Topics on strategic learning

Sylvain Sorin

IMJ-PRG
Université P. et M. Curie - Paris 6
IMJ-PRG
sylvain.sorin@imj-prg.fr

Stochastic Methods in Game Theory
National University of Singapore
Institute for Mathematical Sciences
November 16, 2015

Basic references

CESA-BIANCHI N. AND G. LUGOSI (2006) *Prediction, Learning and Games*, Cambridge University Press.

FUDENBERG D. AND D.K. LEVINE (1998) *Theory of Learning in Games*, M.I.T. Press.

FUDENBERG D. AND DAVID K. LEVINE (2009) *Learning and equilibrium*, Annual Review of Economics, **1**, 385-420.

HART S. AND A. MAS COLELL (2013) *Simple Adaptive Strategies: From Regret-Matching to Uncoupled Dynamics*, World Scientific Publishing.

PERCHET V. (2014) Approachability, regret and calibration: implications and equivalences, *Journal of Dynamics and Games*, **1**, 181-254.

YOUNG P. (1998) *Theory of Learning in Games*, M.I.T. Press.

Topics on strategic learning I:

Unilateral procedures

Presentation

We will consider here an agent acting in discrete time and facing an unknown environment.
At each stage n, he chooses k_n in a finite set K then observes a reward vector $U_n \in \mathscr{U} = [-1,1]^K$ and his payoff is the k_n^{th} component: $\omega_n = U_n^{k_n}$.
We will work in an adversarial framework where no assumption is done on the reward process that is a function of the past history $h_{n-1} = \{k_1, U_1, ..., k_{n-1}, U_{n-1}\}$.
A strategy of the player is a map σ from $H = \cup_{m=0}^{+\infty} H_m$ to $\Delta(K)$ (set of probabilities on K).

A basic tool is approachability theory that we will first cover.

1. Approachability theory

2. No-regret dynamics

3. Calibrating and applications

Deterministic approachability: geometry

All the results are due to Blackwell (1956).

This section presents the basic geometric principle that sustains the approachability property.
Suppose that $x_1, x_2, \ldots$ is a sequence in $\mathbb{R}^K$.
Assume the family uniformly bounded: $\|x_n\|^2 \leq L$.
Denote by $\bar{x}_n$ the average of the first n elements in the sequence:

$$\bar{x}_n = \frac{1}{n} \sum_{m=1}^{n} x_m.$$

For $C \subset \mathbb{R}^K$ closed, $\Pi_C(x)$ is a closest point to x in C.
If C is convex, it is the projection of x on C.
$d(x, C) = \|x - \Pi_C(x)\|$ is the distance from x to C.

Theorem (The geometric principle)

Suppose that $\{x_n\}$ satisfies:

$$\langle x_{n+1} - \Pi_C(\bar{x}_n), \bar{x}_n - \Pi_C(\bar{x}_n)\rangle \leq 0, \tag{1}$$

then $d(\bar{x}_n, C)$ converges to 0.

Proof

Let $y_n = \Pi_C(x_n)$ and $d_n^2 = \|\bar{x}_n - y_n\|^2$. Then:

$$d_{n+1}^2 \leq \|\bar{x}_{n+1} - y_n\|^2 \tag{2}$$

Theorem (The geometric principle)

Suppose that $\{x_n\}$ *satisfies:*

$$\langle x_{n+1} - \Pi_C(\bar{x}_n), \bar{x}_n - \Pi_C(\bar{x}_n)\rangle \le 0, \tag{1}$$

then $d(\bar{x}_n, C)$ *converges to* 0.

Proof

Let $y_n = \Pi_C(x_n)$ and $d_n^2 = \|\bar{x}_n - y_n\|^2$. Then:

$$d_{n+1}^2 \le \|\bar{x}_{n+1} - y_n\|^2 \tag{2}$$

$$d_{n+1}^2 \leq \|\bar{x}_{n+1} - \bar{x}_n\|^2 + \|\bar{x}_n - y_n\|^2 + 2\langle \bar{x}_{n+1} - \bar{x}_n, \bar{x}_n - y_n \rangle. \tag{3}$$

We decompose

$$\begin{aligned}
\langle \bar{x}_{n+1} - \bar{x}_n, \bar{x}_n - y_n \rangle &= (\frac{1}{n+1})\langle x_{n+1} - \bar{x}_n, \bar{x}_n - y_n \rangle \\
&= (\frac{1}{n+1})(\langle x_{n+1} - y_n, \bar{x}_n - y_n \rangle - \|\bar{x}_n - y_n\|^2)
\end{aligned}$$

and we obtain, using property (1):

$$d_{n+1}^2 \leq (1 - \frac{2}{n+1})\, d_n^2 + (\frac{1}{n+1})^2 \|x_{n+1} - \bar{x}_n\|^2. \tag{4}$$

