

## Anticipatory Learning in General Evolutionary Games

Gürdal Arslan and Jeff S. Shamma

**Abstract**— We investigate the problem of convergence to Nash equilibrium for learning in games. Prior work demonstrates how various learning models need not converge to a Nash equilibrium strategy and may even result in chaotic behavior. More recent work demonstrates how the notion of “anticipatory” learning, or, using more traditional feedback control terminology, “lead compensation”, can be used to enable convergence through a simple modification of existing learning models. In this paper, we show that this approach is broadly applicable to a variety of evolutionary game models. We also discuss single population evolutionary models. We introduce “anticipatory” replicator dynamics and discuss the relationship to evolutionary stability.

### I. LEARNING IN GAMES AND FEEDBACK CONTROL

In a multiplayer game, the reward, or utility, that any player receives is a function of that player’s own strategy as well as the strategies of other players. In a Nash equilibrium [14], the collection of player strategies is such that no single player can achieve improved performance by a unilateral change of strategy [3], [15]. As the term implies, a Nash equilibrium describes a fixed, or *static*, situation.

In contrast, the subject of learning in games addresses *dynamic* behavior *away* from equilibrium. There is a large body of literature on the topic of learning in games and the related subject of evolutionary games, including several monographs [8], [12], [18], [24], [25], [26] and recent survey articles [9], [13].

The framework of learning in games is that players continually adjust their strategies in response to observations and outcomes from past experience. A main interest is understanding the limiting behavior of player strategies, and in particular, whether or not they will converge to a Nash equilibrium for different models of learning.

For certain models of learning, strategies need *not* converge to a Nash equilibrium. An early counterexample is due to Shapley [22], who considers learning under “fictitious play” [4], [16]. The recent work of [10] suggests that strategies cannot converge to a Nash equilibrium under the natural incomplete information assumption that players do not know the reward functions of others. Not only can play not converge, but the resulting behavior can be *chaotic*, even for simple two-person games. In particular, the recent paper [19] shows that this is indeed the case for the game of rock-paper-scissors under “replicator dynamics”.

Research supported by NSF grant #ECS-0501394, AFOSR/MURI grant #F49620-01-1-0361, and ARO grant #W911NF-04-1-0316.

G. Arslan is with the Department of Electrical Engineering, University of Hawaii at Manoa, 440 Holmes Hall, 2540 Dole Street, Honolulu, HI 96822, gurdal@hawaii.edu.

J. S. Shamma is with the Department of Mechanical and Aerospace Engineering, University of California, 37-146 Engineering IV, Los Angeles, CA 90095-1597, shamma@ucla.edu.

In all of the aforementioned examples of non-convergence [22], [10], [19], the Nash equilibrium in question is a “mixed strategy”, i.e., a strategy in which player actions are randomly selected from a finite set of choices. One interpretation is that the strategy constitutes the probability of any particular choice. While there are alternative interpretations of mixed strategies [15], the concept has received some scrutiny as being problematic [17]. The lack of convergence of various learning algorithms to mixed strategy Nash equilibrium contributes to such concerns.

In this paper, we present a simple modification to existing learning models based on notions from feedback control. The modification, as well as the accompanying analysis, is applicable to a wide range of learning models. Furthermore, it enables convergence to the mixed strategy Nash equilibrium in several games where the unmodified learning algorithm fails to converge. In particular, we will illustrate that chaos is eliminated in the rock-paper-scissors example of [19]. The connection between feedback control and learning in games is that an individual player’s strategy affects the evolution of other player strategies, which, in turn, affect the original player. In other words, in a learning scenario, players form a feedback loop with each other. It is then natural to consider how concepts from feedback controller design can influence learning in games.

The particular notion used from feedback control is that of “anticipatory” compensation. In traditional feedback control terminology, “lead” compensation reflects responding to an anticipated tracking error. This viewpoint has led to so-called *derivative action* learning for repeated games in [1], [21]. Derivative action can be viewed as a way to respond to approximations of the anticipated strategies of other players.

In a broader sense, introducing derivative action may be viewed as a “dynamic compensation” form of learning, i.e., introducing auxiliary states into the response mechanism. Along these lines, recent work [11] shows that convergence to an approximate Nash equilibrium can be enabled for general multiplayer games by a “bounded recall” learning mechanism, where the length of the required recall grows proportionately to the desired accuracy of the Nash equilibrium approximation.

