# On the Notion of Regret in Infinitely Repeated Games

by

Amir Jafari

B.A., Sharif University, Tehran, Iran, 1995

A.M., The Johns Hopkins University, 1997

Sc.M., Brown University, 2002

Thesis

Submitted in partial fulfillment of the requirements for the Degree of the

Degree of Master of Science in the Department of Computer Sciences at

Brown University

PROVIDENCE, RHODE ISLAND

May 2003

This dissertation by Amir Jafari

is accepted in its present form by the Department of

Computer Sciences as satisfying the

thesis requirements for the degree of Master of Science

Date_____ _____

Amy Greenwald, Ph.D.

Approved by the Graduate Council

Date_____ _____

Karen Newman, Ph.D.

Dean of the Graduate School

# Acknowledgments

First and foremost, I would like to especially thank my advisor, Amy Greenwald, for her help, encouragement, and guidance during the time I was working on this project. The existence of this thesis owes a lot to her.

I also owe a debt of gratitude to my other teachers in the Brown Computer Sciences Department, to Roberto Tamassia, Eli Upfal, John Savage and Anna Lysyanskaya from whom I learned a lot of exciting theories and concepts.

I would especially like to thank my good friend Minh Ha Quang who has always been supportive throughout my time at Brown.

Finally, there are those whose unconditional and constant love, support and understanding have helped me throughout my whole life, especially during the hardest of times, and I cannot express what I owe to them: my mother, my father, and my brother Babak.

# Contents


**1 Introduction** 1

**2 Approachability** 3

**3 On the Notion of No-Regret** 7

**4 The Equilibrium** 10

**5 The Improving Process** 13

5.1 Hedge Method . . . . . 14

5.2 Beating the Gods . . . . . 15

**6 The Limit Behavior** 16

6.1 No-Regret and Nash Equilibrium . . . . . 16

6.2 Convergence of No-Regret Learning . . . . . 19

6.2.1 Almost Convergence . . . . . 19

6.2.2 Non-wandering Sequences . . . . . 21

**7 A Simple NIR Algorithm** 23

**Bibliography** 27

# Chapter 1

# Introduction

Imagine playing a simple game like Rock-Paper-Scissors with an opponent. If you play only one game there will not be any time for learning your opponents strategies, but suppose you repeat the game for a long period of time. One might ask how we can play in a very intelligent way. For example, if you notice that your opponent is playing Rock frequently, it will be a "stupid" move to play Scissors and a "smart" move to play Paper.

How can we compare different players, and say for example player $A$ is smarter than player $B$?

A smart player might use the history of the game to decide what move to make for the next round. Her strategy might be mixed, i.e. a set of weights on the set of her actions.

A **Deterministic Decision Algorithm (DDA)** is a procedure that to the "history" of the game up to now, associates a mixed strategy for the next move. It is called deterministic since the algorithm, for two games with the same history up to round $T$, associates the same strategy for the round $T + 1$. We can generalize this definition to include some random variables as well.

One important question is: Given two different DDA's how can we compare them and say

one is smarter than the other?

Another question that will be discussed in this work is the limit behavior of a game played by "smart" DDA's. Will the game converge to some sort of equilibrium that keeps everybody happy?

Of course to make sense out of questions like these, we have to define exactly what we mean by a "smart"DDA. To do this, we give a precise meaning to the word **Regret**. And we say a DDA is smart if in the long run it feels no regret for the moves it has made. After all isn't someone's life a successful one if he feels a minimum amount of regret when he is dying?

# Chapter 2

# Approachability

An agent with the set of actions $S$ (which is assumed to be a subset of a measure space), is playing against an opponent with the set of actions $S'$. There is a vector-valued pay-off function

$$r : S \times S' \longrightarrow V$$

into a vector space $V$ with an inner-product.

**Definition 2.1.** *A deterministic decision algorithm (DDA), $M = M(r)$ is a set of functions:*

$$q_T = q_T(M, r) : (S \times S')^{T-1} \longrightarrow \Delta(S) \quad \textit{for} \quad T = 1, 2, \ldots$$

*where $\Delta(S)$ is the set of all probability densities on $S$, and $(S \times S')^0$ is defined to be a single point. A non-deterministic decision algorithm denoted by DA, takes its values in $\mathcal{P}(\Delta(S))$, the set of all subsets of $\Delta(S)$.*

As examples of DA's one can consider constant DA, where the value of all the $q_T$'s are a constant element of $\Delta(S)$. This is a history-independent DA. Another example is the best-reply DA that assumes the opponent repeats his last move, $s'_{T-1}$, and for the next

move plays the best action against it.

Following Blackwell [3], we define a notion of approachability for a subset $G$ of $V$.

**Definition 2.2.** *A DDA is said to $r$-approach $G \subseteq V$ if for any $\delta > 0$ and any sequence of elements $s'_1, s'_2, \ldots$ of $S'$ the probability:*

$$\lim_{T\to\infty} P^T\left((s_1,\ldots,s_T)| \quad d(G, \frac{r(s_1,s'_1)+\ldots r(s_T,s'_T)}{T}) > \delta\right) = 0.$$

*Here $P^T$ is the probability measure on $S^T$ defined by the product density:*

$$p^T(s_1,\ldots,s_T) = q_1(s_1)\cdot q_2(s_1,s'_1)(s_2)\ldots q_T(s_1,s'_1,\ldots,s_{T-1},s'_{T-1})(s_T)$$

*This definition can easily be modified to be used for non-deterministic DA's.*

We restrict our attention to the case where $V = \mathbb{R}^N$ and

$$G = R^N_{\leq 0} := \{(x_1,\ldots,x_N)|x_i \leq 0 \quad i = 1,\ldots N\}.$$

Following Hart-Mas Colell [7] we give the following definition:

