

# No- $\Phi$ -Regret: A Connection between Computational Learning Theory and Game Theory

**Amy Greenwald**

*Department of Computer Science  
Brown University, Box 1910  
Providence, RI 02912*

AMY@BROWN.EDU

**Amir Jafari**

*Mathematics Department  
Duke University  
Durham, NC 27708*

AMIR@MATH.DUKE.EDU

**Casey Marks**

*Department of Computer Science  
Brown University, Box 1910  
Providence, RI 02912*

CASEY@CS.BROWN.EDU

## Abstract

This paper explores a fundamental connection between computational learning theory and game theory through a property we call no- $\Phi$ -regret. Given a set of transformations  $\Phi$  (i.e., mappings from actions to actions), a learning algorithm is said to exhibit no  $\Phi$ -regret if an agent experiences no regret for playing the actions the algorithm prescribes, rather than playing the transformed actions prescribed by any of the elements of  $\Phi$ . The existence of no- $\Phi$ -regret learning algorithms is established, for all finite  $\Phi$ .

Analogously, a class of game-theoretic equilibria, called  $\vec{\Phi}$ -equilibria, for  $\vec{\Phi} = (\Phi_i)_{1 \leq i \leq n}$ , is defined (here  $n$  is the number of agents/players). The main contribution of this paper is to show that the empirical distribution of play of no- $\Phi_i$ -regret algorithms converges to the set of  $\vec{\Phi}$ -equilibria. The well-known result that the empirical distribution of play of no-internal-regret learning converges to the set of correlated equilibria follows as an immediate corollary of this general theorem. In addition to providing a sufficient condition, a necessary condition for convergence to the set of  $\vec{\Phi}$ -equilibria is also derived.

This work was originally motivated by an attempt to design a no-regret learning scheme that would converge to a tighter solution concept than the set of correlated equilibria. However, it is argued that the strongest form of no- $\Phi$ -regret learning is no-internal-regret learning. Hence, the tightest game-theoretic solution concept to which any no- $\Phi$ -regret algorithm converges is correlated equilibrium. In particular, Nash equilibrium is not a necessary outcome of learning via any no- $\Phi$ -regret algorithms.

**Keywords** no-regret learning algorithms,  
convergence to equilibrium

## 1. Introduction

We analyze learning among a group of agents<sup>1</sup> playing an infinitely-repeated matrix game. At each stage, each agent chooses among a set of actions. The outcome, which is jointly determined by all the agents' choices, assigns a reward to each agent. A learning algorithm is a mapping from a history of past actions, outcomes, and rewards to a current choice of action. Our goal is to characterize the dynamics of multiple agents abiding by “no-regret” learning algorithms.

In the no-regret framework, the efficacy of learning is determined by comparing the performance of a learning algorithm to the performance of an alternative set of strategies. At each time  $t$ , we compare the action (i.e., pure strategy)  $a_t$  dictated by the learning algorithm with an alternative mixed strategy  $\phi(a_t)$ . The function  $\phi$  is called an action transformation. The agent's regret is the difference between the rewards obtained by playing action  $a_t$  and the rewards it would have expected to obtain had it instead played the transformed action  $\phi(a_t)$ . Given a set  $\Phi$  of action transformations, the  $\Phi$ -regret vector (at time  $t$ ) is the vector of regrets the agent experiences for not having played according to each  $\phi \in \Phi$ . By definition, *no- $\Phi$ -regret* learning algorithms have the property that the time-averaged  $\Phi$ -regret vector approaches the negative orthant.

For example, consider the set of all constant strategies. (A constant strategy always plays action  $a$ , for some action  $a$ .) Learning algorithms that perform at least as well as this strategy set are said to exhibit *no external regret* [16]. As another example, consider a strategy that is identical to the strategy dictated by the learning algorithm, except that every play of action  $a$  suggested by the learning algorithm is replaced by action  $a'$ , for some  $a$  and  $a'$ . Learning algorithms that perform at least as well as all such strategies are said to exhibit *no internal regret* [10]. The following results are well-known (see, for example, Hart and Mas-Colell [17, 18]): In two-player, zero-sum, repeated games, if each player plays using a no-external-regret learning algorithm, then each player's empirical distribution of play converges to his set of minimax strategies.<sup>2</sup> In general-sum, repeated games, if each player plays using a no-internal-regret learning algorithm, then the empirical distribution of play converges to the set of correlated equilibria.

In this paper, we define a general class of no-regret learning algorithms, called no- $\Phi$ -regret learning algorithms, which spans the spectrum from no external regret to no internal regret, and beyond. The set  $\Phi$  describes a set of strategies into which the play of a given learning algorithm is transformed. Such a learning algorithm satisfies no- $\Phi$ -regret if 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 the elements of  $\Phi$ . The existence of no- $\Phi$ -regret learning algorithms is established here (and elsewhere [3, 13, 21]), for all finite  $\Phi$ .

Analogously, we define a class of game-theoretic equilibria, called  $\vec{\Phi}$ -equilibria, for  $\vec{\Phi} = (\Phi_i)_{1 \leq i \leq n}$ . The main contribution of this paper is to show that the empirical distribution of play of no- $\Phi_i$ -regret algorithms converges to the set of  $\vec{\Phi}$ -equilibria. We obtain as corollaries of our theorem the aforementioned results on convergence of no-external-regret learning (no-internal-regret learning) to the set of minimax equilibria (correlated equilibria) in zero-sum (general-sum) games. Furthermore, we establish a necessary condition for convergence to the set of  $\vec{\Phi}$ -equilibria, namely that the time-averaged  $\Phi_i$ -regret experienced by each agent  $i$  approaches the negative orthant.

This work was originally motivated by an attempt to design a no-regret learning scheme that would converge to a tighter solution concept than the set of correlated equilibria (e.g., the convex hull of the set of Nash equilibria). We imagined that by comparing an agent's play to a larger

---

1. In this paper, we use the terms “agent” and “player” interchangeably.

2. In fact, Hart and Mas-Colell [18] establish this convergence result for what they call “better-play” algorithms (see Section 7), but their proof extends immediately to the class of no-external-regret learning algorithms.

set of alternative strategies, we could perhaps design a more powerful algorithm than no-internal-regret learning. Perhaps surprisingly, we find that the strongest form of no- $\Phi$ -regret algorithms are no-internal-regret algorithms. Consequently, the tightest game-theoretic solution concept to which no- $\Phi$ -regret algorithms converge is correlated equilibrium.

This paper is organized as follows. In the next section, we formally define no- $\Phi$ -regret learning and we show that no external regret and no internal regret are special cases of no  $\Phi$ -regret. We prove our existence theorem in Section 3. In Section 4, we define the notion of  $\vec{\Phi}$ -equilibrium; then, in Section 5, we establish necessary and sufficient conditions for convergence to the set of  $\vec{\Phi}$ -equilibria. In Section 6, we show that no internal regret is the strongest form of no  $\Phi$ -regret.

## 2. $\Phi$ -Regret

Our goal in this section is to define no- $\Phi$ -regret learning in the framework of Blackwell’s approachability theory. We start by reviewing Blackwell’s framework, and stating the variant of Blackwell’s theorem on which our existence theorem is based (see Greenwald *et al.* [14] for details). Next, we introduce the notion of an action transformation: a mapping from a pure strategy to a mixed strategy. Finally, for finite sets  $\Phi$  of action transformations, we formally define no- $\Phi$ -regret learning.

### 2.1 Blackwell’s Approachability Theory

Consider an agent with a finite set of actions  $A$  playing a game against a set of opponents who play actions in the (arbitrary) joint action space  $A'$ . (The opponents’ joint action space can be interpreted as the product of independent action sets.) Associated with each possible outcome is a vector given by the function  $\rho : A \times A' \rightarrow V$ , where  $V$  is a vector space over  $\mathbb{R}$  with an inner product  $\cdot$  and a distance metric  $d$  defined by the inner product in the standard manner: i.e.,  $d(x, y) = \|x - y\|_2$ , for all  $x, y \in V$ .

A *vector-valued game* is a 4-tuple  $\Gamma = (A, A', V, \rho)$ . We study *infinitely-repeated* vector-valued games  $\Gamma^\infty$  in which the agent interacts with its opponents repeatedly and indefinitely. Recall that the agent’s action set  $A$  is assumed to be finite. We denote by  $\Delta(A)$  the set of probability distributions over the set  $A$ , and we allow agents to play *mixed strategies*, which means that rather than selecting an action  $a \in A$  to play at each round, the agent selects a probability distribution  $q \in \Delta(A)$ . More specifically, an arbitrary round  $t$  (for  $t \geq 1$ ) proceeds as follows:

1. the agent selects a mixed strategy  $q_t \in \Delta(A)$ ,
2. the agent plays an action  $a_t \in A$  (which is sampled according to the distribution  $q_t$ ); simultaneously, the opponents play action  $a'_t$
3. the agent observes reward vector  $\rho(a_t, a'_t) \in V$

Given an infinitely-repeated vector-valued game  $\Gamma^\infty$  the *set of action histories of length  $t$* , for  $t \geq 0$ , is denoted by  $H_t$ . For  $t \geq 1$ ,  $H_t$  is given by  $(A \times A')^t$ : e.g.,  $h = \{a_\tau, a'_\tau\}_{\tau=1}^t \in H_t$ . The set  $H_0$  is defined to be a singleton. Given an infinitely-repeated vector-valued game  $\Gamma^\infty$ , a *learning algorithm* is a sequence of functions  $\mathcal{L} = \{L_t\}_{t=1}^\infty$ , where  $L_t : H_{t-1} \rightarrow \Delta(A)$ .

Salient examples of learning algorithms include the best-reply heuristic [7] and fictitious play [4, 25]. At time  $t$ , the former returns an element of  $\Delta(A)$  that maximizes the agent’s rewards with respect to only  $a'_{t-1}$ , while the latter returns an element of  $\Delta(A)$  that maximizes the agent’s rewards with respect to the empirical distribution of play through time  $t - 1$ .

