ERC ECONOMICS RESEARCH CENTER

ISCTE
Lisbon University Institute

Learning to Play Nash in Deterministic Uncoupled Dynamics

Vivaldo M. Mendes
Diana A. Mendes
Orlando Gomes

ISCTE - Lisbon University Institute
UNIDE - ECR
AV FORÇAS ARMADAS
1649-126 LISBON-PORTUGAL
http://erc.unide.iscte.pt

Working Paper - 18/08

Learning to Play Nash in Deterministic Uncoupled Dynamics

Vivaldo M. Mendes* Diana A. Mendes† Orlando Gomes‡

Abstract

This paper is concerned with the following problem. In a bounded rational game where players cannot be as super-rational as in Kalai and Leher (1993), are there simple adaptive heuristics or rules that can be used in order to secure convergence to Nash equilibria, or convergence only to a larger set designated by correlated equilibria? Are there games with uncoupled deterministic dynamics in discrete time that converge to Nash equilibrium or not? Young (2008) argues that if an adaptive learning rule follows three conditions — (i) it is uncoupled, (ii) each player's choice of action depends solely on the frequency distribution of past play, and (iii) each player's choice of action, conditional on the state, is deterministic — no such rule leads the players' behavior to converge to Nash equilibrium. In this paper we present a counterexample, showing that there are simple adaptive rules that secure convergence, in fact fast convergence, in a fully deterministic and uncoupled game. We used the Cournot model with nonlinear costs and incomplete information for this purpose and also illustrate that this convergence can be achieved with or without any coordination of the players actions.

Keywords: Uncoupled deterministic dynamics, Nash equilibrium, bounded learning, convergence.

1 Introduction

The particular problem of "learning" in game theory — or the dynamic process through which the players' actions may (or may not) converge to the equilibria of the game — has a long history in economics; in fact, we can say that the former is almost as long as the very history of game theory

*Department of Economics, ISCTE Lisbon, E-mail: vivaldo.mendes@iscte.pt

†Department of Quantitative Methods, IBS Business School, ISCTE, e-mail: diana.mendes@iscte.pt

‡Escola Superior de Comunicação Social - IPL, Lisbon, e-mail: ogomes@escs.ipl.pt

itself. Already in the early 1950s we find seminal contributions to the subject. Brown (1951) presented a dynamic adjustment now widely known as "fictitious play" to be taken as a learning process for computing the equilibria of games, while Hannan (1957) and Blackwell (1956) put forward specific proposals to evaluate the success of convergence of various learning rules to the equilibria.

However, despite such a long history, it seems no exaggeration at all to argue that it was essentially over the last ten years or so that the learning problem took over as one of the crucial elements of modern game theory. The literature has grown and flourished so much that it is totally impossible to acknowledge much of its large volume in a necessarily short space, which includes a significant number of advanced textbooks — see e.g. Fudenberg and Levine (1998), Rubinstein (1998), Young (2004), Sandholm (2008), Camerer (2003) and Cesa-Bianchi and Lugosi (2006) — and more than a hundred papers, see two recent excellent surveys by Hart (2005) and Sandholm (2007).

This paper is concerned with the following problem. In a bounded rational game where players cannot be as super-rational as in Kalai and Leher (1993) — where they were found to be unbounded in what they can remember, compute, or anticipate — are there *simple adaptive heuristics* or rules that can be used in order to secure convergence to Nash equilibria, or convergence only to a larger set designated by correlated equilibria? Do games with uncoupled deterministic dynamics in discrete time converge to Nash equilibrium or not? Does the dynamics in fact converge to, or does it just come close to the Nash equilibria? By uncoupled dynamics, following Hart and Mas-Colell (2003), we mean a game in which the strategy may depend on the actions of the other players but not on their preferences.

This issue has been extensively discussed since the late 1990s. Foster and Vohra (1997,1998), Fudenberg and Levine (1998), Hart and Mas-Colell (2000, 2003) proved that by constructing a calibrated procedure for forecasting opponents' play convergence of time-averaged behavior to the set of *correlated equilibria* can be achieved, independently of the game considered. Notice that we are mentioning correlated equilibria, not the set of *Nash equilibria* which is generally speaking a strict subset of the set of the former equilibria. It was Foster and Young (2003) who tackled the problem of convergence to the set of Nash equilibria and showed that the dynamics are most of the time close to but are not Nash-convergent. The state of the problem, as we currently have it, can be clearly highlighted by a quotation from a very recent paper by Foster and Young (2006):

> "We have repeatedly said that interactive trial and error learning cause behaviors to come close to Nash equilibrium a high proportion of the time. Why not just say that behaviors converge to Nash equilibrium? Because typically they do not

converge. In fact, there are very severe limits to what can be achieved if one insists on convergence to Nash equilibrium. To be specific, suppose that a learning rule has the following properties: (i) it is uncoupled, (ii) each player's choice of action depends solely on the frequency distribution of past play (as in fictitious play), and (iii) each player's choice of action, conditional on the state, is deterministic. Hart and Mas-Colell (2003) show that for a large class of games, no such rule causes the players' period-by-period behavior to converge to Nash equilibrium."(p.7)

Despite the problem being spelled out above with such strong clarity and conviction, from a mere intuitive point of view one may raise some doubts about such generality of the no convergence result to Nash equilibria if certain conditions are considered: if the game has a stationary structure, if one accepts that the game is allowed to be played for a long period of time, and, finally, if players are allowed to learn from the past experience, even if only in a bounded fashion. In fact, the problem of possible no convergence to the Nash equilibrium was already acknowledged by Shapley (1964), but there is a feeling that if the game has an internal dynamic structure that is ergodic (either fully deterministic or stochastically ergodic), it must be subject to some level of prediction or control when the players can use information from past outcomes in order to decide what strategy should be followed. For example, this was exactly what happened with the recent paper by Germano and Lugosi (2007), who showed that the no convergence result of Foster and Young (2003) could be easily reversed if the players were allowed to add experimentation to their learning procedures. Notice that in their approach a rationale of bounded rational players is still adopted, keeping the game far away from the strong rationality hypothesis of Kalai and Lehrer (1993), but the crucial point in their paper is that there is a simple and adaptive process — bounded rational learning — that does in fact deliver convergence to the Nash equilibrium.

But the counterpoint presented by Germano and Lugosi can also be found in many other recent papers. For example, it is well known in game theory that the strategy pair sequence produced by following a gradient ascent algorithm may never converge, see Owen (1995). However Singh, Kearns and Mansour (2000) showed that in general stochastic games if both players follow an Infinitesimal Gradient Ascent (IGA) learning process, then their strategies will converge to a Nash equilibrium or the average payoffs over time will converge in the limit to the expected payoffs of a Nash equilibrium. Their first theorem is extremely interesting because it states one of the first convergence results for a rational multiagent learning algorithm, although the convergence was still somewhat week. This weakness was quickly overcome by Bowling and Veloso (2002), who showed that if the learning process followed a WoLF principle ("Win or Learn Fast") we will obtain a stronger notion of convergence, i.e., players will always converge to a Nash equilibrium. Successful convergence to the Nash equilibrium can also be found in the learning approach proposed by Zinkevich (2003), under the name of Generalized Infinitesimal Gradient Ascent (GIGA), and one finds in Leslie and Collins (2006) an explanation of why bounded rational players might learn to play Nash equilibrium strategies without having any knowledge of the game, or even that they are playing a game.

In this paper we continue on this route of bounded rational learning and the convergence to the Nash equilibrium. In particular, we take the three conditions presented in the above quotation by Young as a delimitation criteria for whether the Nash equilibrium can be learned or not, and put one of the most simple games that has been used in game theory (the Cournot model) to the test. We provide a counterexample that clearly violates those three general conditions above. We take a standard Cournot model in strategic form, with pretty conventional convex cost curves that can be found in any undergraduate microeconomics textbook, and we add bounded learning in order to overcome the extremely high computational requirements needed to achieve the Nash equilibrium of the game. The game is fully deterministic and clearly satisfies the three conditions above: (i) it is uncoupled, because the strategies depend only upon the other player's actions (not upon the opponents profit function); (ii) each player's choice of action depends solely on the past play, and (iii) each player's choice of action is entirely deterministic.

We will show that, under these conditions, a simple adaptive learning rule going back in time as far as $t-1$, concerning information on the actions taken (output in this case), will deliver very fast convergence to the Nash equilibrium, even in a case where we have multiple Nash equilibria. That is, if the game starts to be played relatively near any one of the various Nash equilibria, and players are bounded rational — so that they are not able to compute straight away the level of output correspondent to any of those equilibria — but take decisions for the next period by using information on the output produced at t and $t-1$ by both firms, the dynamics converge very fast to that particular Nash equilibrium. This occurs if both players adopt similar learning procedures, but can also be achieved if just one of the players corrects his mistakes by incorporating in his strategy past information on his own actions and the actions of his rival.

Another interesting point in this bounded rational game consists in the fact that even very simple rules of thumb can be very powerful rules to lead to optimal decision making, because they may render optimal decisions to be achievable by trial and error, in a situation where without them such optimal decisions were hardly feasible, unless a player is equipped with super-rational powers. Such a point was firstly highlighted by Baumol and Quandt (1964) in one of the first papers applying the concepts of bounded rationality to economics:

> "It is easy to jump to the conclusion that the widespread use of rules of thumb is good evidence of sloppy workmanship on the part of business management. We shall argue [...] on the contrary, rules of thumb are among the more efficient pieces of equipment of optimal decision making."(p. 23).