This paper goes beyond the results in [1], [21] by showing that the notion of anticipation can enable convergence to mixed strategy Nash equilibria in a much broader class of learning dynamics than previously considered, including both multiplayer and single population evolutionary models. While the methods presented here are relevant to a variety of evolutionary models, the approach is not universal, i.e., it does not enable convergence for general multiplayer games as in [11].

The remainder of this paper is organized as follows. Section II presents a collection of two-player learning models. Section III presents anticipatory versions of these models. Section IV contains simulation results that illustrate convergence enabled by anticipatory learning. Section V presents an analysis of anticipatory learning for multiplayer games. Section VI presents “counterexamples” where anticipatory learning cannot enable convergence. Section VII shows how the approach can be applied to single population evolutionary models, and in particular, introduces *anticipatory replicator dynamics*. Finally, Section VIII presents some concluding remarks.

## II. MODELS OF LEARNING

We first review selected continuous-time models of learning in games. For simplicity, we present models and simulations of two-player games. However, the forthcoming main result (Theorem 1) is stated for multiplayer games.

In these models, the *strategy* of each player is the probability of playing one of  $n$  possible actions. In particular, the strategy of player 1 is identified as the vector  $p = (p_1, p_2, \dots, p_n)$ , where  $p_i$  represents the probability of playing the  $i^{\text{th}}$  action. Accordingly, we have that  $p_i \geq 0$  and  $\sum_i p_i = 1$ . Similarly, the strategy of player 2 is the vector  $q = (q_1, q_2, \dots, q_n)$ .

The rewards to each player are characterized by matrices,  $A$  and  $B$ . The  $ij^{\text{th}}$  element of  $A$  is the reward player 1 receives when player 1 plays action  $i$  and player 2 plays action  $j$ . In this case, the  $ji^{\text{th}}$  element of  $B$  is the reward that player 2 receives. The reward functions

$$\begin{aligned}\mathcal{U}(p, q) &= p^T A q \\ \mathcal{V}(q, p) &= q^T B p,\end{aligned}$$

can be interpreted as the average reward received by player 1 and 2, respectively, for fixed strategies  $(p, q)$ .

We now specify how players update their strategies. For further background and discussions, see [24], [8].

*Replicator Dynamics:*

$$\begin{aligned}\dot{p}_i &= ((Aq)_i - \mathcal{U}(p, q)) p_i \\ \dot{q}_i &= ((Bp)_i - \mathcal{V}(q, p)) q_i\end{aligned}$$

For the first player, the quantity  $p_i$  represents the probability of playing the  $i^{\text{th}}$  action. Its growth or decay rate is determined by comparing the average reward associated with using the  $i^{\text{th}}$  action, i.e.,  $(Aq)_i$ , to the overall average reward,  $\mathcal{U}(p, q)$ . Successful actions are played with increasing probability, whereas unsuccessful actions are played with decreasing probability. Similar statements hold for the second player.

*Payoff Adjusted Replicator Dynamics:*

$$\begin{aligned}\dot{p}_i &= \frac{1}{\mathcal{U}(p, q)} ((Aq)_i - \mathcal{U}(p, q)) p_i \\ \dot{q}_i &= \frac{1}{\mathcal{V}(q, p)} ((Bp)_i - \mathcal{V}(q, p)) q_i\end{aligned}$$

Here it is assumed that the reward functions,  $\mathcal{U}$  and  $\mathcal{V}$ , are strictly positive. The interpretation is similar to replicator dynamics, except that larger average rewards induce a slower rate of strategy changes.

*Generalized Replicator Dynamics:*

$$\begin{aligned}\dot{p}_i &= \alpha_i(p, q) p_i \\ \dot{q}_i &= \beta_i(q, p) q_i\end{aligned}$$

Both replicator and payoff adjusted replicator dynamics are special cases. Appropriate choices of  $\alpha_i(\cdot)$  and  $\beta_i(\cdot)$  also result in variations of imitation dynamics and contamination dynamics. These strategy-dependent functions determine the growth rate of individual strategy components. They must satisfy

$$\sum_i \alpha_i(p, q) p_i = 0, \quad \sum_i \beta_i(q, p) q_i = 0$$

in order to assure that the sum of the components of  $p$ , and separately  $q$ , remains equal to 1.

Other types of learning dynamics include gradient play and stochastic fictitious play. Anticipatory learning for these dynamics is discussed in [21].

