A General Class of No-Regret Learning Algorithms and Game-Theoretic Equilibria

Amy Greenwald[1] and Amir Jafari[2]

[1] Department of Computer Science
Brown University, Box 1910
Providence, RI 02906
amy@brown.edu
http://www.cs.brown.edu/~amy

[2] Department of Mathematics
Brown University, Box 1917
Providence, RI 02906
amir@math.brown.edu
http://www.math.brown.edu/~amir

Abstract. A general class of no-regret learning algorithms, called Φ-no-regret learning algorithms is defined, which spans the spectrum from no-internal-regret learning to no-external-regret learning, and beyond. Φ describes the set of strategies to which the play of a learning algorithm is compared: a learning algorithm satisfies Φ-no-regret iff no regret is experienced for playing as the algorithm prescribes, rather than playing according to any of the transformations of the algorithm's play prescribed by elements of Φ. Analogously, a class of game-theoretic equilibria, called Φ-equilibria, is defined, and it is shown that the empirical distribution of play of Φ-no-regret algorithms converges to the set of Φ-equilibria. Perhaps surprisingly, the strongest form of no-regret algorithms in this class are no-internal-regret algorithms. Thus, the tightest game-theoretic solution concept to which Φ-no-regret algorithms (provably) converge is correlated equilibrium. In particular, Nash equilibrium is not a necessary outcome of learning via any Φ-no-regret learning algorithms.

1 Introduction

Consider an agent that repeatedly faces some decision problem. The agent is presented with a choice of actions, each with a different outcome, or set of outcomes. After each choice is made, and the corresponding outcome is observed, the agent achieves a reward. In this setting, one reasonable objective for an agent is to maximize its average rewards. If each outcome is deterministic, and if the action set is finite, the agent need only undertake a linear search for an action that yields the maximal reward, and choose that action forever after. But if there is a set of outcomes associated with each choice of action, i.e., if the outcome is nondeterministic, even if the action set is finite, a more complex strategy, or *learning algorithm*, is called for, if indeed the agent seeks to maximize rewards.

No-regret learning algorithms are geared toward maximizing rewards in nondeterministic settings. The efficacy of a no-regret algorithm is determined by comparing the performance of the algorithm with the performance of a set of alternative strategies. For example, one might compare the performance of a learning algorithm with the set of strategies that always choose the same action a, for all actions a. Learning algorithms that outperform this strategy set are said to exhibit no-external-regret [11]. As another example, consider the set of strategies that choose action a' rather than action a, whenever a given learning algorithm chooses a, for all possible actions a and a'. Learning algorithms that outperform all strategies in this set are said to satisfy no-internal-regret [6].

This paper studies the outcome of no-regret learning among a set of agents, or players, playing a repeated game. In a game, each player is presented with a choice of actions, and the outcome of the game is jointly determined by all players' choices. Each outcome assigns a reward to each player. In general, players choose their actions nondeterministically; thus, the outcome of a game can be viewed as nondeterministic, as in the single agent decision problem. Interestingly, in two-player, zero-sum games, if each player plays using a no-external-regret learning algorithm, then the empirical distribution of play converges to the set of minimax equilibria (see, for example, Freund and Schapire [8]). Also of interest, in multi-player games, if each agent plays using a no-internal-regret learning algorithm, then the empirical distribution of play converges to the set of correlated equilibria (see, for example, Hart and Mas-Colell [12]).

In this article, we define a general class of no-regret learning algorithms, called Φ-no-regret learning algorithms, which spans the spectrum from no-internal-regret learning to no-external-regret learning, and beyond. Φ describes the set of strategies to which the play of a learning algorithms is compared: a learning algorithm satisfies Φ-no-regret iff no regret is experienced for playing as the algorithm prescribes, rather than playing according to any of the transformations of the algorithm's play prescribed by elements of Φ. Analogously, we define a class of game-theoretic equilibria, called Φ-equilibria, and we show that the empirical distribution of play of Φ-no-regret algorithms converges to the set of Φ-equilibria. Perhaps surprisingly, no-internal-regret algorithms are the strongest form of no-regret algorithms in this class. Thus, the tightest game-theoretic solution concept to which Φ-no-regret algorithms (provably) converge is correlated equilibrium. In particular, Nash equilibrium is not a necessary outcome of learning via any Φ-no-regret learning algorithms.

This article is organized as follows. In the next section, we present Blackwell's approachability theory, which provides the technology for the proofs that appear throughout this work. In Section 3, we define Φ-no-regret learning, and we show that no-external-regret and no-internal-regret are special cases of Φ-no-regret. We also directly establish the existence of an algorithm that exhibits Φ-no-regret, for an arbitrary choice of Φ. In Section 4, we define Φ-equilibrium, and we prove that Φ-no-regret learning converges to the set of Φ-equilibria. The content of this article is largely based on Jafari's Master's thesis [14].