The paper is organized as follows. In section two the Cournot game with nonlinear costs and incomplete information is presented. Section three introduces the bounded rational process adopted by players in order to overcome the extreme complex computations which are required to play Nash in a one shot game. Section four analyses with some rigor the local and global dynamics associated with the bounded rational game. Section five discusses the simple time-delayed adaptive process to secure convergence to the Nash equilibrium, and the final section presents some concluding remarks.

2 The Cournot game

Take the standard Cournot model of an oligopolistic market (where everything is already known in the literature): 2 firms, homogeneous product, complete information, and constant average costs. Assume that now there is incomplete information and that average production costs are nonlinear. The assumption of incomplete information is necessary in our exercise in order to artificially increase the complexity of the game, such that it would require super levels of rationality from the players to come up with a set of action profiles that would produce the Nash equilibrium in a one shot game. Our fundamental question is whether the Nash–Bayesian Equilibrium (NBE) of such an extremely complex game is learnable or not if played by bounded rational players. Standard results in the literature show that the NBE would not learnable under these conditions. Consider for instance the remark by Cox and Walker (1998, p. 143)

> "But what if marginal costs are not constant? If they are increasing, the reaction functions still look qualitatively as they do in Fig.1 and therefore adaptive learning models continue to predict convergence to the Nash equilibrium. If marginal costs are decreasing, but not decreasing too rapidly (as compared with the demand function), the reaction functions again cross as in Fig.1, and adaptive learning models again predict convergence to the Nash equilibrium. But when one or both firms' marginal costs are declining too rapidly [...] several specific adaptive models, including Cournot's best-reply model [...] all predict that play will not converge to the interior Nash equilibrium but will converge instead to one of the boundary Nash equilibria."

In this paper, we show the opposite: even in a model with an extremely complicated structure, the NBE can be learned over time and the boundary Nash equilibria can be easily ruled out. To prove this we use simple time-delayed feedback learning rules, namely we consider that one single player can render the convergence to the Nash equilibria possible if he uses information on the past actions of play by himself and by the other player, or that convergence can be achieved by both players if they use the same learning process or if there is some form of coordination in learning .

To develop analytically our Cournot game, we begin by presenting some fundamental definitions:

Definition 1 ***Normal form game***. *A finite n-person normal (or strategic) form game is a tuple* (N, A, u) *where*

- N is a finite set of n players, indexed by i;
- $A = A_1 \times ... \times A_n$, where A_i is a finite set of actions available to player i. Each vector $a(a_1, .., a_n) \in A$ is called an action profile;
- $u(u_1, .., u_n)$ where $u_n : A \mapsto \mathbb{R}$ is a real-valued utility (or payoff) function for player i.

Definition 2 ***Cournot oligopoly game*** *is a strategic game among firms where competition is based on quantity adjustment, and includes a set of*
Players *The firms.*
Actions *Each firm's set of actions is the set of its possible positive prodution levels.*
Preferences *Each firm's preferences are represented by the maximization of its profit function.*

Definition 3 ***Nash equilibrium of a strategic game***. *The action profile* a^* *in a normal form game is a Nash equilibrium if, for every player i and every action* a_i*, according to the preferences of player i,* a^* *is at least as good as the action profile* (a_i, a^*_{-i}) *in which player i chooses* a_i *while every other player j chooses* a^*_j*. Equivalently, for every player i, in terms of payoffs we have that*

$$u_i(a^*) \geq u_i(a_i, a^*_{-i}) \text{ for every action } a_i \text{ of player } i$$

where u_i *is a payoff function that represents player i's preferences.*

The major ingredients of the model are trivial and follow any textbook treatment. The market demand function is linear and decreasing

$$P = p(Q) = \begin{cases} a - b(q_i) & , \quad Q \leq a/b \\ 0 & , \quad \text{otherwise} \end{cases}, \tag{1}$$

with $Q = \sum q_i$ the industry total output, $i = 1, 2, ..n$, and $a, b > 0$.For simplicity assume only $i = 1, 2$. The profit functions for each firm (π_i) are given by

$$\pi_1(q_1, q_2) = \begin{cases} (a - b(q_1 + q_2))q_1 - c_1 q_1 & , \quad q_1 + q_2 \leq a/b \\ -c_1 q_1 & , \quad \text{otherwise} \end{cases}$$

$$\pi_2(q_1, q_2) = \begin{cases} (a - b(q_1 + q_2))q_2 - c_2 q_2 & , \quad q_1 + q_2 \leq a/b \\ -c_2 q_2 & , \quad \text{otherwise} \end{cases}$$

A Nash equilibrium in this Cournot game is an action profile q^* with the property that no firm i can be better off by choosing an action different from q_i^*, given that every other player j chooses q_j^*.

$$\begin{cases} q_1^* = \arg\max_{q_1} \ \pi_1(q_1, q_2) \\ q_2^* = \arg\max_{q_2} \ \pi_2(q_1, q_2), \end{cases} \tag{2}$$

so the best response of each firm is given by

$$f_1(q_2) = \begin{cases} \frac{1}{2b}(a - bq_2 - c_1) & , \quad q_2 \leq (a - c_1)/b \\ 0 & , \quad \text{otherwise} \end{cases}$$

$$f_2(q_1) = \begin{cases} \frac{1}{2b}(a - bq_1 - c_2) & , \quad q_2 \leq (a - c_2)/b \\ 0 & , \quad \text{otherwise} \end{cases}$$

This game has a Nash equilibrium that exists and is unique and, due to elementary level of the problem, no proof is required here. We now introduce incomplete information into the model and present a Nash-Bayesian equilibrium.

Denote the probability assigned by the belief of type t_i of player i to state ω by $\Pr(\omega|t_i)$. Denote also the action taken by each type t_j of each player j by $a(j, t_j)$. Player j's signal in state ω is $\tau_j(\omega)$, so his action in state ω is $a(j, \tau_j(\omega))$. For each state ω, denote by $\hat{a}(\omega)$ the action profile in which each player j chooses the action $a(j, \tau_j(\omega))$. Then the expected payoff of type t_i of player i when she chooses the action a_i is $\sum_{\omega \in \Omega} \Pr(\omega|t_i) u_i((a_i, \hat{a}_{-i}(\omega))\omega)$.

Definition 4 ***A Nash equilibrium of a Bayesian game*** *is a Nash equilibrium of the strategic game (with von Neuman–Morgenstern preferences) defined as follows.*
Players*: The set of all pairs (i, t_i) where i is a player in the Bayesian game and t_i is one of the signals that i may receive.*
Actions*: The set of actions of each player (i, t_i) is the set of actions of player i in the Bayesian game.*
Preferences*: The Bernoulli payoff function of each player (i, t_i) is given by $\sum_{\omega \in \Omega} Pr(\omega|t_i) u_i((a_i, \hat{a}_{-i}(\omega))\omega)$*

In the specific case of our system, we assume that firm 2 ($\mathcal{F}_2$) has complete information about both average costs (c_1, c_2). Assume also that firm 1 ($\mathcal{F}_1$) knows its own cost structure (c_1) but is not sure about the average costs of firm 2, believing that firm 2 may have a lower cost structure (c_2^L) with probability θ, and a higher cost structure (c_2^H) with probability $1-\theta$,

$$c_2 = \begin{cases} c_2^L \ , & \text{with probability } \theta \\ c_2^H \ , & \text{with probability } 1-\theta \end{cases}$$

and assume also that $\mathcal{F}_2$ is unsure about whether $\mathcal{F}_1$ knows or not its true cost structure.

A Bayesian game that models the situation is defined as follows, involving players, states, actions, signals, beliefs, and payoff functions. The players are $\mathcal{F}_1$ and $\mathcal{F}_2$. The states are given by the following set $\{L0, H0, L1, H1\}$, corresponding to the states, respectively: $c_2 = c_2^L$ but $\mathcal{F}_1$ is unsure about $\mathcal{F}_2$ cost structure, $c_2 = c_2^H$ but $\mathcal{F}_1$ is unsure about $\mathcal{F}_2$ cost structure, $c_2 = c_2^L$ and $\mathcal{F}_1$ knows this with certainty, $c_2 = c_2^H$ and $\mathcal{F}_1$ knows this with certainty. Each firm's set of actions is the set of its possible outputs $(q_1, q_2 \geq 0)$. As far as the signals are concerned, we have that $\mathcal{F}_1$ gets one of the signals $[0, L, H]$, and her signal function τ_1 satisfies $\tau_1(L0) = \tau_1(H0) = 0$, $\tau_1(L1) = L$, and $\tau_1(H1) = H$. $\mathcal{F}_2$ gets the signal $[L, H]$ and her signal function τ_2 satisfies $\tau_2(L0) = \tau_2(L1) = L$ and $\tau_2(H0) = \tau_2(H1) = H$. On the side of beliefs, we assume that $\mathcal{F}_1$ *type*-0 assigns probability λ to state $L0$ and probability $1-\lambda$ to state $H0$, $\mathcal{F}_1$ *type-L* assigns probability 1 to state $L1$, $\mathcal{F}_1$ *type-H* assigns probability 1 to state H. As far as firm 2 is concerned, $\mathcal{F}_2$ *type-L* assigns probability λ to state $L1$ and probability $1-\lambda$ to state $L0$, $\mathcal{F}_2$ *type-H* assigns probability λ to state $H1$ and probability $1-\lambda$ to state $H0$. The payoff functions are the firms' Bernoulli payoffs which are given by their profits $(\pi_{1,2})$; if the actions chosen are (q_1, q_2), then $\mathcal{F}_1$ profits are π_1 and for $\mathcal{F}_2$ are π_2 in states $L0$ and $L1$, and π_2^L in states $H0$ and $H1$.