## III. ANTICIPATORY LEARNING

All of the aforementioned models can be interpreted as strategy adjustments that are a function of both a player's own strategy and an opponent's strategy, and all can be written as

$$\dot{p} = G(p, q) \tag{5a}$$

$$\dot{q} = H(q, p), \tag{5b}$$

for suitably defined functions,  $G(\cdot)$  and  $H(\cdot)$ . Although not explicit in this general notation, the dynamics of either player does not depend explicitly on the reward matrices of the other player, i.e., the dynamics are “uncoupled” in the terminology of [10].

Now if an opponent's strategy were fixed, then learning dynamics should eventually learn to play a best response. But the opponent's strategy is not fixed, because both players are making strategy adjustments. The recognition that the opponents strategy is non-stationary, and therefore a moving target, motivates introducing an element of “anticipation” into the learning process. In particular, we will introduce a modification that mimics the anticipatory effect of proportional-derivative control through the use of “derivative action”.

Putting derivative action together with the different models of learning (written in the general form of (5)) leads to derivative action learning dynamics

$$\dot{p} = G(p, q + \gamma \dot{w}) \tag{6a}$$

$$\dot{q} = H(q, p + \gamma \dot{v}) \tag{6b}$$

$$\dot{v} = \lambda(p - v) \tag{6c}$$

$$\dot{w} = \lambda(q - w). \tag{6d}$$

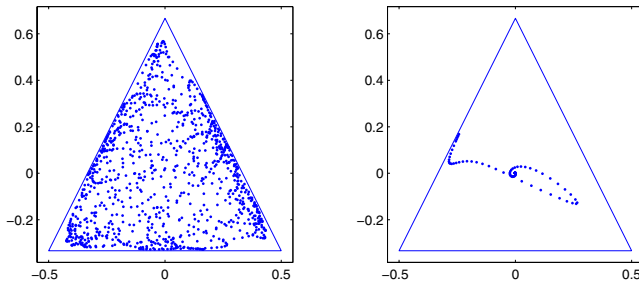


Fig. 1. Strategy of player 1 on the simplex: (a) Standard replicator dynamics (b) With derivative action.

What used to be the opponent's current strategy is replaced with the opponent's *anticipated* strategy through derivative action.

The idea of anticipation also can be found in the work of [6], who considers repeated zero-sum games played in epochs and shows how convergence obstacles presented in [7] can be overcome by generating a projected forecast of the opponents strategy. Likewise, the paper [20] also suggests anticipatory dynamics, but the approach is not "uncoupled" in that each player can employ the reward function of the other. Also related to anticipation is introducing additional memory in learning dynamics, called "relaxation" in [2].

In this paper, we consider (multiplayer) general sum games of imperfect information, where a player does not know the reward function of other players, and give a complete characterization of when the anticipatory model of (6) can lead to stable behavior although the unmodified model (5) does not.

#### IV. SIMULATIONS

We will illustrate derivative action learning on the modified rock-paper-scissors game considered in [19]. The reward matrices are

$$A = \begin{pmatrix} \varepsilon_p & -1 & 1 \\ 1 & \varepsilon_p & -1 \\ -1 & 1 & \varepsilon_p \end{pmatrix}, \quad B = \begin{pmatrix} \varepsilon_q & -1 & 1 \\ 1 & \varepsilon_q & -1 \\ -1 & 1 & \varepsilon_q \end{pmatrix}.$$

The rows and columns can be identified with the actions rock/paper/scissors, respectively. The parameters  $\varepsilon_p$  and  $\varepsilon_q$  satisfy  $-1 < \varepsilon_p < 1$  and  $-1 < \varepsilon_q < 1$  and model rewards in case of a tie.

*Replicator Dynamics:* Standard replicator dynamics leads to chaos [19] for the values  $\varepsilon_p = -\varepsilon_q = 0.5$  and initial conditions  $p(0) = (0.5, 0.01, 0.49)$  and  $q(0) = (0.5, 0.25, 0.25)$ . For these same parameter values and initial conditions, Figure 1 plots the evolving strategy of player 1 on the simplex. The left is standard replicator dynamics, and the right is convergent derivative action replicator dynamics. The strategies converge to the unique Nash equilibrium  $p^* = q^* = (1/3, 1/3, 1/3)$ .

The detailed form of equations (6) for derivative action replicator dynamics is

$$\begin{aligned} \dot{p}_i &= \left( (A(q + \gamma \dot{w}))_i - \mathcal{U}(p, q + \gamma \dot{w}) \right) p_i \\ \dot{q}_i &= \left( (B(p + \gamma \dot{v}))_i - \mathcal{V}(q, p + \gamma \dot{v}) \right) q_i \\ \dot{v}_i &= \lambda(p_i - v_i) \\ \dot{w}_i &= \lambda(q_i - w_i) \end{aligned}$$

The derivative action simulations use  $\gamma = 1$  and  $\lambda = 1$ .