**Definition 2.3.** *Let $\Lambda : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ be a function such that it is zero on $G$. A DDA is said to be $\Lambda$-compatible if there is a constant $C$, such that:*

$$\Lambda\left(\frac{r(s_1,s'_1)+\ldots r(s_T,s'_T)}{T}\right)\cdot r(q_{T+1},s') \leq \frac{C}{T+1}$$

*for all $s' \in S'$. Here dot denotes the dot-product and $r(q,s')$ is the expected value:*

$$\int_S r(s,s')dq(s).$$

Our goal is to give conditions on $\Lambda$ so that a $\Lambda$-compatible DDA $r$-approaches $G$. For example let

$$\Lambda_0 : \mathbb{R}^N \longrightarrow \mathbb{R}^N$$

$$\Lambda_0(x_1,\ldots,x_N) = (x_1^+,\ldots,x_N^+).$$

(where $x^+ = x$ is $x > 0$ and 0 otherwise.)

**Theorem 2.4. (Generalized Blackwell)** *If $r(S \times S')$ is bounded, which is the case if both sets are finite, then a $\Lambda_0$-compatible DDA approaches $G = \mathbb{R}^N_{\leq 0}$.*

We follow the method of Foster and Vohra in [4].

The proof depends on the following general lemma:

**Lemma 2.5.** *Let $M_T$ be a sequence of random variables on $S^T$. Such that:*

- $|M_T - M_{T-1}| \leq f(T)$ *for an increasing function $f$.*
- $M_T$ *is super-Martingle i.e.:*

$$E^T(M_{T+1}) = \int_S M_{T+1}(s_1, \ldots, s_T, s) dq_{T+1}(s_1, s'_1, \ldots, s_T, s'_T)(s) \leq M_T(s_1, \ldots, s_T).$$

*Then for any $\epsilon > 0$, $P^T(M_T > 2\epsilon T f(T)) \leq e^{-\epsilon^2 T}$.*

**Proof.** Let $S_t = \frac{M_t}{f(T)}$ and $X_t = S_t - S_{t-1}$. We have $E^{t-1}(X_t) = \frac{1}{f(T)}(E^{t-1}(M_t) - M_{t-1}) \leq 0$ and $|X_t| \leq \frac{f(t)}{f(T)} \leq 1$. Therefore if we use $e^y \leq 1 + y + y^2$ for $y \leq 1$ we get

$$E^{t-1}(e^{\epsilon X_t}) \leq 1 + \epsilon E^{t-1}(X_t) + \epsilon^2 E^{t-1}(X_t^2) \leq 1 + \epsilon^2$$

Hence:

$$\begin{aligned} P^T(M_T \geq 2\epsilon T f(T)) &= P^T(S_T \geq 2\epsilon T) = P^T(e^{\epsilon S_T} \geq e^{2\epsilon^2 T}) \\ &\leq \frac{E(e^{\epsilon S_T})}{e^{2\epsilon^2 T}} = \frac{\prod_{t=1}^T E^{t-1}(e^{\epsilon X_t})}{e^{2\epsilon^2 T}} \\ &\leq \frac{(1+\epsilon^2)^T}{e^{2\epsilon^2 T}} \leq e^{-\epsilon^2 T} \end{aligned}$$

□

Without loss of generality assume that the diameter of $r(S \times S')$ is 1. Let $A_T = r(s_1, s'_1) + \cdots + r(s_T, s'_T)$. We have:

$$E^T(|A^+_{T+1}|^2) \leq E^T(|A_{T+1} - A^-_T|^2) = |A^+_T + r(q_{T+1}, s'_{T+1})|^2$$

$$\leq |A_T^+|^2 + C + 1$$

Therefore if we let $M_T = |A_T^+|^2 - (C+1)T$ we see easily that $E^T(M_{T+1}) < M_T$ also its easy to check that $|M_{T+1} - M_T| < C'T$ for some explicit constant $C'$. Hence we can use the lemma and get:

$$P^T(|A_T^+|^2 > C_0 \epsilon T^2) \leq e^{-\epsilon^2 T}$$

for some other constant $C_0$. This proves the theorem. □

Finally we mention the following result of Hart-Mas Colell in [7] that we will use only in the last chapter:

**Theorem 2.6.** *With above notation a $\Lambda$-compatible algorithm (with $C = 0$) approaches $G$ if the following properties hold:*

- *$\Lambda$ is continuous on $\mathbb{R}^N - G$.*
- *there is a Lipschitz function $P : \mathbb{R}^N \longrightarrow \mathbb{R}$ such that $\nabla P(x) = \phi(x)\Lambda(x)$ for almost every $x \in \mathbb{R}^N - G$, where $\phi : \mathbb{R}^N - G \longrightarrow \mathbb{R}_{>0}$ is a continuous positive function.*
- *$\Lambda(x) \in \mathbb{R}^N_{\leq 0} - \{0\}$ for all $x \in \mathbb{R}^N - G$.*

□

**Remark.** The last condition is stated in [7] for a general convex set $G$ we stated it only for the simple case $G = \mathbb{R}^N_{\leq 0}$.

Finally as a very special case we mention the following corollary:

**Corollary 2.7.** *A DDA approaches $G = \mathbb{R}^N_{\leq 0}$ if for all $s'$ and $T$:*

$$\left(\frac{r(s_1, s_1') + \cdots + r(s_T, s_T')}{T}\right)^+ \cdot r(q_{T+1}, s') = 0$$

□

# Chapter 3

# On the Notion of No-Regret

Let us consider a real-valued pay-off function:

$$r : S \times S' \longrightarrow \mathbb{R}.$$

Let $\Phi$ be a finite subset of linear maps $\phi : \Delta(S) \longrightarrow \Delta(S)$. Linearity means that for $0 \leq \alpha \leq 1$,

$$\phi(\alpha q_1 + (1-\alpha)q_2) = \alpha\phi(q_1) + (1-\alpha)\phi(q_2).$$

Let

$$r_\Phi : S \times S' \longrightarrow \mathbb{R}^\Phi$$

$$r_\Phi(s, s') = \left(r(\phi(\delta_s), s') - r(s, s')\right)_{\phi \in \Phi}.$$

Here $\delta_s \in \Delta(S)$ is the density concentrated at $s$.