Mathematically we can represent this game with incomplete information by a 5D system. Denote q_0, q_ℓ, q_ℓ the production levels of $\mathcal{F}_1$ for each of the three type signals $[0, L, H]$, and q_L^*, q_H^* the corresponding levels for $\mathcal{F}_2$ for each of the two signals $[L, H]$. A Nash equilibrium is a profile $(q_0^*, q_\ell^*, q_\ell^*, q_L^*, q_H^*)$ for which $q_0^*, q_\ell^*, q_\ell^*$ are best responses to q_L^*, q_H^*, and q_L^*, q_H^* are best responses to $q_0^*, q_\ell^*, q_\ell^*$.

$$\begin{aligned}
&\underset{q_0}{max}[\theta(P(q_0+q_L)-c)q_0+(1-\theta)(P(q_0+q_H)-c)q_0] \\
&\underset{q_\ell}{max}(P(q_\ell+q_L)-c)q_\ell \\
&\underset{q_h}{max}(P(q_h+q_H)-c)q_h \\
&\underset{q_L}{max}[(1-\lambda)\left(P(q_0+q_L)-c_L\right)q_L+\lambda(P(q_\ell+q_L)-c_L)q_L] \\
&\underset{q_H}{max}[(1-\lambda)\left(P(q_0+q_H)-c_H\right)q_H+\lambda(P(q_\ell+q_H)-c_H)q_H]
\end{aligned}$$

As previously mentioned, it has been shown (Cox and Walker,1998) that the Cournot-Bayesian game has the following general properties with constant average costs: (i) there exists a Bayesian–Nash equilibrium if the market demand is linear; (ii) the equilibrium can be learned if the demand is linear; and (iii) if costs are not constant, there is multiple equilibria and learnability is not verified. The third result is the most important one for us here because of the crucial importance that the cost structure imparts upon the type of stability in the Cournot game. What happens to these three results if besides costs being not constant, they have a nonlinear shape as we find in any intermediate microeconomics textbook?

Following the arguments of Cox and Walker, we should consider a fast declining cost structure. In particular, we assume that costs should decline faster than the market demand $P(Q) = a - b\,(q_i)$ .The following parameter values for the market demand are considered $a = 5.58, b = 2$, while the faster declining average cost function is given by

$$c(q_i) = nq_i - m \ln (q_i) \tag{3}$$

with $m = 1.2$ for the three types of cost structures (c_1, c_L, c_H) and n should assume different values in order to differentiate the different cost structures

$$n = \begin{cases} 1.8, \text{ for the } c_2^L \text{ cost structure} \\ 2.0, \text{ for the } c_1 \text{ cost structure} \\ 2.2, \text{ for the } c_2^H \text{ cost structure} \end{cases}$$

In Figure 1 we present the market demand curve and the three cost structures that we have been discussing. The min of $c(q_i)$ for $\{q(c_L), q\,(c_1)\,, q(c_H)\} = \{0.66, 0.6, 0.54\}$.

The Cournot game is a rather simple game leading to rather simple results if the market demand is linear and if the cost structures are also linear — besides obeying the conditions highlighted by Cox and Walker (1998) — even in an incomplete information game in a Bayesian framework. However, when we introduce non linear cost structures into the Bayesian framework, the model becomes extremely complicated. In order to just present some flavour of what one may encounter by such an adventure let us simplify the model by reducing its dimension. Let us assume that $\lambda = 0$ — $\mathcal{F}_2$ does indeed know that $\mathcal{F}_1$ does not know its true cost structure — so that the best response functions are given by

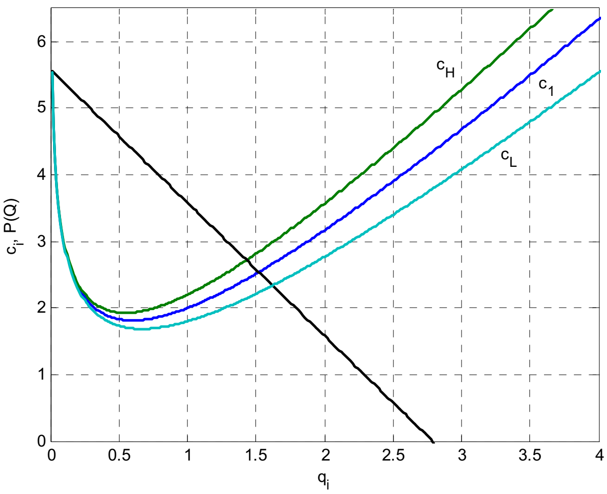

Figure 1: Fast declining cost functions and market demand function

$$
\begin{aligned}
BR_1 &= \begin{cases} a - 2bq_1 - b\left[\theta q_L + (1-\theta)q_H\right] - c_1 & , \quad \theta q_L + (1-\theta)q_H \leq (a - c_1)/b \\ 0 & , \quad \text{otherwise} \end{cases} \\
BR_L &= \begin{cases} (a - bq_1 - 2bq_L - c_L) & , \quad q_1 \leq (a - c_L)/b \\ 0 & , \quad \text{otherwise} \end{cases} \\
BR_H &= \begin{cases} (a - bq_1 - 2bq_H - c_H) & , \quad q_1 \leq (a - c_H)/b \\ 0 & , \quad \text{otherwise} \end{cases}
\end{aligned} \tag{4}
$$

Given the parameter values above presented, these three best response functions lead to a Bayesian-Nash equilibria characterized by eight equilibrium points not extremely far away from each other in the three dimensional space (see next section for further details). If one further restricts the complexity of the game, by assuming for example that firm 1 knows that the true cost structure of firm 2 is c_H $(\theta = 0)$, the system is reduced to two best response functions leading to four Bayesian-Nash equilibrium points (NBE). Now imagine that there is no coordination among the players and also that they are not equipped with the super rational powers to perform all the computations and move directly towards one of the equilibrium points. In this case, it seems questionable that any one of the NBE points would be reached as a simple movement in a one shot game.

But even if one accepts that the agents are endowed with such powerful tools as to be able to perform all the necessary computations and get some coordination on the movement to a particular fixed point, the problem is not so simple because for each individual equilibrium there is a specific payoff which may favour one or the other firm. If both players choose different equilibrium points this leads to a repeated game that can be modelled as a dynamic process. In the next section we present a dynamic adjustment process which overcomes these difficulties and shows two fundamental characteristics: (i) the players are bounded rational in the sense that they do try to maximize profits but are unable to perform all the necessary computations to jump straight to a specific equilibrium point of the game; (ii) they increase/decrease the level of production in response to positive/negative marginal profits; (iii) a very simple learning procedure if adopted by both players (or by just one player) secures convergence to Nash equilibria if the initial conditions of the game are not very far from the equilibrium points under consideration.

3 Bounded learning dynamics

The adaptive process that the two Cournot firms follow in this game has its origins in the the paper by Baumol and Quandt (1964). In the Appendix A of that paper ("Learning Rules of Thumb", the authors argue that a rule of thumb of the type $p_{t+1} = p_t + k(\Delta\pi_t/\Delta p_t)$, where p is the price level, π stands for profits, and k is an adjustment parameter, and t is a time index, as a rule of thumb is "among the more efficient pieces of equipment of optimal decision making" in the maximization of profits by rational firms under incomplete information. Notice that if at time t prices are changing but profits remain constant, one gets a fixed point $p_{t+1} = p_t = p^*$.

But our bounded rational approach to learning is also close to the IGA (Infinitesimal Gradient Ascent) process developed by Singh et al. (2000) and its variants: "WoLF" Win or Learn Fast process by Bowling and Veloso (2002) and the GIGA (Generalized Infinitesimal Gradient Ascent) learning process by Zinkevich (2003). For example, in the WolF process applied in a two-player, two-action, iterated matrix games, it was shown that a bounded rational dynamic process would lead to Nash equilibria. For any two pair of strategies (α, β), the expected payoffs for the row (r) and the column (c) players are given by $V_r(\alpha^k, \beta^k)$ and $V_c(\alpha^k, \beta^k)$.Each player selects an action from $\{1, 2\}$ in the matrix game and let $\alpha \in [0, 1]$ be a strategy for the row player (r), where α corresponds to the probability the player selects the first action and $1 - \alpha$ is the probability the player selects the second action. Similarly, β is a strategy for the column player. Considering the joint strategy (α, β), each player will move its strategy in the direction of the current gradient with some adjustment or step parameter γ. Then the strategies of the kth iteration are given by (α^k, β^k), and these strategies will evolve according to the adaptive rule

$$
\begin{aligned}
\alpha^{k+1} &= \alpha^{k} + \gamma \ell_r^k \frac{\partial V_r(\alpha^k, \beta^k)}{\partial \alpha^k} \\
\beta^{k+1} &= \beta^{k} + \gamma \ell_c^k \frac{\partial V_c(\alpha^k, \beta^k)}{\partial \beta^k}
\end{aligned}
$$

where ℓ represents the variable learning parameter with $\ell_{r,c}^k \in [\ell^{\min}, \ell^{\max}] > 0$.Notice that, once again, if the second term on the right hand side of both equations goes to zero we end up with a stationary point in the game. The interesting issue here is that if ℓ is allowed to change – increase ℓ if you are loosing, decrease it otherwise — such a learning process guarantees converge with probability one to the set of Nash equilibria.

