Approachability of Convex Sets in Games with Partial Monitoring

Vianney Perchet*

July 3, 2018

Abstract

We provide a necessary and sufficient condition under which a convex set is approachable in a game with partial monitoring, i.e. where players do not observe their opponents' moves but receive random signals. This condition is an extension of Blackwell's Criterion in the full monitoring framework, where players observe at least their payoffs. When our condition is fulfilled, we construct explicitly an approachability strategy, derived from a strategy satisfying some *internal consistency* property in an auxiliary game.

We also provide an example of a convex set, that is neither (weakly)-approachable nor (weakly)-excludable, a situation that cannot occur in the full monitoring case.

We finally apply our result to describe an ε-optimal strategy of the uninformed player in a zero-sum repeated game with incomplete information on one side.

Key Words : Repeated Games, Blackwell Approachability, Partial Monitoring, Convex Sets, Incomplete Information

Introduction

Blackwell [4] introduced the notion of approachability in two-person (infinitely) repeated games with vector payoffs in some Euclidian space $\mathbb{R}^d$, as an analogue of Von Neumann's minmax theorem. A player can approach a given set $E \subset \mathbb{R}^d$, if he can insure that, after some stage and with a great probability, the average payoff will always remain close to E. Blackwell [4]

*Équipe Combinatoire et Optimisation, FRE 3232 CNRS, Université Pierre et Marie Curie - Paris 6, 4 place Jussieu, 75005 Paris. vianney.perchet@normalesup.org

proved that if both players observe their payoffs and E satisfies some geometric condition (E is then called a B-set), then Player 1 can approach it. He also deduced that given a convex set C either Player 1 can approach it or Player 2 can exclude it, *i.e.* the latter can approach the complement of a neighborhood of C. As Soulaimani, Quincampoix & Sorin [1] have recently proved that the notions of B-set (in a given repeated game) and discriminating domains (for a suitably chosen differential game) coincide.

We consider the partial monitoring framework, where players do not observe their opponent's moves but receive random signals. We provide in section 1.2 a necessary and sufficient condition under which a convex set is approachable. We also construct an approachability strategy derived from the construction (following Perchet [10]) of a strategy that has no internal regret (internal consistency in this framework has been defined by Lehrer & Solan [8], Definition 9).

Three classical results that hold in the full monitoring case do not extend to the partial monitoring framework. Indeed, in a specific game introduced in section 3.2, there exists a convex set C that is neither approachable by Player 1 nor excludable by Player 2 (see Theorem 3 in Blackwell [4]). Moreover, C is not approachable by Player 1 while every half-space that contains it is approachable by Plater 1 (see Corollary 2 in Blackwell [4]). Finally, C is neither weakly-approachable nor weakly-excludable (see Vieille [12]). We recall that weak-approachability is a weaker notion than approachability, also introduced by Blackwell [4], in finitely repeated games (see Definition 1.2 in section 1).

Kohlberg [7] used the notion of approachability in order to construct optimal strategies of the uninformed player, in the class of zero-sum repeated games with incomplete information on one side (introduced by Aumann & Maschler [2]). Our result can be used in this framework to provide a simple proof of the existence of a value in the infinitely repeated game through the construction of an ϵ-optimal strategy of Player 2.

1 Approachability

Consider a two-person game Γ repeated in discrete time. At stage $n \in \mathbb{N}$, Player 1 (resp. Player 2) chooses an action $i_n \in I$ (resp. $j_n \in J$), where both sets I and J are finite. This generates a vector payoff $\rho_n = \rho(i_n, j_n) \in \mathbb{R}^d$ where ρ is a mapping from $I \times J$ to $\mathbb{R}^d$. Player 1 does not observe j_n nor ρ_n but receives a random signal $s_n \in S$ whose law is $s(i_n, j_n)$ where s is a mapping from $I \times J$ to $\Delta(S)$ (the set of probabilities over the finite set S).

Player 2 observes i_n, j_n and s_n. The choices of i_n and j_n depend only on the past observations of the players and may be random.

Explicitly, a strategy σ of Player 1 is a mapping from H^1 to $\Delta(I)$ where $H^1 = \bigcup_{n\in\mathbb{N}} (I \times S)^n$ is the set of finite histories available to Player 1. After the finite history $h_n^1 \in (I \times S)^n$, $\sigma(h_n^1) \in \Delta(I)$ is the law of i_{n+1}. Similarly, a strategy τ of Player 2 is a mapping from $H^2 = \bigcup_{n\in\mathbb{N}} (I \times S \times J)^n$ to $\Delta(J)$. A couple of strategies (σ, τ) generates a probability, denoted by $\mathbb{P}_{\sigma,\tau}$, over $\mathcal{H} = (I \times S \times J)^{\mathbb{N}}$, the set of plays embedded with the cylinder σ-field.

