Finite-Time Convergence of Gradient-Based Learning in Continuous Games

Benjamin Chasnov,[1] Lillian J. Ratliff,[1] Daniel Calderone,[1] Eric Mazumdar,[2] Samuel A. Burden[1]

[1]Electrical and Computer Engineering
University of Washington, Seattle, WA
{ratliffl,bchasnov,djcal,sburden}@uw.edu

[2]Electrical Engineering and Computer Sciences
University of California, Berkeley, CA
emazumdar@eecs.berkeley.edu

Abstract

We derive convergence guarantees for gradient-based learning algorithms in non-cooperative multi-agent settings. Utilizing the singular values of the game Hessian and of its symmetric part, we obtain a finite-time convergence bound to an ϵ-differential Nash. We support the analysis with several numerical examples including a continuous game that illustrates the analytical convergence guarantee, a linear-quadratic dynamic game with a known Nash equilibrium, and a multi-agent control problem.

1 Introduction

A significant focus in non-cooperative game theory is the characterization and computation of equilibria such as *Nash equilibria* and its refinements. A natural question that arises is how players find or learn such equilibria. With this question in mind, a variety of fields have focused their attention on the problem of learning in games which has lead to a plethora of learning algorithms including gradient play, fictitious play, best response, and multi-agent reinforcement learning among others (Fudenberg and Levine 1998). While convergence has been studied for many of these algorithms, the results tend to asymptotic.

More recently, game theoretic models of algorithm interaction are being adopted in machine learning applications. For instance, game theoretic tools are being used to improve the robustness and generalizability of machine learning algorithms; e.g., generative adversarial networks have become a popular topic of study demanding the use of game theoretic ideas to provide performance guarantees (Daskalakis et al. 2017). In other work from the learning community, game theoretic concepts are being leveraged to analyze the interaction of learning agents—see, e.g., (Heinrich and Silver 2016; Mazumdar and Ratliff 2018; Balduzzi et al. 2018; Tuyls et al. 2018).

Despite this activity, we still lack a complete understanding of the dynamics and limiting behaviors of coupled, competing learning algorithms. In particular, it is important to know when to terminate the algorithms in order to ensure certain performance guarantees or to obtain a finite time bound on the error that can be used to provide guarantees on subsequent control or incentive policy synthesis.

One may imagine that the myraid results on convergence of gradient descent in optimization readily extend to the game setting. Yet, they do not since gradient-based learning schemes in games *do not correspond to gradient flows*. Gradient flows are a very narrow class of flows admitting *nice* convergence guarantees—e.g., almost sure convergence to local minimizers—due to the fact that they preclude flows with the *worst geometries* (Pemantle 2007). In particular, the gradient-based learning dynamics for competitive, multi-agent settings have a *non-symmetric Jacobian* and as a consequence their dynamics may admit complex eigenvalues and non-equilibrium limiting behavior such as periodic orbits. In short, this fact makes it difficut to extend many of the optimization approaches, whose primary technique depends on a cost decreasing at each update, to convergence in single-agent optimization settings to multi-agent settings. In fact, in games, as our examples highlight, a player's cost can increase when they follow the gradient of their own cost. This behavior is due to the coupling between the agents.

In this short paper, we study n-player continuous games in which each player $i \in \mathcal{I} = \{1, \dots, n\}$ wishes to selection an action $x_i \in \mathbb{R}^{d_i}$ that minimizes their cost $f_i(x_i, x_{-i})$ given the actions of all other agents, $x_{-i} = (x_j)_{j \in \mathcal{I}, j \neq i}$. The class of learning algorithms takes the form of simultaneous gradient-based updates given by

$$x_i^+ = x_i - \gamma_i D_i f_i(x_i, x_{-i}), \ \ \forall\, i \in \mathcal{I} \tag{1}$$

where $D_i f_i$ denotes the derivative of f_i with respect to player i's choice variable x_i, and γ_i is player i's learning rate. That is, players myopically update their actions by following the gradient of their cost with respect to their own choice variable. We assume that players have oracle access to their gradient $D_i f_i$ at each time step.

Contributions. Leveraging tools from dynamical systems and optimization, we provide finite-time convergence guarantees for gradient-based learning in non-cooperative games with continuous action spaces. The class of gradient-based learning algorithms we study encompases a wide variety of approaches to learning in games including multi-agent policy gradient and multi-agent gradient-based online optimization where agents have oracle access to their gradients. Specifically, we make the following contributions: (i) we characterize the range of learning rates for which gradient-based learning achieves an asymptotic convergence guaran-

tee, and (ii) assuming players adopt a learning rate γ characterized in terms of the eigenvalues of the Jacobian of the game dynamics, we provide a finite-time convergence guarantee ensuring the players reach an ε-neighborhood of a stable Nash equilibrium.

We support this analysis with several illustrative numerical examples including a continuous game that demonstrates the finite-time analytical convergence guarantee, multi-agent policy gradient applied to a linear-quadratic dynamic game with a known Nash equilibrium that acts as a benchmark for our analytical results, and a minimum-fuel particle intersection problem with unknown equilibria that demonstrates applicability in more complex settings. We conclude with discussion and future work. Specifically, we describe recent extensions to the stochastic setting in which agents only have access to an unbiased estimate of their gradient.