We follow a similar procedure in our deterministic Cournot game. Both players know that in order to maximize profits they have to take into account their best response functions, given respectively by $\frac{\partial \pi_1(t)}{\partial q_1(t)}(q_1(t), q_2(t))$ and $\frac{\partial \pi_2(t)}{\partial q_2(t)}(q_1(t), q_2(t))$. They also know that at the Nash equilibrium the two best responses have to be equal to zero, but while these are negative or positive their profits must be increasing or declining over time. Therefore, when marginal profits are positive, leading to an increase in total profits, each firm has an incentive to increase production, and to reduce production whenever the marginal profits are negative. That is, firms react locally to marginal profits because they can only adjust gradually towards the Nash equilibrium due to the large complexity of the game.

This intuition can be described by the following functions

$$
\begin{aligned}
q_1(t+1) &= \left[1 + \alpha_1 \left(\frac{\partial \pi_1(t)}{\partial q_1(t)}(q_1(t), q_2(t))\right)\right] q_1(t) \\
q_2(t+1) &= \left[1 + \alpha_2 \left(\frac{\partial \pi_2(t)}{\partial q_2(t)}(q_1(t), q_2(t))\right)\right] q_2(t)
\end{aligned} \tag{5}
$$

where α_i represents the speed of adjustment to marginal profits. Notice that if $\frac{\partial \pi_i(t)}{\partial q_i(t)} = 0$,then $q_i(t+1) = q_i(t) = q_i^*$ and we get the Nash equilibrium.

As we will show in the next section, this simple and bounded adaptive process converges to a state close to the Nash equilibrium, cycling around such fixed point, but by itself the full convergence is not secured. However, in a subsequent section we show that if both firms use information from their past actions (going back in time no further then just one period), the convergence is secured and at a very fast rate. Moreover, if one single player wants to guarantee convergence to the Nash equilibrium, while the other player ignores such objective, he can achieve that by using one period lagged information on his actions and on the actions of his rival. Such a simple backward looking information added to the adaptive rule does in fact guarantee successful and fast convergence to the Nash equilibrium.

4 Local and global dynamic analysis

The learning process described by a system like (5) feeded with a nonlinear cost curve as in (3) leads to extremely complicated dynamics. In order to understand the nature of such complexity we need first to present a detailed discussion of the local and global dynamics of our model. For this purpose, we will use some powerful tools from bifurcation analysis. While varying one parameter, a, one may encounter codim 1 bifurcations of fixed points (fold or limit point, flip or period-doubling and Neimark-Sacker), i.e., critical parameter values where the stability of the fixed point changes. Encountering such bifurcation, one may use the formulas for the normal form coefficients derived via the center manifold reduction to analyze the bifurcation (see [**?**]).

The Center Manifold Theorem helps to reduce the dimensionality of the phase space to the dimensionality of the center manifold which in the bifurcation point is tangentially to the eigenspace of the marginal modes of the linear stability analysis. Basically, the center manifold theorem says, that, the dynamics can be projected onto the center manifold without loosing any significant aspect of the dynamics. Moreover, the dynamics projected onto the center manifold can be transformed into the so-called normal forms by a nonlinear transformation of the phase space variables and the normal form coefficients completely describe the codim one bifurcations.

The Limit Point (LP) and Period Doubling (PD) curve for period 1-cycles are computed by Gauss-Newton continuation algorithm applied to minimally extended defining systems, by using a Matlab package, named **cl_ matcont_m**, which was developed by Govaerts et al [**?**]. When a limit point or period doubling point is detected on a curve of fixed points, then the processing of these points includes the computation of the normal form coefficients, giving the complete description of these bifurcations.

4.1 Two-dimensional model

Our two dimensional case corresponds the complete information version of the Cournot game, the simplest version that can be presented for the model, leading nevertheless to not so simple dynamics of the play as we will show here. In the system (4), it corresponds to set $\theta = 0$, that is, firm 1 knows that the cost structure of firm 2 is c_H. In this case the model is given by

$$\begin{cases} q_1(t+1) = q_1(t) + \alpha_1 q_1(t)\left(a - 2bq_1(t) - bq_h(t)\right) - \left(nq_1(t) - m\ln(q_1(t))\right)) \\ q_h(t+1) = q_h(t) + \alpha_2 q_h(t)\left(a - 2bq_h(t) - bq_1(t) - ((n_h q_h(t) - m\ln(q_h(t)))\right)) \end{cases}, \tag{6}$$

where we use the following parameter calibration: $\alpha_1 = 0.7, \alpha_2 = 0.7$, $b = 2.0$, $m = 1.2$, $n = 2$, $n_h = 2.2$ and we let the control parameter a to vary in the interval $[4.0, 8.05]$.

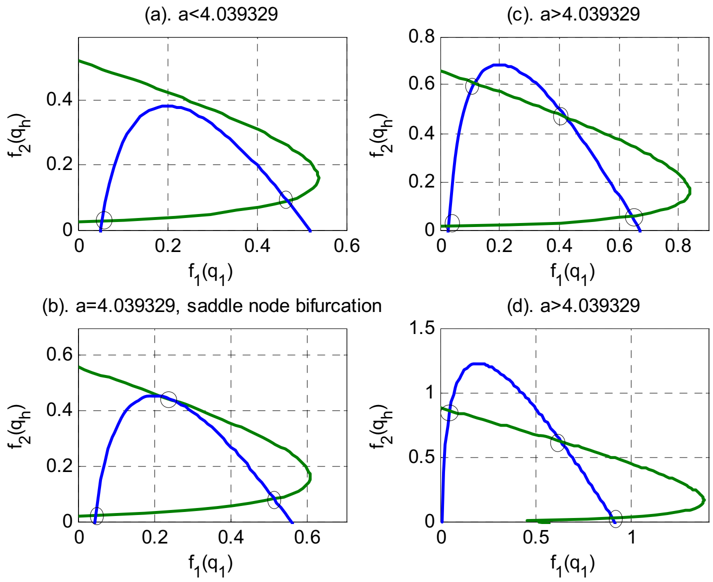

Figure 2: Limit point bifurcation.

Several fixed points, namely 7, 8 or 9, where 2, 3 or 4 are Nash equilibria, can be obtained by applying some numerical algorithm in order to solve the nonlinear system

$$\left(q_1\left(t+1\right), q_h\left(t+1\right)\right) = \left(q_1\left(t\right), q_h\left(t\right)\right).$$

The fixed points that are not Nash equilibria, have no matematical interest, since they are not belonging to the map domain. Initially, for $a = 4$, there are 2 fixed points, all unstable, and by varying the parameter a we obtain the 3^{th} fixed point at $a = 4.03932$, and finally, for any $a > 4.03932$ there are always 4 Nash equilibria. This can be explained by the occurrence of a limit point (saddle-node or fold) bifurcation that take place at the critical value $a_{c1} = 4.03932$ of the parameter a. By computing the normal form coefficient of LP, that is, $C_{LP} = 1.06363 > 0$, we obtain that when $a > 4.03932$ then there are two fixed points: one stable node, (q_1^*, q_h^*), and one unstable saddle,$\left(q_1^{'}, q_h^{'}\right)$; when $a = 4.03932$, there is one critical fixed point with eigenvalue 1, and when $a < 4.03932$, then the saddle and the node collide and disappear. The different scenarios for the existence of Nash equilibria are illustrated in Figure 2.

The fixed point occurring due the limit point bifurcation, is the only one stable, will be denoted by (q_1^*, q_h^*) and represents the nontrivial positive Nash equilibrium that we will study in this paper. By further varying the parameter a the Nash equilibrium loss its stability by a supercritical period doubling bifurcation that takes place at the critical value $a_{c2} = 4.88334$. Thus, (q_1^*, q_h^*) is stable for $4.03932 < a < 4.88334$ and unstable for $a > 4.88334$. At the critical parameter value $a = 4.88334$, the fixed point has multiplier $\lambda = -1$. By verifying the nondegeneracy conditions ([**?**], [**?**]), namely computing the normal form coefficient of PD, that is, $C_{PD} = 1.140270e + 001 \neq 0$, we conclude that a unique and stable period-two cycle bifurcates from $(q_1^*, q_h^*) = (0.52987, 0.47097)$ for $a > 4.88334$. The fate of this period-two cycle can be traced further. It can be verified numerically that this cycle losses stability via another flip bifurcation given rise to a stable period-four cycle, that bifurcates again generating a stable period 8 cycle, and so on generating an infinite sequence of bifurcations values, that finally results in an chaotic attractor at the Misiurewicz point, that is for $a = 5.5285$. The chaotic attractor permanence for any values of a in the interval $[5.5285, 5.59]$ and if we increase a slightly from 5.59 the attractor disappears in a boundary crises bifurcation. Moreover, for $a = 6.60094$, the unstable Nash equilibrium,$(q_1^*, q_h^*) = (0.80843, 0.74757)$, becomes a neutral saddle (NS), that is, a saddle with zero sum of eigenvalues which is as attracting as repelling and that finally bifurcates in a totally unstable fixed point, for $a = 8.03414$, as can be seen in Figure 3.

The saddle fixed point obtained due the limit point bifurcation, $\left(q_1^{'}, q_h^{'}\right) = (0.66254, 0.10519)$, will continue unstable until one of its eigenvalues will cross the unitary circle for $a = 4.679651$, generating a unstable period doubling bifurcation that will diverge to infinity after few time steps.

Figure 3 summarize the dynamical behavior of the two significant Nash equilibria, (q_1^*, q_h^*) and $\left(q_1^{'}, q_h^{'}\right)$. The computation of this equilibrium curves gives the dependence of an equilibrium on the parameter.a. The problem of computing the equilibrium curve is a specific case of the general finite-dimensional continuation problem. When the control parameter crosses the critical value corresponding to a limit point bifurcation, two fixed points of the map collide and disappear and when the control parameter crosses the critical value corresponding to a period doubling bifurcation a cycle of period 2 bifurcates from the fixed point.