2 Approachability

Consider an agent with a set of actions A ($a \in A$) playing a game against a set of opponents with joint action set A' ($a' \in A'$). Associated with each possible outcome is some vector given by the vector-valued function $\rho : A \times A' \to V$. The sets A and A' are σ-algebras, and V is a vector space with an inner product $\cdot$ and a distance metric d defined by the inner product.

Given a game with the vector-valued function ρ, a (deterministic) *learning algorithm* $\mathcal{A}$ is a sequence of functions $q_t = q_t(\rho) : (A \times A')^{t-1} \to \Delta(A)$, for $t = 1, 2, \ldots$, where $\Delta(A)$ is the set of all probability measures on A and $(A \times A')^0$ is defined as a single point: i.e., $q_1 \in \Delta(A)$. Note that a deterministic learning algorithm $\mathcal{A}$ generates nondeterministic actions. A nondeterministic learning algorithm $\mathcal{A}$ assumes values in $\mathcal{P}(\Delta(A))$, the set of all subsets of $\Delta(A)$.

Many examples of learning algorithms appear throughout the literature. A history-independent learning algorithm returns some constant element of $\Delta(A)$. The best-reply learning algorithm [4] returns an element of $\Delta(A)$, at time t, that maximizes the agent's rewards w.r.t. only a'_{t-1}. Fictitious play [3, 15] returns returns an element of $\Delta(A)$, at time t, that maximizes the agent's rewards w.r.t. the empirical distribution of play through time $t-1$.

Following Blackwell [2], we define the notion of approachability as follows.

Definition 1. *Given a game with vector-valued function ρ, a subset $G \subseteq V$ is said to be* ρ-approachable *iff there exists learning algorithm $\mathcal{A} = q_1, q_2, \ldots$* s.t. *for any sequence of opponents' actions $a'_1, a'_2, \ldots$, for all $\epsilon > 0$, there exists t_0* s.t. *for all $t \geq t_0$, $d(G, \bar{\rho}_t) = \inf_{g \in G} d(g, \bar{\rho}_t) < \epsilon$, almost surely, where $\bar{\rho}_t$ denotes the average value of ρ through time t: i.e., $\bar{\rho}_t = \frac{1}{t}\left(\rho(a_1, a'_1) + \ldots + \rho(a_t, a'_t)\right)$.*

Technically, for any sequence of opponents' actions $a'_1, a'_2, \ldots$, for all $\epsilon, \delta > 0$, there exists t_0 s.t. *for all $t \geq t_0$, $P^t\left((a_1, \ldots, a_t) | d(G, \bar{\rho}_t) > \delta\right) < \epsilon$, where P^t is the product measure on A^t induced by $q_1, \ldots, q_t$ as follows:*

$$P^t = q_1 \times q_2(a_1, a'_1) \times \ldots \times q_t((a_1, a'_1), \ldots, (a_{t-1}, a'_{t-1}))$$

In other words, a subset $G \subseteq V$ is ρ-approachable iff there exists a learning algorithm for an agent that generates nondeterministic actions for the agent which ensure that the distance from the set G to the average value of ρ through time t tends to zero as t tends to infinity, almost surely, for any sequence of opponents' actions $a'_1, a'_2, \ldots$.

Throughout the remainder of this paper, we restrict attention to $V = \mathbb{R}^S$ and $G = \mathbb{R}^S_- = \{(x_s)_{s \in S} | x_s \leq 0\}$, for various choices of some finite set S. Ultimately, we interpret the vector-valued function in games as *regrets* (rather than rewards). Thus, we seek learning algorithms that approach the negative orthant: i.e., learning algorithms that achieve no-regret.

Blackwell's seminal approachability theorem [2] gives a sufficient condition on learning algorithms which ensures that $\mathbb{R}^S_- \subseteq \mathbb{R}^S$ (or, more generally, any convex subset $G \subseteq V$) is approachable. We present Blackwell's condition, as well as a generalization of the approachability theorem due to Jafari [14].

Definition 2. *Let $\Lambda : \mathbb{R}^S \to \mathbb{R}^S$ be a function that is zero on $\mathbb{R}^S_-$. Given a game with vector-valued function ρ, a learning algorithm $\mathcal{A} = q_1, q_2, \ldots$ is Λ-compatible iff there exists some constant $c \in \mathbb{R}$ and there exists $t_0 \in \mathbb{N}$* s.t. *for all $t \geq t_0$,*

$$\Lambda(\bar{\rho}_t) \cdot \rho(q_{t+1}, a') \leq \frac{c}{t+1} \tag{1}$$

for all $a' \in A'$. Here $\rho(q, a') = \int_A \rho(a, a') dq(a)$ is the expected value of ρ.