Since $\|x_{n+1} - \bar{x}_n\|^2 \leq 2\|x_{n+1}\|^2 + 2\|\bar{x}_n\|^2 \leq 4L$, we deduce

$$d_{n+1}^2 \leq (\frac{n-1}{n+1})\, d_n^2 + (\frac{1}{n+1})^2\, 4L \tag{5}$$

so that, by induction:

$$d_n^2 \leq \frac{4L}{n}.$$

Minmax theorem

Let A be a $I \times J$ real matrix and $X = \Delta(I), Y = \Delta(J)$.
$xAy = \sum_{ij} x^i A_{ij} y^j$.

Theorem

Assume: $\min_Y \max_X \; xAy \geq 0$. *Then:* $\max_X \min_Y \; xAy \geq 0$.

Proof

Let $C = \mathbb{R}_+^J$ be the set to approach.
Let z_1 be a lign of A.
We construct inductively a sequence $\{z_n\}$ satisfying the previous inequality (1).
If $\bar{z}_n \notin C$, let $\bar{z}_n - \bar{z}_n^+$ be the negative components and $y_{n+1} \in Y$ proportional to $-\bar{z}_n + \bar{z}_n^+$.
By hypothesis, there exists a move i_{n+1} with $e_{i_{n+1}} A y_{n+1} \geq 0$.
Hence with $z_{n+1} = e_{i_{n+1}} A$, one has $\langle z_{n+1}, y_{n+1} \rangle \geq 0$.

Minmax theorem

Let A be a $I \times J$ real matrix and $X = \Delta(I), Y = \Delta(J)$.
$xAy = \sum_{ij} x^i A_{ij} y^j$.

Theorem

Assume: $\min_Y \max_X \; xAy \geq 0$. *Then:* $\max_X \min_Y \; xAy \geq 0$.

Proof

Let $C = \mathbb{R}^J_+$ be the set to approach.
Let z_1 be a lign of A.
We construct inductively a sequence $\{z_n\}$ satisfying the previous inequality (1).
If $\bar{z}_n \notin C$, let $\bar{z}_n - \bar{z}_n^+$ be the negative components and $y_{n+1} \in Y$ proportional to $-\bar{z}_n + \bar{z}_n^+$.
By hypothesis, there exists a move i_{n+1} with $e_{i_{n+1}} A y_{n+1} \geq 0$.
Hence with $z_{n+1} = e_{i_{n+1}} A$, one has $\langle z_{n+1}, y_{n+1} \rangle \geq 0$.

Thus $\langle z_{n+1}, \bar{z}_n - \bar{z}_n^+ \rangle \leq 0$, but:
i) $\bar{z}_n^+ = \Pi_C(\bar{z}_n)$
ii) $\bar{z}_n^+ \perp \bar{z}_n - \bar{z}_n^+$.
Hence the previous inequality implies :

$$\langle z_{n+1} - \Pi_C(\bar{z}_n), \bar{z}_n - \Pi_C(\bar{z}_n) \rangle \leq 0.$$

So that $d(\bar{z}_n, C) \to 0$.
Now $\bar{z}_n = \bar{x}_n A$ where $\bar{x}_n$ is the empirical frequency of moves of player 1.
For any accumulation point $x^* \in X$ of the sequence $\{x_n\}$, one has $x^* A \in C$, which implies $\max_X \min_Y xAy \geq 0$. ∎

Approachability

The framework is as follows:
A is a $I \times J$ matrix with coefficients in $\mathbb{R}^K$.
At each stage n, Player 1 (resp. Player 2) chooses a move i_n in I (resp. j_n in J). The corresponding vector payoff, $g_n = A_{i_n j_n} \in \mathbb{R}^K$ is then announced.
Denote by h_{n-1} the sequence of payoffs before stage n (this is, at least, the information available to both players when playing at stage n) and let $\overline{g}_n = \frac{1}{n}[\sum_{m=1}^{n} g_m]$ be the average payoff up to stage n included.
Let also $\|A\| = \max_{i \in I, j \in J, k \in K} |A_{ij}^k|$.

Definitions

A set C in $\mathbb{R}^K$ is approachable by Player 1 if for any $\varepsilon > 0$ there exists a strategy σ and N such that, for any strategy τ of Player 2 and any $n \geq N$:

$$E_{\sigma,\tau}(d_n) \leq \varepsilon$$

where d_n is the euclidean distance $d(\overline{g}_n, C)$.
A set C in $\mathbb{R}^K$ is excludable by Player 1 if for some $\delta > 0$, the set $C^c_\delta = \{z; d(z,C) \geq \delta\}$ is approachable by him.