#### V. ANALYSIS: CONDITIONS FOR LOCAL CONVERGENCE TO NASH EQUILIBRIUM

We now present an analysis of when derivative action learning models enable a locally stable Nash equilibrium. We will state the main result for multiplayer games and in a unified form that is applicable to a broad class of learning models, including replicator dynamics, payoff adjusted replicator dynamics, and stochastic fictitious play.

Let  $\Delta$  denote the  $n$ -dimensional simplex, i.e., vectors  $v = (v_1, v_2, \dots, v_n)$  that satisfy  $v_i \geq 0$  and  $\sum_i v_i = 1$ . Let  $\text{Int}(\Delta)$  denote the interior subset of vectors with strictly positive elements.

A general structure to represent standard learning models for an  $m$ -player game is

$$\dot{x}^i = F^i(x^i, x^{-i}), \quad x^i(0) \in \Delta, i = 1, \dots, m. \quad (8)$$

The notation " $-i$ " denotes the complement of the set  $\{1, 2, \dots, m\}$ , i.e.,

$$-i = \{1, 2, \dots, i-1, i+1, \dots, m\},$$

which is interpreted as the set of *opponents* of the  $i^{\text{th}}$  player. The learning dynamics (8) indicate that the strategy of the  $i^{\text{th}}$  player,  $x_i$ , evolves as a function of the  $i^{\text{th}}$  player's own strategy and the strategy of other players,  $x_{-i}$ .

We will call the dynamics (8) *admissible* under the following conditions:

- Consider the  $i^{\text{th}}$  component of the dynamics in isolation, i.e.,

$$\dot{x}^i = F^i(x^i, z(t)),$$

where  $z(t)$  is an arbitrary bounded function. Any initial condition  $x^i(0) \in \Delta$  results in  $x^i(t) \in \Delta$  for all  $t \geq 0$ .

- Suppose  $x^* = (x^{1*}, x^{2*}, \dots, x^{m*})$  is an isolated stationary point of (8), i.e.,  $F^i(x^{i*}, x^{-i*}) = 0$ , for all  $i = 1, \dots, m$ . Suppose further that  $x^{i*} \in \text{Int}(\Delta)$ . Then

$$N^T \frac{\partial F^i}{\partial x^i} \Big|_{x^*} N = -kI, \quad (9)$$

for some real number  $k \geq 0$ , where  $N$  is a matrix satisfying

$$N^T N = I \quad \text{and} \quad N^T \mathbf{1} = 0. \quad (10)$$

It is straightforward to verify that replicator dynamics, payoff adjusted replicator dynamics, appropriately defined generalized replicator dynamics, and gradient dynamics all constitute admissible dynamics with  $k = 0$ . In this case

of  $k = 0$ , any completely mixed Nash equilibrium for all of the aforementioned dynamics cannot be asymptotically stable. Stochastic fictitious play also constitutes admissible dynamics, but with  $k = 1$ .

The above admissibility conditions have natural interpretations in the context of learning in games. The first condition states that solutions that start in the simplex stay in the simplex. The second condition concerns behavior near a stationary point—which corresponds to a mixed strategy Nash equilibrium. Namely, if all but the  $i^{\text{th}}$  player use fixed strategies corresponding to a Nash equilibrium, and if the  $i^{\text{th}}$  player's strategy is near the Nash equilibrium, then the  $i^{\text{th}}$  player either will not drift further away from the Nash equilibrium ( $k = 0$ ) or will drift towards the Nash equilibrium ( $k > 0$ ). Note that  $k > 0$  does *not* imply that the Nash equilibrium is locally stable.

A consequence of the admissibility conditions is that solutions to (8) evolve on the simplex and therefore satisfy

$$x^i(t) - x^{i*} = N\delta x^i(t), \quad (11)$$

where  $\delta x^i(t)$  captures the deviation from equilibrium,  $x^{i*}$ . In other words, the deviation from equilibrium slides along the simplex, or equivalently, lies in the span of the columns of  $N$ .