We are interested in the properties of learning algorithms employed by an agent playing an infinitely-repeated vector-valued game  $\Gamma^\infty$ . Given a learning algorithm  $\mathcal{L} = \{L_t\}_{t=1}^\infty$  together with

a sequence of opposing actions  $a'_1, a'_2, \dots \in A'$ , we define a probability space whose universe consists of sequences of the agent's actions and whose measure can be defined inductively:

$$P[a_t = \alpha \mid a_\tau = \alpha_\tau, \forall \tau = 1, \dots, t-1] = L_t((\alpha_1, a'_1), \dots, (\alpha_{t-1}, a'_{t-1}))(\alpha) \quad (1)$$

for all  $\alpha \in A$ . In this probability space, we define two sequences of random variables: cumulative rewards  $R_t = \sum_{\tau=1}^t \rho(a_\tau, a'_\tau)$  and average rewards  $\bar{\rho}_t = \frac{R_t}{t}$ .

Now, following Blackwell, we define the notion of approachability as follows:

**Definition 1 (Approachability)** *Given an infinitely-repeated vector-valued game  $\Gamma^\infty$ , a set  $U \subseteq V$ , and a learning algorithm  $\mathcal{L}$ , the set  $U$  is said to be approachable by  $\mathcal{L}$ , if for all  $\epsilon > 0$ , there exists  $t_0$  such that for any sequence of opposing actions  $a'_1, a'_2, \dots$ ,  $P[\exists t \geq t_0 \text{ s.t. } d(U, \bar{\rho}_t) \geq \epsilon] < \epsilon$ .*

Hence, if a learning algorithm  $\mathcal{L}$  approaches a set  $U \subseteq V$ , then  $d(U, \bar{\rho}_t) \rightarrow 0$  almost surely.

The following theorem [14, 20] gives a sufficient condition for the negative orthant, that is, the set  $\mathbb{R}_-^d = \{x \in \mathbb{R}^d \mid x_i \leq 0, \text{ for all } 1 \leq i \leq d\} \subseteq \mathbb{R}^d$ , to be approachable by a learning algorithm  $\mathcal{L}$  in an infinitely-repeated vector-valued game  $(A, A', \mathbb{R}^d, \rho)^\infty$  where  $d \in \mathbb{N}$  and  $\rho(A \times A')$  is bounded. For  $x \in \mathbb{R}^d$ , define  $x^+$  by  $(x^+)_i = \max\{x_i, 0\}$ , for all  $1 \leq i \leq d$ .

**Theorem 2 (Jafari, 2003)** *Given an infinitely-repeated vector-valued game  $(A, A', \mathbb{R}^d, \rho)^\infty$  with  $d \in \mathbb{N}$  and  $\rho(A \times A')$  bounded and a learning algorithm  $\mathcal{L} = \{L_t\}_{t=1}^\infty$ , the negative orthant  $\mathbb{R}_-^d \subseteq \mathbb{R}^d$  is approachable by  $\mathcal{L}$  if there exists a constant  $c \in \mathbb{R}$  such that for all times  $t \geq 1$ , for all action histories  $h \in H_{t-1}$  of length  $t-1$ , and for all opposing actions  $a'$ ,*

$$(R_{t-1}(h))^+ \cdot \rho(L_t(h), a') \leq c \quad (2)$$

where  $R_t(h) \equiv \sum_{\tau=1}^t \rho(a_\tau, a'_\tau)$  and  $\rho(q, a') \equiv \sum_{a \in A} q(a) \rho(a, a')$ .

Blackwell's seminal approachability theorem provides a sufficient condition to ensure that, in a vector-valued repeated game, a learner's average rewards approach any closed set  $U \subseteq \mathbb{R}^n$  [2, 18]. To prove existence of no- $\Phi$ -regret algorithms, we rely on Theorem 2, a close cousin of Blackwell's theorem. On the one hand, our theorem specializes Blackwell's theorem: it provides a sufficient condition for the negative orthant  $\mathbb{R}_-^n \subseteq \mathbb{R}^n$  to be approachable, rather than an arbitrary closed subset of Euclidean space. On the other hand, our sufficient condition (Equation 2) is weaker than Blackwell's original condition: our condition need only hold for some  $c \in \mathbb{R}$ , rather than precisely for  $c = 0$ . Moreover, in our framework, the opponents (i.e., not the learner) have at their disposal an arbitrary, rather than merely a finite, set of actions.

## 2.2 Action Transformations

An *action transformation* is a function  $\phi : A \rightarrow \Delta(A)$ . Let  $\Phi_{\text{ALL}}(A)$  denote the set of all action transformations over the set  $A$ . Following Blum and Mansour [3], we let  $\Phi_{\text{SWAP}}(A) \subseteq \Phi_{\text{ALL}}(A)$  denote the set of all action transformations that map actions to distributions with all their weight on a single action (i.e., pure strategies).

There are two well-studied subsets of  $\Phi_{\text{SWAP}}(A)$ , namely, external and internal action transformations. Let  $\delta_a \in \Delta(A)$  denote the distribution with all its weight on action  $a$ . An *external* action transformation is simply a constant transformation, so for  $a \in A$ ,

$$\phi_{\text{EXT}}^{(a)} : x \mapsto \delta_a, \quad \text{for all } x \in A \quad (3)$$

An *internal* action transformation behaves like the identity, except on one particular input, so for  $a, b \in A$

$$\phi_{\text{INT}}^{(a,b)} : x \mapsto \begin{cases} \delta_b & \text{if } x = a \\ \delta_x & \text{otherwise} \end{cases} \quad (4)$$

The set of external and internal action transformations are denoted by  $\Phi_{\text{EXT}}(A)$  and  $\Phi_{\text{INT}}(A)$ , respectively. Observe that  $|\Phi_{\text{INT}}(A)| = |A|^2 - |A| + 1$  and  $|\Phi_{\text{EXT}}(A)| = |A|$ .

We can represent an action transformation by a stochastic matrix. Given  $\phi \in \Phi_{\text{ALL}}(A)$  and an enumeration of  $A$ , we define its matrix representation  $[\phi]$  as:

$$[\phi]_{ij} = \phi(a_i)(a_j) \quad (5)$$

where  $a_k$  is the  $k$ th action in the enumeration. For example, for  $A = \{1, 2, 3, 4\}$ , the action transformations  $\phi_{\text{EXT}}^{(2)}$  and  $\phi_{\text{INT}}^{(23)}$  can be represented as:

$$[\phi_{\text{EXT}}^{(2)}] = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad [\phi_{\text{INT}}^{(23)}] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

### 2.3 No $\Phi$ -Regret Learning

A *real-valued* game  $(A, A', R, r)$  is a vector-valued game with  $R \subseteq \mathbb{R}$ . Given a real-valued game  $\Gamma = (A, A', R, r)$  and a set of action transformations  $\Phi \subseteq \Phi_{\text{ALL}}(A)$ , we define the (vector-valued)  $\Phi$ -*regret game* as  $\Gamma^\Phi = (A, A', \mathbb{R}^\Phi, \rho^\Phi)$ , with the vector-valued function  $\rho^\Phi : A \times A' \rightarrow \mathbb{R}^\Phi$  given by:<sup>3</sup>

$$\rho^\Phi(a, a') \equiv \left( \rho^\phi(a, a') \right)_{\phi \in \Phi} \quad (6)$$

where

$$\rho^\phi(a, a') = r(\phi(a), a') - r(a, a') \quad (7)$$

Here,  $r(q, a') = \sum_{a \in A} q(a)r(a, a')$ , for all  $q \in \Delta(A)$ . In words, the  $\phi$ th entry in the regret vector  $\rho^\Phi(a, a')$  describes the difference between the rewards the agent obtains by playing action  $a$  and the rewards the agent would have expected to obtain by playing the mixed strategy  $\phi(a)$  instead, given opposing action  $a'$ .

We now define no- $\Phi$ -regret learning in Blackwell's approachability framework:

**Definition 3 (No- $\Phi$ -Regret Learning)** *Given a real-valued game  $\Gamma = (A, A', R, r)$  and a finite set of action transformations  $\Phi \subseteq \Phi_{\text{ALL}}(A)$ , a no- $\Phi$ -regret learning algorithm  $\mathcal{L}$  is one that approaches the negative orthant  $\mathbb{R}_-^\Phi \subseteq \mathbb{R}^s$  in the infinitely-repeated  $\Phi$ -regret game  $(\Gamma^\Phi)^\infty$ : i.e., for all  $\epsilon > 0$ , there exists  $t_0$  such that for any sequence of opposing actions  $a'_1, a'_2, \dots$ ,*

$$P [\exists t \geq t_0 \text{ s.t. } d(\mathbb{R}_-^\Phi, \bar{\rho}_t^\Phi) \geq \epsilon] < \epsilon \quad (8)$$

In words, if an agent plays an infinitely-repeated game  $\Gamma^\infty$  as prescribed by a no- $\Phi$ -regret learning algorithm, then the time-averaged  $\Phi$ -regret experienced by the agent converges to the negative orthant with probability 1, regardless of the sequence of opposing actions. Moreover, this convergence is uniform.

3. Given two set  $X$  and  $Y$ , the notation  $X^Y$  denotes the set of functions  $\{f : Y \rightarrow X\}$ .

Note that, if  $Y$  is finite, then  $\mathbb{R}^Y$  is isomorphic to  $\mathbb{R}^{|Y|}$ .

### 3. Existence of No- $\Phi$ -Regret Learning Algorithms

Indeed, no- $\Phi$ -regret learning algorithms exist. In particular, no-external-regret algorithms pervade the literature. The earliest date back to Blackwell [2] and Hannan [16]; but, more recently, Littlestone and Warmuth [22], Freund and Schapire [11], Herbster and Warmuth [19], Fudenberg and Levine [12], Foster and Vohra [8], Hart and Mas-Colell [18], and others have studied such algorithms. Foster and Vohra [10] were the first to design an algorithm that exhibits no-internal-regret.