This definition of Λ-compatibility is inspired by Hart and Mas-Colell [13], who introduce the notion of Λ-compatibility for $c = 0$. Blackwell's condition can be stated in terms of Λ-compatibility for a particular choice of Λ, namely Λ_0, which is defined as follows: $\Lambda_0((x_s)_{s \in S}) = (x_s^+)_{s \in S}$, where $x_s^+ = \max\{x_s, 0\}$.

Theorem 1 (Blackwell, 1956). *Given a game with vector-valued function ρ, if A and A' finite, then the set $\mathbb{R}^S_-$ is ρ-approachable iff there exists learning algorithm $\mathcal{A}$ that is Λ_0-compatible, with $c = 0$: i.e., for all times t and for all $a \in A'$, $\bar{\rho}_t^+ \cdot \rho(q_{t+1}, a') \leq 0$. Conversely, if no such learning algorithm exists, then the set $\mathbb{R}^S_-$ is not ρ-approachable.*

The following generalization of Blackwell's theorem, due to Jafari [14], states that Λ_0-compatibility implies ρ-approachability, even for $c \neq 0$.

Theorem 2 (Jafari, 2003). *Given a game with vector-valued function ρ, if $\rho(A \times A')$ is bounded, then the set $\mathbb{R}^S_-$ is ρ-approachable by learning algorithm $\mathcal{A}$ if $\mathcal{A}$ is Λ_0-compatible: i.e., there exists some constant $c \in \mathbb{R}$ and there exists $t_0 \in \mathbb{N}$* s.t. *for all $t \geq t_0$, $\bar{\rho}_t^+ \cdot \rho(q_{t+1}, a') \leq \frac{c}{t+1}$, for all $a' \in A'$. Conversely, if no such learning algorithm exists, then the set $\mathbb{R}^S_-$ is not ρ-approachable.*

3 No-Regret Learning

Let Φ be a finite subset of the set of linear maps $\phi : \Delta(A) \to \Delta(A)$: i.e., for all $0 \leq \alpha \leq 1$, for all $q_1, q_2 \in \Delta(A)$,

$$\phi(\alpha q_1 + (1 - \alpha) q_2) = \alpha \phi(q_1) + (1 - \alpha) \phi(q_2) \tag{2}$$

Thus, each $\phi \in \Phi$ converts one nondeterministic action for an agent into another. Given Φ, and given one distinguished agent's reward function $r : A \times A' \to \mathbb{R}$, we define regret vector $\rho_\Phi : A \times A' \to \mathbb{R}^\Phi$ as follows:

$$\rho_\Phi(a, a') = (r(\phi(\delta_a), a') - r(a, a'))_{\phi \in \Phi} \tag{3}$$

Here δ_a is the Dirac δ function: i.e., all mass is concentrated at a. In words, $\rho_\Phi(a, a')$ is a vector indexed by $\phi \in \Phi$ for which each entry describes the regret the agent feels for choosing action a rather than action $\phi(\delta_a)$, given action a'. Using this framework, we define Φ-no-regret learning.

Definition 3. *A Φ-*no-regret *learning algorithm is one that ρ_Φ-approaches $\mathbb{R}^\Phi_-$.*

Given a game, if an agent with reward function r learns to play in such a way that ρ_Φ approaches the negative orthant, then the agent's learning algorithm is said to exhibit no-regret. There are two well-studied examples of the Φ-no-regret property: *no-external-regret* and *no-internal-regret*.

Given an agent with finite action set A, let $\Phi_{\text{EXT}} = \{\phi_a | a \in A\}$ be the set of constant maps: i.e., $\phi_a(q) = \delta_a$. Thus, $\phi_a(q)$ is a probability density concentrated at a—it ascribes zero probability to any action $b \neq a$. Intuitively, a learning algorithm satisfies Φ_{EXT}-no-regret iff it precludes the agent from experiencing regret relative to any fixed strategy. Φ_{EXT}-no-regret corresponds to no-external-regret, also known as Hannan, or universal, consistency [11].

Our next choice of Φ, once again for A finite, gives rise to the definition of no-internal-regret [6]. Let $\Phi_{\text{INT}} = \{\phi_{ab} | a \neq b \in A\}$, where

$$(\phi_{ab}(q))_c = \begin{cases} q_c & \text{if } c \neq a, b \\ 0 & \text{if } c = a \\ q_a + q_b & \text{if } c = b \end{cases} \tag{4}$$

For $a \neq b$, ϕ_{ab} maps nondeterministic action q into another that ascribes zero probability to a, but instead adds a's probability mass according to q to the probability q ascribes to b. Thus, an algorithm satisfies Φ_{INT}-no-regret if the agent does not feel regret when it plays a instead of b, for all pairs of distinct actions $a \neq b$. Following Foster and Vohra [6], we call this property no-internal-regret, but it is sometimes called conditional universal consistency [9].