A dual definition holds for Player 2.

From the definitions it is enough to consider closed sets C and even their intersection with the closed ball of radius $\|A\|$.
Given x in $X = \Delta(I)$, define $[xA] = co\ \{\sum_i x_i A_{ij};\ j \in J\}$, and similarly $[Ay]$, for y in $Y = \Delta(J)$. If Player 1 uses x, his expected payoff will be in $[xA]$, whatever being the move of player 2.

B-sets and sufficient condition

The first result is a sufficient condition for approachability based on the following notion:

Definition

A closed set C in $\mathbb{R}^K$ is a **B**-set for Player 1 if:
for any $z \notin C$, there exists a closest point $w = w(z)$ in C to z and a mixed move $x = x(z)$ in X, such that the hyperplane trough w orthogonal to the segment $[wz]$ separates z from $[xA]$. Explicitly:

$$\langle z - w, u - w \rangle \leq 0, \forall u \in [xA].$$

Theorem

Let C be a **B***-set for Player 1.*
Then C is approachable by that player.
Explicitly, a strategy satisfying $\sigma(h_{n+1}) = x(\overline{g}_n)$, whenever $\overline{g}_n \notin C$, gives:

$$E_{\sigma\tau}(d_n) \ \leq \ \frac{2\|A\|}{\sqrt{n}}, \quad \forall\tau$$

and d_n converges $P_{\sigma\tau}$ a.s. to 0, more precisely:

$$P(\exists n \geq N; d_n^2 \geq \varepsilon) \leq \frac{8L}{\varepsilon N}$$

Proof

Let Player 1 use a strategy σ as above. Denote $w_n = w(\overline{g}_n)$.

Theorem

Let C be a **B***-set for Player 1.*
Then C is approachable by that player.
Explicitly, a strategy satisfying $\sigma(h_{n+1}) = x(\overline{g}_n)$, whenever $\overline{g}_n \notin C$, gives:

$$E_{\sigma\tau}(d_n) \ \leq \ \frac{2\|A\|}{\sqrt{n}}, \quad \forall\tau$$

and d_n converges $P_{\sigma\tau}$ a.s. to 0, more precisely:

$$P(\exists n \geq N; d_n^2 \geq \varepsilon) \leq \frac{8L}{\varepsilon N}$$

Proof

Let Player 1 use a strategy σ as above. Denote $w_n = w(\overline{g}_n)$.

The property of $x(\overline{g}_n)$ implies that:

$$\langle E(g_{n+1}|h_n) - w_n, \overline{g}_n - w_n \rangle) \leq 0$$

since $E(g_{n+1}|h_n)$ belongs to $[x(\overline{g}_n)A]$.
Hence the previous equation in the deterministic case:

$$d_{n+1}^2 \leq (1 - \frac{2}{n+1})\, d_n^2 + (\frac{1}{n+1})^2 \|x_{n+1} - \overline{x}_n\|^2,$$

gives here by taking conditional expectation with respect to the history h_n:

$$\mathsf{E}(d_{n+1}^2|h_n) \leq (1 - \frac{2}{n+1})\, d_n^2 + (\frac{1}{n+1})^2 \mathsf{E}(\|g_{n+1} - \overline{g}_n\|^2 |h_n). \qquad (6)$$

So that we obtain, like in (5):

$$\mathsf{E}(d_{n+1}^2) \leq (\frac{n-1}{n+1})\ \mathsf{E}(d_n^2) + (\frac{1}{n+1})^2\ 4L$$

and by induction:

$$\mathsf{E}(d_n^2) \leq \frac{4L}{n}.$$

This gives in particular the convergence in probability of d_n to 0. Now introduce the random variable:
$W_n = d_n^2 + L\sum_{m=n+1}^{\infty}(\frac{1}{m^2}E(\|g_m - \overline{g}_m\|^2|h_n)$. We have from (6):

$$E(W_{n+1}|h_n) \leq W_n$$

thus W_n is a positive supermartingale hence converges $P\ a.s.$ to 0. More precisely Doob's maximal inequality gives :

$$P(\exists n \geq N; d_n^2 \geq \varepsilon) \leq \frac{E(W_N)}{\varepsilon} \leq \frac{8L}{\varepsilon N}.$$

In particular one obtains:

Corollary

For any x in S, $[xA]$ is approachable by Player 1, with the constant strategy x.

It follows that a necessary condition for a set C to be approachable by Player 1 is that for any y in Y, $[Ay] \cap C \neq \emptyset$, otherwise C would be excludable by Player 2.