Our first theorem establishes the existence of no- $\Phi$ -regret learning algorithms, for all finite  $\Phi$ .

**Theorem 4 (Existence)** *Given a real-valued game  $\Gamma = (A, A', R, r)$  with  $R \subseteq \mathbb{R}$  bounded, for all finite sets of action transformations  $\Phi \subseteq \Phi_{ALL}(A)$ , there exists a no- $\Phi$ -regret learning algorithm: i.e., one that approaches  $\mathbb{R}_-^\Phi$  in the infinitely-repeated  $\Phi$ -regret game  $\Gamma_\Phi^\infty$ .*

**Proof** By Theorem 2, it suffices to show that there exists a learning algorithm  $\mathcal{L} = \{L_t\}_{t=1}^\infty$  and a constant  $c \in \mathbb{R}$  such that for all times  $t \geq 1$ , for all action histories  $h \in H^{t-1}$  of length  $t-1$ , and for all opposing actions  $a'$ ,

$$(R_{t-1}^\Phi(h))^+ \cdot \rho^\Phi(L_t(h), a') \leq c \quad (9)$$

where  $R_t^\Phi(h) = \sum_{\tau=1}^t \rho^\Phi(a_\tau, a'_\tau)$  and  $\rho^\Phi(q, a') \equiv \sum_{a \in A} q(a) \rho^\Phi(a, a')$ .

Case 1.  $R_{t-1}^\Phi(h) \in \mathbb{R}_-^\Phi$ : If  $R_{t-1}^\Phi(h) \in \mathbb{R}_-^\Phi$ , so that  $(R_{t-1}^\Phi(h))^+ = 0$ , Equation 9 holds for  $c = 0$ .

Case 2.  $R_{t-1}^\Phi(h) \notin \mathbb{R}_-^\Phi$ : We show that for all  $x^\Phi \notin \mathbb{R}_-^\Phi$  there exists  $q \equiv q(x^\Phi) \in \Delta(A)$  such that for all  $a' \in A'$ ,  $(x^\Phi)^+ \cdot \rho^\Phi(q, a') = 0$ . Then, letting  $L_t(h) = q(R_{t-1}^\Phi(h))$ , Equation 9 holds for  $c = 0$ .

Let  $q$  be the (row) vector representation of a mixed strategy.

$$0 = (x^\Phi)^+ \cdot \rho^\Phi(q, a') \quad (10)$$

$$= \sum_{\phi \in \Phi} (x^\phi)^+ (r(q[\phi], a') - r(q, a')) \quad (11)$$

$$= \sum_{\phi \in \Phi} (x^\phi)^+ \left( \sum_{a \in A} r(a, a') (q[\phi])_a - \sum_{a \in A} r(a, a') q_a \right) \quad (12)$$

$$= \sum_{a \in A} r(a, a') \left[ \left( q \sum_{\phi \in \Phi} (x^\phi)^+ [\phi] \right)_a - \left( q \sum_{\phi \in \Phi} (x^\phi)^+ \right)_a \right] \quad (13)$$

Equation 11 follows from the definitions of the inner product and  $\rho^\Phi$ . Equation 12 follows from the definition of expectation. Equation 13 follows via algebra.

Now it suffices to show the following:

$$q \sum_{\phi \in \Phi} (x^\phi)^+ [\phi] = q \sum_{\phi \in \Phi} (x^\phi)^+ \quad (14)$$

Define the matrix  $M$  as follows:

$$M = \frac{\sum_{\phi \in \Phi} (x^\phi)^+ [\phi]}{\sum_{\phi \in \Phi} (x^\phi)^+} \quad (15)$$

Since  $M$  is a convex combination of stochastic matrices,  $M$  itself is a stochastic matrix with at least one fixed point with non-negative entries that sum to 1. Any algorithm for computing such a fixed point of  $M$  gives rise to a no- $\Phi$ -regret learning algorithm.  $\blacksquare$

---

**Algorithm 1** No-Regret Learning Algorithm  $((A, A', R, r), \Phi \subseteq \Phi_{\text{ALL}}(A))$ 

---

```
1: initialize  $x_0 = 0$ 
2: for  $t = 1, 2, \dots$ , do
3:   sample pure action  $a \sim q_t$ 
4:   choose opposing actions  $a'_t \in A'$ 
5:   observe reward vector  $r_t = r(\cdot, a'_t) \in R^A$ 
6:   for all  $\phi \in \Phi$  do
7:     compute instantaneous regret  $y_t^\phi = r_t \cdot e_a[\phi] - r_t \cdot e_a$ 
8:     update cumulative regret vector  $x_t^\phi = x_{t-1}^\phi + y_t^\phi$ 
9:   end for
10:  if  $(x_t^\Phi)^+ = 0$  then
11:    set  $q_{t+1} \in \Delta(A)$  arbitrarily
12:  else
13:    let  $M_t = \sum_{\phi \in \Phi} (x_t^\phi)^+ [\phi] / \sum_{\phi \in \Phi} (x_t^\phi)^+$ 
14:    solve for a fixed point  $q_{t+1} = q_{t+1} M_t$ 
15:  end if
16: end for
```

---

Algorithm 1 lists the steps in the no- $\Phi$ -regret learning algorithm derived in the proof of the existence theorem. At time  $t$ , the agent plays the mixed strategy  $q_t$  by sampling a pure action  $a$  according to the distribution  $q_t$ , after which it observes an  $|A|$ -dimensional reward vector  $r_t$ , where  $(r_t)_a = r(a, a'_t)$ , assuming  $a'_t$  is the opponents' pure action vector at time  $t$ . Given this reward vector, the agent computes its instantaneous regret in all dimensions  $\phi \in \Phi$ : specifically,  $\rho^\phi(a_t, a'_t) = r(\phi(a_t), a'_t) - r(a_t, a'_t)$ , which, since  $r(q, a')$  is an expectation, we compute via dot products in Step 7. The cumulative regret vector is then updated accordingly, after which its positive part is extracted. If this quantity is zero, then the algorithm outputs an arbitrary mixed strategy. Otherwise, a fixed point of the stochastic matrix  $M$  derived in Equation 15 is returned.

**Complexity** Each iteration of Algorithm 1 has time complexity  $O(\max\{|\Phi||A|^2, |A|^3\})$ . Updating the cumulative regret vector in steps 6–9 takes time  $O(|\Phi||A|)$ , since computing instantaneous regret for each  $\phi \in \Phi$  (step 7) is an  $O(|A|)$  operation. Computing the stochastic matrix  $M$  in step 13 takes time  $O(|\Phi||A|^2)$ , since each matrix  $[\phi]$  has dimensions  $|A| \times |A|$ . Finding the fixed point of an  $n \times n$  stochastic matrix (step 14), which can be accomplished, for example, via Gaussian elimination, takes  $O(n^3)$  time.

If, however,  $\Phi \subseteq \Phi_{\text{SWAP}}(A)$ , then the time complexity reduces to  $O(\max\{|\Phi||A|, |A|^3\})$ , since in this case, (i) computing instantaneous regret for each  $\phi \in \Phi$  (step 7) takes constant time so that updating the cumulative regret vector takes time  $O(|\Phi|)$ ; and (ii) computing the stochastic matrix  $M$  in step 13 is only an  $O(|\Phi||A|)$  operation, since there are only  $|A|$  nonzero entries in each  $\phi \in \Phi$ . In particular, if  $\Phi = \Phi_{\text{INT}}(A)$ , then the time complexity reduces to  $O(|A|^3)$ , because  $|\Phi_{\text{INT}}(A)| = O(|A|^2)$ . Moreover, if  $\Phi = \Phi_{\text{EXT}}(A)$ , then the time complexity reduces even further to  $O(|A|)$ , because matrix manipulation is not required in the special case of no-external-regret learning. The rows of  $M$  are constant: each is a copy of the (normalized) cumulative regret vector, which is precisely the fixed point of  $M$ .

The space complexity of Algorithm 1 is  $O(|\Phi||A|^2) = O(\max\{|\Phi||A|^2, |A|^2\})$  because it is necessary to store  $|\Phi|$  matrices, each with dimensions  $|A| \times |A|$ , and computing the fixed point of an  $|A| \times |A|$  stochastic matrix (via Gaussian elimination) requires  $O(|A|^2)$  space. If, however,

$\Phi \subseteq \Phi_{\text{SWAP}}(A)$ , then the space complexity reduces to  $O(\max\{|\Phi||A|, |A|^2\})$ , since, in this case, there are only  $|A|$  nonzero entries in each  $\phi \in \Phi$ . In particular, if  $\Phi = \Phi_{\text{INT}}(A)$  then the space complexity reduces to  $O(|A|^2)$ , since it suffices to store cumulative regrets in a matrix of size  $|A| \times |A|$ . Similarly, if  $\Phi = \Phi_{\text{EXT}}(A)$ , then the space complexity reduces to  $O(|A|)$ , since it suffices to store cumulative regrets in a vector of size  $|A|$ . Our discussion of the time and space complexity of Algorithm 1 is summarized in Table 1.

	Time	Space
$\Phi \subseteq \Phi_{\text{ALL}}$	$O(\max\{ \Phi  A ^2,  A ^3\})$	$O( \Phi  A ^2)$
$\Phi \subseteq \Phi_{\text{SWAP}}$	$O(\max\{ \Phi  A ,  A ^3\})$	$O(\max\{ \Phi  A ,  A ^2\})$
$\Phi = \Phi_{\text{INT}}$	$O( A ^3)$	$O( A ^2)$
$\Phi = \Phi_{\text{EXT}}$	$O( A )$	$O( A )$

Table 1: Complexity of No- $\Phi$ -Regret Learning

#### 4. $\vec{\Phi}$ -Equilibria

In this section, we define the notion of  $\vec{\Phi}$ -equilibria, of which correlated, Nash, and minimax equilibria are all special cases. We show that the set of  $\vec{\Phi}$ -equilibria is convex, for all  $\vec{\Phi}$ .