Organization. The remainder of the paper is organized as follows. In Section 2, we provide a breif overview of the relevant game theoretic concepts. We provide the main results on convergence of gradient-based learning in games in Section 3. The convergence results are followed by illustrative examples in Section 4 and we conclude with discussion and a description of extensions to the presented framework in Section 5.

2 Mathematical Preliminaries

Consider the game $(f_1, \dots, f_n)$ on $X = X_1 \times \cdots \times X_n$ where $f_i : X \to \mathbb{R}$ is player i's cost function and $X_i = \mathbb{R}^{d_i}$ is their action space. We use the term *player* and *agent* interchangably. Let $x = (x_i, x_{-i}) \in X$ denote the joint strategy where $x_i \in X_i$ is player i's choice variable and $x_{-i} \in X_{-i}$ is the vector of choice variables of all players excluding i. If each f_i is differentiable, then the *differential game form* (Ratliff, Burden, and Sastry 2016) is given by $\omega(x) = [D_1 f_1(x) \ \cdots \ D_n f_n(x)]$ where $D_i f_i$ is the partial derivative of f_i with respect to x_i.

Assumption 1. *For each $i \in \mathcal{I}$, $f_i \in C^r(X, \mathbb{R})$ for some $r \geq 2$. Moreover, $\omega(x)$ is L–Lipschitz—i.e., $\|\omega(x) - \omega(x')\| \leq L\|x - x'\|$.*

The *game Hessian* is given by

$$D\omega(x) = \begin{bmatrix} D_{11} f_1(x) & \cdots & D_{1n} f_1(x) \\ \vdots & \ddots & \vdots \\ D_{n1} f_n(x) & \cdots & D_{nn} f_n(x) \end{bmatrix}.$$

The entries of the above matrix are dependent on x, however, we will drop this dependence where obvious. Note that each $D_{ii} f_i$ is symmetric under Assumption 1, yet $D\omega$ is in general *not* symmetric. This is an important point and leads to a cruicial difference between the analysis of gradient-based learning in games versus typical analysis of gradient-based approaches to optimization of a single objective.

One of the most common characterizations of the limiting behavior in games is a Nash equilibrium.

Definition 1. *An $x \in X$ 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 $W_i \subset X_i$ on which $f_i(x_i, x_{-i}) \leq f_i(x_i', x_{-i})$ for all $x_i' \in W_i$ and such that $x_i \in W_i$. If the above inequalities are strict, x is a strict local Nash equilibrium.*

Local Nash equilibria can be characterized in terms of first and second order conditions on player cost functions $(f_1, \dots, f_n)$.

Definition 2. *An $x \in X$ is said to be a critical point for the game if $\omega(x) = 0$.*

As shown in (Ratliff, Burden, and Sastry 2016), $\omega(x) = 0$ and $D_{ii}^2 f_i(x) \geq 0$ for each $i \in \mathcal{I}$ are necessary conditions for x to be a local Nash equilibrium. Sufficient conditions give rise to the following definition of a *differential Nash equilibrium*.

Definition 3 ((Ratliff, Burden, and Sastry 2016))**.** *An $x \in X$ is a differential Nash equilibrium if $\omega(x) = 0$ and $D_{ii}^2 f_i(x)$ is positive–definite for each $i \in \mathcal{I}$.*

Differential Nash need not be isolated; yet, for a differential Nash x, if $D\omega(x)$ is non-degenerate (i.e., $\det(D\omega(x)) \neq 0$), then x is an *isolated strict local Nash equilibrium*. Non-degenerate differential Nash are *generic* amongst local Nash equilibria and they are *structurally stable* (Ratliff, Burden, and Sastry 2014) which ensures they persist under small perturbations. This also implies an asymptotic convergence result: if the spectrum of $D\omega$ is strictly in the right-half plane (i.e. $\mathrm{spec}(D\omega(x)) \subset \mathbb{C}_+^\circ$), then a differential Nash equilibrium x is (exponentially) attracting under the flow of $-\omega$ (Ratliff, Burden, and Sastry 2016, Proposition 2).

Definition 4. *A local Nash equilibrium $x \in X$ is stable if $\mathrm{spec}(D\omega(x)) \subset \mathbb{C}_+^\circ$.*

3 Convergence Guarantees

The multi-agent learning framework we analyze is such that each agent's rule for updating their choice variable consists of the agent modifying their action x_i in the direction of their individual gradient $D_i f_i$. Collectively, the dynamics of gradient-based learning can be written as

$$x_{k+1} = x_k - \gamma \odot \omega(x_k) \tag{2}$$

where $\gamma = (\gamma_1, \dots, \gamma_n)$ is a vector of learning rates, one for each agent, and where the $\odot$ notation means block-entry-wise multiplication—i.e.,

$$\gamma \odot \omega(x) = \mathrm{diag}(\gamma_1 I_{d_1}, \dots, \gamma_n I_{d_n})\omega(x)$$

with I_{d_i} a $d_i \times d_i$ identity matrix.