**Definition 3.1.** *A DA is called $\Phi$-No Regret ($\Phi$-NR) if it $r_\Phi$-approaches $G := R^\Phi_{\leq 0}$. (Refer to the definition 2.2.)*

In concrete terms it means that in the long run the agent feels regret if instead of playing the recommended strategy $q_t$, he plays $\phi(q_t)$ for any fixed element $\phi \in \Phi$.

**Examples.** If $S$ is finite and $\Phi$ is the set of constant maps $\phi_s$:

$$\phi_s(q) = \delta_s$$

then we arrive at the definition of Hanna consistency [6], which is also called **No External Regret (NER)** algorithms. If we take $\Phi$ to be the set of the maps $\phi_{s_1,s_2}$ for distinct couples $s_i \in S$ defined by $\phi_{s_1,s_2}(q)(s) = q(s)$ if $s \neq s_1, s_2$, $\phi_{s_1,s_2}(q)(s_1) = 0$ and $\phi_{s_1,s_2}(q)(s_2) = q(s_1) + q(s_2)$, we get the definition of **No Internal Regret (NIR)** of Vohra and Foster [4].

**Lemma 3.2.** *If a DA is $\Phi$-NR then it is $\Phi'$-NR for any finite subset $\Phi'$ of the convex-hull of $\Phi$.*

**Proof.** Note that $r_{\Phi'} = A \cdot r_\Phi$ where $A$ is a matrix with non-negative entries whose each row sum to 1. Now since $A(\mathbb{R}^\Phi_{\leq 0}) \subseteq \mathbb{R}^{\Phi'}_{\leq 0}$ the result follows. □

**Lemma 3.3.** *An NIR algorithm is $\Phi$-NR for any finite subset of linear maps on $\Delta(S)$.*

**Proof.** Let $S = \{1, \ldots, k\}$. Let $A(n_1, \ldots n_k)$ for $1 \leq n_i \leq k$ be the stochastic matrix with 1's on the entries $(i, n_i)$ and 0's elsewhere. But since:

$$A(n_1, \ldots, n_k) = \phi_{1,n_1} + \cdots + \phi_{k,n_k} - (k-1)Id$$

the algorithm is $A(n_1, \ldots, n_k)$-NR. Because the set of the stochatoc matrices is the convex hull of $A(n_1, \ldots, n_k)$'s the lemma follows from the previous one. □

Finally following the method of the proof of Foster and Vohra [4] we prove the following result:

**Theorem 3.4.** *Let $S$ and $S'$ be a compact spaces and $r : S \times S' \longrightarrow \mathbb{R}$ be continuous. Let $\Phi$ be a finite subset of continuous linear maps on $\Delta(S)$. Then there exists a $\Phi$-NR algorithm.*

**Proof.** We will use Blackwell's theorem, corollary 2.7. We have to show for any $x \notin \mathbb{R}^{\Phi}_{\leq 0}$ there is $q \in \Delta(S)$ such that for all $s' \in S'$:

$$0 = r_{\Phi}(q, s') \cdot x^{+} = \sum_{\phi}(r(\phi q, s') - r(q, s')).x^{+}_{\phi} = r((\sum_{\phi} x^{+}_{\phi}\phi)(q), s') - r((\sum_{\phi} x^{+})q, s')$$

For this it is enough to have:

$$(\sum_{\phi\in\Phi} x^{+}_{\phi}\phi)(q) = (\sum_{\phi\in\Phi} x^{+}_{\phi})q.$$

But since by Brouwer fixed point theorem

$$\frac{\sum_{\phi\in\Phi} x^{+}_{\phi}\phi}{\sum_{\phi\in\Phi} x^{+}_{\phi}}$$

has a fixed point, the theorem is proved. If $S$ is a finite set there is no continuity assumption needed and the existence of the fixed point follows from the fact that any stochastic matrix has a positive fixed point. □

**Remark.** There might be many solutions for $q$ so the algorithm above is not deterministic. If the set of actions are finite we can make it deterministic by taking the solution given by the least-square method.

# Chapter 4

# The Equilibrium

In this chapter we give a definition of equilibrium states of a game that has the correlated equilibrium of Aumann [2] and MiniMax as its special cases.

A simple game with $N$-players is given by a function

$$r : S = S_1 \times \cdots \times S_N \longrightarrow \mathbb{R}^N.$$

Here $S_i$'s are the set of actions for player $i$ and $r_i : S_1 \times \cdots \times S_N \longrightarrow \mathbb{R}$ is the payoff function for player $i$. Let $\Phi^i$ be a subset of linear maps $\phi_i : \Delta(S_i) \longrightarrow \Delta(S_i)$. (recall that $\Delta(S_i)$ is the set of probability densities on the set $S_i$.) Such a map extends to a linear map on $\Delta(S_1 \times \cdots \times S_N)$, which by abuse of notation we still denote it by $\phi_i$:

$$\phi_i(q)(s_1, \ldots, s_N) := \phi(q(s_1, \ldots, s_{i-1}, ?, s_{i+1}, \ldots, s_N))(s_i).$$

**Definition 4.1.** *An element $q \in \Delta(S_1 \times \cdots \times S_N)$ is a $(\Phi^1, \ldots, \Phi^N)$-equilibrium if for all $i$ and all $\phi_i \in \Phi^i$:*

$$r_i(q) \geq r_i(\phi_i(q))$$

*Here $r_i(q)$ is the expected value as usual.*

**Examples.** If $\Phi^i$ is the set of linear maps $\phi_{s_1^i, s_2^i}$ ,defined after definition 3.1. of chapter 3, we arrive at the definition of the correlated equilibrium. If we assume furthermore that $q = q_1 \times \dots q_N$ is independent we arrive at the definition of the Nash equilibrium.