In the particular case that we will use in the next section (where we apply a delayed feedback control technique), that is for $\alpha_1 = 0.7; \alpha_2 = 0.7$; $b = 2.0$; $m = 1.2$; $n = 2$; $n_h = 2.2$; and $a = 5.58$; the two-dimensional map takes the more simple form

$$\begin{cases} q_1(t+1) = q_1(t) + 0.7q_1(t)(5.58 - 6.0q_1(t) - 2.0q_h(t) + 1.2\ln(q_1(t))) \\ q_h(t+1) = q_h(t) + 0.7q_h(t)(5.58 - 6.2q_h(t) - 2.0q_1(t) + 1.2\ln(q_h(t))) \end{cases},$$

and by solving the nonlinear system $(q_1(t+1), q_h(t+1)) = (q_h(t), q_1(t))$ we obtain the following 4 Nash equilibria, where 3 are unstable sources with two eigenvalues with modulus greater than one and one is a saddle:

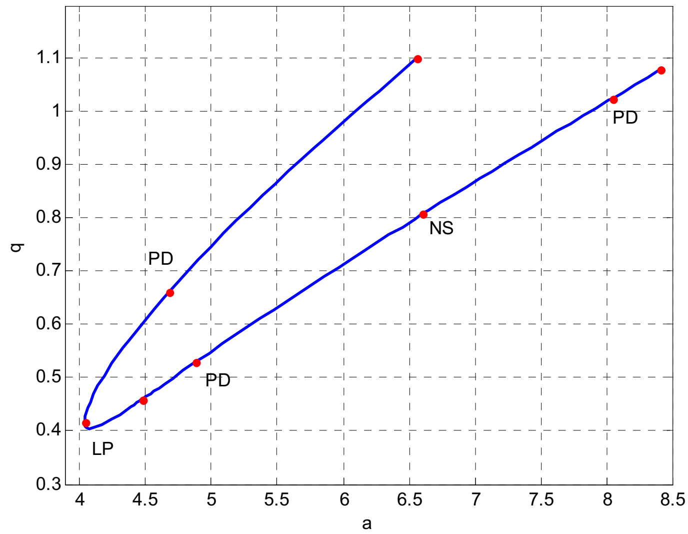

Figure 3: Continuation curve

$$\begin{cases} (q_1, q_h) = (0.01023, 0.01025)\,, & (q_1, q_h) = (0.05113, 0.85265) \\ (q_1^*, q_h^*) = (0.64636, 0.58906)\,, & (q_1', q_h') = (0.88742, 0.05606) \end{cases}$$

The asymptotic behavior of the system is given by the chaotic attractor presented in Figure 4. As we can easily observe the time series for both actions q1 and qH do not show full convergence to the Nash equilibrium point $(q_1^*, q_h^*) = (0.64636, 0.58906)$, neither do they exhibit divergence from that point, gravitating instead around this Nash equilibrium.

We obtain the values of the equilibria and proceed to the local stability analysis of each one of the fixed points by using advanced numerical algorithms available in Matlab software, since there is no possibility to obtain specific formulas in this case.

4.2 Three-dimensional model

The three dimensional case corresponds to the system (4), where for simplicity it assumed that $\mathcal{F}_2$ does indeed know that $\mathcal{F}_1$ does not know its true cost structure ($\lambda = 0$).

In this case, we use the following parameter calibration: $\alpha_1 = 0.7; \alpha_2 = 0.7; \alpha_3 = 0.7$; $b = 2.0$; $m = 1.2$; $n = 2$; $n_h = 2.2$; $n_l = 1.8$, and we let the

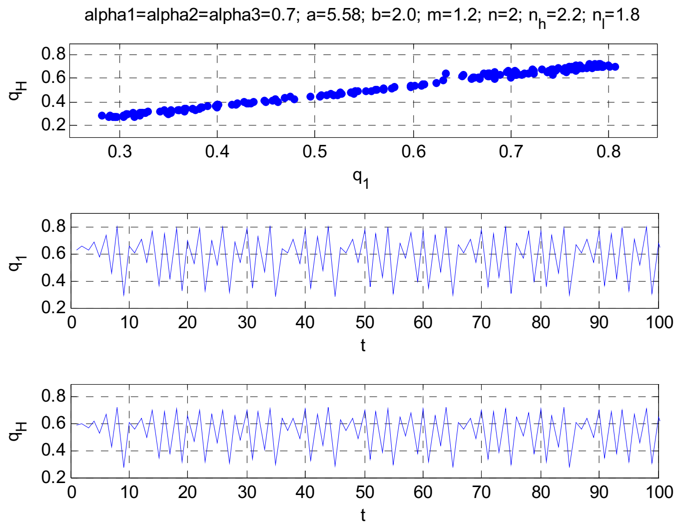

Figure 4: Chaotic attractor and time series for $a = 5.58$.

control parameter a to vary in the interval $[3.9, 6]$.The three-dimensional model takes the form

$$\begin{cases} q_1(t+1) = q_1(t) + 0.7q_1(t)(5.58 - 6q_1(t) - 1.6q_l(t) - 0.4q_h(t) + 1.2\ln(q_1(t))) \\ q_h(t+1) = q_h(t) + 0.7q_h(t)(5.58 - 6.2q_h(t) - 2q_1(t) + 1.2\ln(q_h(t))) \\ q_h(t+1) = q_l(t) + 0.7q_l(t)(5.58 - 5.8q_l(t) - 2q_1(t) + 1.2\ln(q_l(t))) \end{cases}.$$

Several fixed points, namely 25, 26 or 27, where 6, 7 or 8 are Nash equilibria,can be obtain by applying some numerical algorithm to solve the nonlinear systems

$$(q_1(t+1), q_h(t+1), q_l(t+1)) = (q_1(t), q_h(t), q_l(t)).$$

The fixed points that are not Nash equilibria, have no matematical interest, since they are not belonging to the map domain. Initially, for $a = 3.9$, there are 6 fixed points, all unstable, and by varying the parameter a we can obtain the 7^{th} fixed point for $a = 3.99103$. For any $a > 3.99103$ there are always 8 Nash equilibria. This is explained by the occurrence of a limit point (LP) bifurcation that take place at the critical value $a_{c1} = 3.99103$ of the control parameter a. By computing the normal form coefficient of the limit point bifurcation, that is, $C_{LP} = 7.830251e - 001 > 0$, we obtain that for $a > 3.991039$ there are two fixed points, one stable node,(q_1^*, q_h^*, q_l^*), and one unstable saddle,$\left(q_1^{'}, q_h^{'}, q_l^{'}\right)$, at $a = 3.99103$, there is one tangential critical

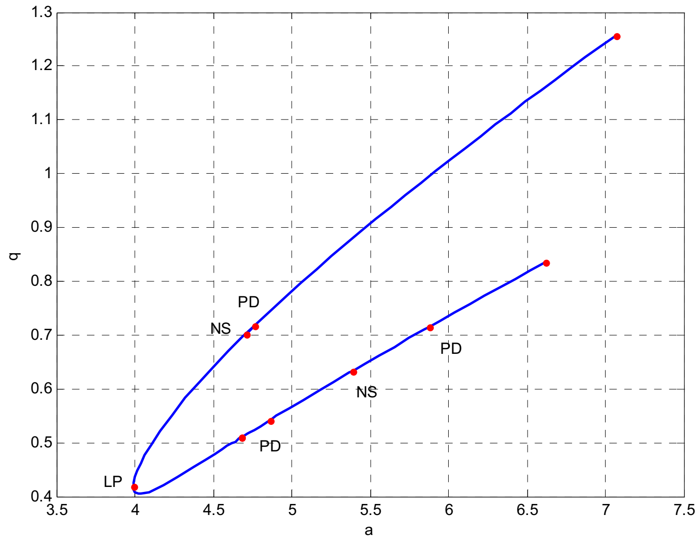

Figure 5: Continuation curve

fixed point with eigenvalue 1, and for $a < 3.99103$, the saddle and the node points collide and disappear. The different scenarios for the existence of Nash equilibria are illustrated Figure 5.

The fixed point occurring due the limit point bifurcation is the only one stable, represents the positive significative Nash equilibrium that we will study in this paper and will be denoted by (q_1^*, q_h^*, q_l^*). By further varying the parameter a the Nash equilibrium loss its stability by a supercritical period doubling (flip) bifurcation that takes place at the critical value $a_{c2} = 4.859617$. Thus, (q_1^*, q_h^*, q_l^*) is stable for $3.99103 < a < 4.85961$ and unstable for $a > 4.85961$. At the critical parameter value $a = 4.85961$, the fixed point has multiplier $\lambda = -1$. By verifying the nondegeneracy conditions ([?], [?]), namely the normal form coefficient of PD, that is, $C_{PD} = 7.73667 \neq 0$, we conclude that a unique and stable period-two cycle bifurcates from $(q_1^*, q_h^*, q_l^*) = (0.49060, 0.48582, 0.54195)$ for $a > 4.88334$. The fate of this period-two cycle can be traced further. It can be verified numerically that this cycle losses stability via another flip bifurcation given rise to a stable period-four cycle, that bifurcates again generating a stable period 8 cycle, and so on generating an infinite sequence of bifurcations values, that finally results in an chaotic attractor after the Misiurewicz point, that is for $a = 5.514$. The chaotic attractor permanence for any values of a in the interval $[5.514, 5.59]$ and if we increase a slightly from 5.59 the attractor disappears in a boundary crises bifurcation. Moreover, for $a = 5.38771$,