In fact this condition is also sufficient for convex sets.

In particular one obtains:

Corollary

For any x in S, $[xA]$ is approachable by Player 1, with the constant strategy x.

It follows that a necessary condition for a set C to be approachable by Player 1 is that for any y in Y, $[Ay] \cap C \neq \emptyset$, otherwise C would be excludable by Player 2.

In fact this condition is also sufficient for convex sets.

In particular one obtains:

Corollary

For any x in S, $[xA]$ is approachable by Player 1, with the constant strategy x.

It follows that a necessary condition for a set C to be approachable by Player 1 is that for any y in Y, $[Ay] \cap C \neq \emptyset$, otherwise C would be excludable by Player 2.

In fact this condition is also sufficient for convex sets.

Convex case

Theorem

Assume C closed and convex in $\mathbb{R}^K$.
*C is a **B**-set for Player 1 iff*

$$(*) \qquad [Ay] \cap C \neq \emptyset, \qquad \forall y \in Y.$$

*In particular a set is approachable iff it is a **B**-set.*

Proof

By the previous Corollary, it is enough to prove that $(*)$ implies that C is a **B**-set.
The idea is to reduce by projection the problem to the one-dimentional case and to use the minmax theorem.

Convex case

Theorem

Assume C closed and convex in $\mathbb{R}^K$.
*C is a **B**-set for Player 1 iff*

$$(*) \qquad [Ay] \cap C \neq \emptyset, \qquad \forall y \in Y.$$

*In particular a set is approachable iff it is a **B**-set.*

Proof

By the previous Corollary, it is enough to prove that $(*)$ implies that C is a **B**-set.
The idea is to reduce by projection the problem to the one-dimentional case and to use the minmax theorem.

In fact, let $z \notin C$, $w = \Pi_C(z)$, and consider the game with real payoff matrix $B = \langle w - z, A \rangle$. Since $[Ay] \cap C \neq \emptyset$ for all $y \in Y$, this implies that its value is at least $\min_{c \in C} \langle w - z, c \rangle = \langle w - z, w \rangle$. Hence there exists an optimal strategy $x \in X$ of player 1 such that $\langle w - z, \sum_i x_i A_{ij} \rangle \;\geq\; \langle w - z, w \rangle$ for any $j \in J$, which shows that xA is on the opposite side of the hyperplane to z, and the result follows. ∎

The previous proof gives also the following practical criteria:

Corollary

A closed convex set C is a **B***- set for Player 1 iff, for any α in $\mathbb{R}^K$:*

$$\mathtt{val}\langle \alpha, A \rangle \geq \min_{c \in C} \langle \alpha, c \rangle,$$

where `val` *is the $\max_X \min_Y$ operator.*

Extensions

1. In dimension 1, any set is either approachable or excludable. There exist sets that are neither approachable nor excludable. Extension to random payoffs, uniformly bounded in L^2 (Blackwell, 1956).
2. Any set is either weakly approachable or weakly excludable (strategy adapted to the duration) [first link with differential games] (Vieille, 1992).
3. Any approachable set contains a **B**-set (Spinat, 2002).
4. Extension to infinite dimension (Lehrer, 2002).
5. General active states (Lehrer, 2003).

6. Idea of a potential (convex case)
Write

$$\langle x_{n+1} - \Pi_C(\bar{x}_n), \bar{x}_n - \Pi_C(\bar{x}_n) \rangle \leq 0, \tag{7}$$

as

$$\langle x_{n+1} - \bar{x}_n, \nabla P_C(\bar{x}_n) \rangle \leq -2\, P_C(\bar{x}_n), \tag{8}$$

with $P_C(x) = \|x - \Pi_C(x)\|^2$ and $\nabla P_C(x) = 2[x - \Pi_C(x)]$
7. Geometric condition and proximal normal (dual approach) (As Soulaimani, Quincampoix and Sorin, 2009)
8. Approachability and viability [second link with differential games] (As Soulaimani, Quincampoix and Sorin, 2009)

1. Approachability theory

2. No-regret dynamics

3. Calibrating and applications

External regret

Back to the on-line decision problem, we introduce the regret given $k \in K$ and $U \in \mathscr{U} \subset \mathbb{R}^K$ as the vector $R(k,U) \in \mathbb{R}^K$ defined by:

$$R(k,U)^\ell = U^\ell - U^k, \; \ell \in K.$$

Evaluation = regret at stage n = $R_n = R(k_n, U_n)$ with $\omega_n = U_n^{k_n}$ and:

$$R_n^\ell = U_n^\ell - \omega_n, \; \ell \in K.$$

Average external regret vector at stage n, $\bar{R}_n$ with

$$\overline{R}_n^\ell = \overline{U}_n^\ell - \overline{\omega}_n, \; \ell \in K.$$

Compare the actual (average) payoff to the payoff corresponding to the choice of a constant component, see Hannan (1957), Foster and Vohra (1999), Fudenberg and Levine (1995).

