the maximum absolute value of the eigenvalues, so that the above implies that all eigenvalues of $\Gamma D\omega(x)$ have absolute value less than 1.

Since g is injective by the preceding argument, its inverse is well-defined and since g is a local diffeomorphism on $\mathbb{R}^m$, it follows that g^{-1} is smooth on $\mathbb{R}^m$. Thus, g is a diffeomorphism.

(b) *Application of the stable manifold theorem.* Consider all critical points to the game, i.e., $\mathcal{X}_c = \{x \in X | \ \omega(x) = 0\}$. For each $p \in \mathcal{X}_c$, let B_p be the open ball derived from Theorem A.1 and let $\mathcal{B} = \cup_p B_p$. Since $X \subseteq \mathbb{R}^m$, Lindelõf's lemma [20]—every open cover has a countable subcover—gives a countable subcover of $\mathcal{B}$. That is, for a countable set of critical points $\{p_i\}_{i=1}^{\infty}$ with $p_i \in \mathcal{X}_c$, we have that $\mathcal{B} = \cup_{i=1}^{\infty} B_{p_i}$.

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 $g^t(x_0) \in B_{p_i}$ for all $t \geq t_0$. Again, applying Theorem A.1 and using that $g(X) \subset X$—which we note is obviously true if $X = \mathbb{R}^m$—we get that $g^t(x_0) \in W^{cs}_{\text{loc}} \cap X$.

Using the fact that g is invertible, we can iteratively construct the sequence of sets defined by $W_1(p_i) = g^{-1}(W^{cs}_{\text{loc}} \cap X)$ and $W_{k+1}(p_i) = g^{-1}(W_k(p_i) \cap X)$. Then we have that $x_0 \in W_t(p_i)$ for all $t \geq t_0$. The set $\mathcal{X}_0 = \cup_{i=1}^{\infty} \cup_{t=0}^{\infty} W_t(p_i)$ contains all the initial points in X such that gradient-based learning converges to a strict saddle. Since p_i is a strict saddle, $I - \Gamma D\omega(p_i)$ has an eigenvalue greater than 1. This implies that the codimension of E^u is strictly less than m. (i.e., $\dim(W^{cs}_{\text{loc}}) < m$). Hence, $W^{cs}_{\text{loc}} \cap X$ has Lebesgue measure zero in $\mathbb{R}^m$.

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

The proof of Corollary 4.2 follows from the symmetry of $D\omega$ in potential games, and our observations in section 3.

Proof of Corollary 4.2. Since the game admits a potential function ϕ, there is a transformation of coordinates such that agents following the dynamics $x_{t+1} = x_t - \gamma \odot \omega(x_t)$ converge to the same equilibria as the gradient dynamics $x_{t+1} = x_t - \gamma \odot D\phi(x_t)$. Hence, the analysis of the gradient-based learning scheme reduces to analyzing gradient-based optimization of ϕ. Moreover, existence of a potential function also implies that $D_{ij}f_j \equiv D_{ji}f_i$ so that $D\omega$ is symmetric. Indeed, writing $\omega(x)$ as the differential form $\sum_{i=1}^n D_i f_i(x) \mathrm{d}x_i$ and noting that $\mathrm{d} \circ \mathrm{d} = 0$ for the differential operator d, we have that $\mathrm{d}(\omega) = \sum_i \mathrm{d}(D_i f_i) \wedge \mathrm{d}x_i = \sum_{i,j:j>i} (D_{ij}f_j - D_{ji}f_i)\, \mathrm{d}x_i \wedge \mathrm{d}x_j = 0$, where $\wedge$ is the standard exterior product [21]. Symmetry of $D\omega$ implies that all periodic orbits are equilibria, i.e., the dynamics do not possess any limit cycles. By Theorem 4.1, the set of initial points that converge to strict saddle points is of measure zero. Since all the stable critical points of the dynamics are equilibria, with the assumption that $\lim_{t\to\infty} g^t(x)$ exists for all $x \in X$, we have that $P_\nu \left[\lim_{t\to\infty} g^t(x) = x^*\right] = 1$, where x^* is a nondegenerate differential Nash equilibrium which is generically a local Nash equilibrium [35]. ■

A.3. Classical results from dynamical systems.

The remaining results use the following classical result from dynamical systems theory. 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.