**Lemma 4.2.** *The set of* $(\Phi^1, \dots, \Phi^N)$*-equilibria is a convex set.*

**Proof.** This follows from linearity of $r_i$ on $\Delta(S)$. □

**Lemma 4.3.** *If* $q \in \Delta(S)$ *is a* $(\Phi^1, \dots, \Phi^N)$*-equilibrium then it is* $(\widehat{\Phi^1}, \dots, \widehat{\Phi^N})$*-equilibrium. Here hat denotes the convex hull.*

**Proof.** Same comment as the previous lemma. □

**Remark.** Note that the convex hull of the set of Nash equilibrium is inside the set of $(\Phi^1, \dots, \Phi^N)$-equilibrium for any choice of the subsets $\Phi^i$. However the converse of this result fails even for simple games such as chicken.

**Theorem 4.4.** *If player* $i$ *uses a* $\Phi^i$*-NR algorithm to play for all* $i$ *then the joint empirical distribution almost surely converges to the set of* $(\Phi^1, \dots, \Phi^N)$*-equilibria.*

**Proof.** The empirical distribution is defined by:

$$z^T(s) = \frac{\text{Number of } s\text{'s appeared up to round } T}{T}.$$

Therefore:

$$r_i(z^T) = \frac{1}{T} \sum_{t=1}^{T} r_i(s_i^t, s_{-i}^t).$$

By definition of $\Phi^i$-NR for any $\epsilon > 0$ and $\phi_i \in \Phi^i$, almost surely:

$$\sum_{t=1}^{T} \frac{r_i(\phi_i(\delta_{s_i^t}), s_{-i}^t) - r_i(s_i^t, s_{-i}^t)}{T} < \epsilon$$

for $T$ large enough. This means that:

$$r_i(\phi_i(z^T)) - r_i(z^T) < \epsilon$$

which means that $z^T$ almost surely converges to the set of $(\Phi^1, \ldots, \Phi^N)$-equilibriums. $\square$

# Chapter 5

# The Improving Process

The best DA is the one that knows the "future", I call this GOD. At round $T$ the machine knows what his opponent will play and plays the best response against it. In this chapter we will show that if a player has access to the all the past moves but not the future, and plays against a DDA that has finite past memory then in the long run he can play as good as GOD!

To do this we need to use a well-know result: Given DDA's $M_1, \ldots, M_n$ one can construct a DDA that does no worse than any single one, in the long run. Clearly by an induction argument we only need to consider the case where $n = 2$. There are two different approaches to solve this problem. One is called Hedge and is due to Auer et al [1], and the second one is to use an NIR algorithm and is due to Foster and Vohra [4].

**Remark.** Throughout this chapter we assume that the payoff functions takes its values in the interval $[0, 1]$ and the set $S$ of our agent's actions and $S'$ for out opponent's are finite.

## 5.1 Hedge Method

Assume that $M_1$ and $M_2$ are two DDA'a. At round $T$ they recommend mixed strategies $q_T^1$ and $q_T^2$ respectively. Take $\alpha > 0$. Let $H(M_1, M_2, \alpha)$ be an algorithm that at round $T$ recommends:

$$q_T := \frac{q_T^1(1+\alpha)^{R^1(T-1)} + q_T^2(1+\alpha)^{R^2(T-1)}}{(1+\alpha)^{R^1(T-1)} + (1+\alpha)^{R^2(T-1)}}$$

where

$$R^i(T-1) = \sum_{t=1}^{T-1} r(q_t^i, s_t')$$

**Theorem 5.1.** *The algorithm $M$ has the following property. For any sequence $s_1', s_2', \ldots$ of $S'$, for $i = 1, 2$ we have:*

$$\frac{\sum_{t=1}^{T}(r(q_t^i, s_t') - r(q_t, s_t'))}{T} \leq \frac{\alpha}{2} + \frac{\log|S|}{\alpha T}.$$

**Proof.** See [1] section 8. □

**Theorem 5.2.** *Assume $M(n) = H(M_1, M_2, \frac{1}{\sqrt{n}})$. Let $M$ be an algorithm that for the periods between $1 + \cdots + (n-1)$ and $1 + \cdots + n$ uses the machine $M(n)$ to recommend its strategy $q_T$. Then for any sequence $s_1', s_2', \ldots$ of $S'$ and any $i = 1, 2$*

$$\lim_{T\to\infty} \frac{\sum_{t=1}^{T}(r(q_t^i, s_t') - r(q_t, s_t'))}{T} = 0$$

**Proof.** By construction and the above theorem we have:

$$\frac{\sum_{t=1+\cdots+(n-1)}^{1+\cdots+n}(r(q_t^i, s_t') - r(q_t, s_t'))}{n} \leq \frac{1}{2\sqrt{n}} + \frac{\log|S|}{\sqrt{n}}$$

Therefore:

$$\sum_{t=1}^{1+\cdots+n}(r(q_t^i, s_t') - r(q_t, s_t')) \leq (\frac{1}{2} + \log|S|)\sum_{k=1}^{n}\sqrt{k}$$

which after dividing by $1 + \cdots + n$ and letting $n \to \infty$ proves the theorem. □

So we have constructed a DDA that in the long run does no worse than any of the original DDA'a $M_1$ and $M_2$. As we mention before an easy induction implies the existence of a machine that does no worse than any single of the $n$ machines $M_1, \ldots, M_n$. For a different method look at the paper by Foster and Vohra [4] section 3, theorem 2.