The two functions ρ and s are extended multilinearly to $\Delta(I) \times \Delta(J)$ by $\rho(x, y) = \mathbb{E}_{x,y}[\rho(i, j)] \in \mathbb{R}^d$ and $s(x, y) = \mathbb{E}_{x,y}[s(i, j)] \in \Delta(S)$.

The following notations will be used: for any sequence $a = \{a_m \in \mathbb{R}^d\}_{m\in\mathbb{N}}$, the average of a up to stage n is denoted by $\overline{a}_n := \sum_{m=1}^{n} a_m/n$ and for any set $E \subset \mathbb{R}^d$, the distance to E is denoted by $d_E(z) := \inf_{e\in E} \|z - e\|$, where $\|\cdot\|$ is the Euclidian norm.

Definition 1.1 (Blackwell [4]) *i) A closed set $E \subset \mathbb{R}^d$ is approachable by Player 1 if for every $\varepsilon > 0$, there exist a strategy σ of Player 1 and $N \in \mathbb{N}$ such that for every strategy τ of Player 2 and every $n \geq N$:*

$$\mathbb{E}_{\sigma,\tau}[d_E(\overline{\rho}_n)] \leq \varepsilon \quad \text{and} \quad \mathbb{P}_{\sigma,\tau}\left(\sup_{n\geq N} d_E(\overline{\rho}_n) \geq \varepsilon\right) \leq \varepsilon.$$

Such a strategy σ_ε is called an ε-approachability strategy of E.

ii) A set E is excludable by Player 2, if there exists $\delta > 0$ such that the complement of E^δ is approachable by Player 2, where $E^\delta = \{z \in \mathbb{R}^d; d_E(z) \leq \delta\}$.

In words, a set $E \subset \mathbb{R}^d$ is approachable by Player 1, if he can insure that the average payoff converges almost surely to E, uniformly with respect to the strategies of Player 2. Obviously, a set E cannot be both approachable by Player 1 and excludable by Player 2.

Definition 1.2 *i) A closed set E is weakly-approachable by Player 1 if for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for every $n \geq N$, there is some strategy σ_n of Player 1 such that for every strategy τ of Player 2:*

$$\mathbb{E}_{\sigma_n,\tau}[d_E(\overline{\rho}_n)] \leq \varepsilon.$$

ii) A set E is weakly-excludable by Player 2, if there exists $\delta > 0$ such that the complement of E^δ is weakly-approachable by Player 2.

We emphasize the fact that in the definition of weak-approachability, the strategy of Player 1 might depend on n, the length of the game, which was not the case in the definition of approachability.

1.1 Full monitoring case

A game satisfies full monitoring if Player 1 observes the moves of Player 2, thus if $S = J$ and $s(i,j) = j$. Blackwell [4] gave a sufficient geometric condition under which a closed set E is approachable by Player 1. He also provided a full characterization for convex sets. Stating his condition requires the following notations: $\Pi_E(z) = \{e \in E; d_E(z) = \|z - e\|\}$ is the set of closest points to $z \in \mathbb{R}^d$ in E, and $P^1(x) = \{\rho(x,y); y \in \Delta(J)\}$ (resp. $P^2(y) = \{\rho(x,y); x \in \Delta(I)\}$) is the set of expected payoffs compatible with $x \in \Delta(I)$ (resp. $y \in \Delta(J)$).

Definition 1.3 *A closed set E of $\mathbb{R}^d$ is a B-set, if for every $z \in \mathbb{R}^d$, there exist $p \in \Pi_E(z)$ and $x\,(= x(z)) \in \Delta(I)$ such that the hyperplane through p and perpendicular to $z - p$ separates z from $P^1(x)$, or formally:*

$$\forall z \in \mathbb{R}^d, \exists p \in \Pi_E(z), \exists x \in \Delta(I), \langle \rho(x,y) - p, z - p\rangle \leq 0, \quad \forall y \in \Delta(J). \tag{1}$$

Condition (1) and therefore Theorem 1.4 do not require that Player 1 observes Player 2's moves, but only his own payoffs (which was Blackwell's assumption).

Theorem 1.4 (Blackwell [4]) *A B-set E is approachable by Player 1.*

Moreover, consider the strategy σ of Player 1 defined by $\sigma(h_n) = x(\overline{\rho}_n)$. Then for every strategy τ of Player 2 and every $\eta > 0$:

$$\mathbb{E}_{\sigma,\tau}[d_E^2(\overline{\rho}_n)] \leq \frac{4B}{n} \quad \text{and} \quad \mathbb{P}_{\sigma,\tau}\left(\sup_{n \geq N} d_E(\overline{\rho}_n) \geq \eta\right) \leq \frac{8B}{\eta^2 N}, \tag{2}$$

with $B = \sup_{i,j} \|\rho(i,j)\|^2$.