Theorem A.2 (see [33, Theorem 1]). *Suppose γ_t is $\mathcal{F}_t$-measurable and $\mathbb{E}[w_t|\mathcal{F}_t] = 0$. Let the stochastic process $\{x_t\}_{t\geq 0}$ be defined as above for some sequence of random variables $\{\epsilon_t\}$ and $\{\gamma_t\}$. Let $p \in X$ with $h(p) = 0$ and let W be a neighborhood of p. Assume that there are constants $\eta \in (1/2, 1]$ and $c_1, c_2, c_3, c_4 > 0$ for which the following conditions are satisfied whenever $x_t \in W$ and t sufficiently large:* (i) *p is a linear unstable critical point;* (ii) *$c_1/t^\eta \leq \gamma_t \leq c_2/t^\eta$;* (iii) *$\mathbb{E}[(w_t \cdot v)^+|\mathcal{F}_t] \geq c_3/t^\eta$ for every unit vector $v \in TX$; and* (iv) *$\|w_t\|_2 \leq c_4/t^\eta$. Then $P(x_t \to p) = 0$.*

Appendix B. Expanded results in the stochastic setting.

In this appendix , we provide extended results in the stochastic setting that require more mathematical formalism than the main body of the paper. In addition, we introduce a new class of games that generalizes potential games and has stronger convergence guarantees than the broader class of general-sum continuous games.

B.1. Avoidance of repelling sets.

To show that stochastic gradient-based learning avoids more general limiting behaviors than saddle points, we need further assumptions on our underlying space, i.e., we need the underlying decision spaces of each agent, i.e., X_i for each

$i \in \mathcal{I}$, to be *smooth, compact manifolds without boundary.*[4] The stochastic process $\{x_n\}$ which follows (4.2) is *defined on* X, that is, $x_n \in X$ for all $n \geq 0$. As before, it is natural to compare sample points $\{x_n\}$ to solutions of $\dot{x} = -\omega(x)$, where we think of (4.2) as a noisy approximation. The asymptotic behavior of $\{x_n\}$ can indeed be described by the asymptotic behavior of the flow generated by ω.

We also need a formal notion of *cycles.* A nonstationary periodic orbit of ω is called a cycle. Let $\xi \subset X$ be a cycle of period $T > 0$. Denote by Φ_T the flow corresponding to ω. For any $x \in \xi$, $\mathrm{spec}(D\Phi_T(x)) = \{1\} \cup C(\xi)$, where $C(\xi)$ is the set of characteristic multipliers. We say ξ is *hyperbolic* if no element of $C(\xi)$ is on the complex unit circle. Further, if $C(\xi)$ is strictly inside the unit circle, ξ is called *linearly stable* and, on the other hand, if $C(\xi)$ has at least one element on the outside of the unit circle, that is, $D\Phi_T(x)$ for $x \in \xi$ has an eigenvalue with real part strictly greater than 1, then ξ is called *linearly unstable.* The latter is the analog of strict saddle points in the context of periodic orbits. We denote by $\{x_t\}$ sample paths of the process (4.2) and $L(\{x_t\})$ is the *limit set* of any sequence $\{x_t\}_{t\geq 0}$ which is defined in the usual way as all $p \in X$ such that $\lim_{k\to\infty} x_{t_k} = p$ for some sequence $t_k \to \infty$. It was shown in [3] that under less restrictive assumptions than Assumptions 2 and 3, $L(\{x_t\})$ is contained in the *chain recurrent set* of ω and $L(\{x_t\})$ is a nonempty, compact and connected set invariant under the flow of ω.

Theorem B.1. *Consider a game $(f_1, \ldots, f_n)$, where each X_i is a smooth, compact manifold without boundary. Suppose each agent $i \in \mathcal{I}$ adopts a stochastic gradient-based learning algorithm that satisfies Assumptions* 2 *and* 3 *and is such that sample points $x_t \in X$ for all $t \geq 0$. Further, suppose that for each $i \in \mathcal{I}$, there exist a constant $b_i > 0$ such that $\mathbb{E}[(w_{i,t}\cdot v)^+|\mathcal{F}_{i,t}] \geq b_i$ for every unit vector $v \in \mathbb{R}^{m_i}$. Then competitive stochastic gradient-based learning converges to linearly unstable cycles on a set of measure zero, i.e., $P(L(x_t) = \xi) = 0$, where $\{x_t\}$ is a sample path.*

As we noted, periodic orbits are not necessarily excluded from the limiting behavior of gradient-based learning in games. We leave out the proof of Theorem B.1 since after some algebraic manipulation, it is a direct application of [4, Theorem 2.1] which is stated below.