Perhaps surprisingly, no-internal-regret is the strongest form of Φ-no-regret, as Proposition 1 demonstrates. To prove this claim, we rely on the next lemma.

Lemma 1. *If learning algorithm $\mathcal{A}$ satisfies Φ-no-regret, then $\mathcal{A}$ also satisfies Φ'-no-regret, for all finite subsets $\Phi' \subseteq \text{SCH}(\Phi)$, the* super *convex hull of Φ, defined as follows:*

$$\text{SCH}(\Phi) = \left\{ \sum_{i=1}^{k+1} \alpha_i \phi_i \mid \phi_i \in \Phi, \textit{ for } 1 \le i \le k,\ \phi_{k+1} = I, \textit{ the identity map}, \right.$$
$$\left. \alpha_i \ge 0, \textit{ for } 1 \le i \le k,\ \alpha_{k+1} \in \mathbb{R}, \textit{ and } \sum_{i=1}^{k+1} \alpha_i = 1 \right\}$$

Proof. Let $\phi \in \text{SCH}(\Phi)$. By the calculation below, $\rho_\phi = \sum_{i=1}^{k} \alpha_i \rho_{\phi_i}$. Thus, $\rho_{\Phi'} = M\rho_\Phi$, for some matrix M with non-negative entries. If learning algorithm $\mathcal{A}$ satisfies Φ-no-regret (i.e., if $\mathbb{R}_-^\Phi$ is ρ_Φ-approachable), then $d(\mathbb{R}_-^\Phi, \overline{\rho}_{\Phi,t}) \to 0$, as $t \to \infty$, almost surely. But then, by the continuity of M, $d(M\mathbb{R}_-^\Phi, M\overline{\rho}_{\Phi,t}) = d(M\mathbb{R}_-^\Phi, \overline{\rho}_{\Phi',t}) \to 0$, as $t \to \infty$, almost surely. In other words, $M\mathbb{R}_-^\Phi$ is $\rho_{\Phi'}$-approachable. Now since all entries in M are non-negative, it follows that $M\mathbb{R}_-^\Phi \subseteq \mathbb{R}_-^{\Phi'}$. Therefore, $\mathbb{R}_-^{\Phi'}$ is $\rho_{\Phi'}$-approachable: i.e., $\mathcal{A}$ satisfies Φ'-no-regret.

Finally,

$$\rho_\phi(a, a') = r(\phi(\delta_a), a') - r(a, a')$$

$$
\begin{aligned}
&= r\left(\left(\sum_{i=1}^{k+1} \alpha_i \phi_i\right)(\delta_a), a'\right) - r(a,a') \\
&= \sum_{i=1}^{k+1} \alpha_i r(\phi_i(\delta_a), a') - r(a,a') \\
&= \sum_{i=1}^{k+1} \alpha_i r(\phi_i(\delta_a), a') - \sum_{i=1}^{k+1} \alpha_i r(a,a') \\
&= \sum_{i=1}^{k+1} \alpha_i \left(r(\phi_i(\delta_a), a') - r(a,a')\right) \\
&= \sum_{i=1}^{k+1} \alpha_i \rho_{\phi_i}(a,a') \\
&= \sum_{i=1}^{k} \alpha_i \rho_{\phi_i}(a,a')
\end{aligned}
$$

The last step follows from the fact that $\rho_{\phi_{k+1}}(a,a') = \rho_I(a,a') = r(I(\delta_a),a') - r(a,a') = r(a,a') - r(a,a') = 0$.

Corollary 1. *If learning algorithm $\mathcal{A}$ satisfies Φ-no-regret, then $\mathcal{A}$ also satisfies Φ'-no-regret, for all finite subsets $\Phi' \subseteq \text{CH}(\Phi)$, the convex hull of Φ.*

Proof. Choose $\alpha_{k+1} = 0$.

Proposition 1. *If learning algorithm $\mathcal{A}$ satisfies no-internal-regret, then $\mathcal{A}$ also satisfies Φ-no-regret for all finite subsets Φ of the set of stochastic matrices.*

Proof. An elementary matrix is one with exactly one 1 per row, and 0's elsewhere. By Corollary 1, if an algorithm is Φ-no-regret for the set Φ of elementary matrices, then the algorithm satisfies Φ'-no-regret for all finite subsets Φ' of the set of stochastic matrices, since the set of stochastic matrices is the convex hull of the set of elementary matrices. Thus, it suffices to show that an algorithm that satisfies no-internal-regret is Φ-no-regret for the set Φ of elementary matrices.