The above learning rule can be thought of as a discretized numerical scheme approximating the continuous time dynamics

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

As we remarked in the preceding section, with a judicious choice of learning rate γ, (2) will converge (at an exponential rate) to a locally stable equilibrium of the dynamics $\dot{x} = -\omega(x)$.

For a stable differential Nash x^*, let $B_r(x^*)$ be a ball of radius $r > 0$ around the equilibrium x^* that is contained in

the region of attraction for x^*[1].

Proposition 1. *Consider an n–player continuous game $(f_1, \ldots, f_n)$ satisfying Assumption 1. Let $x^* \in X$ be a stable differential Nash equilibrium. Suppose agents use the gradient-based learning rule $x_{k+1} = x_k - \gamma \odot \omega(x_k)$ with learning rates $0 < \gamma_i < \tilde{\gamma}$ for each $i \in \mathcal{I}$ and where $\tilde{\gamma}$ is the smallest positive h such that $\max_j |1 - h\lambda_j(D\omega(x^*))| = 1$. Then, for $x_0 \in B_r(x^*)$, $x_k \to x^*$.*

The above result provides a range for the possible learning rates for which (2) converges to a stable differential Nash equilibrium x^* of $(f_1, \ldots, f_n)$ assuming agents initialize in a ball contained in the region of attraction of x^*. Note that the usual assumption in gradient-based approaches to single-objective optimization problems (in which case $D\omega$ is symmetric) is that $\gamma < 1/L$. This is sufficient to guarantee convergence since the spectral radius of a matrix is always less than any operator norm which, in turn, ensures that $|1 - \gamma\lambda_j| < 1$ for each $\lambda_j \in \operatorname{spec}(D\omega(x^*))$.

The convergence guarantee in Proposition 1 is asymptotic in nature. It is often useful, from both an analysis and synthesis perspective, to have non-asymptotic or finite-time convergence results. Such results can be used to provide guarantees on decision-making processes wrapped around the coupled learning processes of the otherwise autonomous agents. The next result, provides a finite-time convergence guarantee for gradient-based learning where agents uniformly use a fixed step size.

Let $B_r(x^*)$ be defined as before with the added condition that it be defined to be the largest ball in the region of attraction such that on $B_r(x^*)$ the symmetric part of $D\omega$—i.e., $S = \frac{1}{2}(D\omega + D\omega^T)$—is positive definite. For a given symmetric matrix $A \in \mathbb{R}^{d\times d}$ (where $d = \sum_{i\in\mathcal{I}} d_i$), let $\lambda_d(A) \leq \cdots \leq \lambda_1(A)$ be its eigenvalues and define

$$\alpha = \min_{x\in B_r(x^*)} \lambda_d\left(S(x)^T S(x)\right)$$

and

$$\beta = \max_{x\in B_r(x^*)} \lambda_1(D\omega(x)^T D\omega(x)).$$

Theorem 1. *Consider a game $(f_1, \ldots, f_N)$ on $X = X_1 \times \cdots \times X_n$ satisfying Assumption 1. Let $x^* \in X$ be a stable differential Nash equilibrium. Suppose $x_0 \in B_r(x^*)$ and that $\alpha < \beta$. Then, given $\varepsilon > 0$, the gradient-based learning dynamics with learning rate $\gamma = \sqrt{\alpha}/\beta$ obtains an ε–differential Nash such that $x_t \in B_\varepsilon(x^*) \subset B_r(x^*)$ for all*

$$t \geq \left\lceil 2\frac{\beta}{\alpha}\log\frac{r}{\varepsilon}\right\rceil.$$

Before we proceed to the proof, let us remark on the assumption that $\alpha < \beta$. First, $\alpha \leq \beta$ is always true; indeed, supressing the dependence on x,

$$\lambda_d(S^T S) \leq \lambda_1(S^T S) \leq \sigma_{\max}(D\omega)^2 = \lambda_1(D\omega^T D\omega)$$

where $\sigma_{\max}(\cdot)$ denotes the largest singular value of its argument. Thus, the condition that $\alpha < \beta$ is generally true, for equality to hold, the symmetric part of $D\omega(x)$ would have *repeated* eigenvalues, which is not generic. Hence, we include this assumption in Theorem 1, but note that it is not restrictive and is fairly benign.

Proof of Theorem 1. It suffices to show that for the choice of γ, the eigenvalues of $I - \gamma D\omega(x)$ are in the unit circle. Indeed, since $\omega(x^*) = 0$, we have that

$$\begin{aligned}\|x_{t+1} - x^*\|_2 &= \|x_t - x^* - \gamma(\omega(x_t) - \omega(x^*))\|_2 \\ &\leq \sup_{x\in B_r(x^*)} \|I - \gamma D\omega(x)\|_2 \|x_t - x^*\|_2\end{aligned}$$