Theorem B.2 (see [4, Theorem 2.1]). *Let $\xi \subset X$ be a hyperbolic linearly unstable cycle of h. Assume the following:* (i) $h \in C^2$*;* (ii) $c_1/t^\eta \leq \gamma_t \leq c_2/t^\eta$ *with* $0 < c_1 \leq c_2$ *and* $0 < \eta \leq 1$*; and* (iii) *there exists $b \geq 0$ such that for all unit vectors $v \in \mathbb{R}^m$, $\mathbb{E}[(w_t \cdot v)^+|\mathcal{F}_t] \geq b$. Then $P(L(\{x_t\}) = \xi) = 0$.*

B.2. Morse–Smale games.

For a class of games admitting *gradient-like* vector fields we can go beyond nonconvergence results and give convergence guarantees. Following [4], we introduce a new class of games, which we call *Morse–Smale games*, that are a generalization of potential games. Such games represent an important class since the vector field of ω corresponds to a Morse–Smale vector field which is known to be generic in $\mathbb{R}^2$ and are otherwise structurally stable [18, 31].

[4] The torus $\mathbb{T} = \mathbb{S}^1 \times \mathbb{S}^1$ is an example. The interested reader can consult, e.g., [21] for more details on differential geometry.

Definition B.3. *A game $(f_1, \dots, f_n)$ with $f_i \in C^r$ for some $r \geq 3$ and where the strategy spaces X_i are smooth, compact manifolds without boundaries for each $i \in \mathcal{I}$ is a Morse–Smale game if the vector field corresponding to the differential ω is Morse–Smale, that is, the following hold:* (i) *all periodic orbits ξ (i.e., equilibria and cycles) are hyperbolic and $W^s(\xi) \pitchfork W^u(\xi)$ (i.e., the stable and unstable manifolds of ξ intersect transversally);* (ii) *every forward and backward ω limit set is a periodic orbit;* (iii) *and ω has a global attractor.*

The conditions of Morse–Smale in the above definition ensure that there are only finitely many periodic orbits. The dynamics of games with more general vector fields, on the other hand, can admit chaos (e.g., the classic Lorentz attractor can be cast as gradient-play in a three-player game). Hyperbolic equilibria and periodic orbits are the only types of limiting behavior that have been shown to correspond to strategies relevant to the underlying game [5]. The simplest example of a Morse–Smale vector field is a gradient flow. However, not all Morse–Smale vector fields are gradient flows and, hence, not all Morse–Smale games are potential games.

Example 1. Consider the n-player game with $X_i = \mathbb{R}$ for each $i \in \mathcal{I}$ and $f_n(x) = x_n (x_1^2 - 1)$, $f_i(x) = x_i x_{i+1}$ $\forall i \in \mathcal{I}/\{n\}$ This is a Morse–Smale game that is not a potential game. Indeed, $\dot{x} = -\omega(x)$, where $\omega = [x_2, x_3, \dots, x_{n-1}, x_1^2 - 1]$ is a dynamical system with a Morse–Smale vector field that is not a gradient vector field [10].

Essentially, in a neighborhood of a critical point for a Morse–Smale game, the game behavior can be described by a Morse function ϕ such that near critical points ω can be written as $D\phi$ and away from critical points ω points in the same direction as $D\phi$—i.e., $\omega \cdot D\phi > 0$. Specializing the class of Morse–Smale games, we have stronger convergence guarantees.

Theorem B.4. *Consider a Morse–Smale game $(f_1, \dots, f_n)$ on smooth boundaryless compact manifold X. Suppose Assumptions 2 and 3 hold and that $\{x_t\}$ is defined on X. Let $\{\xi_i,\ i = 1, \dots, l\}$ denote the set of periodic orbits in X. Then $\sum_{i=1}^{l} P(L(\{x_t\}) = \xi_i) = 1$ and $P(L(\{x_t\}) = \xi_i) > 0$ implies ξ_i is linearly stable. Moreover, if the periodic orbit ξ_i with $P(L(\{x_t\}) = \xi_i) > 0$ is an equilibrium, then it is either a nondegenerate differential Nash equilibrium—which is generically a local Nash—or a non-Nash locally asymptotically stable equilibrium.*

The proof of Theorem B.4 follows by invoking Corollary B.5 which is stated below.