It will be convenient to analyze the dynamics of  $\delta x$  rather than  $x$ . Accordingly, we can write the “reduced” dynamics

$$\delta \dot{x}^i = F_{\text{red}}^i(\delta x^i, \delta x^{-i}), \quad (12)$$

which admits the stationary point  $\delta x^i = 0$ . A consequence of being admissible dynamics is that the Jacobian matrix of partial derivatives of  $F_{\text{red}} = (F_{\text{red}}^1, \dots, F_{\text{red}}^m)$  evaluated at  $\delta x = (\delta x^1, \dots, \delta x^m) = 0$  has the form

$$\left( \begin{array}{ccc} \partial F_{\text{red}}^1 / \partial \delta x^1 & \dots & \partial F_{\text{red}}^1 / \partial \delta x^m \\ \vdots & \ddots & \vdots \\ \partial F_{\text{red}}^m / \partial \delta x^1 & \dots & \partial F_{\text{red}}^m / \partial \delta x^m \end{array} \right) \Big|_{\delta x=0} = -kI + \mathcal{D}, \quad (13)$$

where  $\mathcal{D}$  is a matrix whose diagonal blocks all equal zero.

Now define the derivative action version of (8), namely, for  $i = 1, \dots, m$ ,

$$\dot{x}^i = F^i(x^i, x^{-i} + \gamma \dot{y}^{-i}), \quad x^i(0) \in \Delta, \quad (14a)$$

$$\dot{y}^i = \lambda(x^i - y^i), \quad y^i(0) \in \Delta, \quad (14b)$$

As before, the strategies of other players in the adjustment dynamics is replaced with anticipated strategies through the use of derivative action.

*Theorem 1:* Let  $\{F^i(\cdot)\}_{i=1}^m$  belong to the class of admissible dynamics. Let  $x^*$  be an isolated stationary point of (8), with  $x^{i*} \in \text{Int}(\Delta)$ ,  $i = 1, \dots, m$ . Denote the eigenvalues of the Jacobian matrix (13) by  $a_i \pm jb_i$ . Then the stationary point  $(x^*, x^*)$  of the derivative action dynamics (14) is locally asymptotically stable for sufficiently large  $\lambda > 0$  if and only if  $\gamma \geq 0$  satisfies

$$\text{T1: } \max_i a_i < \frac{1-\gamma k}{\gamma}, \text{ if } \max_i a_i < 0;$$

$$\text{T2: } \max_i \frac{a_i}{a_i^2 + b_i^2} < \frac{\gamma}{1-\gamma k} < \frac{1}{\max_i a_i}, \text{ if } \max_i a_i \geq 0.$$

The proof is omitted for the sake of brevity.

The important consequence of this theorem is that it characterizes when introducing derivative action can render a mixed strategy Nash equilibrium locally asymptotically stable. Since derivative action introduces auxiliary states, the class of dynamics is broader than those considered in [10], and so the obstacles to convergent dynamics presented there are not applicable.

Statement T1 gives conditions in the case where standard dynamics (8) are locally stable near a Nash equilibrium. In this case, any  $\gamma$  satisfying  $0 < \gamma < 1/k$  results in local stability of the derivative action dynamics (8). For dynamics with  $k = 0$ , any  $\gamma > 0$  works.

Statement T2 gives conditions on  $\gamma$  in case the standard dynamics (8) are unstable near a Nash equilibrium. For example, in the case of a single complex conjugate pair of eigenvalues with positive real parts, condition T2 is equivalent to  $|b_i| > 0$ , i.e., not pure real. In the case of a single real positive eigenvalue, this condition cannot be satisfied by any  $\gamma$ .

Since the theorem is based on an eigenvalue analysis, the conclusions are restricted to local stability. As seen in the simulations, it may well be that the dynamics exhibit a large basin of attraction. Characterizing global stability remains an open question.

The following is an application of the main theorem to the rock-paper-scissors example.

*Replicator Dynamics:* For parameter values  $\varepsilon_p = -\varepsilon_q = 0.5$ , the reduced dynamics Jacobian matrix (13) is

$$\begin{pmatrix} 0 & 0 & 0.1667 & -0.5774 \\ 0 & 0 & 0.5774 & 0.1667 \\ -0.1667 & -0.5774 & 0 & 0 \\ 0.5774 & -0.1667 & 0 & 0 \end{pmatrix},$$

which has the form  $-kI + \mathcal{D}$  with  $k = 0$  and the block diagonal elements of  $\mathcal{D}$  equal to zero. Note that this matrix depends on the specific  $N$  used in (11) (although the conclusions of the Theorem are independent of  $N$ ). In this case,

$$N = \begin{pmatrix} -0.5774 & -0.5774 \\ 0.7887 & -0.2113 \\ -0.2113 & 0.7887 \end{pmatrix}.$$

The eigenvalues of the Jacobian are repeated  $\pm 0.6009i$ . Since these are purely imaginary, and since  $k = 0$  for replicator dynamics, any  $\gamma > 0$  will result in locally asymptotically stable derivative action replicator dynamics (for sufficiently large  $\lambda$ ).