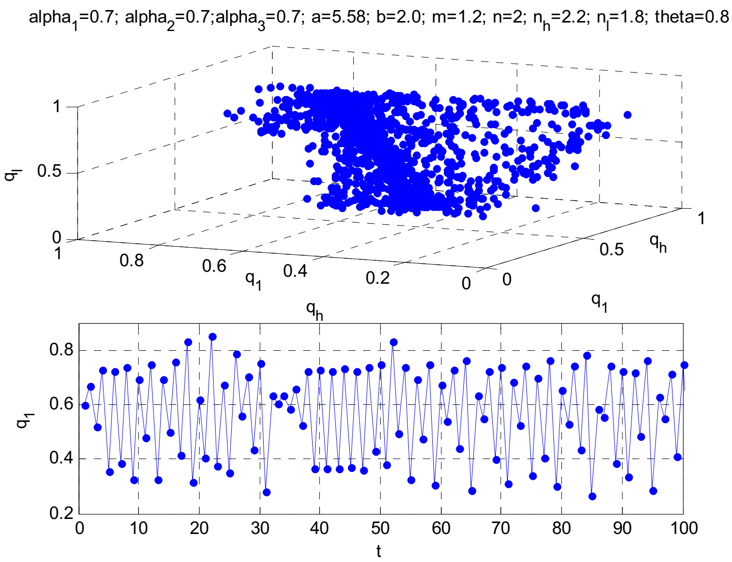

Figure 6: Chaotic attractor and time series for $a = 5.58$

the unstable Nash equilibrium,$(q_1^*, q_h^*, q_l^*) = (0.58283, 0.57329, 0.63349)$, becomes a neutral saddle (NS), that is, a saddle with zero sum of eigenvalues which is as attracting as repelling and finally bifurcates in an totally unstable fixed point, for $a = 5.87829$, as can be seen in Figure 5. Finally a new neutral saddle take place at $(0.78522, 0.76911, 0.84052)$ for $a = 6.65396$.

The saddle fixed point obtained by the limit point bifurcation, $\left(q_1^{'}, q_h^{'}, q_l^{'}\right)$, becomes a neutral saddle $(0.10574, 0.63890, 0.70264)$ for $a = 4.71030$ and finally for $a = 4.76266$ one of its eigenvalues will cross the unitary circle, generating a unstable period doubling bifurcation that will diverge to infinity after few time steps.

Figure 5 summarize the dynamical behavior of the two significant Nash equilibria, (q_1^*, q_h^*, q_l^*) and $\left(q_1^{'}, q_h^{'}, q_l^{'}\right)$. The computation of this equilibrium curves gives the dependence of an equilibrium on the parameter.a. The problem of computing the equilibrium curve is a specific case of the general finite-dimensional continuation problem. When the control parameter crosses the critical value corresponding to a limit point bifurcation, two fixed points of the map collide and disappear and when the control parameter crosses the critical value corresponding to a period doubling bifurcation a cycle of period 2 bifurcates from the fixed point.

We fix the control parameter a to assume the value $a = 5.58$ for which the asymptotic behavior of the system is given by the chaotic attractor presented in Figure 6. Now, by solving the nonlinear system

$$5.58 - 6q_1(t) - 1.6q_l(t) - 0.4q_h(t) + 1.2\ln(q_1(t))) = 0$$
$$5.58 - 6.2q_h(t) - 2q_1(t) + 1.2\ln(q_h(t))) = 0$$
$$5.58 - 5.8q_l(t) - 2q_1(t) + 1.2\ln(q_l(t))) = 0$$

we obtain the following 8 Nash equilibria, where 5 are unstable sources and 3 are saddle fixed points :

$$(q_1, q_h, q_l) = (0.0408, 0.0108, 0.9337), (q_1, q_h, q_l) = (0.0102, 0.0102, 0.0102)$$
$$(q_1^*, q_h^*, q_l^*) = (0.6148, 0.6040, 0.6659), \left(q_1', q_h', q_l'\right) = (0.0584, 0.8496, 0.9260)$$
$$(q_1, q_h, q_l) = (0.0138, 0.8681, 0.0102), (q_1, q_h, q_l) = (0.8878, 0.0561, 0.0547)$$
$$(q_1, q_h, q_l) = (0.8522, 0.4850, 0.0505), (q_1, q_h, q_l) = (0.6829, 0.0359, 0.6314)$$

The values of the equilibria and the study of the local stability of each one of the fixed points was obtained by using advanced numerical algorithms available in Matlab software. The Nash equilibrium (q_1^*, q_h^*, q_l^*) represents the main fixed point in term of stabilization by delay feedback control, since is the most relevant from dyamical point of view, as has been showed above.

5 Convergence to the Nash equilibrium

As we saw in the previous section, in both the two and three dimensional cases, the dynamics of play do not show full convergence to the Nash equilibrium points, neither do they diverge from such points. In fact, they gravitate around those points. We can exploit this property of the dynamics to force the system to move towards the Nash equilibrium by a small perturbation to the initial bounded adaptive process. All that is required is that either both players make use of current and of a one period lagged information about their own actions — or just one player can do so by using information concerning his own actions and the actions of his rival — in order to smooth his reaction to the ups and downs of the outputs of both firms over time. That is, player i can achieve convergence by using an adaptive real coefficient p for that purpose, such that $p(q_i(t) - q_i(t-1))$ turns out to be the perturbation required to secure convergence to the Nash equilibrium.

This method follows Pyragas [27] who proposed a process called delayed feedback control (DFC) in which the control input is fed by the difference between the current state and the delayed state. The delay time is determined as the period of the unstable periodic orbit to be stabilized. Hence, the control input vanishes when the unstable periodic orbit is stabilized. In addition, this method requires no preliminary calculation of the unstable periodic orbit.

However, it has been reported that discrete-time DFC has a limitation [33]. That is, DFC never stabilizes an unstable fixed point of a chaotic discrete-time system (continuous-time system), if the Jacobian matrix of the linearized system around the unstable fixed point has an odd number of real eigenvalues greater than unity. This property is called the odd number limitation.

To overcome the odd number property, several methods have been proposed (see for example, [23], [34], [17], between others). In this paper we closely follows the necessary and sufficient condition for stabilizability of discrete-time systems via delayed feedback control presented by Zhu and Tian (Theorem 2 in [37]).

For the nonlinear system

$$x(t+1) = f(x(t), u(t)), \; u(t) \in \mathbb{R}, \; x(t) \in \mathbb{R}^n, \; f: \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$$

let $x^* \in \mathbb{R}^n$ be a fixed point, that is a solution to the system $x^* = f(x^*, 0)$. We denote by $A = J_x f(x^*, 0)$ and $B = J_u f(x^*, 0)$ the Jacobian matrices of the map f in order to the variables x and u.

The linearized system of the map f around x^* is given by

$$y(t+1) = Ay(t) + Bu(t) \tag{7}$$

where $y(t) = x(t) - x^*$ and

$$u(t) = p(y(t) - y(t-1)) \tag{8}$$

Theorem 5 *Assume (A, B) is controllable. There exists delayed feedback control (8) such that the closed-loop system composed by (7) and (8) is asymptotically stable if and only if*

$$0 < \det(I_n - A) < 2^{n+1}$$

where I_n is the identity matrix of order n.

An important question that lies at the heart of control using state space models is whether we can steer the state via the control input to certain locations in the state space, or if the system is controllable. We recall that the pair (A, B) is controllable if

$$C = \begin{bmatrix} B \; AB \; A^2B \; ...A^{n-1}B \end{bmatrix}$$

has full rank n (that is $\det(C) \neq 0$), where n is the order of the matrix A.

We also have to specify that all these results are presented for matrices A and B given in the following controllable canonical form

$$A_C = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ a_1 & a_2 & a_3 & \dots & a_{n-1} & a_n \end{bmatrix}, \; B_C = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

where $a_i, \; i = 1, ..., n$ are the coefficients of the characteristic polynomial of initial Jacobian matrix A (for more details see [7]). The above model has a special form and any completely controllable system can be expressed in this way.

5.1 Case 1: Two-dimensional model

Let us now concentrate on the unstable chaotic Nash equilibrium

$$(q_1^*, q_h^*) = (0.64636, 0.58906),$$

with the goal to stabilize it by delay feedback control. In Figure 4 we showed the chaotic attractor and the time series associated with the variable q_1 when the parameter $a = 5.58$. We will demonstrate that the irregular behavior of the play can be easily changed and forced to converge to the Nash equilibrium point.

The Jacobian matrix of the system calculated for the Nash equilibrium is given by

$$A = J_x f\,(q_1^*, q_h^*, 0) = \begin{bmatrix} -0.87474 & -0.90491 \\ -0.82469 & -0.71654 \end{bmatrix}$$

and the canonical form of the linearized closed-loop system is

$$A_c = \begin{bmatrix} 0 & 1 \\ 0.11948 & -1.59128 \end{bmatrix}, \; B_c = J_u f\,(q_1^*, q_h^*, 0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

It is easy to check that the pair (A_c, B_c) is controllable since the matrix

$$[B_c \;\; A_c B_c] = \begin{bmatrix} 0 & 1 \\ 1 & -1.59128 \end{bmatrix}$$

has maximum rank $r = 2$. We can also obtain immediately that the eigenvalues of matrix A_c, are given by $\lambda_1 = -1.66312, \; \lambda_2 = 0.07184$ which means that we are in the presence of the odd number limitation (if we are not in the presence of the odd number limitation the system can be controlled by simple one-period feedback delay control). This implies that we can not stabilize the fixed point by the simpler delay feedback control, but since

$$\begin{aligned} 0 &< \det(I_2 - A) = \begin{vmatrix} 1.87474 & 0.90491 \\ 0.82469 & 1.71654 \end{vmatrix} = \\ &= \det(I_2 - A_c) = \begin{vmatrix} 0 & 1 \\ 0.11948 & -1.59128 \end{vmatrix} = 2.4718 < 2^3 = 8 \end{aligned}$$

we can apply the necessary and sufficient condition of stabilizability from [37] in order to obtain DFC for the Nash equilibrium point in our model.