Corollary B.5 (see [4, Corollary 2.2]). *Assume that there exists $\delta \geq 1$ such that $\sum_{n \geq 0} \gamma_n^{1+\delta} < \infty$ and that h is a Morse–Smale vector field. If we denote by $\{\xi_i,\ i = 1, \dots, l\}$ the set of periodic orbits in X, then $\sum_{i=1}^{l} P(L(\{x_t\}) = \xi_i) = 1$. Further, if conditions* (i)–(iii) *of Theorem* B.2 *hold, then $P(L(\{x_t\}) = \xi_i) > 0$ implies ξ_i is linearly stable.*

Thus, in Morse–Smale games, with probability one, the limit sets of competitive gradient-based learning with stochastic updates are attractors (i.e., periodic orbits, which include limit cyles and equilibria) of $\dot{x} = -\omega(x)$ and if any attractor has positive probability of being a limit set of the players' collective update rule, then it is (linearly) stable. Moreover, attractors that are equilibria are either nondegenerate differential Nash equilibria (generically local Nash equilbiria) or non-Nash locally asymptotically stable equilibria, but not saddle points.

If we further restrict the class of games to potential games, the results for Morse–Smale games imply convergence to Nash almost surely, a particularly strong convergence guarantee.

Corollary B.6. *Consider the game $(f_1, \dots, f_n)$ on smooth boundaryless compact manifold $X = X_1 \times \cdots \times X_n$ admitting potential function ϕ. Suppose each agent $i \in \mathcal{I}$ adopts a stochastic gradient-based learning algorithm that satisfies Assumptions 2 and 3 and such that $\{x_t\}$ evolves on X. Further, suppose that for each $i \in \mathcal{I}$, there exists a constant $b_i > 0$ such that $\mathbb{E}[(w_{i,t}\cdot v)^+|\mathcal{F}_{i,t}] \geq b_i$ for every unit vector $v \in \mathbb{R}^{m_i}$. Then, competitive stochastic gradient-based learning converges to a nondegenerate differential Nash equilibrium almost surely.*

The proof of Corollary B.6 follows from the fact that potential games are trivially Morse–Smale games that admit no periodic cycles as we showed in the proof of Corollary 4.2.

Proof of Corollary B.6. Consider a potential game $(f_1, \dots, f_n)$, where each X_i is a smooth, compact boundaryless manifold. Then $\omega = D\phi$ for some $\phi \in C^r$ which implies that ω is a gradient flow and, hence, does not admit limit cycles. Let $\{\xi_i,\ i = 1, \dots, l\}$ be the set of equilibrium points in X. Under the assumptions of Theorem B.4, $\sum_{i=1}^{l} P(L(\{x_t\}) = \xi_i) = 1$ and, if $P(L(\{x_t\}) = \xi_i) > 0$, then ξ_i is a linearly stable equilibrium point which is a nondegenerate differential Nash equilibrium of the game due to the fact that $D\omega(x)$ is symmetric in potential games. Hence, a sample path $\{x_t\}$ converges to a nondegenerate differential Nash equilibrium with probability one. Moreover, by [35], we know it is generically a local Nash. ■

We note, that even though a potential function is enough to guarantee convergence to a local Nash equilibrium, potential games can still admit local Nash equilibria that are strict saddle points as shown in section 3. Thus, even this relatively well-behaved class of games has fundamental problems when applying a gradient-based learning scheme.

Appendix C. Classes of gradient-based learning algorithms.

In this section, we provide derivation of the gradient-based learning rules provided in Table 1. We note that the derivation of gradient-based approaches for multiarmed bandits can be found in [46] among other classic references on reinforcement learning.

C.1. Online optimization: Gradient play in noncooperative games.

We first show that classical online optimization algorithms fit into the framework we describe. In this case, each agent is directly trying to minimize their own function $f_i(x_i, x_{-i})$, which can depend on the current iterate of the other agents. There are many examples in the optimization literature of this type of setup. We note that in the full information case, the competitive gradient-based learning framework we describe here is simply *gradient-play* [16], a very well-studied game theoretic learning rule.