If $\sup_{x\in B_r(x^*)} \|I - \gamma D\omega(x)\|_2$ is less than one, where the norm is the operator 2–norm, then the dynamics are contracting. For notational convenience, we drop the explicit dependence on x. Since

$$\begin{aligned}&(I - \gamma D\omega)^T(I - \gamma D\omega) \\ &\qquad \leq (1 - 2\gamma\lambda_d(S) + \gamma^2\lambda_1((D\omega)^T D\omega))I \\ &\qquad \leq (1 - \tfrac{\alpha}{\beta})I\end{aligned}$$

where the last inequality holds for $\gamma = \sqrt{\alpha}/\beta$ and we note that $\lambda_d(S) \geq \sqrt{\alpha}$ on $B_r(x^*)$. Hence,

$$\begin{aligned}\|x_{t+1} - x^*\|_2 &\leq \sup_{x\in B_r(x^*)} \|I - \gamma D\omega(x)\|_2 \|x_t - x^*\|_2 \\ &\leq (1 - \tfrac{\alpha}{\beta})^{1/2}\|x_t - x^*\|_2.\end{aligned}$$

Since $\alpha < \beta$, $(1 - \alpha/\beta) < \exp(-\alpha/\beta)$ so that

$$\|x_T - x^*\|_2 \leq \exp(-T\alpha/(2\beta))\|x_0 - x^*\|_2.$$

This, in turn, implies that $x_t \in B_\varepsilon(x^*)$ for all $t \geq T = \lceil 2\frac{\beta}{\alpha}\log(r/\varepsilon)\rceil$. □

Note that $\gamma = \sqrt{\alpha}/\beta$ is selected to minimize $1 - 2\gamma\lambda_1(S) + \gamma^2\lambda_d((D\omega)^T D\omega)$. Hence, this is the fastest learning rate given the worst case eigenstructure of $D\omega$ over the ball $B_r(x^*)$ for the choice of operator norm $\|\cdot\|_2$. We note, however, that faster convergence is possible as indicated by Proposition 1 and observed in the examples in Section 4. Indeed, we note that the spectral radius $\rho(\cdot)$ of a matrix is always less than its maximum singular value—i.e. $\rho(I - \gamma D\omega) \leq \|I - \gamma D\omega\|_2$—so it is possible to contract at a faster rate to the equilibrium perhaps given another choice of norm and γ. We remark that if $D\omega$ was symmetric (i.e., in the case of a potential game (Monderer and Shapley 1996) or a single-agent optimization problem), then $\rho(I - \gamma D\omega) = \|I - \gamma D\omega\|_2$ and in this case $\sqrt{\alpha}/\beta$ is the largest rate one could select. In games, however, $D\omega$ is not symmetric.

4 Numerical Experiments

We consider three numerical examples that show convergence to local Nash equilibria. The first is a simple warm-up example to illustrate the main theorem. The second is a more complicated numerical example that serves as a benchmark problem since its Nash equilibrium is completed characterizable and computable. The third is a multi-agent control problem with initially unknown Nash equilibria.

[1]Many techniques exists for approximating the region of attraction; e.g., given a Lyapunov function, its largest invariant level set can be used as an approximation (Sastry 1999). Since $\operatorname{spec}(D\omega(x^*)) \subset \mathbb{C}_\circ^+$, the converse Lyapunov theorem guarantees the existence of a local Lyapunov function.

4.1 Two-player game with quadratic costs

Consider two players indexed by $i = 1, 2$ with strategies $x_i \in \mathbb{R}$ and coupled quadratic costs given by

$$f_1(x_1, x_2) = 0.5x_1^2 - 20x_1x_2,$$

and

$$f_2(x_1, x_2) = 15x_2x_1 + 1.5x_2^2.$$

The point $(x_1, x_2) = (0, 0)$ is a Nash equilibrium for this problem. The agents perform simultaneous gradient descent using the update rules given by

$$x_1^+ = x_1 - \gamma D_1 f_1(x_1, x_2),$$
$$x_2^+ = x_2 - \gamma D_2 f_2(x_1, x_2),$$

where the gradients for each agent are

$$D_1 f_1(x_1, x_2) = x_1 - 20x_2,$$
$$D_2 f_2(x_1, x_2) = 15x_1 + 3x_2,$$

and γ is determined using Theorem 1. We compute the maximum singular value of the game hessian $D\omega$ and minimum singular value of its symmetric part, $S = \frac{1}{2}(D\omega^T + D\omega)$, and square them to obtain α and β. With

$$D\omega = \begin{bmatrix} 1 & -20 \\ 15 & 3 \end{bmatrix} \text{ and } S = \begin{bmatrix} 1 & -2.5 \\ -2.5 & 3 \end{bmatrix},$$

we determine that $\alpha = 0.4898$ and $\beta = 412.3$. Using a learning rate $\gamma = \sqrt{\alpha}/\beta = 1.68 \times 10^{-3}$ we show that the algorithm converges in Figure 1 as the blue curve, and it is upper bounded by the result of Theorem 1 shown as the dashed curve. We note that since gradient-based learning schemes in games do not correspond to gradient flows, the cost does not necessarily decrease at each step.

4.2 Policy Gradient in LQ Dynamic Games

We now consider a linear quadratic dynamic game with two players in the space of linear feedback policies. This game serves as a useful benchmark since it has a unique global equilibrium in linear feedback strategies (Basar and Olsder 1998) that we can compute via a set of coupled algebraic Riccati equations.