## 5.2 Beating the Gods

Assume the opponent is a DDA with finite past memory. This means that The functions

$$q'_T : (S \times S')^{T-1} \longrightarrow \Delta(S')$$

depend only on the last $N$ coordinates for a fixed (and for simplicity known) $N$. Let

$$s^0_T = \arg.\max_{s \in S} r(s, q'_T)$$

In this section we prove

**Theorem 5.3.** *For any $\epsilon > 0$ there is a DDA $M$ such that for $T$ large enough:*

$$\frac{\sum_{t=1}^{T}(r(s^0_t, q'_t) - r(q_t(M), q'_t))}{T} \leq \epsilon.$$

A simple doubling argument, proves the existence of a DDA $M$ such that:

$$\lim_{T\to\infty} \frac{\sum_{t=1}^{T}(r(s^0_t, q'_t) - r(q_t(M), q'_t))}{T} = 0.$$

**Proof.** Since the set of functions $(S \times S')^N \longrightarrow \Delta(S')$ is compact we can find functions $q(1), \ldots, q(K)$ in this set such that for any function $q$ in this space, there is $1 \leq i \leq K$ such that $|q - q(i)| \leq \epsilon$. Let $M(i)$ be th DDA that plays the best response against the DDA with finite past memory defined by $q(i)$. Let $M$ be a DDA that does no worse than any of the $M(i)$'s in the long run (such machine exists because of the results of the previous section), then since the strategy of the opponent falls in an $\epsilon$ neighborhood of the $q(i)$'s, $M$ satisfies the required property. □

# Chapter 6

# The Limit Behavior

## 6.1 No-Regret and Nash Equilibrium

Assume we have a $N$-person game with payoff function:

$$r : S_1 \times \cdots \times S_N \longrightarrow [0,1]^N.$$

In this section we study the behavior of a repeated game for players that exhibit No-Regret learning.

**Definition 6.1.** *A sequence of mixed strategies $q^t$ in $\Delta(S_1 \times \cdots \times S_N)$ is called a No-Regret model if for all $i$ and all $q_i \in \Delta(S_i)$:*

$$\lim_{T\to\infty} \sup \frac{\sum_{t=1}^{T} \rho_i(q_i, q_i^t | q_{-i}^t)}{T} \leq 0$$

*Here $\rho_i(q_i, q_i^t | q_{-i}^t) := r_i(q_i, q_{-i}^t) - r_i(q^t)$.*

A mixed strategy profile $\hat{q}^*$ is an $\epsilon$-Nash equilibrium iff all players play $\epsilon$-best-responses: i.e., $\rho_i(q_i, \hat{q}_i^* | \hat{q}_{-i}^*) < \epsilon$, for all players $i$, for all strategies $q_i$, and for some $\epsilon > 0$. Given sequence $\{q^t\}$ of mixed strategy profiles, define sequence $\{q_i^t\}$ of mixed strategies for player

$i$ to be *almost $\epsilon$-best-response* w.r.t. $\{q_{-i}^t\}$ iff the set of times for which $q_i^t$ is not an $\epsilon$-best-response has density zero: i.e.,

$$\lim_{T\to\infty} \frac{\#\{t < T \,|\, \exists q_i \;\rho_i(q_i, q_i^t|q_{-i}^t) \geq \epsilon\}}{T} = 0 \tag{6.1}$$

The sequence $\{q_i^t\}$ is *almost best-response* for player $i$ iff it is almost $\epsilon$-best-response for all $\epsilon > 0$. Lastly, the sequence $\{q^t\}$ of mixed strategies is *almost Nash* iff $\{q_i^t\}$ is an almost best-response for all players $i$.

**Theorem 6.2.** *If a sequence $\{q_i^t\}$ is almost best-response for player $i$ w.r.t. to some opposing sequence $\{q_{-i}^t\}$, then it satisfies no-regret w.r.t. model $\{q_{-i}^t\}$.*

*Proof.* For $\epsilon > 0$, let $A_{i,\epsilon} = \{t|\, \exists q_i \;\rho_i(q_i, q_i^t|q_{-i}^t) \geq \epsilon\}$ and $A_{i,\epsilon}^T = \{t < T|\, t \in A_{i,\epsilon}\}$. By assumption, the sequence $\{q_i^t\}$ is almost best-response: i.e., $d(A_{i,\epsilon}) = 0$. Now given $T$, $\rho_i(q_i, q_i^t|q_{-i}^t) < \epsilon$ for all $t \notin A_{i,\epsilon}^T$. Therefore, since regrets are bounded,

$$\begin{aligned}
& \lim_{T\to\infty} \sup \frac{1}{T} \sum_{t=1}^{T} \rho_i(q_i, q_i^t|q_{-i}^t) \\
= \; & \lim_{T\to\infty} \sup \frac{1}{T} \left( \sum_{t\in A_{i,\epsilon}^T} \rho_i(q_i, q_i^t|q_{-i}^t) + \sum_{t\notin A_{i,\epsilon}^T} \rho_i(q_i, q_i^t|q_{-i}^t) \right) \\
= \; & \lim_{T\to\infty} \sup \frac{1}{T} \sum_{t\notin A_{i,\epsilon}^T} \rho_i(q_i, q_i^t|q_{-i}^t) \\
< \; & \epsilon
\end{aligned}$$

for all strategies $q_i$. Since $\epsilon$ was arbitrary, the sequence $\{q_i^t\}$ exhibits no-regret w.r.t. model $\{q_{-i}^t\}$. □

**Corollary 6.3.** *If a sequence $\{q^t\}$ is almost Nash, then it satisfies no-regret.*

**Definition 6.4.** *A sequence $\{q^t\}$ is said to be* regular for player $i$ *iff for $t \gg 0$, there exists $q_i^*$ such that $r_i(q_i^*, q_{-i}^t) \geq r_i(q_i, q_{-i}^t)$, for all strategies $q_i$. As usual, $\{q^t\}$ is* regular *iff it is regular for all players $i$.*

A strategy $q_i^*$ is *(weakly) dominant* iff $r_i(q_i^*, q_{-i}) \geq r_i(q_i, q_{-i})$, for all $q_i, q_{-i}$. An equilibrium in (weakly) dominant strategies is a Nash equilibrium that consists of only (weakly) dominant strategies.