## VI. COUNTEREXAMPLES

Derivative action does not guarantee that for all games, convergence to a mixed Nash equilibrium is possible.

Consider the two-player/two-move identical interest game:

$$A = B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

A mixed strategy Nash equilibrium is  $p^* = q^* = (1/2, 1/2)$ . None of the aforementioned learning dynamics are stable at this equilibrium. Furthermore, the instability stems from a purely real eigenvalue, and so, according to Theorem 1, derivative action does not remedy this instability. For this game, there are also two pure equilibria,  $p^* = q^* = (1, 0)$  or  $(0, 1)$ , and the various learning dynamics do converge to either of these equilibria.

It is also possible to construct a counterexample with a unique Nash equilibrium. Consider a three-player/two-move game, where  $x^i$  denotes the strategy for the  $i^{\text{th}}$  player. The reward function for the first player is

$$(x^1)^T(M_{12}x^2 + M_{13}x^3),$$

where

$$M_{12} = \begin{pmatrix} 0.2 & 1.8 \\ 1.3 & 0.7 \end{pmatrix} \quad M_{13} = \begin{pmatrix} 1.3 & 0.8 \\ 1.5 & 0.6 \end{pmatrix}.$$

Similarly, the reward functions of the second and third players are

$$(x^2)^T(M_{21}x^1 + M_{23}x^3) \\ (x^3)^T(M_{31}x^1 + M_{32}x^2),$$

respectively, where

$$M_{21} = \begin{pmatrix} 0.3 & 1.3 \\ 1.2 & 1.1 \end{pmatrix} \quad M_{23} = \begin{pmatrix} 1.4 & 0.6 \\ 0.4 & 0.9 \end{pmatrix} \\ M_{31} = \begin{pmatrix} 1.9 & 0.7 \\ 1.1 & 0.1 \end{pmatrix} \quad M_{32} = \begin{pmatrix} 0.8 & 1.3 \\ 1.5 & 1.9 \end{pmatrix}$$

This game has a unique Nash equilibrium  $x^{1*} = (0.268, 0.732)$ ,  $x^{2*} = (0.536, 0.464)$ , and  $x^{3*} = (0.304, 0.696)$ . The stability condition of Theorem 1 is not satisfied since the  $3 \times 3$  matrix

$$\begin{pmatrix} N^T & & \\ & N^T & \\ & & N^T \end{pmatrix} \begin{pmatrix} 0 & M_{12} & M_{13} \\ M_{21} & 0 & M_{23} \\ M_{31} & M_{32} & 0 \end{pmatrix} \begin{pmatrix} N & & \\ & N & \\ & & N \end{pmatrix},$$

where  $N = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$ , has a purely real positive eigenvalue. Therefore, derivative action gradient play is unstable for any selection of  $\gamma \geq 0$ .

## VII. SINGLE POPULATION DYNAMICS

An interpretation of the previous models of two player games is that each player represents a “population”, and each player’s strategy represents the relative proportions of the subpopulations.

A more common use of replicator dynamics is to model the evolution of a *single* population with

$$\dot{p}_i = ((Ap)_i - \mathcal{U}(p, p)) p_i$$

where  $\mathcal{U}(p, q) = p^T A q$  for some specified reward matrix,  $A$ . The structure of the dynamics assures that the state evolves in the simplex, i.e.,  $p(t) \in \Delta$  for all time. As compared to the two-player replicator equations, it appears as though a single player is competing with itself.

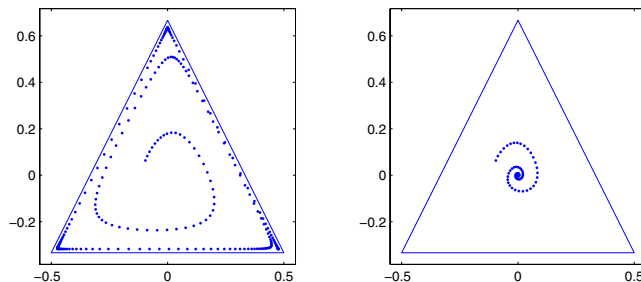


Fig. 2. Subpopulation proportions on the simplex: (a) Single population replicator dynamics (b) Anticipatory single population replicator dynamics.

