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

Amy Greenwald
Brown University

and

Amir Jafari
Northwestern University

Computational Learning Theory 2003

Background

No-external-regret learning converges to the set of minimax equilibria.
[e.g., Freund and Schapire 1996]

No-internal-regret learning converges to the set of correlated equilibria.
[e.g., Foster and Vohra 1997]

Three Theorems

1. Existence Theorem
   - Φ-no-regret learning algorithms exist, for all Φ.

2. Convergence Theorem
   - Φ-no-regret learning converges to the set of Φ-equilibria.

3. Negative Result
   - No-internal-regret is the strongest form of Φ-no-regret.

Single Agent Model

- a set of agent's actions A ($a \in A$)
- a set of opponents' actions A' ($a' \in A'$)
- vector-valued outcome function $\rho : A \times A' \to V$
- bounded reward function $r : A \times A' \to \mathbb{R}$

A learning algorithm $\mathcal{A}$ is a sequence of functions $q_t : (A \times A')^{t-1} \to \Delta(A)$, for $t = 1, 2, \ldots$, where $\Delta(A)$ is the set of all probability measures on A

Regret

Let Φ be a finite subset of the set of linear maps $\{\phi : \Delta(A) \to \Delta(A)\}$.
The regret vector $\rho_\Phi : A \times A' \to \mathbb{R}^\Phi$ is defined as follows:

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

Here δ_a is the Dirac δ function: i.e., all mass is concentrated at a.

No-External-Regret

$\Phi_{\text{EXT}} = \{\phi_a | a \in A\}$ be the set of constant maps: i.e., $\phi_a(q) = \delta_a$.

If $|A| = 4$, and if $a = 2$, then

$$\phi_a = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

Thus, $\langle q_1, q_2, q_3, q_4 \rangle \phi_a = \langle 0, 1, 0, 0 \rangle$, for all $\langle q_1, q_2, q_3, q_4 \rangle \in \Delta(A)$.

No-Internal-Regret

$\Phi_{\text{INT}} = \{\phi_{ab} | a \neq b \in A\}$, where

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

If $|A| = 4$, and if $a = 2$ and $b = 3$, then

$$\phi_{ab} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Thus, $\langle q_1, q_2, q_3, q_4 \rangle \phi_{ab} = \langle q_1, 0, q_2 + q_3, q_4 \rangle$, for all $\langle q_1, q_2, q_3, q_4 \rangle \in \Delta(A)$.

Approachability

$G \subseteq V$ 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 $a'_1, a'_2, \ldots$,

$$\lim_{t\to\infty} d(G, \bar{\rho}_t) = \lim_{t\to\infty} \inf_{g\in G} d(g, \bar{\rho}_t) = 0$$

almost surely, where $\bar{\rho}_t$ denotes the average value of ρ through time t: i.e., $\bar{\rho}_t = \frac{1}{t}\Big(\rho(a_1, a'_1) + \ldots + \rho(a_t, a'_t)\Big)$.