Definition

A strategy σ satisfies external consistency (or exhibits no external regret) if, for every process $\{U_m\} \in \mathscr{U}$:

$$\max_{k \in K} [\overline{R}_n^k]^+ \longrightarrow 0 \text{ a.s., as } n \to +\infty$$

or, equivalently $\quad \sum_{m=1}^{n} (U_m^k - \omega_m) \leq o(n), \qquad \forall k \in K.$

We will prove the existence of a strategy satifying EC by showing that the negative orthant $D = \mathbb{R}^K_-$ is approachable by the sequence of regret $\{R_n\}$.

Lemma

$\forall x \in \Delta(K), \forall U \in \mathscr{U}$:

$$\langle x, \mathsf{E}_x[R(., U)]\rangle = 0.$$

Proof

One has

$$\mathsf{E}_x[R(., U)] = \sum_{k \in K} x_k\, R(k, U) = \sum_{k \in K} x_k (U - U^k \mathbf{1}) = U - \langle x, U\rangle \mathbf{1}$$

($\mathbf{1}$ is the K-vector of ones), thus $\langle x, \mathsf{E}_x[R(., U)]\rangle = 0$. ∎

We will prove the existence of a strategy satifying EC by showing that the negative orthant $D = \mathbb{R}_-^K$ is approachable by the sequence of regret $\{R_n\}$.

Lemma

$\forall x \in \Delta(K), \forall U \in \mathscr{U}$:

$$\langle x, \mathsf{E}_x[R(.,U)] \rangle = 0.$$

Proof

One has

$$\mathsf{E}_x[R(.,U)] = \sum_{k \in K} x_k \, R(k,U) = \sum_{k \in K} x_k (U - U^k \mathbf{1}) = U - \langle x, U \rangle \mathbf{1}$$

($\mathbf{1}$ is the K-vector of ones), thus $\langle x, \mathsf{E}_x[R(.,U)] \rangle = 0$. ■

We will prove the existence of a strategy satifying EC by showing that the negative orthant $D = \mathbb{R}^K_-$ is approachable by the sequence of regret $\{R_n\}$.

Lemma

$\forall x \in \Delta(K), \forall U \in \mathscr{U}$:

$$\langle x, \mathsf{E}_x[R(.,U)] \rangle = 0.$$

Proof

One has

$$\mathsf{E}_x[R(.,U)] = \sum_{k \in K} x_k \, R(k,U) = \sum_{k \in K} x_k (U - U^k \mathbf{1}) = U - \langle x, U \rangle \mathbf{1}$$

($\mathbf{1}$ is the K-vector of ones), thus $\langle x, \mathsf{E}_x[R(.,U)] \rangle = 0.$ ∎

Define, if $\bar{R}_n^+ \neq 0$, $\sigma(h_n)$ to be proportional to this vector. Then

$$\langle \mathsf{E}(R_{n+1}|h_n) - \Pi_D(\bar{R}_n), \bar{R}_n - \Pi_D(\bar{R}_n)\rangle = 0$$

since again $\langle \Pi_D(\bar{R}_n), \bar{R}_n - \Pi_D(\bar{R}_n)\rangle = 0$ and

$$\begin{array}{rcl} \langle \mathsf{E}(R_{n+1}|h_n), \bar{R}_n - \Pi_D(\bar{R}_n)\rangle & = & \langle \mathsf{E}(R_{n+1}|h_n), \bar{R}_n^+\rangle \\ & \div & \langle \mathsf{E}(R_{n+1}|h_n), \sigma(h_n)\rangle \\ & = & \langle \mathsf{E}_x[R(.,U_{n+1})], x\rangle, \quad \text{for } x = \sigma(h_n) \\ & = & 0 \end{array}$$

Thus the $(*)$ condition is satisfied, so D is *approachable* hence $d(\bar{R}_n, \mathbb{R}_-^K)$ goes to 0 and $\max_{k\in K}[\overline{R}_n^k]^+ \longrightarrow 0$.