Consider the discrete time linear dynamical system

$$z(t+1) = Az(t) + B_1u_1(t) + B_2u_2(t)$$

where $z(t) \in \mathbb{R}^2$ and $u_i(t) = -K_iz(t) \in \mathbb{R}$ is a linear feedback control policy for player i. Each player seeks to minimize their cost

$$f_i(z_0, u_i) = \sum_{t=0}^{\infty} z(t)^T Q_i z(t) + u_i(t)^T R_i u_i(t)$$

with respect to u_i and subject to the state equation constraints which couple the players. We let the parameters be defined as

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0.4 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$
$$Q_{1,2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad R_1 = [4], \quad R_2 = [1].$$

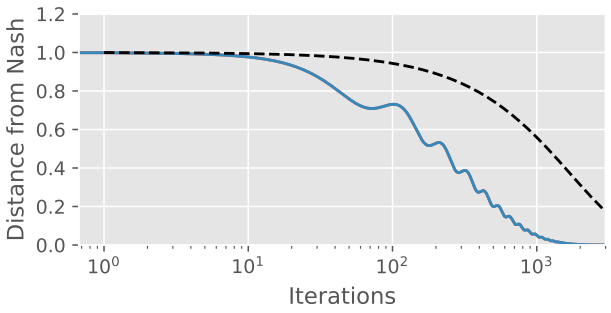
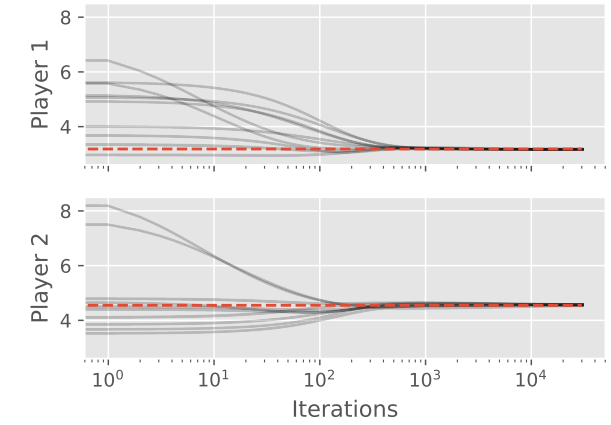
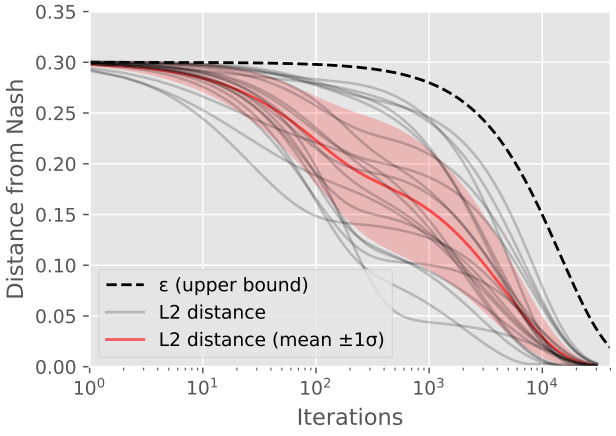


Figure 1: Quadratic toy example showing convergence to the Nash equilibrium through gradient-play. Blue curve plots $\|[x_1, x_2]\|_2^2$ with learning rate $\gamma = \sqrt{\alpha}/\beta$ and the dashed curve is the upper bound from Theorem 1.



(a)



(b)

Figure 2: Convergence of linear quadratic policy gradient to the Nash equilibrium: (a) Each player's cost converges to optimal cost shown by the dotted red lines. (b) The dotted black line shows the number of iterations required to converge within a value of ε as per Theorem 1. Convergence occurs faster than the theorized bound for various initializations of K_1 and K_2 when using $\gamma = \sqrt{\alpha}/\beta$.

We note that the open loop system is unstable, so the players must stabilize the system while minimizing their respective costs.

We compute the game form $\omega(K_1, K_2)$ for costs f_1 and f_2 using initial state $z_0 = [1 \quad 1]^T$. An explicit expression for the game form is given by $\omega(K_1, K_2) = [\omega_1(K_1, K_2) \;\; \omega_2(K_1, K_2)]$ where

$$\omega_i(K_1, K_2) = (R_i K_i + B_i^T P_i (B_i K_i + B_{-i} K_{-i}) - B_i^T P_i A) \sum_{t=0}^{\infty} z(t) z(t)^T$$

We compute P_1 and P_2 by solving the Riccati equations for a given K_1 and K_2.

$$P_i = (A - B_1 K_1 - B_2 K_2)^T P_i (A - B_1 K_1 - B_2 K_2) + K_i^T R_i K_i + Q_i, \quad i = 1, 2.$$

In order to compute the appropriate step size, γ, we compute the optimal Nash feedback gains for the infinite horizon game, K_1^*, K_2^* via a set of coupled Riccati equations using a well-known iterative Lyapunov algorithm (Li and Gajic 1995).