In what follows we have to determine the expression of the DFC that stabilize $(q_1^*, q_h^*) = (0.64636, 0.58906)$. We choose some $\eta = 3.5$ such that

$$2.4718 = \det(I_2 - A_c) < \eta < 2^3 = 8$$

holds. We compute

$$\theta = 1 - \sqrt[n]{\frac{\eta}{2}} = -0.32288$$

and

$$\beta = 1 - \frac{2\det(I_2 - A_c)}{\eta} = -0.41246$$

in order to obtain the following polynomial

$$\begin{aligned} d(s) &= (s-\theta)^n (s-\beta) = (s+0.32288)^2 (s+0.41246) = \\ &= s^3 + 1.0582s^2 + 0.3706s + 0.04300 \\ &= s^3 + d_2 s^2 + d_1 s^1 + d_0 \end{aligned}$$

Now, we can write down the DFC rule, that is

$$\begin{aligned} p_j &= -\sum_{i=j}^{n} (a_i + d_i), \; j = 1, 2, ..., n \\ p_1 &= -(a_1 + a_2 + d_1 + d_2) = 0.043 \\ p_2 &= -(a_2 + d_2) = 0.53308 \\ u(t) &= \sum_{i=1}^{n} p_i [x_i(t) - x_i(t-1)] = p_1 y_1(t) + p_2 y_2(t) = \\ &= 0.043 (q_1(t) - q_1(t-1)) + 0.53308 (q_h(t) - q_h(t-1)) \end{aligned}$$

Finally the closed loop controlled system has the form

$$\begin{cases} q_1(t+1) = q_1(t) + 0.7q_1(t)(5.58 - 6.0q_1(t) - 2.0q_h(t) + 1.2\ln(q_1(t))) + \\ \quad +0.043(q_1(t) - q_1(t-1)) + 0.53308(q_h(t) - q_h(t-1)) \\ q_h(t+1) = q_h(t) + 0.7q_h(t)(5.58 - 6.2q_h(t) - 2.0q_1(t) + 1.2\ln(q_h(t))) \end{cases}.$$

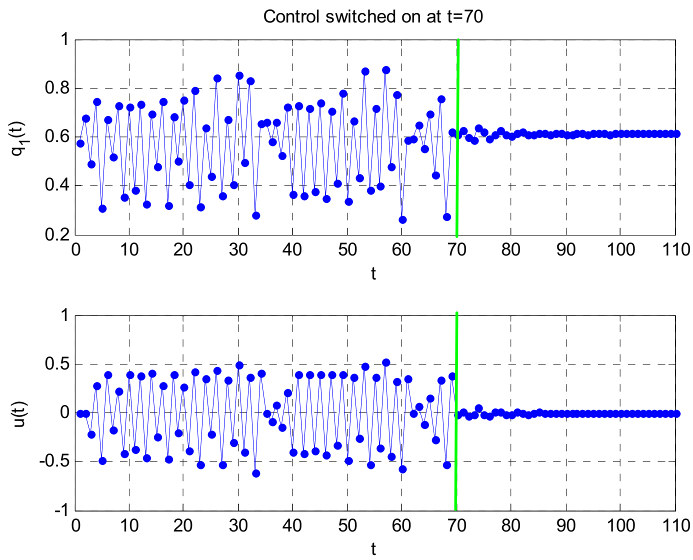

Figure 7: Learning to converge to Nash equilibium.

Since stabilization is guaranteed only in a neighborhood of the fixed point, we adopt the following small control law proposed by Pyragas

$$u_s(t) = \begin{cases} u(t) & \text{if } \; u(t) < \varepsilon \\ 0 & \text{otherwise} \end{cases}$$

where ε is a small positive number. Figure 7 shows the stabilization of the Nash equilibrium point (for the variable q_1) and the convergence to zero of the control input term. The delay feedback control was switched on at $t = 70$ and it takes around 10 time steps to steer down at the fixed point value ($q_1^* = 0.64636$).

We see that controllability and convergence to the Nash equilibrium is a black and white issue: a model either is completely controllable or is not. Clearly, to know that something is uncontrollable is a valuable piece of information. However, to know that something is controllable really tells us nothing about the degree of controllability, i.e., about the difficulty that might be involved in achieving a certain objective. If a system is not completely controllable, it can be decomposed into a controllable and a completely uncontrollable subsystem.

5.2 Case 2: Three-dimensional model

We have the following linearization for the original system in the neighborhood of the Nash equilibrium point $(q_1^*, q_h^*, q_l^*) = (0.61488, 0.60409, 0.66591)$

$$\begin{aligned}
y(t+1) &= Ay(t) + Bu(t) \\
A &= \begin{bmatrix} -0.7425147 & -0.1721676 & -0.6886705 \\ -0.8457374 & -0.7817862 & 0 \\ -0.9322852 & 0 & -0.8636272 \end{bmatrix}, \ B = B_c = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \\
y(t) &= \begin{bmatrix} q_1(t) - q_1^* \\ q_h(t) - q_h^* \\ q_l(t) - q_l^* \end{bmatrix}, \ u(t) = [p_1 y_1(t) + p_2 y_2(t) + p_3 y_3(t)]
\end{aligned}$$

and now we consider the canonical decomposition for the matrix A, denoted by A_c

$$A_c = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.1263625 & -1.109269 & -2.387928 \end{bmatrix}$$

It is easy to check that (A_c, B_c) is controllable, since

$$C = [B_c \ \ A_c B_c \ \ A_c^2 B_c] = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & -2.38792 \\ 1 & -2.38792 & 4.59293 \end{bmatrix}$$

is triangular and has full rank 3.

We compute the eigenvalues of the matrix A_c and we obtain that $\lambda_1 = -1.68518$, $\lambda_2 = 0.09410$ and $\lambda_3 = -0.79684$ which means that we are in the presence of the odd number limitation (one real eigenvalue of modulus greater than one). This implies that we can not stabilize the fixed point by the simple delay feedback control, but since

$$0 < \det(I_3 - A_c) = \begin{vmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -0.12636 & 1.10926 & 3.38792 \end{vmatrix} = 4.3708 < 2^3 = 8$$

we can apply the necessary and sufficient condition of stabilizability from [37] in order to obtain DFC.

In what follows we have to determine the expression of the DFC that stabilize (q_1^*, q_h^*, q_l^*). We choose some $\eta = 5$ such that

$$4.3708 = \det(I_3 - A_c) < \eta < 2^4 = 16$$

holds and we compute

$$\theta = 1 - \sqrt[n]{\frac{\eta}{2}} = -0.35721$$

and

$$\beta = 1 - \frac{2\det\left(I_3 - A_c\right)}{\eta} = -0.74832$$

in order to get a stable polynomial defined as

$$\begin{aligned} d\left(s\right) &= \left(s-\theta\right)^n\left(s-\beta\right) = \left(s+0.35721\right)^3\left(s+0.74832\right) = \\ &= s^4 + 1.8200s^3 + 1.1847s^2 + 0.33203s + 0.03410 \\ &= s^4 + d_3s^3 + d_2s^2 + d_1s^1 + d_0 \end{aligned}$$

Finally, we can obtain the DFC

$$\begin{aligned} p_j &= -\sum_{i=j}^{n}\left(a_i + d_i\right),\ j = 1, 2, ..., n \\ p_1 &= -\left(a_1 + a_2 + a_3 + d_1 + d_2 + d_3\right) = 0.034105 \\ p_2 &= -\left(a_2 + a_3 + d_2 + d_3\right) = 0.49250 \\ p_3 &= -\left(a_3 + d_3\right) = 0.56793 \\ u\left(t\right) &= p_1y_1\left(t\right) + p_2y_2\left(t\right) + p_3y_3\left(t\right) = \\ &= p_1\left(q_1\left(t\right) - q_1\left(t-1\right)\right) + p_2\left(q_h\left(t\right) - q_h\left(t-1\right)\right) + p_l\left(q_3\left(t\right) - q_l\left(t-1\right)\right) = \\ &= 0.03410\left(q_1\left(t\right) - q_1\left(t-1\right)\right) + 0.49250\left(q_h\left(t\right) - q_h\left(t-1\right)\right) + \\ &\quad +0.56793\left(q_l\left(t\right) - q_l\left(t-1\right)\right) \end{aligned}$$

and the closed loop controlled system takes the form

$$\left\{\begin{array}{l} q_1\left(t+1\right) = q_1\left(t\right) + 0.7q_1\left(t\right)\left(5.58 - 6q_1\left(t\right) - 1.6q_l\left(t\right) - 0.4q_h\left(t\right)\right. \\ \left.+1.2\ln(q_1\left(t\right))\right) + 0.03410\left(q_1\left(t\right) - q_1\left(t-1\right)\right) + 0.49250\left(q_h\left(t\right) - q_h\left(t-1\right)\right) \\ +0.56793\left(q_l\left(t\right) - q_l\left(t-1\right)\right) \\ q_h\left(t+1\right) = q_h\left(t\right) + 0.7q_h\left(t\right)\left(5.58 - 6.2q_h\left(t\right) - 2q_1\left(t\right) + 1.2\ln(q_h\left(t\right))\right) \\ q_h\left(t+1\right) = q_l\left(t\right) + 0.7q_l\left(t\right)\left(5.58 - 5.8q_l\left(t\right) - 2q_1\left(t\right) + 1.2\ln(q_l\left(t\right))\right) \end{array}\right. .$$

Since stabilization is guaranteed only in a neighborhood of the fixed point, we adopt the following small control law proposed by Pyragas

$$u_s\left(t\right) = \left\{\begin{array}{cl} u\left(t\right) & \text{if } \ u\left(t\right) < \varepsilon \\ 0 & \text{otherwise} \end{array}\right.$$

where ε is a small positive number. Figure 8 shows the stabilization of the Nash equilibrium point (for the variable q_1 and q_h). The delay feedback control was switched on at $t = 70$ and it takes less than 10 time steps to steer down at the fixed point value $(q_1^*, q_h^*, q_l^*) = (0.61488, 0.60409, 0.66591)$.