**Remark 6.5.** *If a game $\Gamma$ has an equilibrium in (weakly) dominant strategies, then every sequence of play of $\Gamma^\infty$ is regular.*

**Lemma 6.6.** *Given a regular sequence $\{q^t\}$ for player $i$ with select strategy $q_i^*$ such that $r_i(q_i^*, q_{-i}^t) \geq r_i(q_i, q_{-i}^t)$, for all strategies $q_i$. The following hold true:*

1. *$\rho_i(q_i^*, q_i^t | q_{-i}^t) \geq 0$, for $t \gg 0$*
2. *$\rho_i(q_i^*, q_i^t | q_{-i}^t) \geq \rho_i(q_i, q_i^t | q_{-i}^t)$, for all strategies $q_i$, for $t \gg 0$*

*Proof of 1.* It follows from the definition of regularity that $r_i(q_i^*, q_{-i}^t) \geq r_i(q_i^t, q_{-i}^t)$. Therefore, $\rho_i(q_i^*, q_i^t | q_{-i}^t) = r_i(q_i^*, q_{-i}^t) - r_i(q_i^t, q_{-i}^t) \geq 0$. [Proof of 2] By the definition of regularity, $r_i(q_i^*, q_{-i}^t) - r_i(q_i^t, q_{-i}^t) \geq r_i(q_i, q_{-i}^t) - r_i(q_i^t, q_{-i}^t)$, for all $q_i$. Therefore, $\rho_i(q_i^*, q_i^t | q_{-i}^t) \geq \rho_i(q_i, q_i^t | q_{-i}^t)$, for all $q_i$. □

**Theorem 6.7.** *Given a regular sequence $\{q^t\}$ for player $i$, if the sequence $\{q_i^t\}$ exhibits no-regret w.r.t. model $\{q_{-i}^t\}$, then it is almost best-response w.r.t. $\{q_{-i}^t\}$.*

*Proof.* Suppose not: i.e., suppose the sequence $\{q_i^t\}$ is not almost best-response given $\{q_{-i}^t\}$. Define $B_{i,\epsilon} = \{t \mid \exists q_i \ \rho_i(q_i, q_i^t | q_{-i}^t) \geq \epsilon\}$ and $B_{i,\epsilon}^T = \{t < T \mid t \in B_{i,\epsilon}\}$, for some $\epsilon > 0$. Since $\{q_i^t\}$ is not almost best-response, there exists $\delta > 0$ such that $d(B_{i,\epsilon}) \geq \delta$. Now the following

holds true of the strategy $q_i^*$:

$$\begin{aligned}
& \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \rho_i(q_i^*, q_i^t | q_{-i}^t) \\
= & \limsup_{T\to\infty} \frac{1}{T} \left( \sum_{t\in B_{i,\epsilon}^T} \rho_i(q_i^*, q_i^t | q_{-i}^t) + \sum_{t\notin B_{i,\epsilon}^T} \rho_i(q_i^*, q_i^t | q_{-i}^t) \right) \\
\geq & \limsup_{T\to\infty} \frac{1}{T} \sum_{t\in B_{i,\epsilon}^T} \rho_i(q_i^*, q_i^t | q_{-i}^t) \\
\geq & \limsup_{T\to\infty} \frac{1}{T} \sum_{t\in B_{i,\epsilon}^T} \epsilon \\
\geq & \ \delta\epsilon
\end{aligned}$$

The third step follows from Lemma 6.6(1). The fourth step follows from Lemma 6.6(2) and the fact that given $T$, for $t \in B_{i,\epsilon}^T$, $\rho_i(q_i, q_i^t | q_{-i}^t) \geq \epsilon$. Finally, since $\epsilon, \delta > 0$, the sequence $\{q_i^t\}$ does not satisfy no-regret w.r.t. model $\{q_{-i}^t\}$. Contradiction. □

**Corollary 6.8.** *Given a game $\Gamma$ with an equilibrium in (weakly) dominant strategies. If the sequence $\{q^t\}$ of play of $\Gamma^\infty$ exhibits no-regret, then it is almost Nash.*

## 6.2 Convergence of No-Regret Learning

In this section, we prove that if multi-agent no-regret learning generates weights that converge, then those weights must converge to a Nash equilibrium. Moreover, we show that in games for which there exists a *unique* equilibrium in dominant strategies, no-regret sequences (almost) converge to Nash equilibrium.

### 6.2.1 Almost Convergence

Recall that a sequence $\{a_n\}$ converges to $a$ iff the number of $n$'s for which $|a_n - a| \geq \epsilon$ is finite, for all $\epsilon > 0$. We weaken the definition of convergence slightly to arrive at a notion

of almost convergence.