In a (real-valued)  $n$ -player game  $\Gamma_n = \langle (A_i, r_i)_{1 \leq i \leq n} \rangle$ , each player  $i$  chooses an action from the finite set  $A_i$ , and the rewards earned by player  $i$  are determined by the function  $r_i : A_1 \times \dots \times A_n \rightarrow \mathbb{R}$ . We abbreviate action profile  $(a_1, \dots, a_n)$  by  $(a_i, a_{-i}) \in A_i \times \prod_{j \neq i} A_j$ .

Given an  $n$ -player game  $\Gamma_n$ , an action transformation  $\phi \in \Phi_{\text{ALL}}(A_i)$  can be extended to a linear map  $\tilde{\phi} : \Delta(A_1 \times \dots \times A_n) \rightarrow \Delta(A_1 \times \dots \times A_n)$  as follows:

$$\tilde{\phi}(q)(a_i, a_{-i}) \equiv \sum_{b_i \in A_i} q(b_i, a_{-i}) \phi(b_i)(a_i) \quad (16)$$

It is easily verified that  $\tilde{\phi}$  is indeed a probability distribution.

**Definition 5 ( $\vec{\Phi}$ -Equilibrium)** *Given an  $n$ -player game  $\Gamma_n = \langle (A_i, r_i)_{1 \leq i \leq n} \rangle$  and a vector  $\vec{\Phi} = (\Phi_i)_{1 \leq i \leq n}$  such that  $\Phi_i \subseteq \Phi_{\text{ALL}}(A_i)$  for  $1 \leq i \leq n$ , an element  $q \in \Delta(A_1 \times \dots \times A_n)$  is called a  $\vec{\Phi}$ -equilibrium if  $r_i(q) \geq r_i(\tilde{\phi}(q))$ , for all players  $i$  and for all  $\phi \in \Phi_i$ .*

If for all players  $i$ , each  $\Phi_i$  is of the same type, e.g.,  $\Phi_i = \Phi_{\text{EXT}}(A_i)$ , then we refer to the  $\vec{\Phi}$ -equilibrium accordingly, e.g.,  $\Phi_{\text{EXT}}$ -equilibrium.

##### 4.1 Examples of $\vec{\Phi}$ -Equilibria

Correlated, Nash, and minimax equilibria are all special cases of  $\vec{\Phi}$ -equilibria.

**Correlated Equilibrium** Given an  $n$ -player game  $\Gamma_n = \langle (A_i, r_i)_{1 \leq i \leq n} \rangle$  let  $\Phi_i = \Phi_{\text{INT}}(A_i)$  for all players  $i$ . For  $q \in \Delta(A_1 \times \dots \times A_n)$ , the expression  $r_i(\tilde{\phi}_{\text{INT}}^{(\alpha\beta)}(q))$  simplifies as follows: for  $\alpha, \beta \in A_i$ ,

$$r_i(\tilde{\phi}_{i,\text{INT}}^{(\alpha\beta)}(q)) \quad (17)$$

$$= \sum_{a_i, a_{-i}} r_i(a_i, a_{-i}) \sum_{b_i \in A_i} q(b_i, a_{-i}) \phi_{\text{INT}}^{(\alpha\beta)}(b_i)(a_i) \quad (18)$$

$$= \sum_{a_i, a_{-i}} r_i(a_i, a_{-i}) \sum_{b_i \in A_i} q(b_i, a_{-i}) \left\{ \begin{array}{l} \delta_\beta \quad \text{if } b_i = \alpha \\ \delta_{b_i} \quad \text{otherwise} \end{array} \right\} (a_i) \quad (19)$$

$$= \sum_{a_i, a_{-i}} r_i(a_i, a_{-i}) \sum_{b_i \in A_i} q(b_i, a_{-i}) \left\{ \begin{array}{l} \mathbf{1}_{a_i=\beta} \quad \text{if } b_i = \alpha \\ \mathbf{1}_{a_i=b_i} \quad \text{otherwise} \end{array} \right. \quad (20)$$

$$= \sum_{a_i, a_{-i}} r_i(a_i, a_{-i}) \left[ \sum_{b_i=\alpha} q(b_i, a_{-i}) \mathbf{1}_{a_i=\beta} + \sum_{b_i \neq \alpha} q(b_i, a_{-i}) \mathbf{1}_{a_i=b_i} \right] \quad (21)$$

$$= \sum_{a_i, a_{-i}} r_i(a_i, a_{-i}) [q(\alpha, a_{-i}) \mathbf{1}_{a_i=\beta} + q(a_i, a_{-i}) \mathbf{1}_{a_i \neq \alpha}] \quad (22)$$

$$= \sum_{a_i, a_{-i}} r_i(a_i, a_{-i}) q(\alpha, a_{-i}) \mathbf{1}_{a_i=\beta} + \left( \sum_{a_i, a_{-i}} r_i(a_i, a_{-i}) q(a_i, a_{-i}) - \sum_{a_i, a_{-i}} r_i(a_i, a_{-i}) q(a_i, a_{-i}) \mathbf{1}_{a_i=\alpha} \right)$$

$$= \sum_{a_{-i}} r_i(\beta, a_{-i}) q(\alpha, a_{-i}) + \sum_{a_i, a_{-i}} r_i(a_i, a_{-i}) q(a_i, a_{-i}) - \sum_{a_{-i}} r_i(\alpha, a_{-i}) q(\alpha, a_{-i}) \quad (23)$$

Therefore, since

$$r_i(q) = \sum_{a_i, a_{-i}} r_i(a_i, a_{-i}) q(a_i, a_{-i}) \quad (24)$$

it follows that

$$r_i(q) - r_i(\tilde{\phi}_{\text{INT}}^{(\alpha\beta)}(q)) \geq 0 \quad \forall \alpha, \beta \in A_i \quad (25)$$

$$\text{iff } \sum_{a_{-i}} q(\alpha, a_{-i}) (r_i(\alpha, a_{-i}) - r_i(\beta, a_{-i})) \geq 0 \quad \forall \alpha, \beta \in A_i \quad (26)$$

Equation 26, holding for all players  $i$ , is precisely the definition of correlated equilibrium [1].

**Coarse Correlated Equilibrium** Given an  $n$ -player game  $\Gamma_n = \langle (A_i, r_i)_{1 \leq i \leq n} \rangle$ , let  $\Phi_i = \Phi_{\text{EXT}}(A_i)$  for all players  $i$ . For  $q \in \Delta(A_1 \times \dots \times A_n)$ , the expression  $r_i(\tilde{\phi}_{\text{EXT}}^{(\alpha)}(q))$  simplifies as follows: for  $\alpha \in A_i$ ,

$$r_i(\tilde{\phi}_{i,\text{EXT}}^{(\alpha)}(q)) = \sum_{a_i, a_{-i}} r_i(a_i, a_{-i}) \sum_{b_i \in A_i} q(b_i, a_{-i}) \phi_{\text{EXT}}^{(\alpha)}(b_i)(a_i) \quad (27)$$

$$= \sum_{a_i, a_{-i}} r_i(a_i, a_{-i}) \sum_{b_i \in A_i} q(b_i, a_{-i}) \delta_\alpha(a_i) \quad (28)$$

$$= \sum_{a_i, a_{-i}} r_i(a_i, a_{-i}) \mathbf{1}_{\alpha=a_i} \sum_{b_i \in A_i} q(b_i, a_{-i}) \quad (29)$$

$$= \sum_{a_{-i}} r_i(\alpha, a_{-i}) \sum_{b_i \in A_i} q(b_i, a_{-i}) \quad (30)$$

$$= \sum_{a_{-i}} r_i(\alpha, a_{-i}) q_{-i}(a_{-i}) \quad (31)$$

$$= r_i(\alpha, q_{-i}) \quad (32)$$

Here,  $q_{-i} \in \Delta\left(\prod_{j \neq i} A_j\right)$ . Thus,  $q \in \Delta(A_1 \times \dots \times A_n)$  is a  $\Phi_{\text{EXT}}$ -equilibrium if and only if  $r_i(q) \geq r_i(\alpha, q_{-i})$  for all players  $i$  and for all  $\alpha \in A_i$ , which is the definition of coarse correlated equilibrium (also called weak correlated equilibrium) [23].

**Nash Equilibrium** Given an  $n$ -player game  $\Gamma_n = \langle (A_i, r_i)_{1 \leq i \leq n} \rangle$ , a Nash equilibrium [24] is an independent element  $q \in \Delta(A_1 \times \dots \times A_n)$  such that  $r(q) \geq r(q_1, \dots, a_i, \dots, q_n)$ , for all players  $i$  and for all actions  $a_i \in A_i$ . An element  $q \in \Delta(A_1 \times \dots \times A_n)$  is called *independent* if it can be written as the product of  $n$  independent elements  $q_i \in \Delta(A_i)$ : i.e.,  $q = q_1 \times \dots \times q_n$ . Thus, by definition, a Nash equilibrium is an independent coarse correlated equilibrium. However, a Nash equilibrium is also an independent correlated equilibrium. Therefore, the set of independent coarse correlated equilibria and independent correlated equilibria coincide.

In general, however, a coarse correlated equilibrium need not be a correlated equilibrium. This observation is intuitive for general-sum games, but perhaps less so for zero-sum games. In the following zero-sum game, with row as maximizer and column as minimizer, the joint distribution with half its weight on (T,L) and the other half on (B,M) is a coarse correlated equilibrium, but not a correlated equilibrium. It is a coarse correlated equilibrium because row has no incentive to deviate from its marginal distribution (half its weight on T and half on B), and column has no incentive to deviate from its marginal distribution (half its weight on L and half on M). If column were to deviate to R, it would expect to lose  $\frac{1}{2}$  instead of 0. It is not, however, a correlated equilibrium: if column is advised to play L, then row is playing T, in which case column actually prefers to play R, where it would win 1 instead of 0.