Internal regret

The internal regret evaluation given (k,U) is the $K \times K$ matrix $S(k,U)$ with components: $S^{j\ell}(k,U) = (U^\ell - U^j)\,\mathbf{I}_{\{j=k\}}$.
The evaluation at stage n is $S_n = S(k_n, U_n)$ hence defined by:

$$S_n^{k\ell} = \begin{cases} U_n^\ell - U_n^k & \textit{for } k = k_n \\ 0 & \textit{otherwise.} \end{cases}$$

Average internal regret matrix:

$$\overline{S}_n^{k\ell} = \frac{1}{n} \sum_{m=1, k_m = k}^{n} (U_m^\ell - U_m^k)$$

Comparison for each component k, of the average payoff obtained on the dates where k was played, to the payoff for an alternative choice ℓ.
See Foster and Vohra (1999), Fudenberg and Levine (1999).

Definition

A strategy σ satisfies internal consistency (or exhibits no internal regret) if, for every process $\{U_m\} \in \mathscr{U}$ and every couple k, ℓ:

$$[\overline{S}_n^{k\ell}]^+ \longrightarrow 0 \text{ a.s., as } n \to +\infty$$

Given a $K \times K$ real matrix A with nonnegative coefficients, let $Inv[A]$ be the non-empty set of invariant measures for A, namely vectors $\mu \in \Delta(K)$ satisfying:

$$\sum_{k \in K} \mu^k A^{k\ell} = \mu^\ell \sum_{k \in K} A^{\ell k} \qquad \forall \ell \in K.$$

(This follows from the existence of an invariant measure for a Markov chain- which is itself a consequence of the minmax theorem).

Lemma

Given $A \in \mathbb{R}_+^{K^2}$, let $\mu \in Inv[A]$ then:

$$\langle A, \mathsf{E}_\mu(S(.,U)) \rangle = 0, \quad \forall U \in \mathscr{U}.$$

Given a $K \times K$ real matrix A with nonnegative coefficients, let $Inv[A]$ be the non-empty set of invariant measures for A, namely vectors $\mu \in \Delta(K)$ satisfying:

$$\sum_{k \in K} \mu^k A^{k\ell} = \mu^\ell \sum_{k \in K} A^{\ell k} \qquad \forall \ell \in K.$$

(This follows from the existence of an invariant measure for a Markov chain- which is itself a consequence of the minmax theorem).

Lemma

Given $A \in \mathbb{R}_+^{K^2}$, let $\mu \in Inv[A]$ then:

$$\langle A, \mathsf{E}_\mu(S(.,U)) \rangle = 0, \quad \forall U \in \mathscr{U}.$$

Proof

$$\langle A, \mathsf{E}_\mu(S(.,U))\rangle = \sum_{k,\ell} A^{k\ell}\mu^k(U^\ell - U^k)$$

and the coefficient of each U^ℓ is

$$\sum_{k\in K}\mu^k A^{k\ell} - \mu^\ell \sum_{k\in K} A^{\ell k} = 0$$

$\blacksquare$

To prove the existence of a strategy satisfying internal consistency, we show that $\Delta = \mathbb{R}_-^{K\times K}$ is approachable by the sequence of internal regret $\{S_n\}$.

Define, if $B = \bar{S}_n^+ \neq 0$, $\sigma(h_n)$ to be an invariant measure of B. Then:

$$\langle \mathsf{E}(S_{n+1}|h_n) - \Pi_\Delta(\bar{S}_n), \bar{S}_n - \Pi_\Delta(\bar{S}_n)\rangle = 0$$

since again $\langle \Pi_\Delta(\bar{S}_n), \bar{S}_n - \Pi_\Delta(\bar{S}_n)\rangle = 0$ and

$$\begin{aligned}
\langle \mathsf{E}(S_{n+1}|h_n), \bar{S}_n - \Pi_\Delta(\bar{S}_n)\rangle &= \langle \mathsf{E}(S_{n+1}|h_n), \bar{S}_n^+\rangle \\
&= \langle \mathsf{E}(S_{n+1}|h_n), B\rangle \\
&= \langle \mathsf{E}_\mu[S(.,U_{n+1})], B\rangle, \quad \text{for } \mu = \sigma(h_n) \\
&= 0
\end{aligned}$$

Then Δ is approachable hence $\max_{k,\ell}[\overline{S}_n^{k,\ell}]^+ \longrightarrow 0$. ∎

To prove the existence of a strategy satisfying internal consistency, we show that $\Delta = \mathbb{R}_-^{K\times K}$ is approachable by the sequence of internal regret $\{S_n\}$.
Define, if $B = \bar{S}_n^+ \neq 0$, $\sigma(h_n)$ to be an invariant measure of B.
Then:

$$\langle \mathsf{E}(S_{n+1}|h_n) - \Pi_\Delta(\bar{S}_n), \bar{S}_n - \Pi_\Delta(\bar{S}_n)\rangle = 0$$

since again $\langle \Pi_\Delta(\bar{S}_n), \bar{S}_n - \Pi_\Delta(\bar{S}_n)\rangle = 0$ and

$$\begin{aligned}
\langle \mathsf{E}(S_{n+1}|h_n), \bar{S}_n - \Pi_\Delta(\bar{S}_n)\rangle &= \langle \mathsf{E}(S_{n+1}|h_n), \bar{S}_n^+\rangle \\
&= \langle \mathsf{E}(S_{n+1}|h_n), B\rangle \\
&= \langle \mathsf{E}_\mu[S(.,U_{n+1})], B\rangle, \quad \text{for } \mu = \sigma(h_n) \\
&= 0
\end{aligned}$$