**Definition 6.9.** *Given $\epsilon > 0$, a sequence $\{a_n\}$* almost $\epsilon$-converges *to a (notation $a_n \rightsquigarrow_\epsilon a$) iff the set of $n$'s for which $|a_n - a| \geq \epsilon$ has zero density: i.e.,*

$$\lim_{N\to\infty} \frac{\#\{n < N \,|\, |a_n - a| \geq \epsilon\}}{N} = 0 \tag{6.2}$$

*A sequence $\{a_n\}$* almost converges *to a (notation $a_n \rightsquigarrow a$) iff for all $\epsilon > 0$, $a_n \rightsquigarrow_\epsilon a$.*

**Theorem 6.10.** *Given mixed strategy sequences $\{q_i^t\}$ for all players $i$ satisfying no-regret. If $q_i^t \rightsquigarrow \bar{q}_i$ for all players $i$, then $\bar{q} = (\bar{q}_1, \ldots, \bar{q}_n)$ is a Nash equilibrium.*

*Proof.* By assumption, $q_i^t \rightsquigarrow \bar{q}_i$, for arbitrary player $i$. For $\epsilon > 0$, define $C_{i,\epsilon} = \{t \,|\, |q_i^t - \bar{q}_i| \geq \epsilon\}$, and $C_{i,\epsilon}^T = \{t < T \,|\, t \in C_{i,\epsilon}\}$. Now, since $d(C_{i,\epsilon}) = 0$ and regrets are bounded, it follows that for all strategies $q_i$,

$$\begin{aligned}
& \rho_i(q_i, \bar{q}_i | \bar{q}_{-i}) \\
= & \lim_{T\to\infty} \frac{1}{T} \sum_{t \notin C_{i,\epsilon}^T} \rho_i(q_i, q_i^t | q_{-i}^t) \\
= & \lim_{T\to\infty} \frac{1}{T} \left( \sum_{t \in C_{i,\epsilon}^T} \rho_i(q_i, q_i^t | q_{-i}^t) + \sum_{t \notin C_{i,\epsilon}^T} \rho_i(q_i, q_i^t | q_{-i}^t) \right) \\
= & \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \rho_i(q_i, q_i^t | q_{-i}^t) \\
< & \ \epsilon
\end{aligned}$$

for all $\epsilon > 0$. Since $i$ was arbitrary, this conclusion holds for all players. Therefore, $\bar{q}$ is a Nash equilibrium. □

### 6.2.2 Non-wandering Sequences

No-regret learning does not in general imply convergence or even almost convergence in games for which there exist multiple equilibria, since it is possible for no-regret sequences to wander through the space of Nash equilibria without ever converging. We propose the following definition of non-wandering to exclude this possibility in some cases.

**Definition 6.11.** *Given set $A$. A sequence $\{a_n\}$ is* non-$\epsilon$-wandering about $a^* \in A$ *iff for all $a \neq a^* \in A$,*

$$\lim_{N\to\infty} \frac{\#\{n < N \,|\, |a_n - a| < \epsilon\}}{N} = 0 \tag{6.3}$$

*A sequence $\{a_n\}$ is* non-wandering about $a^*$ *iff it is non-$\epsilon$-wandering about $a^*$, for all $\epsilon > 0$.*

**Theorem 6.12.** *An almost Nash sequence $\{q^t\}$ that is non-wandering about $q^*$ almost converges to $q^*$.*

*Proof.* Suppose not: i.e., suppose that for some player $i$ the sequence $\{q_i^t\}$ is not almost convergent. For some $\epsilon > 0$, define $D_{i,\epsilon} = \{t \,|\, |q_i^t - q_i^*| \geq \epsilon\}$; by assumption, there exists $\delta > 0$ such that $d(D_{i,\epsilon}) \geq \delta$. Also define $E_{i,\epsilon} = \{t \,|\, \exists q_i \neq q_i^* \; |q_i^t - q_i| < \epsilon\}$; since $\{q_i^t\}$ is non-wandering about $q_i^*$, we can choose $\epsilon$ sufficiently small such that $d(E_{i,\epsilon}) \leq \delta/2$. Thus, $D_{i,\epsilon} \setminus E_{i,\epsilon} = \{t \,|\, \forall q_i \; |q_i^t - q_i| \geq \epsilon\}$, and $d(D_{i,\epsilon} \setminus E_{i,\epsilon}) \geq \delta/2$.

Now since the sequence $\{q_i^t\}$ is almost best-response, the set of times for which $q_i^t$ is not a $1/n$-best-response has density zero. Thus, there must exist a sequence of times $\{t_n\}$ such that $\{q_i^{t_n}\}$ is a sequence of $1/n$-best-responses, and moreover, this latter sequence must have a convergent subsequence that converges to a best-response, say $q_i^{**}$. But then the set $\{t \,|\, |q_i^t - q_i^{**}| < \epsilon\}$ has positive density, which contradicts the fact that $d(D_{i,\epsilon} \setminus E_{i,\epsilon}) \geq \delta/2$. □

**Corollary 6.13.** *Given a sequence $\{q^t\}$ that is non-wandering about $q^*$ and regular, if the sequence exhibits no-regret, then it almost converges to $q^*$.*

**Remark 6.14.** *In games for which there exist unique Nash equilibria, say $q^*$, all almost Nash sequences are non-wandering about $q^*$.*

**Corollary 6.15.** *If the sequence $\{q^t\}$ of plays of $\Gamma^\infty$ exhibits no-regret, then it almost converges to $q^*$, if $q^*$ is the unique dominant strategy equilibrium of $\Gamma$.*

The conclusion that no-regret sequences of play converge to Nash equilibrium as specified by the above corollary is guaranteed to hold only in games for which such equilibria are unique and consist only of dominant strategies. Simulation experiments (refer to [5]) suggest that this convergence result can be strengthened to the case of all games for which pure strategy Nash equilibria exist. Moreover, in constant-sum games, no-regret learning converges in empirical distributions to Nash equilibrium. In general-sum games, however, no-regret learning need not imply convergence (even in empirical distributions) to Nash equilibrium.