	L	M	R
T	0	0	-1
B	0	0	2

Figure 1: Sample Zero-Sum Game.

**Zero-Sum Games** In the case of two-player, zero-sum games, we obtain the following results for coarse correlated equilibria (and consequently correlated equilibria):

**Proposition 6** *Given a two-player, zero-sum game  $\Gamma$  with reward function  $r$  and value  $v$ . If  $q$  is a coarse correlated equilibrium, then (i)  $r(q) = v$  and (ii) each player's marginal distribution is an optimal strategy (i.e., optimal for the maximizing player means: guarantees he wins at least  $v$ ; optimal for the minimizing player means: guarantees he loses at most  $v$ ).*

**Proof** Let  $q_1$  and  $q_2$  denote the marginal distributions of the maximizer and the minimizer in  $q$ , respectively. First,  $r(q) \geq \max_{\alpha \in A_1} r(\alpha, q_2) \geq v$  since  $q$  is a coarse correlated equilibrium and  $v$  is the value of the game. Symmetrically,  $r(q) \leq \max_{\beta \in A_2} r(q_1, \beta) \leq v$ . Hence,  $r(q) = v$ .

Second, applying the definition of coarse correlated equilibrium again together with the above result,  $v = r(q) \geq \max_{\alpha \in A_1} r(\alpha, q_2)$ , so by playing  $q_2$ , player 2 loses at most  $v$ . Symmetrically,  $v = r(q) \leq \max_{\beta \in A_2} r(q_1, \beta)$ , so by playing  $q_1$ , player 1 wins at least  $v$ . ■

Note that the sets of coarse correlated equilibria and minimax equilibria need not coincide, since the former allows for correlations while the latter does not. For this reason, we refer to coarse correlated equilibria in two-player, zero-sum games as *generalized* minimax equilibria.

## 4.2 Properties of $\vec{\Phi}$ -Equilibrium

Next we discuss two convexity properties of the set of  $\vec{\Phi}$ -equilibria.

**Proposition 7** *Given an  $n$ -player game  $\Gamma_n = \langle (A_i, r_i)_{1 \leq i \leq n} \rangle$  and a vector  $\vec{\Phi} = (\Phi_i)_{1 \leq i \leq n}$  such that  $\Phi_i \subseteq \Phi_{ALL}(A_i)$  for  $1 \leq i \leq n$ , the set of  $\vec{\Phi}$ -equilibria is convex.*

**Proof** If  $q, q' \in \Delta(A_1 \times \dots \times A_n)$  are both  $\vec{\Phi}$ -equilibria, then  $r_i(q) \geq r_i(\tilde{\phi}(q))$  and  $r_i(q') \geq r_i(\tilde{\phi}(q'))$ , for all players  $i$  and for all  $\tilde{\phi} \in \Phi_i$ . Because  $r_i$  and  $\tilde{\phi}$  are linear, it follows that

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

for all  $\alpha \in [0, 1]$  and for all players  $i$ . Thus,  $\alpha q + (1 - \alpha)q'$  is a  $\vec{\Phi}$ -equilibrium.  $\blacksquare$

Given a set of actions  $A$ , let  $I$  denote the identity map: i.e.,  $I(a) = \delta_a$  for all  $a \in A$ .

**Definition 8** *Given a set of actions  $A$  and a set of action transformations  $\Phi \subseteq \Phi_{ALL}(A)$ , we define the super convex hull of  $\Phi$ , denoted  $\text{SCH}(\Phi)$ , as follows:*

$$\text{SCH}(\Phi) = \left\{ \left( \sum_{j=1}^k \alpha_j \phi_j \right) + \beta I \mid k \in \mathbb{N}, \phi_j \in \Phi, \alpha_j \geq 0, \beta \in \mathbb{R}, \text{ and } \sum_{j=1}^k \alpha_j + \beta = 1 \right\} \quad (33)$$

**Proposition 9** *Given an  $n$ -player game  $\Gamma_n = \langle (A_i, r_i)_{1 \leq i \leq n} \rangle$  and a vector  $\vec{\Phi} = (\Phi_i)_{1 \leq i \leq n}$  such that  $\Phi_i \subseteq \Phi_{ALL}(A_i)$  for  $1 \leq i \leq n$ , if  $q$  is a  $\vec{\Phi}$ -equilibrium, then  $q$  is also a  $\vec{\Phi}'$ -equilibrium, where  $\vec{\Phi}' = (\text{SCH}(\Phi_i))_{1 \leq i \leq n}$ .*

**Proof** Let  $i$  be an arbitrary player and let  $\phi^*$  be an arbitrary element of  $\text{SCH}(\Phi_i)$ . Since  $q$  is a  $\vec{\Phi}$ -equilibrium,  $r_i(q) \geq r_i(\tilde{\phi}_i(q))$ , for all players  $i$  and for all  $\tilde{\phi}_i \in \Phi_i$ . Choose  $k \in \mathbb{N}$ ,  $\phi_j \in \Phi$  and  $\alpha_j \geq 0$  for all  $1 \leq j \leq k$ , and  $\beta \in \mathbb{R}$  such that  $\phi^* = \left( \sum_{j=1}^k \alpha_j \phi_j \right) + \beta I$  and  $\sum_{j=1}^k \alpha_j + \beta = 1$ . Because  $r_i$  and  $\tilde{\phi}_j$  are linear, it follows that

$$\begin{aligned} r_i(q) &= \sum_{j=1}^k \alpha_j r_i(q) + \beta r_i(q) \\ &\geq \sum_{j=1}^k \alpha_j r_i(\tilde{\phi}_j(q)) + \beta r_i(q) \\ &= r_i \left( \sum_{j=1}^k \alpha_j \tilde{\phi}_j(q) + \beta I(q) \right) \\ &= r_i(\tilde{\phi}^*(q)) \end{aligned}$$

$\blacksquare$

## 5. Convergence of No- $\vec{\Phi}$ -Regret Learning Algorithms

In this section, we establish a fundamental relationship between no-regret learning algorithms and game-theoretic equilibria. We prove that learning algorithms that satisfy no- $\vec{\Phi}$ -regret converge to the set of  $\vec{\Phi}$ -equilibria. We derive as corollaries of this theorem the following two specific results: no- $\Phi_{\text{EXT}}$ -regret algorithms (i.e., no-external-regret algorithms) converge to the set of  $\Phi_{\text{EXT}}$ -equilibria, which correspond to generalized minimax equilibria in zero-sum games; and no- $\Phi_{\text{INT}}$ -regret algorithms (i.e., no-internal-regret algorithms) converge to the set of  $\Phi_{\text{INT}}$ -equilibria, which correspond to correlated equilibria in general-sum games. This latter result is well-known [17]. By Proposition 6, we arrive at another known result, namely, in two-player, zero-sum games, if each player plays using a no-external-regret learning algorithm, then each player's empirical distribution of play converges to his set of minimax strategies [18].

In addition to giving sufficient conditions for convergence to the set of  $\vec{\Phi}$ -equilibria, we also give *necessary* conditions. We show that multiagent learning converges to the set of  $\vec{\Phi}$ -equilibria only if the time-averaged  $\Phi_i$ -regret experienced by each player  $i$  converges to the negative orthant.

Given an infinitely-repeated  $n$ -player game  $\Gamma_n^\infty$ , a *run* of the game is a sequence of action vectors  $\{\vec{a}_\tau\}_{\tau=1}^\infty$  with each  $\vec{a}_\tau \in A_1 \times \dots \times A_n$ . Given a run  $\{\vec{a}_\tau\}_{\tau=1}^\infty$  of  $\Gamma_n^\infty$ , the *empirical distribution of play through time  $t$* , denoted  $z_t$ , is the element of  $\Delta(A_1 \times \dots \times A_n)$  given by:

$$z_t(\vec{b}) = \frac{1}{t} \sum_{\tau=1}^t \mathbf{1}_{\vec{a}_\tau = \vec{b}} \quad (34)$$

where  $\mathbf{1}_{x=y}$  denotes the indicator function, which equals 1 whenever  $x = y$ , and 0 otherwise.

The results in this section rely on a technical lemma, the statement and proof of which appear in Appendix A. We apply this lemma via the following corollary, which relates the empirical distribution of play at equilibrium to the players' rewards at equilibrium.

**Corollary 10** *Given an  $n$ -player game  $\Gamma_n$  and a vector of sets of action transformations  $\vec{\Phi} = (\Phi_i)_{1 \leq i \leq n}$  such that  $\Phi_i \subseteq \Phi_{\text{ALL}}(A_i)$  for  $1 \leq i \leq n$ . If  $Z$  is the set of  $\vec{\Phi}$ -equilibria of  $\Gamma_n$ , then  $d(z_t, Z) \rightarrow 0$  as  $t \rightarrow \infty$  if and only if  $r_i(\tilde{\phi}_i(z_t)) - r_i(z_t) \rightarrow \mathbb{R}_-$  as  $t \rightarrow \infty$ , for all players  $i$  and for all action transformations  $\phi_i \in \Phi_i$ .*

**Proof** For all players  $i$  and action transformations  $\phi_i \in \Phi_i$ , let  $f_i^{\phi_i}(q) = r_i(\tilde{\phi}_i(q)) - r_i(q)$  and  $Z_i^{\phi_i} = \{q \in \Delta(A_1 \times \dots \times A_n) \mid f_i^{\phi_i}(q) \leq 0\}$ , for all  $q \in \Delta(A_1 \times \dots \times A_n)$ . The set of  $\vec{\Phi}$ -equilibria is thus  $Z = \bigcap_{1 \leq i \leq n} \bigcap_{\phi_i \in \Phi_i} Z_i^{\phi_i}$ . For each  $i$  and  $\phi_i$ , apply Lemma 17 to  $f_i^{\phi_i}$  and  $Z_i^{\phi_i}$  so that  $d(z_t, Z_i^{\phi_i}) \rightarrow 0$  as  $t \rightarrow \infty$  if and only if  $r_i(\tilde{\phi}_i(z_t)) - r_i(z_t) \rightarrow \mathbb{R}_-$  as  $t \rightarrow \infty$ . ■

In words, Corollary 10 states that the empirical distribution of play converges to the set of  $\vec{\Phi}$ -equilibria if and only if the rewards each player  $i$  obtains exceed the rewards player  $i$  could have expected to obtain by playing according to any of the action transformations  $\tilde{\phi}_i \in \Phi_i$  of the empirical distribution of play.