But by Lemma 1, it further suffices to show that any elementary matrix M can be expressed as follows: $M = \alpha_1\phi_1 + \ldots + \alpha_k\phi_k - \alpha_{k+1}I$, where $\phi_i \in \Phi_{\text{INT}}$ and $\alpha_i \geq 0$, for $1 \leq i \leq k$, $\alpha_{k+1} \in \mathbb{R}$, $\sum_{i=1}^{k+1} \alpha_i = 1$, and I is the identity map.

Let $A = \{1,\ldots,m\}$. Let $M(n_1,\ldots,n_m)$ denote the elementary matrix with 0's everywhere except 1's at entries (i,n_i) for $1 \leq i \leq m$. Now the linear map $\phi_{i,j}$ defined in Equation 4 is represented by the elementary matrix with 1's on the diagonal except in row i, where the 1 appears in column j. Thus, $\phi_{1n_1}+\ldots+\phi_{mn_m}$ corresponds to the matrix with 0's everywhere, except 1's at entries (i,n_i) for $1 \leq i \leq m$, and $m-1$'s on the diagonal. It follows that

$$M(n_1,\ldots,n_m) = \phi_{1n_1} + \ldots + \phi_{mn_m} - (m-1)I \tag{5}$$

Indeed, Φ-no-regret learning algorithms exists. In particular, no-external-regret algorithms pervade the literature. The earliest date back to Blackwell [2] and Hannan [11]; but, more recently, Foster and Vohra [5] Freund and Schapire [7], Fudenberg and Levine [10], Hart and Mas-Colell [13], and others have studied such algorithms. To our knowledge, only Foster and Vohra [6] have proposed an algorithm that satisfies no-internal-regret.

The next theorem establishes the existence of Φ-no-regret algorithms, for all sets Φ. The method of proof is related to that of Foster and Vohra [6].

Theorem 3. *Given reward function $r : A \times A' \to \mathbb{R}$. If $r(A \times A')$ is bounded, then there exists a learning algorithm that satisfies Φ-no-regret, for all finite subsets Φ of the set of continuous, linear maps on $\Delta(A)$.*

Proof. It suffices to show that for all $x \in \mathbb{R}^{\Phi} \setminus \mathbb{R}^{\Phi}_{-}$, there exists $q = q(x) \in \Delta(A)$ *s.t.* $x^{+} \cdot \rho_{\Phi}(q, a') \leq 0$, for all $a' \in A$. Letting $x = \overline{\rho}_t$, for $t = 1, 2, \ldots$, the result follows from Blackwell's approachability theorem. In fact, we show equality:

$$\begin{aligned}
0 &= x^{+} \cdot \rho_{\Phi}(q, a') \\
&= \sum_{\phi \in \Phi} x^{+}_{\phi} (r(\phi(q), a') - r(q, a')) \\
&= \sum_{\phi \in \Phi} x^{+}_{\phi} r(\phi(q), a') - \sum_{\phi \in \Phi} x^{+}_{\phi} r(q, a') \\
&= r\left(\left(\sum_{\phi \in \Phi} x^{+}_{\phi} \phi\right)(q), a'\right) - r\left(\left(\sum_{\phi \in \Phi} x^{+}_{\phi}\right) q, a'\right)
\end{aligned}$$

Now it suffices to show the following:

$$\left(\sum_{\phi \in \Phi} x^{+}_{\phi} \phi\right)(q) = \left(\sum_{\phi \in \Phi} x^{+}_{\phi}\right) q \tag{6}$$

Define $M : \Delta(A) \to \Delta(A)$ as follows:

$$M = \frac{\sum_{\phi \in \Phi} x^{+}_{\phi} \phi}{\sum_{\phi \in \Phi} x^{+}_{\phi}} \tag{7}$$

The function M maps a compact space, namely $\Delta(A)$, into itself. Moreover, it is continuous, since all $\phi \in \Phi$ are continuous, by assumption. Therefore, by Brouwer's fixed point theorem, M has a fixed point.

Technically speaking, this theorem does not establish the existence of an "algorithm," as stated, because the proof of Brouwer's fixed point theorem is not constructive. Moreover, the solution to Equation 6 need not be unique. Thus, at best, this theorem establishes the existence of a *nondeterministic*, Φ-no-regret "algorithm." But if A is finite, then the function M defined in Equation 7 is a stochastic matrix, and a solution to Equation 6 arises from the fact that any stochastic matrix has a positive fixed point. In this case, the least squares method of solving systems of equations yields a deterministic, Φ-no-regret algorithm.