The unique feedback Nash is given by $(K_1^*, K_2^*) = ([0.0019 \quad 0.2301], [0.6825 \quad 0.3605])$. We then compute the game Hessian at the optimal infinite horizon gains $D\omega(K_1^*, K_2^*)$ and $\gamma = \sqrt{\alpha}/\beta = 3.60 \times 10^{-4}$ with $\alpha = 1.49 \times 10^{-1}$ and $\beta = 1.07 \times 10^3$.

We initialize the descent algorithm from a variety of gains K_1, K_2 sampled around a ball of radius 0.3. Figure 2 shows the convergence of the descent algorithm to the Nash policies. Observe that by choosing a learning rate determined by Theorem 1 using the game hessian $D\omega(K_1^*, K_2^*)$, the iterations required to converge to an ε–differential Nash conforms to the theoretical bound given in Theorem 1; the dashed black line in Figure 2 shows the curve of (ε, T)–pairs where $T = \lceil 2\beta/\alpha \log(r/\varepsilon) \rceil$. However, this learning rate is not optimal, as choosing a larger γ empirically results in quicker convergence. We note also this is a worst case bound over all initializations in the ball $B_r(x^*)$.

Furthermore, individual players' costs can increase despite performing gradient descent. Figure 2a illustrates several samples where the players begin at a cost lower than optimal. This is characteristic of a game where players may converge to a stable Nash with higher cost for each player.

4.3 Minimum-fuel particle avoidance

We present an example with $n = 4$ collision-avoiding particles traversing across the unit circle. Each particle follows discrete-time linear dynamics

$$z_i(t+1) = A z_i(t) + B u_i(t)$$

for $t = 1, \cdots, N$ where

$$A = \begin{bmatrix} I & hI \\ 0 & I \end{bmatrix} \in \mathbb{R}^{4\times4} \quad \text{and} \quad B = \begin{bmatrix} h^2 I \\ I \end{bmatrix} \in \mathbb{R}^{4\times2}.$$

The identity matrix is I and constant $h = 0.1$. These dynamics represent a typical discretized version of the continuous dynamics $\ddot{r} = u$ in which u represents a $\mathbb{R}^2$ force vector used to accelerate the particle, and the state $z = [r, \dot{r}]$ represents its position and velocity. Each particle has cost

$$J_i(\mathbf{u}_1, \cdots, \mathbf{u}_i, \cdots \mathbf{u}_n) = \sum_{t=1}^{N} \|u_i(t)\|_R^2 + \sum_{t=1}^{N+1} \|z_i(t) - \bar{z}_i\|_Q^2 + \sum_{j \neq i} \sum_{t=1}^{N+1} \rho e^{-\sigma \|z_i(t) - z_j(t)\|_S^2}$$

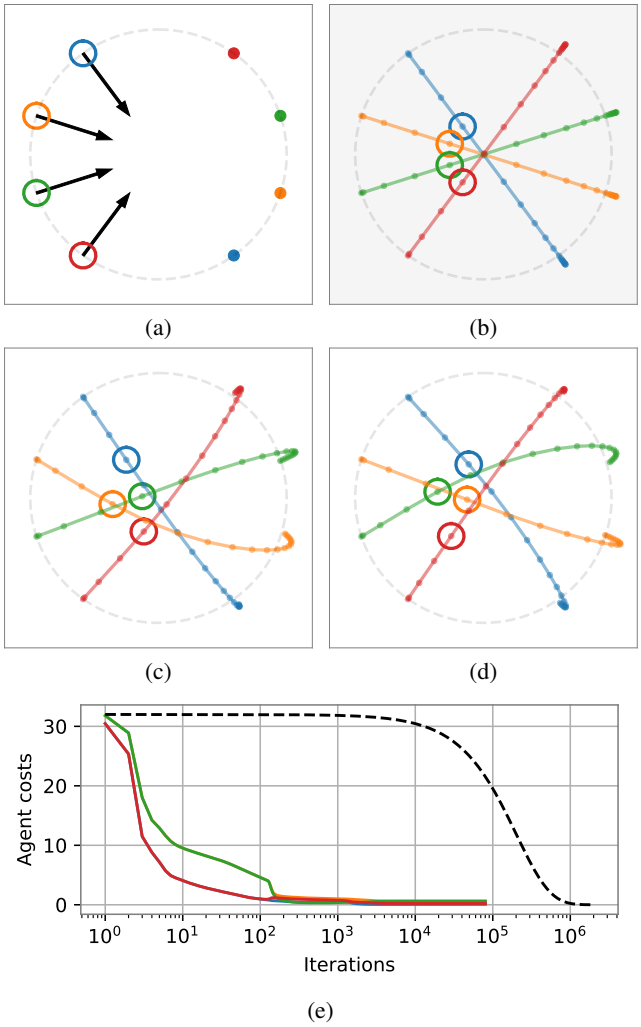


Figure 3: Minimum-fuel particle avoidance control example. (a) Each particle seeks to reach the opposite side of the circle using minimum fuel while avoiding each other. (b) The joint strategy $x = (\mathbf{u}_1, \cdots, \mathbf{u}_4)$ is initialized to the minimum fuel solution ignoring interaction between particles. (c) and (d) Two different equilibrium solutions achieved using slightly randomized initial condition. The circles represent the approximate boundaries around each particle at time $t = 5$. (e) Convergence of the descent method compared to the upper bound.