**Theorem 11** *Given an  $n$ -player game  $\Gamma_n$  and a vector of sets of action transformations  $\vec{\Phi} = (\Phi_i)_{1 \leq i \leq n}$  such that  $\Phi_i \subseteq \Phi_{\text{ALL}}(A_i)$  is finite for  $1 \leq i \leq n$ . As  $t \rightarrow \infty$ , the average  $\Phi_i$ -regret experienced by each player  $i$  through time  $t$  converges to the negative orthant if and only if the empirical distribution of play converges to the set of  $\vec{\Phi}$ -equilibria of  $\Gamma_n$ .*

**Proof** By Corollary 10, it suffices to show that, as  $t \rightarrow \infty$ , the average  $\Phi_i$ -regret through time  $t$  experienced by each player  $i$  converges to the negative orthant if and only if for all players  $i$  and for all  $\phi_i \in \Phi_i$ ,  $r_i(\tilde{\phi}_i(z_t)) - r_i(z_t) \rightarrow \mathbb{R}_-$  as  $t \rightarrow \infty$ .

First, for arbitrary player  $i$ ,

$$r_i(z_t) = \frac{1}{t} \sum_{\tau=1}^t r_i(a_{i,\tau}, a_{-i,\tau}) \quad (35)$$

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

$$r_i(\tilde{\phi}_i(z_t)) = \sum_{a_i, a_{-i}} \tilde{\phi}_i(z_t)(a_i, a_{-i}) r_i(a_i, a_{-i}) \quad (36)$$

$$= \sum_{a_i, a_{-i}} \sum_{b_i \in A_i} z_t(b_i, a_{-i}) \phi_i(b_i)(a_i) r_i(a_i, a_{-i}) \quad (37)$$

$$= \sum_{b_i, a_{-i}} z_t(b_i, a_{-i}) r_i(\phi_i(b_i), a_{-i}) \quad (38)$$

$$= \frac{1}{t} \sum_{\tau=1}^t r_i(\phi_i(a_{i,\tau}), a_{-i,\tau}) \quad (39)$$

In Equation 36, we expand the definition of expected rewards, whereas in Equation 38 we collapse this definition. Equation 37 relies on the definition of  $\tilde{\phi}_i$ , the extension of  $\phi_i$ . Equation 39 follows from the definition of the empirical distribution  $z_t$  in Equation 34.

Therefore,

$$r_i(\tilde{\phi}_i(z_t)) - r_i(z_t) = \frac{1}{t} \sum_{\tau=1}^t (r_i(\phi_i(a_{i,\tau}), a_{-i,\tau}) - r_i(a_{i,\tau}, a_{-i,\tau})) \quad (40)$$

$$= \frac{1}{t} \sum_{\tau=1}^t \rho^{\phi_i}(a_{i,\tau}, a_{-i,\tau}) \quad (41)$$

From this equivalence, the conclusion follows immediately.  $\blacksquare$

By Theorem 11, if the time-averaged  $\Phi_i$ -regret experienced by each player  $i$  converges to the negative orthant with probability 1, then empirical distribution of play converges to the set of  $\vec{\Phi}$ -equilibria with probability 1. But if each player  $i$  plays according to a no- $\Phi_i$ -regret learning algorithm, then the time-averaged  $\Phi_i$ -regret experienced by each player  $i$  converges to the negative orthant with probability 1, regardless of the sequence of opposing actions: i.e., on any run of the game. From this discussion, we draw the following general conclusion:

**Theorem 12** *Given an  $n$ -player game  $\Gamma_n$  and a vector of sets of action transformations  $\vec{\Phi} = (\Phi_i)_{1 \leq i \leq n}$  such that  $\Phi_i \subseteq \Phi_{ALL}(A_i)$  is finite for  $1 \leq i \leq n$ . If all players  $i$  play no- $\Phi_i$ -regret learning algorithms, then the empirical distribution of play converges to the set of  $\vec{\Phi}$ -equilibria of  $\Gamma_n$  with probability 1.*

Thus, we see that if all players abide by no-internal-regret algorithms, then the distribution of play converges to the set of correlated equilibria. Moreover, in two-player, zero-sum games if all players abide by no-external-regret algorithms, then the distribution of play converges to the set of generalized minimax equilibria, that is, the set of minimax-valued joint distributions. Again, by Proposition 6, this latter result implies that each player's empirical distribution of play converges to his set of minimax strategies, under the stated assumptions.

## 6. The Power of No Internal Regret

Perhaps surprisingly, no internal regret is the strongest form of no  $\Phi$ -regret, for finite  $\Phi$ . It follows from the results in Section 5 that the tightest game-theoretic solution concept to which no- $\Phi$ -regret learning algorithms converge is correlated equilibrium. In particular, Nash equilibrium is not a necessary outcome of learning via no- $\Phi$ -regret algorithms.

**Theorem 13** *Given a real-valued game  $(A, A', R, r)$ , if a learning algorithm  $\mathcal{L}$  satisfies no internal regret, then  $\mathcal{L}$  also satisfies no  $\Phi$ -regret for all finite sets  $\Phi \subseteq \Phi_{ALL}(A)$ .*

The proof of this theorem follows immediately from the following lemmas.

**Lemma 14** *Given a real-valued game  $(A, A', R, r)$ , for all  $\Phi \subseteq \Phi_{ALL}(A)$  and  $\Phi' \subseteq \text{SCH}(\Phi)$ , there exists a constant  $c > 0$  such that  $d(\mathbb{R}_-^{\Phi'}, \bar{\rho}_t^{\Phi'}) \leq c d(\mathbb{R}_-^{\Phi}, \bar{\rho}_t^{\Phi})$ , for all  $t$ .*

**Proof** For each  $\phi' \in \Phi' \subseteq \text{SCH}(\Phi)$  there exist  $k \in \mathbb{N}$ ,  $\phi_j \in \Phi$  and  $\alpha_j \geq 0$  for all  $1 \leq j \leq k$ , and  $\beta \in \mathbb{R}$  such that  $\phi' = \left(\sum_{j=1}^k \alpha_j \phi_j\right) + \beta I$  and  $\sum_{j=1}^k \alpha_j + \beta = 1$ . Now

$$\rho^{\phi'}(a, a') = r(\phi'(a), a') - r(a, a') \quad (42)$$

$$= r\left(\left(\sum_{j=1}^k \alpha_j \phi_j + \beta I\right)(a), a'\right) - r(a, a') \quad (43)$$

$$= \sum_{j=1}^k \alpha_j r(\phi_j(a), a') + (\beta - 1)r(a, a') \quad (44)$$

$$= \sum_{j=1}^k \alpha_j (r(\phi_j(a), a') - r(a, a')) \quad (45)$$

$$= \sum_{j=1}^k \alpha_j \rho^{\phi_j}(a, a') \quad (46)$$

Line 45 follows because  $\sum_{j=1}^k \alpha_j = 1 - \beta$ .

Thus, we can define a linear transformation  $F : \mathbb{R}^{\Phi} \rightarrow \mathbb{R}^{\Phi'}$  such that  $F(\rho^{\Phi}(a, a')) = \rho^{\Phi'}(a, a')$ , for all  $a, a'$ . Because the  $\alpha_i$  are all non-negative,  $F$  exhibits the following property:

$$\rho^{\Phi}(a, a') \in \mathbb{R}_-^{\Phi} \Rightarrow F(\rho^{\Phi}(a, a')) = \rho^{\Phi'}(a, a') \in \mathbb{R}_-^{\Phi'} \quad (47)$$

i.e.,  $F(\mathbb{R}_-^{\Phi}) \subseteq \mathbb{R}_-^{\Phi'}$ . Further, because  $F$  is linear,

$$d(\mathbb{R}_-^{\Phi'}, \bar{\rho}_t^{\Phi'}) = d(\mathbb{R}_-^{\Phi'}, F(\bar{\rho}_t^{\Phi})) \quad (48)$$

$$\leq d(F(\mathbb{R}_-^{\Phi}), F(\bar{\rho}_t^{\Phi})) \quad (49)$$

$$\leq c d(\mathbb{R}_-^{\Phi}, \bar{\rho}_t^{\Phi}) \quad (50)$$

where  $c > 0$  is the operator norm of  $F$ . ■

**Lemma 15** *Given a real-valued game  $(A, A', R, r)$ , if a learning algorithm  $\mathcal{L}$  satisfies no  $\Phi$ -regret for some finite set  $\Phi \subseteq \Phi_{ALL}(A)$ , then  $\mathcal{L}$  also satisfies no  $\Phi'$ -regret, for all finite sets  $\Phi' \subseteq \text{SCH}(\Phi)$ .*

**Proof** By Lemma 14, there exists a constant  $c > 0$  such that  $d(\mathbb{R}_-^{\Phi'}, \bar{\rho}_t^{\Phi'}) \leq c d(\mathbb{R}_-^{\Phi}, \bar{\rho}_t^{\Phi})$ , for all  $t$ . Now for any  $\epsilon > 0$ , let  $\delta = \min\{\frac{\epsilon}{c}, \epsilon\}$ . Since  $\delta \leq \frac{\epsilon}{c}$ ,

$$d(\mathbb{R}_-^{\Phi'}, \bar{\rho}_t^{\Phi'}) \geq \epsilon \Rightarrow c d(\mathbb{R}_-^{\Phi}, \bar{\rho}_t^{\Phi}) \geq \epsilon \quad (51)$$

$$\Rightarrow d(\mathbb{R}_-^{\Phi}, \bar{\rho}_t^{\Phi}) \geq \delta \quad (52)$$

Because  $\mathcal{L}$  satisfies no  $\Phi$ -regret, we can choose  $t_0$  such that for any  $a'_1, a'_2, \dots$ ,

$$P[\exists t \geq t_0 \text{ s.t. } d(\mathbb{R}_-^{\Phi}, \bar{\rho}_t^{\Phi}) \geq \delta] < \delta \quad (53)$$

But then, by Equation 52, and since  $\delta \leq \epsilon$ ,

$$P[\exists t \geq t_0 \text{ s.t. } d(\mathbb{R}_-^{\Phi'}, \bar{\rho}_t^{\Phi'}) \geq \epsilon] < \epsilon \quad (54)$$

Therefore, no  $\Phi$ -regret implies no  $\Phi'$ -regret.  $\blacksquare$

**Lemma 16** For any (finite) set of actions,  $A$ , the super convex hull of  $\Phi_{\text{INT}}(A)$  is  $\Phi_{\text{ALL}}(A)$ .