4 Φ-Equilibrium

In this section, we define the notion of Φ-equilibrium, and we prove that learning algorithms that satisfy Φ-no-regret converge to Φ-equilibria. In particular, Φ_{EXT}-no-regret algorithms (i.e., no-external-regret algorithms) converge to (generalized) minimax equilibria; and Φ_{INT}-no-regret algorithms (i.e., no-internal-regret algorithms) converge to correlated equilibria.

Consider an n-player game where each player i chooses an action from the set A_i, and rewards are determined by the function $r : A_1 \times \ldots \times A_n \to \mathbb{R}^n$. Let Φ_i be a finite subset of the set of linear maps $\phi_i : \Delta(A_i) \to \Delta(A_i)$. A linear map ϕ_i extends to a linear map $\phi_i : \Delta(A_1 \times \ldots \times A_n) \to \Delta(A_1 \times \ldots \times A_n)$ as follows:

$$\begin{aligned}
\phi_i(q)(b_i, a_{-i}) &\equiv \phi_i((q(a_i, a_{-i}))_{a_i \in A_i})(b_i) \\
&= \phi_i\left(\sum_{a_i \in A_i} q(a_i, a_{-i})\delta_{a_i}\right)(b_i) \\
&= \sum_{a_i \in A_i} q(a_i, a_{-i})\phi_i(\delta_{a_i})(b_i) \qquad (8)
\end{aligned}$$

An element $q \in \Delta(A_1 \times \ldots \times A_n)$ is called independent iff it can be written as the product $q = q_1 \times \ldots \times q_n$ of n independent elements $q_i \in \Delta(A_i)$. For independent $q \in \Delta(A_1 \times \ldots \times A_n)$,

$$\begin{aligned}
&\phi_i(q)(a_1, \ldots, a_{i-1}, b_i, a_{i+1}, \ldots, a_n) \\
&= \sum_{a_i \in A_i} (q_1 \times \ldots \times q_i \times \ldots \times q_n)(a_1, \ldots, a_i, \ldots, a_n)\phi_i(\delta_{a_i})(b_i) \\
&= \sum_{a_i \in A_i} q_1(a_1) \ldots q_i(a_i) \ldots q_n(a_n)\phi_i(\delta_{a_i})(b_i) \\
&= q_1(a_1) \ldots q_{i-1}(a_{i-1})q_{i+1}(a_{i+1}) \ldots q_n(a_n) \sum_{a_i \in A_i} q_i(a_i)\phi_i(\delta_{a_i})(b_i) \\
&= q_1(a_1) \ldots q_{i-1}(a_{i-1})q_{i+1}(a_{i+1}) \ldots q_n(a_n)\phi_i\left(\sum_{a_i \in A_i} q_i(a_i)\delta_{a_i}\right)(b_i) \\
&= q_1(a_1) \ldots q_{i-1}(a_{i-1})q_{i+1}(a_{i+1}) \ldots q_n(a_n)\phi_i(q_i)(b_i)
\end{aligned}$$

The stated definition of the extended map ϕ_i applies to all $q \in \Delta(A_1 \times \ldots \times A_n)$, with the property that $\phi(q) = q_1 \times \ldots \times \phi_i(q_i) \times \ldots \times q_n$, for independent q. Note also that this definition yields an extension that is indeed a probability measure, since

$$\begin{aligned}
\sum_{b_i, a_{-i}} \phi_i(q)(b_i, a_{-i}) &= \sum_{b_i, a_{-i}} \left(\sum_{a_i \in A_i} q(a_i, a_{-i})\phi_i(\delta_{a_i})(b_i)\right) \\
&= \sum_{a_i, a_{-i}} q(a_i, a_{-i}) \sum_{b_i} \phi_i(\delta_{a_i})(b_i) \\
&= 1
\end{aligned}$$

Definition 4. *Given a game (i.e., given reward function r), and given vector $\Phi = (\Phi_i)_{1 \leq i \leq n}$, an element $q \in \Delta(A_1 \times \ldots \times A_n)$ is called a Φ-equilibrium iff $r_i(q) \geq r_i(\phi_i(q))$, for all players i and for all $\phi_i \in \Phi_i$.*

Generalized minimax and correlated equilibrium are both special cases of Φ-equilibrium. We define generalized minimax equilibria as Φ-equilibria, with $\Phi_i = \Phi_{\text{EXT}}$ for all players i. Correlated equilibria are Φ-equilibria, where $\Phi_i = \Phi_{\text{INT}}$ for all players i. Next we discuss two convexity properties of the set of Φ-equilibria, and the relationship between Φ-equilibria and Nash equilibria.