where $\|\cdot\|_P$ is the quadratic norm, i.e. $\|z\|_P^2 = z^T P z$ with P positive semi-definite, and boldface $\mathbf{u}_i$ is a concatenated vector of control vectors for all time, i.e $\mathbf{u}_i = (u_i(1), \cdots, u_i(N))$. The first two terms of the cost correspond to the minimum fuel objective and quadratic cost from desired final state $\bar{z}_i$, a typical setup for optimal control problems. We use $R = \text{diag}(0.1, 0.1)$ and $Q = \text{diag}(1, 1, 0, 0)$.

The final term of the cost function is the sum of all pairwise interaction terms between the particles, modeled after the shape of a gaussian with scaling constants $\rho = 10$ and $\sigma = 100$. This gaussian-shaped pairwise cost encodes smooth boundaries around the particles.

To find an equilibrium solution between all particles, we use simultaneous gradient descent on the agents' respective cost function. Each agents' strategy x_i is vector $\mathbf{u}_i$ and the gradient $D_i f_i$ is $\frac{\partial J_i}{\partial \mathbf{u}_i}$ of size $2N$.

Figure 3(a) visualizes the problem setup. Each particles' initial position $z_i(0)$ is located on the left side of a unit circle, separated by $\pi/5$, and their desired final positions $\bar{z}_i$ are located directly opposite. The particles begin with zero velocity and must solve for a minimum control solution that also avoids collision with other particles.

We first initialize the problem with the optimal solution for each agent ignoring the pairwise interaction terms, shown in Figure 3(b) . This can be computed using classical discrete-time LQR methods or by gradient descent. Then we randomly perturb this initialization to allow for convergence to different equilibria. We find two of such equilibria by running simultaneous gradient descent with $\gamma = 1 \times 10^{-4}$ until the gradients reach machine precision zero. These equilibria are shown in Figure 3(c) and 3(d). The block diagonal entries of the game Hessian are positive-definite for each $i = 1, \cdots, 4$, ($D_{ii} f_i \geq 0$), and therefore the solutions are differential Nash equilibria for the problem.

5 Discussion

We provide asymptotic and finite-time convergence guarantees for gradient-based learning in general-sum, continuous games. Specifically, we leverage the limiting continuous-time dynamical system and its Jacobian to construct a learning rate γ such that if the agents uniformly adopt this learning rate, they will be guaranteed to converge to a neighborhood of a stable local Nash equilibrium in finite-time. Despite γ not being an optimal learning rate, this method shines light on the theoretical basis of interaction of gradient-based learning dynamics in multi-agent settings where agents have their own individual objective that depends on the actions of others. Beyond analysis, the results are also useful for synthesis. One can use them to design games with desirable properties; this includes incentive design, control theory, and even machine learning, e.g., where game theoretic techniques are being employed to learn robust neural networks such as generative adversarial networks.

We empirically verify the theoretical bounds by testing them with a toy continuous game and an LQ dynamic game, both with known Nash equilibria. The experiments verify that, with a learning rate defined by Theorem 1, agents converge to an $\varepsilon-$differential Nash in finite time. We also present a particle collision avoidance example to demonstrate convergence to multiple equilibria, both not known *a priori*.

5.1 Extensions

Though not included in this short paper, we have extended the results to the stochastic setting. In particular, we provide finite-time, high-probability convergence guarantees for gradient-based learning in games in which the agents do not have oracle access to their gradients, but rather only have access to an unbiased estimator of their individual gradients (Ratliff et al. 2019). Just as in the deterministic setting, we leverage the continuous time limiting dynamics and argue that sample points from the stochastic gradient-based learning update are asymptotic pseudo-trajectories of the semi-flow corresponding to $-\omega$. This allows us to analyze the behavior of the limiting system and argue that sample path representing the sequence of updates made by players does not deviate "far" from the continuous time trajectory. In the stochastic setting, agents do not have a constant learning rate. Instead they each possess a sequence of learning rates. The results apply to both the case where agents have uniform learning rates and the case where agents have distinct learning rates. A direction we are actively pursuing is non-asymptotic convergence guarantees in the stochastic setting. Such results will be particularly useful for providing guarantees on the design of control or incentive policies to coordinate agents in finite time.

5.2 Future Work

The work in this short paper simply scratches the surface; there are a number of avenues to pursue for future work. For instance, the results as stated apply to continuous games with Euclidean strategy spaces. An interesting avenue to pursue is the study of learning in games where the agents decision spaces are constrained sets or Riemannian manifolds. The latter arises in a number of robotics applications and in this case, the update rule will need to be modified by the appropriately defined retraction such as $x_{k+1} = \exp_{x_k}(\gamma_k(\omega(x_k)))$ (Shah 2017). The former arises in a variety of applications where the learning rules are abstractions of agents learning in, e.g., physically constrained environments. The update rule in this case will also need to be defined in terms of the appropriate proximal map thereby leading to potentially non-smooth dynamics (Borkar 2008; Kushner and Yin 2003) which is even more challenging in the stochastic setting. Yet, such extensions will lead to a framework and set of analysis tools that apply to a broader class of multi-agent learning algorithms.