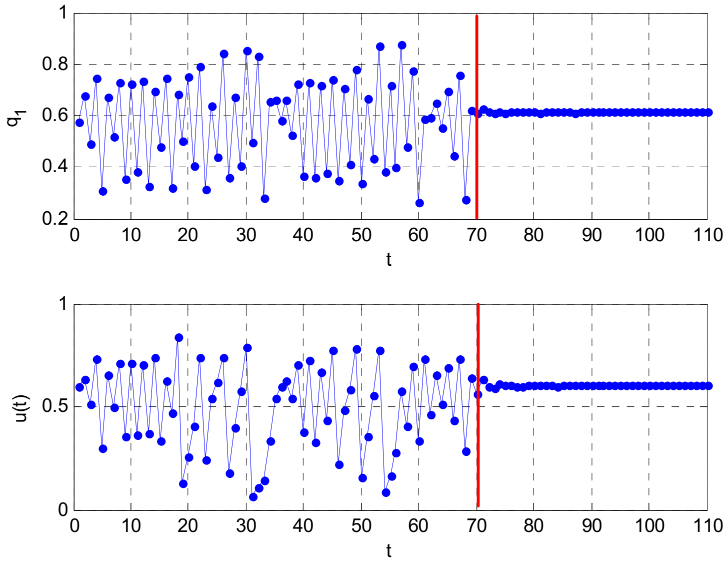

Figure 8: Learning to converge to Nash equilibium.

6 Concluding remarks

In this paper we provided a counterexample to the well accepted results concerning the lack of convergence of simple adaptive rules to Nash equilibria in a certain class of games. As it is widely known, it has been argued that if a simple adaptive rule obeys three basic conditions — (i) it is uncoupled, (ii) each player's choice of action depends solely on the frequency distribution of past play (as in fictitious play), and (iii) each player's choice of action, conditional on the state, is deterministic — there seems to exist no such rule that renders the actions of the players to secure, period by period, convergence to the Nash equilibrium.

This argument has been substantiated in some recent very influential papers, even if the former may seem to go against common intuition. Building upon this basic intuition, we take the three conditions presented as a delimitation criteria for whether the Nash equilibrium can be learned or not by using simple adaptive rules, and put one of the most simple games that has been used in game theory (the Cournot model) to the test. We provide a counterexample that clearly violates those three general conditions above; the game is uncoupled, because the strategies depend only upon the other player's actions (not upon the opponents profit function); each player's choice of action depends solely on the past play, and each player's choice of action is entirely deterministic. Nevertheless, convergence to the Nash equilibria is secured and a very fast rate, and in order to do so players have to go back in time collecting information about their own actions no further than at $t-1$.

Finally, we should add a note of caution. We claim no generality in the results presented here. It is a mere counterexample, but as a conjecture we reason that if the dynamics of any game are ergodic, we expect similar kind of results to apply in these circumstances.

Acknowledgement 6 *Financial support from the Fundação Ciência e Tecnologia, Lisbon, is grateful acknowledged, under the contract No POCTI/ ECO /48628/ 2002, partially funded by the European Regional Development Fund (ERDF).*

References

[1] Baumol, W. and Quandt, R. H. (1964). Rules of Thumb and Optimally Imperfect Decisions, *American Economic Review*, Vol. 54, No. 2, 23-46

[2] Blackwell, D. (1956). Controlled random walks. In *Proceedings of the International Congress of Mathematicians*, volume 3, p.336–338. North-Holland.

[3] Bowling, M. and Veloso, M. (2002). Multiagent Learning Using a VariableLearning Rate, *Artificial Intelligence* , 136:215–250, 2002

[4] Brown, G. (1951). Iterative solution of games by fictitious play. In *Activity Analysis of Production and Allocation.* John Wiley & Sons, New York.

[5] Camerer, C. (2003). *Behavioral Game Theory*, Princeton University Press.

[6] Cesa-Bianchi, N. and Lugosi, G. (2006). *Prediction and Learning in Games*, Cambridge University Press.

[7] Chen, C-T. (1999). *Linear System Theory and Design*, 3rd edition, Oxford University Press.

[8] Cox, J. C. and Walker, M. (1998). Learning to play Cournot duopoly strategies, *Journal of Economic Behavior & Organization*, Vol. 36 (1998) 141-161

[9] Foster, D. P. and Vohra, R. (2006). Regret testing: learning to play Nash equilibrium without knowing you have an opponent, *Theoretical Economics,* 1 , 341–367

[10] Foster, D. P. and Vohra, R. (1999). Regret in the on-line decision problem. *Games and Economic Behavior*, 29:7–36.

[11] Foster, D. P. and Vohra, R. (1997). Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21:40–55.

[12] Foster, D. P. and Vohra, R. (1998). Asymptotic calibration. *Biometrika*, 85:379–390.

[13] Foster, D. P. and Young, H. P. (2003). Learning, hypothesis testing, and Nash equilibrium. *Games and Economic Behavior*, 45:73–96.

[14] Fudenberg, D. and Levine, D. K. (1998). *The Theory of Learning in Games*. MIT Press, Cambridge, MA.

[15] Germano, F. and Lugosi, G., 2007. Global convergence of Foster and Young´s regret testing. *Games and Economic Behavior* 60, 135-154.

[16] Govaerts, W., R. K. Ghaziani, YU. A. Kuznetsov and H. G. E. Meijer, Numerical methods for two-parameter local bifurcation analysis of maps, *SIAM J. Sci. Comput*, Vol. 29, No. 6 (2007), pp. 2644–2667

[17] Jiang, Guo-Ping and Zheng, Wei Xing (2005). A simple method of chaos control for a class of chaotic discrete-time systems, *Chaos, Solitons & Fractals*, Volume 23, Issue 3, February , Pages 843-849.

[18] Hart, S. (2005). Adaptive Heuristics, *Econometrica*, Vol. 73, No. 5 (September, 2005), 1401–1430

[19] Hart, Sergiu and Andreu Mas-Colell (2003), "Uncoupled dynamics do not lead to nash equilibrium." *American Economic Review*, 93, 1830–1836.

[20] Hart, S. and Mas-Colell, A. (2000). A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68:1127–1150.

[21] Hannan, J. F. (1957). Approximation to Bayes risk in repeated plays. *Contributions to the Theory of Games*, 3:97–13

[22] Kalai, E. and Lehrer, E. (1993). Rational learning leads to Nash equilibrium. *Econometrica*, 61:1019–1045.

[23] Konishi, K. and Kokame, H. (1998). "Observer-based delayed-feedback control for discrete-time chaotic systems," *Phys. Lett. A*, 248, pp.359–368.

[24] Kuznetsov, Y. A. *Elements of applied bifurcation theory*. Second edition. Applied Mathematical Sciences, 112. Springer-Verlag, New York, (1998).

[25] Leslie, D. and Collins, E. (2006). Generalised weakened fictitious play, *Games and Economic Behavior*, Volume 56, Issue 2, August 2006, Pages 285-298

[26] Owen, G. (1995). *Game Theory*, Academic Press, London.

[27] Pyragas, K. (1992). Continuous Control of Chaos by Self-Controlling Feedback, *Phys. Lett. A* 170, 421-427..

[28] Rubinstein, A. (1998). *Modelling Bounded Rationality*, MIT Press

[29] Shapley, L. S. (1964). Some topics in two person games. In Dresher, M., Shapley, L. S., and Tucker, A.W., editors, *Advances in Game Theory*, volume 52 of Annals of Mathematics Studies, pages 1–28. Princeton University Press

[30] Sandholm, W. (2008). *Population Games and Evolutionary Dynamics*, MIT Press.

[31] Sandholm, W. (2007). Evolution and Learning in Games: Overview, mimeo, University of Wisconsin

[32] Singh, s., M. Kearns, Y. Mansour, Nash convergence of gradient dynamics in general-sum games, in: *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufman, 2000, pp. 541-548.

[33] Ushio, T. (1996). "Limitation of delayed feedback control in nonlinear discrete-time systems," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol.43, no.9, pp.815–816.

[34] Yamamoto, S., T. Hino and T. Ushio (2001). "Dynamic delayed feedback controllers for chaotic discrete-time systems," *IEEE Trans. Circuits Syst. I*, vol.46, no.6, pp.785–789.

[35] Young, H. P. (2008). Learning by Trial and Error, mimeo, University of Oxford.

[36] Young, H. P. (2004). *Strategic Learning and Its Limits*. Oxford University Press.

[37] Zhu, Jiandong; Tian, Yu-Ping (2005). "Necessary and sufficient conditions for stabilizability of discrete-time systems via delayed feedback control", *Physics Letters A*, Volume 343, Issue 1-3, p. 95-107.

[38] Zinkevich, M. (2003). "Online Convex Programming and Generalized Infinitesimal Gradient Ascent", Carnegie Mellon University Technical Report CMU-CS-03-110.