Then Δ is approachable hence $\max_{k,\ell}[\bar{S}_n^{k,\ell}]^+ \longrightarrow 0$. ∎

Calibrating

One considers a sequence of random variables X_m with values in a finite set Ω (that will be written as a basis of $\mathbb{R}^{\Omega}$).
Obviously any deterministic prediction algorithm ϕ_m - where the loss is measured by $\|X_m - \phi_m\|$ - will have a worst loss 1 and any random predictor a loss at least $1/2$ (take $X_m = 1$ iff $\phi_m(1) \leq 1/2$).
We consider here a predictor with values in a finite discretization V of $D = \Delta(\Omega)$ with the following interpretation: "$\phi_m = v$" means that the anticipated probability that $X_m = \omega$ (or $X_m^{\omega} = 1$) is v^{ω}.
Definition:
ϕ is ε-calibrated if, for any $v \in V$:

$$\lim_{n \to +\infty} \frac{1}{n} \| \sum_{\{m \leq n, \phi_m = v\}} (X_m - v) \| \leq \varepsilon$$

This says that if the average number of times v is predicted does not vanish, the average value of X_m on these dates is close to v.
More precisely let B_n^v the set of stages before n where v is announced, let N_n^v be its cardinal and $\bar{X}_n(v)$ the empirical average of X_m on these stages.
Then the condition writes

$$\lim_{n\to+\infty} \frac{N_n^v}{n} \|\bar{X}_n(v) - v\| \leq \varepsilon, \qquad \forall v \in V.$$

From internal consistency to calibrating

Foster and Vohra (1997) We consider the online algorithm where the choice set of the forecaster is V and the outcome given v and X_m is

$$U_m^v = \|X_m - v\|^2$$

(where we use the L^2 norm).
Given an internal consistent procedure ϕ one obtains (the outcome is here a loss)

$$\frac{1}{n} \sum_{m \in B_n^v} (U_m^v - U_m^w) \leq o(n), \qquad \forall w \in V,$$

which is

$$\frac{1}{n} \sum_{m \in B_n^v} (\|X_m - v\|^2 - \|X_m - w\|^2) \leq o(n), \qquad \forall w \in V,$$

hence implies:

$$\frac{N_n^v}{n}(\|\bar{X}_n(v)-v\|^2-\|\bar{X}_n(v)-w\|^2)\leq o(n), \qquad \forall w\in V.$$

In particular by chosing a point w closest to $\bar{X}_n(v)$

$$\frac{N_n^v}{n}(\|\bar{X}_n(v)-v\|^2)\leq \delta^2+o(n)$$

where δ is the L^2 mesh of V, from which calibration follows.

From calibrating to approachability

Foster and Vohra (1997)
We use calibrating to prove approachability of convex sets.

Assume that C satisfies: $\forall y \in Y, \exists x \in X$ such that $xAy \in C$.
Consider a δ-grid of Y defined by $\{y_v, v \in V\}$.
A stage is of type v if player 1 predicts y_v and then plays a mixed move x_v such that $x_v A y_v \in C$.
By using a calibrated procedure, the average move of player 2 on the stages of type v will be δ close to y_v.
By a martingale argument the average payoff will then be ε close to $x_v A y_v$ for δ small enough and n large enough.
Finally the total average payoff is a convex combination of such amounts hence is close to C by convexity.

From calibrating to approachability

Foster and Vohra (1997)
We use calibrating to prove approachability of convex sets.

Assume that C satisfies: $\forall y \in Y, \exists x \in X$ such that $xAy \in C$.
Consider a δ-grid of Y defined by $\{y_v, v \in V\}$.
A stage is of type v if player 1 predicts y_v and then plays a mixed move x_v such that $x_v A y_v \in C$.
By using a calibrated procedure, the average move of player 2 on the stages of type v will be δ close to y_v.
By a martingale argument the average payoff will then be ε close to $x_v A y_v$ for δ small enough and n large enough.
Finally the total average payoff is a convex combination of such amounts hence is close to C by convexity.