Of more interest are some gradient-free online optimization algorithms that also fit into the framework we describe. The game can be described as follows. At each iteration, t of the game, every player publishes their current iterate $x_{i,t}$. Player i, implementing this algorithm, then updates their iterate by taking a random unit vector u, and querying $f_i(x_i + \delta_i u, x_{-i})$. The update map is given by $x_{i,t+1} = x_{i,t} - \gamma_i f_i(x_i + \delta_i u, x_{-i})u$. It is shown in [14] that $f_i(x_i + \delta_i u, x_{-i})u$ is an unbiased estimate of the gradient of a smoothed version of f_i, i.e., $\hat{f}_i(x_i, x_{-i}) = \mathbb{E}_v[f_i(x + \delta v, x_{-i})]$. Thus the loss function being minimized by the agent is $\hat{f}_i$. In this case, the results on characterizing limiting behavior presented in section 4.2 apply.

C.2. Generative adversarial networks. GANs take a game theoretic approach to fitting a generative model in complex structured spaces. Specifically, they approach the problem of fitting a generative model from a data set of samples from some distribution $Q \in \Delta(Y)$ as a zero-sum game between a *generator* and a *discriminator*. In general, both the generator and the discriminator are modeled as deep neural networks. The generator network outputs a sample $G_\theta(z) \in Y$ in the same space Y as the sampled data set given a random noise signal $z \sim F$ as an input. The discriminator $D_w(y)$ tries to discriminate between a true sample and a sample generated by the generator, that is, it takes as input a sample y drawn from Q or the generator and tries to determine if its *real* or *fake*. The goal, is to find a Nash equilibrium of the zero-sum game under which the generator will learn to generate samples that are indistinguishable from the true samples, i.e., in equilibrium, the generator has learned the underlying distribution.

To prevent instabilities in the training of GANs with zero-one discriminators, the Wasserstein GAN attempts to approximate the Wasserstein-1 metric between the true distribution and the distribution of the generator. In this setting, $D_w(\cdot)$ is a 1-Lipschitz function leading to the problem

$$\inf_\theta \sup_w \mathbb{E}_{y\sim Q}[D_w(y)] - \mathbb{E}_{z\sim F}[D_w(G_\theta(z))]$$

which has corresponding dynamics $w_{t+1} = w_t + \gamma\nabla_w L(\theta_t, w_t)$ and $\theta_{t+1} = \theta_t - \gamma\nabla_\theta L(\theta_t, w_t)$, where $L(\theta, w) = \mathbb{E}_{y\sim Q}[D_w(y)] - \mathbb{E}_{z\sim F}[D_w(G_\theta(z))]$ and where γ is the learning rate.

GANs are notoriously difficult to train. The typical approach is to allow each player to perform (stochastic) gradient descent on the derivative of their cost with respect to their own choice variable. There are two important observations about gradient-based learning approaches to GANs relevant to this paper. First, the equilibrium that is sought is generally a saddle point and, second, the dynamics of GANs are complex enough to admit limit cycles [25]. Nonetheless, training GANs with gradient descent is still very common. We note that our results suggest that, on top of periodic orbits and oscillations, training GANs with gradient descent can result in convergence to non-Nash equilibria.

C.3. Multi-agent reinforcement learning algorithms. Consider a setting in which all agents are operating in a Markov decision process (MDP). There is a shared state space $\mathcal{S}$. Each agent, indexed by $\mathcal{I} = \{1, \dots, n\}$ has their own action space U_i and reward function $R_i : \mathcal{S} \times U \to \Delta_{\mathbb{R}}$, where $U = U_1 \times \cdots \times U_n$. We note the reward functions could themselves be random, but for illustrative purposes we suppose they are deterministic. Finally, the dynamics of the MDP are described by a state transition kernel $P : \mathcal{S} \times U \to \Delta_{\mathcal{S}}$ and an initial state distribution P_0. Each agent i also has a policy, π_i, that returns a distribution over U_i for each state $s \in \mathcal{S}$. We define a trajectory of the MDP, τ, as $\tau = \{(s_t, u_{i,t}, u_{-i,u})\}_{t=0}^{T-1}$. Thus, a trajectory is a finite sequence of states, the actions of each player in that state, and the reward agent i received in that state, where T is the time horizon. Given fixed policies we can define a distribution over the space of all trajectories Γ, namely, $P_\Gamma(\pi)$, by

$$P_\Gamma(\tau;\pi) = P_0(s_0)\prod_{i\in\mathcal{I}}\pi_i(u_{i,0}|s_0)\cdots P(s_t|s_{t-1}, u_{t-1})\prod_{i\in\mathcal{I}}\pi_i(u_{i,t}|s_t)\cdots.$$