# Chapter 7

# A Simple NIR Algorithm

Let us consider two special class of subsets of stochastic matrices $\mathbb{R}^S \longrightarrow \mathbb{R}^S$:

1. $\Phi_{ext}$ is the set of matrices

$$\phi_j(e_k) = e_j$$

where $e_k$ is the standard basis for $\mathbb{R}^S$.

2. $\Phi_{int}$ is the set of matrices $\phi_{ab}$ for distinct $a$ and $b$ in $S_i$:

$$\phi_{ab}(e_k) = 0 \text{ if } k \neq a, b$$

$$\phi_{ab}(e_a) = 0$$

$$\phi_{ab}(e_b) = e_a + e_b$$

**Definition 7.1.** *A $\Phi_{ext}$-NR machine is also called an NER (no external regret) machine, a $\Phi_{int}$-NR machine is also called an NIR (no internal regret) machine.*

**Corollary 7.2.** *An NIR machine is also an NER machine.*

**Proof.** Follows from lemma 3.3. □

**Corollary 7.3.** *If all players play using an internal NR machine then the joint empirical distribution converges almost surely to the set of the correlated equilibriums.*

**Proof.** Follows from Theorem 4.4. and the definition of correlated equilibrium.

**Remark.** The converse of this theorem is false. Consider a symmetric game between three players with two actions $a$ and $b$ the rewards are as follows:

$$r(a,a,a) = 1,\ \ r(a,b,a) = r(a,a,b) = r(b,a,a) = 0$$

$$r(b,b,a) = r(b,a,b) = r(a,b,b) = r(b,b,b) = 1000$$

If a decision machine recommends to play $a$ all the time, and all players use it then the outcome is always Nash, but this machine is not internal or even external NR.

**Theorem 7.4.** *(Hart-Mas Colell [7]) The following decision machine is NER.*

*At period $T+1$ play with the probability:*

$$q^{T+1}(j) = \frac{R^T(j)^+}{\sum_k R^T(k)^+}$$

*where* $R^T(j) = \sum_{t=1}^{T} r_i(j, s_{-i}^t) - r_i(s_i^t, s_{-i}^t)$.

**Proof.** Let $R : S_i \times S_{-i} \longrightarrow \mathbb{R}^{S_i}$ be defined by:

$$R(s_i, s_{-i}) = (r_i(j, s_{-i}) - r(s_i, s_{-i}))_{j \in S_i}$$

According to the Blackwell's theorem we need to show that for $s \notin \mathbb{R}^{S_i}_{\leq 0}$ we have:

$$R(x^+, s_{-i}).x^+ = 0$$

which is

$$\sum_{j \neq k} r_i(j, s_{-i}).x_k^+ x_j^+ - r(k, s_{-i}) x_k^+ x_j^+ = 0$$

by symmetry. □

**Theorem 7.5.** *The following decision machine is NIR:*

*at time $T+1$ play with the probability:*

$$q^{T+1}(j) = \frac{\sum_{k\neq j} R^T(j,k)^+}{\sum_{j'k} R^T(j',k)^+}$$

*here*

$$R^T(j,k) = \sum_{t=1, s_i^t=k}^{T} (r_i(j, s_{-i}^t) - r_i(k, s_{-i}^t))$$

**Proof.** Let $|S| = N$ define the map

$$P : \mathbb{R}^{N(N-1)} \longrightarrow \mathbb{R}$$

$$P((x_{ij})_{i\neq j}) = \frac{1}{2}\sum_j (\sum_{i\neq j} x_{ij}^+)^2$$

Let $\Lambda = \nabla(P)$ its $ij$'th component is:

$$\Lambda_{ij} = \sum_{k\neq j} x_{kj}^+$$

Let

$$r_\Phi : S \times S' \longrightarrow \mathbb{R}^{N(N-1)}$$

where

$$r_\Phi(s,s')_{ij} = r(j,s') - r(i,s') \quad \text{if} \quad s = i \quad \text{and zero otherwise}$$

To prove the result according to theorem 2.6. we need to show for any $x \notin \mathbb{R}^{N(N-1}_{\leq 0}$:

$$r_\Phi(q(x), s') \cdot \Lambda(x) = 0$$

where

$$q(x)(i) = \frac{\sum_{j\neq i} x_{ij}^+}{\sum_{k\neq j} x_{kj}^+}.$$

This reduces to:

$$\sum_{ij}(\sum_{k\neq j} x^+_{kj})(\sum_{l\neq i} x^+_{il})(r(j,s') - r(i,s')) = 0$$

which is true due to anti-symmetry with respect to $i$ and $j$. □

# Bibliography


[1] Auer P., Cesa-Bianchi N., Freund Y. and Schapire R.: Gambling in a rigged casino: the adversarial multi-armed bandit problem, Proceedings of the 36th annual symposium on foundations of computer science, 322-331, (1995).

[2] Aumann, R.: Subjectivity and correlation in randomized strategies, Journal of Mathematical Economics 1, 67-96 (1974).

[3] Blackwell, D: An analog of the minimax theorem for vector payoffs, Pacific Journal of Mathematics 6, 1-8 (1956).

[4] Foster D. and Vohra R. : Regret in the on-line decision problem. games and Economic Behavior, 21, 40-55 (1997).

[5] Greenwald A., Jafari A., Ercal G, Gondek D: On No-Regret learning, ficticious play and Nash equilibrium, Proceedings of the 18th international conference on machine learning (2001)

[6] Hannan, J. : Approximation to Bayes risk in repeated plays. In M.Dresher, A.W. Tucker and P.Wolf, editors, Contributions to the theory of games, volume 3, 97-139, Princeton Univ. Press, (1957).

[7] Hart S. and Mas Colell A.: A general class of adaptive strategies. Technical report, Center for rationality and interactive decision theory, (2000).