The “single population” interpretation is that the  $i^{\text{th}}$  component of  $p$  represents the relative proportion of the  $i^{\text{th}}$  subpopulation. The subpopulation growth rate is proportional to its success, measured by  $(Ap)_i$ , as compared to the overall population performance, measured  $\mathcal{U}(p, p)$ .

This view also leads to an alternative derivative action version,

$$\dot{p}_i = ((A(p + \gamma v))_i - \mathcal{U}(p, p + \gamma v)) p_i, \\ \dot{v}_i = \lambda(p_i - v_i).$$

The interpretation is that the  $i^{\text{th}}$  subpopulation’s growth rate now depends on its *anticipated* success measured against how the current population performs against the *anticipated* population.

The introduction of anticipatory dynamics once again can enable a stationary point to become locally stable. Consider the single population rock-scissors-paper game studied in [24, Section 3.1.5],

$$A = \begin{pmatrix} 1 & 2+a & 0 \\ 0 & 1 & 2+a \\ 2+a & 0 & 1 \end{pmatrix}.$$

Standard replicator dynamics has a stationary point at  $p^* = (1/3, 1/3, 1/3)$ . This stationary point is asymptotically stable for  $a > 0$ , but unstable for  $a < 0$ .

Figure 2 shows the subpopulation evolution for the unstable case  $a = -0.5$ . On the left is standard replicator dynamics, which spiral outwards towards the simplex boundary. On the right is anticipatory replicator dynamics ( $\gamma = \lambda = 1$ ), which spirals inwards towards the equilibrium.

This example shows that anticipatory replicator dynamics can enable asymptotic stability for a strategy that is not evolutionarily stable. It is known that non-evolutionarily stable strategies can be stable under replicator dynamics [24], but such a phenomenon is dependent on the choice of coordinates (actually, the choice of “replicators”). In this case, for the same set of replicators, unstable behavior becomes stable behavior through the addition of derivative action.

We can follow the prior analysis for multiplayer games to derive conditions under which derivative action replicator dynamics render an interior stationary point stable.

The starting general structure, which included replicator dynamics as a special case, is:

$$\dot{x} = F(x, z), \quad (16)$$

where  $F(x, z)$  satisfies the following admissibility requirements:

- Consider the dynamics

$$\dot{x} = F(x, z(t)),$$

where  $z(t)$  is an arbitrary bounded function. Any initial condition  $x(0) \in \Delta$  results in  $x(t) \in \Delta$  for all  $t \geq 0$ .

- Suppose  $x^* \in \text{Int}(\Delta)$  is an isolated interior stationary point of (16), i.e.,  $F(x^*, x^*) = 0$ . Then

$$N^T \left. \frac{\partial F(x, z)}{\partial x} \right|_{(x^*, x^*)} N = -kI,$$

for some real number  $k \geq 0$ , where  $N$  satisfies (10).

One can verify that these conditions are met by replicator dynamics with  $k = 0$ .

The derivative action version of (16) is

$$\dot{x} = F(x, x + \gamma \dot{y}), \quad (17a)$$

$$\dot{y} = \lambda(x - y). \quad (17b)$$

*Theorem 2: Let  $F(\cdot, \cdot)$  belong to the class of admissible dynamics. Let  $x^* \in \text{Int}(\Delta)$  be an isolated stationary point of (16). Define*

$$D = N^T \left. \frac{\partial F(x, z)}{\partial z} \right|_{(x^*, x^*)} N,$$

and let  $a_i \pm jb_i$  denote of the eigenvalues of  $-kI + D$ . Then the stationary point  $(x^*, x^*)$  of the derivative action dynamics (17) is locally asymptotically stable for sufficiently large  $\lambda > 0$  if and only if  $\gamma \geq 0$  satisfies

$$T1: \max_i a_i < \frac{1-\gamma k}{\gamma}, \text{ if } \max_i a_i < 0;$$

$$T2: \max_i \frac{a_i}{a_i^2 + b_i^2} < \frac{\gamma}{1-\gamma k} < \frac{1}{\max_i a_i}, \text{ if } \max_i a_i \geq 0.$$

As in Theorem 1, the eigenvalues of a Jacobian matrix extracted from the original (16) can be used to infer the stability of the derivative action form of (17). The proof follows the same arguments as those of Theorem 1.