Lemma 2. *Given a game (i.e., given reward function r), and given vector $\Phi = (\Phi_i)_{1 \leq i \leq n}$, the set of Φ-equilibria is convex.*

Proof. If q and q' are both Φ-equilibria, then $r_i(q) \geq r_i(\phi_i q)$ and $r_i(q') \geq r_i(\phi_i q')$, for all players i and for all $\phi_i \in \Phi_i$. Since r_i and ϕ_i are linear on $\Delta(A_1 \times \ldots \times A_n)$, it follows that

$$\begin{aligned}
& r_i(\alpha q + (1-\alpha) q') \\
&= \alpha r_i(q) + (1-\alpha) r_i(q') \\
&\geq \alpha r_i(\phi_i q) + (1-\alpha) r_i(\phi_i q') \\
&= r_i(\alpha \phi_i q + (1-\alpha) \phi_i q') \\
&= r_i(\phi_i(\alpha q + (1-\alpha) q'))
\end{aligned}$$

Lemma 3. *Given a game (i.e., given reward function r), and given vector $\Phi = (\Phi_i)_{1 \leq i \leq n}$, if $q \in \Delta(A_1 \times \ldots \times A_n)$ is a Φ-equilibrium, then it is also a Φ'-equilibrium, where $\Phi' = (\Phi'_i)_{1 \leq i \leq n}$ and Φ'_i is the convex hull of Φ_i, for $1 \leq i \leq n$.*

Proof. Since q is a Φ-equilibrium, $r_i(q) \geq r_i(\phi_i(q))$, for all players i and for all $\phi_i \in \Phi_i$. In particular, $r_i(q) \geq r_i(\phi_{i_1}(q))$ and $r_i(q) \geq r_i(\phi_{i_2}(q))$. Since r_i, ϕ_{i_1}, and ϕ_{i_2} are linear on $\Delta(A_1 \times \ldots \times A_n)$, it follows that

$$\begin{aligned}
& r_i(q) \\
&= \alpha r_i(q) + (1-\alpha) r_i(q) \\
&\geq \alpha r_i(\phi_{i_1}(q)) + (1-\alpha) r_i(\phi_{i_2}(q)) \\
&= r_i(\alpha \phi_{i_1}(q)) + (1-\alpha) \phi_{i_2}(q)) \\
&= r_i((\alpha \phi_{i_1} + (1-\alpha) \phi_{i_2})(q))
\end{aligned}$$

A Nash equilibrium $q \in \Delta(A_1 \times \ldots \times A_n)$ is an independent Φ-equilibrium: if $r_i(q) \geq r_i(\phi_i(q)) = r(q_1, \ldots, q_{i-1}, \phi_i(q_i), q_{i+1}, \ldots, q_n)$, for all players i and for all $\phi_i \in \Phi_i$, then $r(q) \geq r(q_1, \ldots, q_{i-1}, q'_i, q_{i+1}, \ldots, q_n)$, for all players i and for all $q'_i \in \Delta(A_i)$. In other words, the set of Φ-equilibria contains the set of Nash equilibria. Moreover, since the set of Φ-equilibria is convex, this set also contains the convex hull of the set of Nash equilibria. But the convex hull of the set of Nash equilibria need not contain even the smallest set of Φ-equilibria: in particular, the convex hull of the set of Nash equilibria need not contain the set of correlated equilibria [1].

Theorem 4. *Given a game described by reward function r, if all players i play via some Φ_i-no-regret learning algorithm, then the joint empirical distribution of play converges to the set of Φ-equilibrium, almost surely.*

Proof. Define the empirical distribution z_t of play through time t as follows:

$$z_t(a_i, a_{-i}) = \frac{1}{t} \sum_{\tau=1}^{t} \mathbf{1}_{a_{i,\tau}=a_i} \mathbf{1}_{a_{-i,\tau}=a_{-i}} \tag{9}$$

for all actions $a_i \in A_i$ and $a_{-i} \in A_{-i} \equiv \prod_{j \neq i} A_j$. (The notation $\mathbf{1}_{a_{i,\tau}=a_i}$ denotes the indicator function, which equals 1 whenever $a_{i,\tau} = a_i$, and 0 otherwise.) It suffices to show that for all players i and for all $\phi_i \in \Phi_i$, $r_i(\phi_i(z_t)) - r_i(z_t) \to 0$, as $t \to \infty$, almost surely.