Φ-No-Regret

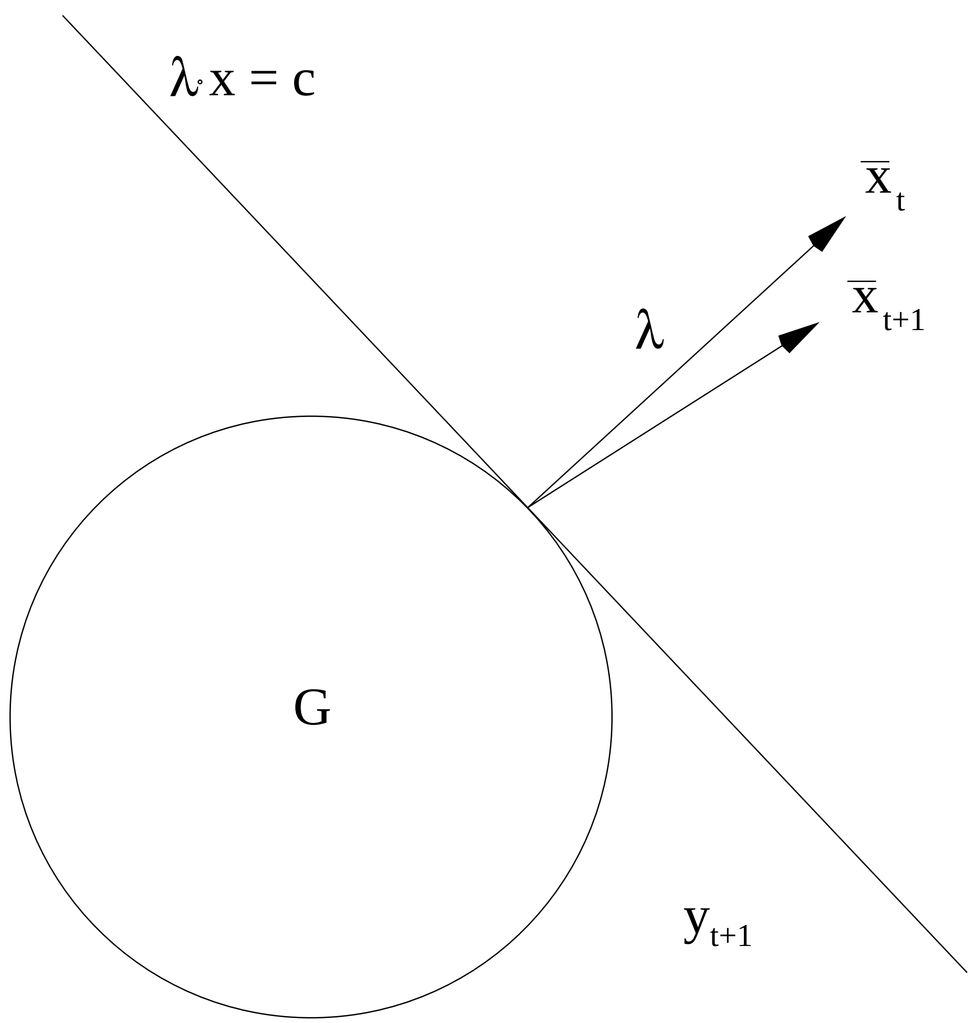
A Φ-no-regret learning algorithm is one that ρ_Φ-approaches $\mathbb{R}^\Phi_-$: i.e.,

$$\begin{aligned}
&\lim_{t\to\infty} d(\mathbb{R}^\Phi_-, \bar\rho_{\Phi,t}) = 0 \\
\text{iff}\quad &\limsup_{t\to\infty} \bar\rho_{\Phi,t} \le 0 \\
\text{iff}\quad &\limsup_{t\to\infty} \frac{1}{t}\sum_{\tau=1}^{t} \Big(r(\phi(\delta_{a_\tau}), a'_\tau) - r(a_\tau, a'_\tau) \Big) \le 0, \text{ for all } \phi \in \Phi
\end{aligned}$$

A no-external-regret algorithm ρ_Φ-approaches $\mathbb{R}^\Phi_-$ for $\Phi = \Phi_{\mathrm{EXT}}$.

A no-internal-regret algorithm ρ_Φ-approaches $\mathbb{R}^\Phi_-$ for $\Phi = \Phi_{\mathrm{INT}}$.

Blackwell's Theorem



Blackwell's Theorem

Any convex subset $G \subseteq V$ is ρ-approachable if there exists a learning algorithm $\mathcal{A}$ *s.t.* for all times t and for all $a' \in A'$, $\lambda(\bar{\rho}_t) \cdot \rho(q_{t+1}, a') \leq c$, where $\lambda(x)$ is the vector between $\bar{\rho}_t$ and the closest point to $\bar{\rho}_t$ in G.

Moreover, the following procedure can be used to approach the set G:

- if $\bar{\rho}_t \in G$, play arbitrarily; but if $\bar{\rho}_t \in V \setminus G$, play according to $\mathcal{A}$.