**Proof** Let  $\phi^*$  be an arbitrary element of  $\Phi_{\text{ALL}}(A)$ . Define  $\hat{\phi} \in \text{SCH}(\Phi_{\text{INT}}(A))$  by

$$\hat{\phi} = \sum_{a,b \in A} \phi^*(a)(b) \phi_{\text{INT}}^{(ab)} + (1 - |A|)I \quad (55)$$

For any  $x \in A$ ,

$$\hat{\phi}(x) = \sum_{a,b \in A} \phi^*(a)(b) \phi_{\text{INT}}^{(ab)}(x) + (1 - |A|)\delta_x \quad (56)$$

$$= \sum_{a,b \in A} \phi^*(a)(b) \left\{ \begin{array}{ll} \delta_b & \text{if } x = a \\ \delta_x & \text{otherwise} \end{array} \right\} + (1 - |A|)\delta_x \quad (57)$$

$$= \sum_{a,b \in A} \phi^*(a)(b) \left( \sum_{x=a} \delta_b + \sum_{x \neq a} \delta_x \right) + (1 - |A|)\delta_x \quad (58)$$

$$= \sum_{b \in A} \phi^*(x)(b) \delta_b + \sum_{x \neq a} \sum_{b \in A} \phi^*(x)(b) \delta_x + (1 - |A|)\delta_x \quad (59)$$

$$= \sum_{b \in A} \phi^*(x)(b) \delta_b + (|A| - 1)\delta_x + (1 - |A|)\delta_x \quad (60)$$

$$= \sum_{b \in A} \phi^*(x)(b) \delta_b \quad (61)$$

Further, for any  $y \in A$ ,

$$\hat{\phi}(x)(y) = \sum_{b \in A} \phi^*(x)(b) \delta_b(y) \quad (62)$$

$$= \phi^*(x)(y) \quad (63)$$

Therefore,  $\phi^* = \hat{\phi} \in \text{SCH}(\Phi_{\text{INT}}(A))$ .  $\blacksquare$

## 7. Related Work

In this section, we relate our theorems on the existence of no- $\Phi$ -regret algorithms (Theorem 4), the convergence of no- $\Phi$ -regret learning to game-theoretic equilibria (Theorem 12), and the power of no-internal-regret (Theorem 13) to results published elsewhere.

### 7.1 On the Existence of No-Regret Algorithms

In Theorem 4, we rely on Theorem 2 to establish the existence of no- $\Phi$ -regret learning algorithms, for all finite  $\Phi \subseteq \Phi_{\text{ALL}}$ . The algorithms presented are, in the terminology of Greenwald *et al.* [15], *action*-regret-based learning algorithms, which means that regret at time  $t$  is computed with respect to the action  $a_t$  as in Equations 6 and 7, as opposed *distribution*-regret-based learning algorithms, in which regret at time  $t$  is calculated with respect to the distribution  $q_t$  as follows:

$$\rho^\Phi(q, a') \equiv \left( \rho^\phi(q, a') \right)_{\phi \in \Phi} \quad (64)$$

where

$$\rho^\phi(q, a') = \sum_{a \in A} q(a) [r(\phi(a), a') - r(a, a')] \quad (65)$$

Many well-known no-regret learning algorithms arise as instances, or close cousins, of action- or distribution-regret-based variants of Algorithm 1:

1. The no-external-regret algorithm of Hart and Mas-Colell [17] (Theorem B) is the special case of Algorithm 1 (action-regret-based) when  $\Phi = \Phi_{\text{EXT}}(A)$ .
2. The no-internal-regret algorithm of Foster and Vohra [10] is closely related to Algorithm 1 (distribution-regret-based) when  $\Phi = \Phi_{\text{INT}}(A)$ . Foster and Vohra calculate the fixed points of a stochastic matrix that is derived from the internal regret vector. Their matrix is identical to  $M$  (computed in terms of distribution-based regrets) up to normalization. Consequently, both their matrix and  $M$  have the same set of fixed points.<sup>4</sup>
3. By replacing  $+$  operation (i.e.,  $(x_t^\phi)^+$ ) in steps 10 and 13 of Algorithm 1 (distribution-regret-based) with  $e^{x_t^\phi}$ , we arrive at Freund and Schapire's Hedge algorithm [11] when  $\Phi = \Phi_{\text{EXT}}(A)$ , and an instance of an algorithm discussed by Cesa-Bianchi and Lugosi [5] when  $\Phi = \Phi_{\text{INT}}(A)$ .

Lehrer [21] derives a “wide range no-regret theorem” analogous to our existence result. Lehrer's approach combines “replacing schemes,” functions from  $\mathcal{H} \times A$  to  $A$ , with “activeness functions” from  $\mathcal{H} \times A$  to  $\{0, 1\}$ . Given a replacing scheme  $g$  and an activeness function  $I$ , Lehrer's framework compares the agent's rewards to the rewards that could have been obtained by playing action  $g(h_t, a_t)$ , but only if  $I(h_t, a_t) = 1$ , yielding a general form of action regret. Lehrer establishes the existence of *action*-regret-based no-regret algorithms whose regret with respect to any countable set of pairs of replacing schemes and activeness functions, averaged over the number of times each pair is “active,” approaches the negative orthant.

---

4. Let  $R_t^{(ij)}$  denote the cumulative  $\phi_{\text{INT}}^{(ij)}$  regret at time  $t$ . Define the matrix  $Q_t$  by  $(Q_t)_{ii} = -\sum_j (R_t^{(ij)})^+$  and  $(Q_t)_{ij} = (R_t^{(ij)})^+$  for  $i \neq j$ . Our algorithm plays the fixed point of a matrix  $A$  which can be written as  $A = I + \frac{1}{\sum_i |(Q_t)_{ii}|} Q_t$ . Foster and Vohra's algorithm plays the fixed point of a matrix  $A'$  which can be written as  $A' = I + \frac{1}{\max_i |(Q_t)_{ii}|} Q_t$ . It can be shown that  $A$  and  $A'$  have the same set of fixed points: If  $q$  is a fixed point of  $A$ , then  $qA = q \Rightarrow q + \frac{1}{\sum_i |(Q_t)_{ii}|} qQ = q \Rightarrow qQ = 0 \Rightarrow qA' = qI = q$ . And similarly in the other direction.

Because Lehrer deals with *countable* sets of replacing schemes, whereas we restrict attention to *finite* sets of transformations, it may seem that Lehrer’s theorem immediately subsumes ours. However, in Lehrer’s framework, the co-domain of each transformation is  $A$ , not  $\Delta(A)$ , as it is in ours. In other words, in our framework, actions are transformed into mixed, rather than pure, strategies. This turns out to yield no additional power however. By Theorem 13, it suffices to consider the set of internal action transformations, each element of which can be expressed simply as a function from  $A \rightarrow A$ . Hence, in light of Theorem 13, Lehrer’s theorem can in fact be viewed as subsuming our existence theorem.

Blum and Mansour’s [3] framework is similar to Lehrer’s, but yields results about *distribution* regret. Their “modification rules” are the same as replacing schemes, but instead of activeness functions, they pair modification rules with “time selection functions,” which are functions from  $\mathbb{N}$  to the interval  $[0, 1]$ . The rewards an agent could have obtained under each modification rule are weighted according to how “awake” the rule is, as indicated by the corresponding time selection function. They present a method that, given a collection of algorithms whose external distribution regret is bounded above by  $f(t)$  at time  $t$ , generates an algorithm whose swap distribution regret (and hence, internal distribution regret) is bounded above by  $|A|f(t)$ .

Cesa-Bianchi and Lugosi [5] develop a framework of “generalized” *distribution* regret. They rely on a notion of “experts,” which they define as functions from  $\mathbb{N}$  to  $A$ , and following Lehrer, they pair experts  $f_1, \dots, f_N$  with activation functions  $I_i : A \times \mathbb{N} \rightarrow \{0, 1\}$ . At time  $t$ , for each  $i$ , if  $I_i(a_t, t) = 1$ , they compare the agent’s rewards to the rewards the agent could have obtained by playing  $f_i(t)$ . This approach is more general than our action-transformation framework in that alternatives may depend on time. At the same time, it is more limited in that it does not naturally represent swap regret (but this is not necessarily a shortcoming, in light of Theorem 13). Cesa-Bianchi and Lugosi’s calculations yield bounds on generalized distribution regret.

From the bounds derived by Blum and Mansour and Cesa-Bianchi and Lugosi, one can infer the existence of *distribution*-regret-based no-regret learning algorithms, by applying the Hoeffding-Azuma lemma (see, for example, the Appendix of Cesa-Bianchi and Lugosi [6]).