The goal of each single agent in this setup is to maximize their cumulative expected reward over a time horizon T. That is, the agent is trying to find a policy π_i so as to maximize some function, which in keeping with our general formulation in section 2, we write as $-f_i$ since this problem is a maximization. When an agent is employing policy gradient in this MARL setup, we assume that their policy comes from a parametric class of policies parameterized by $x_i \in X_i \subset \mathbb{R}^{m_i}$. To simplify notation, we write the parametric policy as $\pi_i(x_i)$, where, for each x_i, given a state s, $\pi_i(x_i)$ is a probability distribution on actions u_i which we denote by $\pi_i(x_i)(\cdot|s)$.

The policy gradient MARL algorithm can be reformulated in the competitive gradient-based learning framework. An agent i using policy gradient is trying to tune the parameters x_i of their policy to maximize their expected reward over a trajectory of length T. We define the reward of agent i over a trajectory of the MDP, $\tau \in \Gamma$, to be $\boldsymbol{R}_i(\tau) = \sum_{t=0}^{T-1} R_i(s_{t,i,t}, u_{-i,t})$. Thus, each agent's loss function f_i, in keeping with our notation, is given by $f_i(x_i, x_{-i}) = -J_i(\pi_i(x_i), \pi_{-i}) = -\mathbb{E}_{\tau \sim P_\Gamma(\pi)}[\boldsymbol{R}_i(\tau))]$. The actions of agent i in the continuous game framework described in previous sections are the parameters of their policy, and thus their action space is $X_i \subset \mathbb{R}^{m_i}$. We note that we have made no assumptions on the other player's actions x_{-i}. That is, they do not need to be employing the same parameterized policy class or exactly the same gradient-based update procedure; the only requirement is that they also be using a gradient based multiagent learning algorithm, and that their actions give rise to a set of policies π_{-i} that govern the way they choose their actions in the MDP.

In the full information case, at each round, t of the game, a player plays according to $\pi_i(x_{i,t})$ for a time horizon T, and then performs a gradient update on their parameters, where $D_i f_i(x_i, x_{-i}) = D_i J_i(\pi_i(x_i), \pi_{-i,t})$ is given by

$$D_i J_i(\pi_i(x_i), \pi_{-i}) = \mathbb{E}_{\tau \sim P_\Gamma(\pi)}\Big[\sum_{t=0}^{T-1} R_i(s_t, u_t) \sum_{j=0}^{t} \nabla_{x_i} \log \pi_i(x_i)(u_{i,j}|s_j)\Big]. \tag{C.1}$$

The derivation of this gradient is exactly the same as that of the classic policy gradient. From (C.1) it is clear that an unbiased estimate of the gradient can be constructed. At each time t in the policy gradient update procedure, agent i receives a T horizon rollout, say $z_{i,t} = \{(s_k, u_{i,k}, r_{i,k})\}_{k=0}^{T-1}$, and constructs the unbiased estimate of the gradient, i.e., $\widehat{D_i J_i} = \sum_{k=0}^{T-1} r_{i,k}(\sum_{j=0}^{k} \nabla_{x_i} \log \pi_i(x_{i,t})(u_{i,j}|s_j))$. We note that in this case, the agent does not need to know the policies of the other agents, or anything about the dynamics of the MDP. The agent can construct the estimator solely from the sequence of states, the reward they received in those states, and their own actions. With these two derivations of the gradient for the full information and gradient-free cases, the policy gradient for MARL conforms to the competitive gradient-based learning framework and, hence, the results of section 4 apply under appropriate assumptions.

REFERENCES

[1] S. Abdallah and V. Lesser, *A multiagent reinforcement learning algorithm with non-linear dynamics*, JAIR, 33 (2008), pp. 521–549, https://doi.org/10.1613/jair.2628.

[2] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory*, 2nd ed., Classics Appl. Math., SIAM, Philadelphia, 1998, https://doi.org/10.1137/1.9781611971132.