Bibliography I

- AUER P., CESA-BIANCHI N., FREUND Y. AND R.E. SHAPIRE (2002) The nonstochastic multiarmed bandit problem, *SIAM J. Comput.*, **32**, 48-77.
- BENAIM M., (1999) *Dynamics of Stochastic Algorithms* , Séminaire de Probabilités, XXIII, Azéma J. and alii eds, LNM **1709**, Springer, 36-72.
- BENAIM M., J. HOFBAUER AND S. SORIN (2005) Stochastic approximations and differential inclusions, *SIAM J. Opt. and Control*, **44**, 328-348.
- BENAIM M., J. HOFBAUER AND S. SORIN (2006) Stochastic approximations and differential inclusions. Part II: applications, *Mathematics of Operations Research*, **31**, 673-695.
- BLACKWELL D. (1956) An analog of the minimax theorem for vector payoffs, *Pacific Journal of Mathematics*, **6**, 1-8.

Bibliography II

- BLUM A. AND Y. MANSOUR (2007) From external to internal regret, *Journal of Machine Learning Reserach*, **8**, 1307-1324.
- BROWN G. W.(1951) Iterative solution of games by fictitious play, in Koopmans T.C. (ed.) *Activity Analysis of Production and Allocation* , Wiley, 374-376.
- CESA-BIANCHI N. AND G. LUGOSI (2003) Potential-based algorithms in on-line prediction and game theory, *Machine Learning*, **51**, 239-261.
- CESA-BIANCHI N. AND G. LUGOSI (2006) *Prediction, Learning and Games*, Cambridge University Press.
- FOSTER D. AND R. VOHRA (1999) Regret in the on-line decision problem, *Games and Economic Behavior*, **29**, 7-35.

Bibliography III

- FREUND Y. AND R.E. SCHAPIRE (1999) Adaptive game playing using multiplicative weights, *Games and Economic Behavior*, **29**, 79-103.
- FUDENBERG D. AND D. K. LEVINE (1995) Consistency and cautious fictitious play, *Journal of Economic Dynamics and Control*, **19**, 1065-1089.
- FUDENBERG D. AND D. K. LEVINE (1998) *The Theory of Learning in Games*, MIT Press.
- FUDENBERG D. AND D. K. LEVINE (1999) Conditional universal consistency *Games and Economic Behavior*, **29**, 104-130.
- HANNAN J. (1957) Approximation to Bayes risk in repeated plays, *Contributions to the Theory of Games, III*, Drescher M., A.W. Tucker and P. Wolfe eds., Princeton University Press, 97-139.

Bibliography IV

- HART S. (2005) Adaptive heuristics, *Econometrica*, **73**, 1401-1430.
- HART S. AND A. MAS-COLELL, (2000) A simple adaptive procedure leading to correlated equilibrium, *Econometrica*, **68**, 1127-1150.
- HART S. AND A. MAS-COLELL (2001) A general class of adaptive strategies, *Journal of Economic Theory*, **98**, 26-54.
- HART S. AND A. MAS-COLELL (2003) Regret-based continuous time dynamics, *Games and Economic Behavior*, **45**, 375-394.
- HOFBAUER J. AND K. SIGMUND, (1998) *Evolutionary Games and Population Dynamics*, Cambridge U.P.
- HOFBAUER J. AND K. SIGMUND, (2003) Evolutionary Games Dynamics, *Bulletin A.M.S.*, **40**, 479-519.

Bibliography V

- HOFBAUER J. AND S. SORIN, (2006) Best response dynamics for continuous zero-sum games, *Discrete and Continuous Dynamical Systems -Series B*, **6**, 215-224.
- HOFBAUER J., S. SORIN AND Y. VIOSSAT (2009) Time average replicator and best reply dynamics, *Mathematics of Operations Research*, **34**, 263-269.
- LITTLESTONE N. AND M.K. WARMUTH (1994) The weighted majority algorithm, *Information and Computation*, **108**, 212-261.
- LUGOSI G., S. MANNOR AND G. STOLTZ (2008) Strategies for prediction under imperfect monitoring, *Mathematics of Operations Research*, **33**, 513-528.
- ROBINSON J., (1951) An iterative method of solving a game, *Annals of Mathematics*, **54**, 296-301.

Bibliography VI

- SORIN S. (2009) Exponential weight algorithm in continuous time, *Mathematical Programming*, Ser. B **116**, 513-528.
- STOLTZ G. AND G. LUGOSI (2005) Internal Regret in On-Line Portfolio Selection, *Machine Learning*, **59**, 125-159.
- VIOSSAT Y. (2007) The replicator dynamics does not lead to correlated equilibria, *Games and Economic Behavior*, **59**, 397-407