### VIII. CONCLUDING REMARKS

We have shown how a simple modification of learning dynamics, namely the introduction of approximate anticipation via derivative action, can enable qualitative changes in the stability of learning dynamics. In the case of general-sum multiplayer games, this means convergence to mixed strategy Nash equilibrium—and possibly elimination of chaos—even in the “uncoupled” imperfect information case where players cannot access the reward functions of other players. In the case of single population dynamics, this means enabling stability of non-evolutionarily stable strategies. The potential role of such dynamics as models of players in games (cf., [5]) or as models of natural evolutionary phenomena remains to be addressed. Some results along these lines are reported in [23].

### IX. REFERENCES

- [1] G. Arslan and J.S. Shamma. Distributed convergence to Nash equilibria with local utility measurements. In *43rd IEEE Conference on Decision and Control*, pages 1538–1543, 2004.
- [2] T. Basar. Relaxation techniques and asynchronous algorithms for on-line computation of non-cooperative equilibria. *Journal of Economic Dynamics and Control*, **11**:531–549, 1987.
- [3] K. Binmore. *Fun and Games: A Text on Game Theory*. D.C. Heath, 1991.
- [4] G.W. Brown. Iterative solutions of games by fictitious play. In T.C. Koopmans, editor, *Activity Analysis of Production and Allocation*, pages 374–376. Wiley, New York, 1951.
- [5] C.F. Camerer. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press, 2003.
- [6] J. Conlisk. Adaptation in games: Two solutions to the Crawford puzzle. *Journal of Economic Behavior and Organization*, **22**:25–50, 1993.
- [7] V.P. Crawford. Learning behavior and mixed strategy Nash equilibria. *Journal of Economic Behavior and Organization*, **6**:69–78, 1985.
- [8] D. Fudenberg and D.K. Levine. *The Theory of Learning in Games*. MIT Press, Cambridge, MA, 1998.
- [9] S. Hart. Adaptive heuristics. *Econometrica*, **73**(5):1401–1430, 2005.
- [10] S. Hart and A. Mas-Colell. Uncoupled dynamics do not lead to Nash equilibrium. *American Economic Review*, **93**(5):1830–1836, 2003.
- [11] S. Hart and A. Mas-Colell. Stochastic uncoupled dynamics and Nash equilibrium. Preprint, <http://www.ma.huji.ac.il/~hart/abs/uncoupled.html>, 2004.
- [12] J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge, UK, 1998.
- [13] J. Hofbauer and K. Sigmund. Evolutionary game dynamics. *Bulletin of the American Mathematical Society*, **40**(4):479–519, 2003.
- [14] J.F. Nash. Equilibrium points in  $N$ -person games. *Proceedings of the National Academy of Sciences*, **36**:48–49, 1950.
- [15] M.J. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, Cambridge, MA, 1994.
- [16] J. Robinson. An iterative method of solving a game. *Ann. Math.*, **54**:296–301, 1951.
- [17] A. Rubinstein. Comments on the interpretation of game theory. *Econometrica*, **59**(4):909–924, 1991.
- [18] L. Samuelson. *Evolutionary Games and Equilibrium Selection*. MIT Press, Cambridge, MA, 1997.
- [19] S. Sato, E. Akiyama, and J.D. Farmer. Chaos in learning a simple two person game. *Proceedings of the National Academy of Sciences*, **99**(7):4748–4751, 2002.
- [20] R. Selten. Anticipatory learning in two-person games. In R. Selten, editor, *Game Equilibrium Models, Vol. I: Evolution and Game Dynamics*, pages 98–153. Springer-Verlag, 1991.
- [21] J.S. Shamma and G. Arslan. Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria. *IEEE Transactions on Automatic Control*, **50**(3):312–327, 2005.
- [22] L.S. Shapley. Some topics in two-person games. In L.S. Shapley M. Dresher and A.W. Tucker, editors, *Advances in Game Theory*, pages 1–29. University Press, Princeton, NJ, 1964.
- [23] F.-F. Tang. Anticipatory learning in two-person games: some experimental results. *Journal of Economic Behavior & Organization*, **44**:221–232, 2001.
- [24] J.W. Weibull. *Evolutionary Game Theory*. MIT Press, Cambridge, MA, 1995.
- [25] H. P. Young. *Individual Strategy and Social Structure*. Princeton University Press, Princeton, NJ, 1998.
- [26] H.P. Young. *Strategic Learning and its Limits*. Oxford University Press, 2006.