Finally, it has been observed that swap regret is bounded above by  $|A|$  times internal regret. Hence, no-internal-regret implies no-swap-regret,<sup>5</sup> which in turn implies no- $\Phi$ -regret, for all  $\Phi \subseteq \Phi_{\text{ALL}}$ , because the elements of any  $\Phi$  can be constructed as a convex combination of the elements of  $\Phi_{\text{SWAP}}$ . It follows from this observation that every no-internal-regret algorithm—such as the algorithms of Foster and Vohra [10], Hart and Mas-Colell [18], Cesa-Bianchi and Lugosi [5], and Young [27]—is a no- $\Phi$ -regret algorithm, for all  $\Phi \subseteq \Phi_{\text{ALL}}$ . In other words, the existence of (action- or distribution-based) no-internal-regret algorithms implies the existence of (action- or distribution-based) no- $\Phi$ -regret algorithms, for all  $\Phi \subseteq \Phi_{\text{ALL}}$ .

## 7.2 On the Connection between Learning and Games

Several authors before us have explored the connection between no-regret (and related) learning algorithms and game-theoretic equilibria. This literature originates with Foster and Vohra [9], who present a (calibrated) learning algorithm such that if all players play according to it, the empirical distribution of play converges to the set of correlated equilibria.

Hart and Mas-Colell [17] exhibit a simple adaptive procedure such that if all players follow this procedure, then the time-averaged internal regret vector of each player converges to zero almost surely, and the empirical distribution of play converges to the set of correlated equilibrium almost surely. Their algorithm is not no-internal-regret, however; it is not regret-minimizing against an

---

5. This observation also follows from Theorem 13. Choose  $\Phi = \Phi_{\text{SWAP}}$ .

arbitrary opponent. More fundamentally, they show that a necessary and sufficient condition for the empirical distribution of play to converge to the set of correlated equilibria is that all players’ internal regrets converge to zero.<sup>6</sup> In Theorem 11, we generalize this result beyond  $\Phi_{\text{INT}}$ .

Hart and Mas-Colell [18] present a class of no-external-regret learning algorithms. They show that in repeated two-player zero-sum games, if both players play according to algorithms in this class, then each player’s empirical distribution of play converges to his set of minimax strategies and the players’ average rewards converge to the minimax value of the game, almost surely. They also discuss a class of algorithms which are no-internal-regret. They argue that if each player plays according to an algorithm in this class, then the empirical distribution of play converges to the set of correlated equilibria almost surely, which is an immediate consequence of their earlier result.

## 8. Summary

In this article, we defined a general class of no-regret learning algorithms, called no- $\Phi$ -regret learning algorithms, which spans the spectrum from no-external-regret learning to no-internal-regret learning and beyond. Analogously, we defined a general class of game-theoretic equilibria, called  $\vec{\Phi}$ -equilibria, and we showed that the empirical distribution of play of no- $\Phi_i$ -regret algorithms converges to the set of  $\vec{\Phi}$ -equilibria. Moulin and Vial [23] also define a general class of game-theoretic equilibria, ranging from pure strategy Nash equilibria to coarse correlated equilibria. To our knowledge, their generalizations have not been widely applicable in practice. Similarly, our generalized notions of equilibria may not be of significant practical value—at present, we know of no other interesting classes of  $\vec{\Phi}$ -equilibria besides  $\Phi_{\text{EXT}}$ - and  $\Phi_{\text{INT}}$ -equilibria. Still, we believe it is of theoretical interest to observe that no-external-regret and no-internal-regret can be viewed along the same continuum, and moreover, that they correspond to game-theoretic equilibria along an analogous continuum.

## Acknowledgments

We gratefully acknowledge Dean Foster for sparking our interest in this topic, and for ongoing discussions that helped to clarify many of the technical points in this paper. We also thank Dave Seaman for providing a proof of the harder direction of Lemma 17 and anonymous reviewers for their constructive criticism. The content of this article is based on Greenwald and Jafari [13], which in turn is based on Jafari’s Master’s thesis [20]. This research was supported by NSF Career Grant #IIS-0133689 and NSF IGERT Grant #9870676.

## 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] A. Blum and Y. Mansour. From external to internal regret. In *Proceedings of the 2005 Computational Learning Theory Conferences*, pages 621–636, June 2005.
- [4] G. Brown. Iterative solutions of games by fictitious play. In T. Koopmans, editor, *Activity Analysis of Production and Allocation*. Wiley, New York, 1951.

---

6. Stoltz and Lugosi [26] generalize this result to the case where the set of actions is a compact and convex subset of a normed space and the reward function is continuous.

- [5] N. Cesa-Bianchi and G. Lugosi. Potential-based algorithms in on-line prediction and game theory. *Machine Learning*, 51(3):239–261, 2003.
- [6] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.
- [7] A. Cournot. *Recherches sur les Principes Mathématiques de la Théorie de la Richesse*. Hachette, 1838.
- [8] D. Foster and R. Vohra. A randomization rule for selecting forecasts. *Operations Research*, 41(4):704–709, 1993.
- [9] D. Foster and R. Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21:40–55, 1997.
- [10] D. Foster and R. Vohra. Regret in the on-line decision problem. *Games and Economic Behavior*, 29:7–35, 1999.
- [11] Y. Freund and R. E. 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.
- [12] D. Fudenberg and D. K. Levine. Universal consistency and cautious fictitious play. *Journal of Economic Dynamics and Control*, 19:1065–1090, 1995.
- [13] A. Greenwald and A. Jafari. A general class of no-regret algorithms and game-theoretic equilibria. In *Proceedings of the 2003 Computational Learning Theory Conference*, pages 1–11, August 2003.
- [14] A. Greenwald, A. Jafari, and C. Marks. Blackwell’s approachability theorem: A generalization in a special case. Technical Report CS-06-01, Brown University, Department of Computer Science, January 2006.
- [15] A. Greenwald, Z. Li, and C. Marks. Bounds for regret-matching algorithms. Technical Report CS-06-10, Brown University, Department of Computer Science, June 2006.
- [16] 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.
- [17] S. Hart and A. Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68:1127–1150, 2000.
- [18] S. Hart and A. Mas-Colell. A general class of adaptive strategies. *Journal of Economic Theory*, 98(1):26–54, 2001.
- [19] M. Herbster and M. K. Warmuth. Tracking the best expert. *Machine Learning*, 32(2):151–178, 1998.
- [20] A. Jafari. *On the Notion of Regret in Infinitely Repeated Games*. Master’s Thesis, Brown University, Providence, May 2003.

- [21] E. Lehrer. A wide range no-regret theorem. *Games and Economic Behavior*, 42(1):101–115, 2003.
- [22] N. Littlestone and M. K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108:212 – 261, 1994.
- [23] H. Moulin and J. P. Vial. Strategically zero-sum games. *International Journal of Game Theory*, 7:201–221, 1978.
- [24] J. Nash. Non-cooperative games. *Annals of Mathematics*, 54:286–295, 1951.
- [25] J. Robinson. An iterative method of solving a game. *Annals of Mathematics*, 54:298–301, 1951.
- [26] G. Stoltz and G. Lugosi. Learning correlated equilibria in games with compact sets of strategies. *Games and Economic Behavior*, To Appear.
- [27] P. Young. *Strategic Learning and its Limits*. Oxford University Press, Oxford, 2004.

## Appendix A. Proof of Lemma 17

**Lemma 17** *Let  $(X, d_X)$  be a compact metric space and let  $(Y, d_Y)$  be a metric space. Let  $\{x_t\}$  be an  $X$ -valued sequence, and let  $S$  be a nonempty, closed subset of  $Y$ . If  $f : X \rightarrow Y$  is continuous and if  $f^{-1}(S)$  is nonempty, then  $d_X(x_t, f^{-1}(S)) \rightarrow 0$  as  $t \rightarrow \infty$  if and only if  $d_Y(f(x_t), S) \rightarrow 0$  as  $t \rightarrow \infty$ .*

**Proof** We write  $d = d_X$  and  $d = d_Y$ , since the appropriate choice of distance metric is always clear from the context. To prove the forward implication, assume  $d(x_t, f^{-1}(S)) \rightarrow 0$  as  $t \rightarrow \infty$ . Choose  $t_0$  s.t. for all  $t \geq t_0$ ,  $d(x_t, f^{-1}(S)) < \frac{\delta}{2}$ . Observe that for all  $x_t$  and for all  $\gamma > 0$ , there exists  $q_t^{(\gamma)} \in f^{-1}(S)$  s.t.  $d(x_t, q_t^{(\gamma)}) < d(x_t, f^{-1}(S)) + \gamma$ . Now, since  $d(x_t, q_t^{(\frac{\delta}{2})}) < \frac{\delta}{2} + \frac{\delta}{2} = \delta$ , by the continuity of  $f$ ,  $d(f(x_t), f(q_t^{(\frac{\delta}{2})})) < \epsilon$ , for all  $\epsilon > 0$ . Therefore,  $d(f(x_t), S) < \epsilon$ , since  $f(q_t^{(\frac{\delta}{2})}) \in S$ .

To prove the reverse implication, assume  $d(f(x_t), S) \rightarrow 0$  as  $t \rightarrow \infty$ . We must show that for all  $\epsilon > 0$ , there exists a  $t_0$  s.t. for all  $t \geq t_0$ ,  $d(x_t, f^{-1}(S)) < \epsilon$ . Define  $T = \{x \in X \mid d(x, f^{-1}(S)) \geq \epsilon\}$ . If  $T = \emptyset$ , the claim holds. Otherwise, observe that  $T$  can be expressed as the complement of the union of open balls, so that  $T$  is closed and thus compact. Define  $g : X \rightarrow \mathbb{R}$  as  $g(x) = d(f(x), S)$ . By assumption  $S$  is closed; hence,  $g(x) > 0$ , for all  $x$ . Because  $T$  is compact,  $g$  achieves some minimum value, say  $L > 0$ , on  $T$ . Choose  $t_0$  s.t.  $d(f(x_t), S) < L$  for all  $t \geq t_0$ . Thus, for all  $t \geq t_0$ ,  $g(x_t) < L \Rightarrow x_t \notin T \Rightarrow d(x_t, f^{-1}(S)) < \epsilon$ . ■