[3] M. Benaim, *A dynamical system approach to stochastic approximations*, SIAM J. Control Optim., 34 (1996), pp. 437–472, https://doi.org/10.1137/S0363012993253534.

[4] M. Benaim and M. Hirsch, *Dynamics of Morse-Smale urn processes*, Ergodic Theory Dynam. Systems, 15 (1995), pp. 1005–1030, https://doi.org/10.1017/S0143385700009767.

[5] M. Benaim and M. Hirsch, *Learning Processes, Mixed Equilibria and Dynamical Systems Arising from Repeated Games*, manuscript, 1997.

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

[7] L. Bottou, *Large-scale machine learning with stochastic gradient descent*, Proceedings in Computational Statistics, Springer, Berlin, 2010, pp. 177–186.

[8] M. Bravo, D. Leslie, and P. Mertikopoulos, *Bandit learning in concave n-person games*, in Proceedings of the 32nd International Conference on Neural Information Processing Systems, Curran, Red Hook, NY, 2018, pp. 5666–5676.

[9] A. S. Chivukula and W. Liu, *Adversarial learning games with deep learning models*, International Joint Conference on Neural Networks, IEEE, Piscataway, NJ, 2017, pp. 2758–2767.

[10] C. Conley, *Isolated Invariant Sets and the Morse Index*, CBMS Reg. Conf. Ser. Math., AMS, Providence, RI, 1978.

[11] C. Daskalakis, A. Ilyas, V. Syrgkanis, and H. Zeng, *Training GANs with Optimism*, preprint, arXiv:1711.00141, 2017.

[12] C. Daskalakis and I. Panageas, *The limit points of (optimistic) gradient descent in min-max optimization*, in Proceedings of the 32nd International Conference on Neural Information Processing Systems, Curran, Red Hook, NY, 2018, pp. 9256–9266.

[13] M. Fazel, R. Ge, S. M. Kakade, and M. Mesbahi, *Global convergence of policy gradient methods for the linear quadratic regulator*, in Proceedings of the 35th International Conference on Machine Learning, Curran, Red Hook, NY, 2018.

[14] A. D. Flaxman, A. T. Kalai, and H. B. McMahan, *Online convex optimization in the bandit setting: Gradient descent without a gradient*, in Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, Philadelphia, 2005, pp. 385–394.

[15] J. Foerster, R. Y. Chen, M. Al-Shedivat, S. Whiteson, P. Abbeel, and I. Mordatch, *Learning with opponent-learning awareness*, in Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, International Foundation for Autonomous Agents and Multiagent Systems, Curran, Red Hook, NY, 2018, pp. 122–130.

[16] D. Fudenberg and D. K. Levine, *The Theory of Learning in Games*, Vol. 2, MIT Press, Cambridge, MA, 1998.

[17] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, *Generative adversarial nets*, in Proceedings of the 27th International Conference on Neural Information Processing Systems, Curran, Red Hook, NY, 2014, pp. 2672–2680.

[18] M. W. Hirsch, *Differential Topology*, Springer, New York, 1976.

[19] C. H. Hommes and M. I. Ochea, *Multiple equilibria and limit cycles in evolutionary games with logit dynamics*, Games Econom. Behav., 74 (2012), pp. 434–441, https://doi.org/10.1016/j.geb.2011.05.014.

[20] J. Kelley, *General Topology*, Van Nostrand, New York, 1955.

[21] J. Lee, *Introduction to Smooth Manifolds*, Springer, New York, 2012.

[22] J. D. Lee, M. Simchowitz, M. I. Jordan, and B. Recht, *Gradient descent only converges to minimizers*, in 29th Annual Conference on Learning Theory, Proc. Mach. Learn. Res. (PMLR), JMLA, Cambridge, MA, Vol. 49, 2016, pp. 1246–1257.

[23] D. S. Leslie and E. J. Collins, *Individual Q-learning in normal form games*, SIAM J. Control Optim., 44 (2005), pp. 495–514.

[24] T.-Y. Li and Z. Gajic, *Lyapunov iterations for solving coupled algebraic Riccati equations of Nash differential games and algebraic Riccati equations of zero-sum games*, in New Trends in Dynamic Games and Applications, Birkhäuser, Boston, 1995, pp. 333–351.

[25] P. Mertikopoulos, C. Papadimitriou, and G. Piliouras, *Cycles in adversarial regularized learning*, in Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, Philadelphia, 2018, pp. 2703–2717.