While we present the work in the context of gradient-based learning in games, there is nothing that precludes the results from applying to update rules that conform to our setting in the sense that agents are myopically updating their decisions in time using a process of the form $x_{k+1} = x_k - \gamma \odot g(x_k)$. In particular, it is not necessary that the dynamics g correspond to a game form $\omega(x) = (D_i f_i(x))_{i \in \mathcal{I}}$. For instance, in the stochastic setting, variants of multi-agent Q-learning conform to this setting since Q-learning can be

written as a stochastic approximation update. Exploring and characterizing which commonly used multi-agent learning approaches conform to this setting and, more interestingly, discovering what their limiting dynamics are is a particularly exciting avenue of research.

As pointed out in (Mazumdar and Ratliff 2018), *not all* critical points of the dynamics $\dot{x} = -\omega(x)$ that are attracting are necessarily Nash equilibria; some of the stable equilibria of the dynamics are such that in the context of the game, one or more agent has an incentive to deviate despite the equilibrium being attracting under the flow of $-\omega$. In particular, any equilibrium x such that $\mathrm{spec}(D\omega(x)) \subset \mathbb{C}^{\circ}_{-}$ and at least one of the $D_{ii}f_i(x)$ has a non-positive eigenvalue is a candidate. The higher the dimension of the game, the more likely that there are spurious equilibria of the gradient-based dynamics that do not correspond to Nash equilibria. Understanding this phenomena and developing computational techniques to avoid them is an interesting avenue of future research. Recent work has explored this in the context of zero-sum games (Mazumdar, Jordan, and Sastry 2019), yet the proposed approach requires coordination amongst the learning agents, which is not a problem in settings where the goal is to compute Nash (e.g., for the purpose of training generative adversarial networks). However, in settings where the objective is to study the learning behavior of autonomous agents seeking an equilibrium, an alternative approach or perspective is needed.

Lastly, it was also shown in (Mazumdar and Ratliff 2018) that gradient-based learning in games converges on a set of measure zero to strict saddle points of the continuous time dynamics $\dot{x} = -\omega(x)$, which may or may not correspond to Nash equilibria of the game. Another direction of interest for future research is in quantifying the difficulty of escaping saddle points in games perhaps using the geometry of the underlying space and the graphs of the functions.

6 Acknowledgements

This material is based upon work supported by Computational Neuroscience Graduate Training Program at the University of Washington.

References

Balduzzi, D.; Racaniere, S.; Martens, J.; Foerster, J.; Tuyls, K.; and Graepel, T. 2018. The mechanics of n-player differentiable games. *CoRR* abs/1802.05642.

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

Borkar, V. 2008. *Stochastic Approximation: A Dynamical Systems Viewpoint*. Springer.

Daskalakis, C.; Ilyas, A.; Syrgkanis, V.; and Zeng, H. 2017. Traning GANs with Optimism. *arxiv:1711.00141*.

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

Heinrich, J., and Silver, D. 2016. Deep reinforcement learning from self-play in imperfect-information games. *arxiv:1603.01121*.

Kushner, H. J., and Yin, G. G. 2003. *Stochastic Approximation and Recursive Algorithms and Applications*. Springer, 2nd edition.

Li, T.-Y., and Gajic, Z. 1995. Lyapunov iterations for solving coupled algebraic riccati equations of nash differential games and algebraic riccati equations of zero-sum games. In Olsder, G. J., ed., *New Trends in Dynamic Games and Applications*, 333–351. Boston, MA: Birkhäuser Boston.

Mazumdar, E., and Ratliff, L. J. 2018. On the convergence of competitive, multi-agent gradient-based learning algorithms. *arxiv:1804.05464*.

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

Monderer, D., and Shapley, L. S. 1996. Potential games. *Games and Economic Behavior* 14(1):124–143.

Pemantle, R. 2007. A survey of random processes with reinforcement. *Probability Surveys* 4(1–79).

Ratliff, L. J.; Chasnov, B.; Mazumdar, E.; and Burden, S. A. 2019. Convergence guarantees for gradient-based learning in continuous games. *working paper*.

Ratliff, L. J.; Burden, S. A.; and Sastry, S. S. 2014. Generictiy and Structural Stability of Non–Degenerate Differential Nash Equilibria. In *Proc. 2014 Amer. Controls Conf.*

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

Sastry, S. 1999. *Nonlinear Systems*. Springer New York.

Shah, S. M. 2017. Stochastic approximation on riemannian manifolds. *arXiv*.

Tuyls, K.; Pérolat, J.; Lanctot, M.; Ostrovski, G.; Savani, R.; Leibo, J. Z.; Ord, T.; Graepel, T.; and Legg, S. 2018. Symmetric decomposition of asymmetric games. *Scientific Reports* 8(1):1015.