Existence Theorem

For all finite subsets Φ of the set of continuous, linear maps on $\Delta(A)$, there exists a learning algorithm that satisfies Φ-no-regret.

Proof

By Blackwell's theorem, it suffices to show that for all $x \in \mathbb{R}^{\Phi} \setminus \mathbb{R}^{\Phi}_{-}$, there exists $q \in \Delta(A)$ *s.t.* for all $a' \in A$, $x^{+} \cdot \rho_{\Phi}(q, a') \leq 0$.

Proof of Existence Theorem

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

Now it suffices to show

$$\left(\sum_{\phi \in \Phi} x_\phi^+ \phi\right)(q) = \left(\sum_{\phi \in \Phi} x_\phi^+\right) q$$

But by Brouwer's fixed point theorem, $M = \frac{\sum_{\phi \in \Phi} x_\phi^+ \phi}{\sum_{\phi \in \Phi} x_\phi^+}$ has a fixed point.

Multiagent Model

- a set of players I ($i \in I$)
- for all players i,
  - a set of actions A_i ($a_i \in A_i$, $a_{-i} \in \prod_{j \neq i} A_j$)
  - a reward function $r_i : \prod_{i \in I} A_i \to \mathbb{R}$
  - a set Φ_i ($\phi_i \in \Phi_i$)

Φ-Equilibrium

An element $q \in \Delta(\prod_{i \in I} A_i)$ is a Φ-equilibrium iff $r_i(\phi_i(q)) \leq r_i(q)$, for all players i and for all $\phi_i \in \Phi_i$.

Examples

Correlated Equilibrium: $\Phi_i = \Phi_{\mathsf{INT}}$, for all players i

Generalized Minimax Equilibrium: $\Phi_i = \Phi_{\mathsf{EXT}}$, for all players i

Convergence Theorem

If all players i play via some Φ_i-no-regret algorithm, then the joint empirical distribution of play converges to the set of Φ-equilibria, almost surely.

Proof

For all players i, for all $\phi_i \in \Phi_i$,

$$\begin{aligned}
& \limsup_{t\to\infty} r_i(\phi_i(z_t)) - r_i(z_t) \\
= \; & \limsup_{t\to\infty} \frac{1}{t}\sum_{\tau=1}^{t} r_i(\phi_i(\delta_{a_{i,\tau}}), a_{-i,\tau}) - \frac{1}{t}\sum_{\tau=1}^{t} r_i(a_{i,\tau}, a_{-i,\tau}) \\
\leq \; & 0
\end{aligned}$$

almost surely.

Negative Result

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.

Lemma

If learning algorithm $\mathcal{A}$ satisfies Φ-no-regret, then $\mathcal{A}$ also satisfies Φ'-no-regret, for all finite subsets $\Phi' \subseteq \mathrm{SCH}(\Phi)$, the super CH of Φ:

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

Proof of Negative Result

Proof

An elementary matrix is one with one 1 per row, and 0's elsewhere.

Let $M(n_1,\ldots,n_m)$ denote the elementary matrix with 0's everywhere except 1's at entries (i,n_i) for $1 \le i \le m$. If $\phi_{in_i} \in \Phi_{\mathrm{INT}}$, for $1 \le i \le m$,

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

Applications of Lemma

1. If $\mathcal{A}$ is Φ-NR for $\Phi = \Phi_{\mathrm{INT}}$, then $\mathcal{A}$ is Φ-NR for $\Phi = \mathrm{EM}$.
2. If $\mathcal{A}$ is Φ-NR for $\Phi = \mathrm{EM}$, then $\mathcal{A}$ is Φ-NR for all $\Phi \subseteq \mathrm{SM}$.

Summary and Conclusions

No-external- and no-internal-regret can be defined along one continuum.

No-internal-regret learning is the strongest form of Φ-no-regret learning.

Therefore, Nash equilibrium cannot be learned via Φ-no-regret learning.