[26] P. Mertikopoulos and M. Staudigl, *On the convergence of gradient-like flows with noisy gradient input*, SIAM J. Optim., 28 (2018), pp. 163–197, https://doi.org/10.1137/16M1105682.

[27] P. Mertikopoulos and Z. Zhou, *Learning in games with continuous action sets and unknown payoff functions*, Math. Program., 173 (2019), pp. 465–507, https://doi.org/10.1007/s10107-018-1254-8.

[28] D. Monderer and L. S. Shapley, *Potential games*, Games Econom. Behav., 14 (1996), pp. 124–143, https://doi.org/10.1006/game.1996.0044.

[29] S. Omidshafiei, J. Pazis, C. Amato, J. P. How, and J. Vian, *Deep Decentralized Multi-task Multi-Agent Reinforcement Learning under Partial Observability*, preprint, arXiv:17403.06182, 2017.

[30] M. Osborne, *A Course in Game Theory*, MIT Press, Cambridge, MA, 1994.

[31] J. Palis and S. Smale, *Structural stability theorems*, in Global Analysis, Proc. Symp. Pure Math., 1970, pp. 223–232.

[32] I. Panageas and G. Piliouras, *Gradient descent only converges to minimizers: Non-isolated critical points and invariant regions*, in Innovations in Theoretical Computer Science, Wadern, Germany, Schloss, Dagstuhl, 2016.

[33] R. Pemantle, *Nonconvergence to unstable points in urn models and stochastic approximations*, Ann. Probab., 18 (1990), pp. 698–712.

[34] L. J. Ratliff, S. A. Burden, and S. S. Sastry, *Characterization and computation of local Nash equilibria in continuous games*, in Proceedings of the 51st Annual Allerton Conference on Communication, Control, and Computing, Curran, Red Hook, NY, 2013, pp. 917–924.

[35] L. J. Ratliff, S. A. Burden, and S. S. Sastry, *Genericity and structural stability of non–degenerate differential Nash equilibria*, in Proceedings of the American Control Conference, Piscataway, NJ, 2014, pp. 3990–3995.

[36] L. J. Ratliff, S. A. Burden, and S. S. Sastry, *On the characterization of local Nash equilibria in continuous games*, IEEE Trans. Automat. Control, 61 (2016), pp. 2301–2307, https://doi.org/10.1109/TAC.2016.2583518.

[37] J. W. Robbin, *A structural stability theorem*, Ann. of Math. (2), 94 (1971), pp. 447–493.

[38] H. Robbins and D. Siegmund, *A convergence theorem for non negative almost supermartingales and some applications*, in Herbert Robbins Selected Papers, Springer, New York, 1985, pp. 111–135.

[39] R. Pemantle, *A survey of random processes with reinforcement*, Probab. Surv., 4 (2007), pp. 1–79.

[40] J. B. Rosen, *Existence and uniqueness of equilibrium points for concave n-person games*, Econometrica, 33 (1965), pp. 520–534.

[41] S. Sastry, *Nonlinear Systems*, Springer, New York, 1999.

[42] D. Scieur, V. Roulet, F. Bach, and A. d'Aspremont, *Integration methods and optimization algorithms*, in Proceedings of the 31st International Conference on Neural Information Processing Systems, Curran, Red Hook, NY, 2017, pp. 1109–1118.

[43] M. Shub, *Global Stability of Dynamical Systems*, Springer, New York, 1987.

[44] S. P. Singh, M. J. Kearns, and Y. Mansour, *Nash convergence of gradient dynamics in general-sum games*, in Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, San Francisco, 2000, pp. 541–548.

[45] S. Smale, *Differentiable dynamical systems*, Bull. Amer. Math. Soc., 73 (1967), pp. 747–817, https://doi.org/10.1090/S0002-9904-1967-11798-1.

[46] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, MA, 2017.

[47] E. Wesson and R. H. Rand, *Hopf bifurcations in delayed rock-paper-scissors replicator dynamics*, Dyn. Games Appl., 6 (2016), pp. 139–156.

[48] A. C. Wilson, B. Recht, and M. I. Jordan, *A Lyapunov analysis of momentum methods in optimization*, preprint, arXiv:1611.02635, 2016.

[49] C. Zhang and V. Lesser, *Multi-agent learning with policy prediction*, in Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI Press, Menlo Park, CA, 2010.