First, for arbitrary player i and for arbitrary $\phi_i \in \Phi_i$,

$$\begin{aligned} r_i(\phi_i(z_t)) &= \sum_{b_i, a_{-i}} \phi_i(z_t)(b_i, a_{-i}) r_i(b_i, a_{-i}) \\ &= \sum_{b_i, a_{-i}} \sum_{a_i \in A_i} z_t(a_i, a_{-i}) \phi_i(\delta_{a_i})(b_i) r_i(b_i, a_{-i}) \\ &= \sum_{a_i, a_{-i}} z_t(a_i, a_{-i}) r_i(\phi_i(\delta_{a_i}), a_{-i}) \\ &= \frac{1}{t} \sum_{\tau=1}^{t} r_i(\phi_i(\delta_{a_{i,\tau}}), a_{-i,\tau}) \end{aligned}$$

In the first step, we expand the definition of the expectation $r_i(q, a_{-i})$; and in the third step we collapse this definition. The second step relies on the extended definition of $\phi_i : \Delta(A_1 \times \ldots \times A_n) \to \Delta(A_1 \times \ldots \times A_n)$—see Equation 8. The last step follows from the definition of the empirical distribution z_t.

Second, for arbitrary player i,

$$r_i(z_t) = \frac{1}{t} \sum_{\tau=1}^{t} r_i(a_{i,\tau}, a_{-i,\tau})$$

Now, by assumption all players i play according to some Φ_i-no-regret learning algorithm. Thus, for all players i, for all $\phi_i \in \Phi_i$,

$$\begin{aligned} &\limsup_{t \to \infty} r_i(\phi_i(z_t)) - r_i(z_t) \leq 0 \\ \text{iff } &\limsup_{t \to \infty} \frac{1}{t} \sum_{\tau=1}^{t} r_i(\phi_i(\delta_{a_{i,\tau}}), a_{-i,\tau}) - \frac{1}{t} \sum_{\tau=1}^{t} r_i(a_{i,\tau}, a_{-i,\tau}) \leq 0 \\ \text{iff } &\limsup_{t \to \infty} \bar{\rho}_{\Phi_i,t} \leq 0 \\ \text{iff } &\lim_{t \to \infty} d(\mathbb{R}_{-}^{\Phi}, \bar{\rho}_{\Phi_i,t}) = 0 \end{aligned}$$

almost surely. In other words, the joint empirical distribution of play converges to the set of Φ-equilibrium, almost surely.

5 Conclusion

In this article, we defined a general class of no-regret learning algorithms, called Φ-no-regret learning algorithms, which spans the spectrum from no-internal-regret learning to no-external-regret Analogously, we defined a general class of game-theoretic equilibria, called Φ-equilibria, and we showed that the empirical distribution of play of Φ-no-regret algorithms converges to the set of Φ-equilibria. But the set Φ was restricted: it contained only linear maps. In future work, we plan to generalize this framework to include nonlinear, as well as, linear maps. Perhaps by doing so, we can obtain convergence results to tighter solution concepts than correlated equilibrium.

References

1. R. Aumann. Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, 1:67–96, 1974.
2. D. Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6:1–8, 1956.
3. G. Brown. Iterative solutions of games by fictitious play. In T. Koopmans, editor, *Activity Analysis of Production and Allocation.* Wiley, New York, 1951.
4. A. Cournot. *Recherches sur les Principes Mathematics de la Theorie de la Richesse.* Hachette, 1838.
5. D. Foster and R. Vohra. A randomization rule for selecting forecasts. *Operations Research*, 41(4):704–709, 1993.
6. D. Foster and R. Vohra. Regret in the on-line decision problem. *Games and Economic Behavior*, 21:40–55, 1997.
7. Y. Freund and R. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *Computational Learning Theory: Proceedings of the Second European Conference*, pages 23–37. Springer-Verlag, 1995.
8. Y. Freund and R. Schapire. Game theory, on-line prediction, and boosting. In *Proceedings of the 9th Annual Conference on Computational Learning Theory*, pages 325–332. ACM Press, May 1996.
9. D. Fudenberg and D. K. Levine. Conditional universal consistency. *Games and Economic Behavior*, Forthcoming.
10. D. Fudenberg and D.K. Levine. Universal consistency and cautious fictitious play. *Journal of Economic Dyanmics and Control*, 19:1065–1090, 1995.
11. J. Hannan. Approximation to Bayes risk in repeated plays. In M. Dresher, A.W. Tucker, and P. Wolfe, editors, *Contributions to the Theory of Games*, volume 3, pages 97–139. Princeton University Press, 1957.
12. S. Hart and A. Mas Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68:1127–1150, 2000.
13. S. Hart and A. Mas Colell. A general class of adaptive strategies. *Economic Theory*, 98:26–54, 2001.
14. A. Jafari. *On the Notion of Regret in Infinitely Repeated Games.* Master's Thesis, Brown University, Providence, May 2003.
15. J. Robinson. An iterative method of solving a game. *Annals of Mathematics*, 54:298–301, 1951.