For a closed convex set C, a full characterization is available:

Corollary 1.5 (Blackwell [4]) *A closed convex set $C \subset \mathbb{R}^d$ is approachable by Player 1 if and only if:*

$$P^2(y) \cap C \neq \emptyset, \quad \forall y \in \Delta(J). \tag{3}$$

Using a minmax argument, Blackwell [4] proved that condition (3) implies condition (1), therefore the B-set C is approachable by Player 1. This characterization implies the following properties on convex sets:

Corollary 1.6 (Blackwell [4]) *1. A closed convex set C is either approachable by Player 1 or excludable by Player 2.*

2. A closed convex set C is approachable by Player 1 if and only if every half-space that contains C is approachable by Player 1.

If condition (3) is not fulfilled for some $y_0 \in \Delta(J)$, then (by the law of large numbers) Player 2 just has to play accordingly to y_0 at each stage to exclude C. If every half-space that contains C is approachable, then C is a B-set. Conversely any set that contains an approachable set is approachable.

Blackwell also conjectured the following result on weak-approachability, proved by Vieille:

Theorem 1.7 (Vieille [12]) *A closed set is either weakly-approachable by Player 1 or weakly-excludable by Player 2.*

Vieille [12] constructed a differential game $\mathcal{D}$ (in continuous time and with finite length) such that the finite repetitions of Γ can be seen as a discretization of $\mathcal{D}$. The existence of the value for $\mathcal{D}$ implies the result.

1.2 Partial monitoring case

The main objective of this section is to provide a simple necessary and sufficient condition under which a convex set C is approachable in the partial monitoring case.

Before stating it, we introduce the following notations: the vector of probabilities over S defined by $\mathbf{s}(y) = (s(i,y))_{i \in I} \in \Delta(S)^I$ is called the flag generated by $y \in \Delta(J)$. This flag is not observed by Player 1 since if he plays $i \in I$ he only observes a signal s which is the realization of the i-th component of $\mathbf{s}(y)$. However, it is theoretically the maximal information available to him about $y \in \Delta(J)$. Indeed, Player 1 will never be able to distinguish between any two mixed action y and y' that generate the same flag, *i.e.* such that $\mathbf{s}(y) = \mathbf{s}(y')$.

Given a flag μ in $\mathcal{S}$, the range of $\mathbf{s}$, $\mathbf{s}^{-1}(\mu) = \{y \in \Delta(J); \mathbf{s}(y) = \mu\}$ is the set of mixed actions of Player 2 compatible with μ. $P(x,\mu) = \{\rho(x,y); y \in \mathbf{s}^{-1}(\mu)\}$ is the set of expected payoffs compatible with $x \in \Delta(I)$ and $\mu \in \mathcal{S}$.

Our main result is:

Theorem 1.8 *A closed convex set $C \subset \mathbb{R}^d$ is approachable by Player 1 if and only if:*

$$\forall \mu \in \mathcal{S}, \exists x \in \Delta(I), P(x,\mu) \subset C. \tag{4}$$

$P(x,\cdot)$ can be extended to $\Delta(S)^I$ (without changing condition (4)) by defining, for every $\mu \notin \mathcal{S}$, either $P(x,\mu) = \emptyset$ or $P(x,\mu) = P(x,\Pi_{\mathcal{S}}(\mu))$, where $\Pi_{\mathcal{S}}(\cdot)$ is the projection onto $\mathcal{S}$.

In the full monitoring case, condition (4) is exactly condition (3). Indeed, if Player 1 observes Player 2's action then $S = J$, $\mathcal{S} = \{(y,\ldots,y) \in \Delta(J)^I; y \in \Delta(J)\}$ and given $\mathbf{y} = (y,\ldots,y) \in \mathcal{S}$, $P(x,\mathbf{y}) = \{\rho(x,y)\}$. Condition (4) implies that for every $y \in \Delta(J)$ there exists $x \in \Delta(I)$ such that $\rho(x,y) \in C$, or equivalently $P^2(y) \cap C \neq \emptyset$.

An other important result is that Corollary 1.6 and Theorem 1.7 do not extend:

Proposition 1.9

1. *There exists a closed convex set that is neither approachable by Player 1 nor excludable by Player 2*
2. *An half-space is either approachable by Player 1 or excludable by Player 2*
3. *There exists a closed convex set that is not approachable by Player 1 while every half-space that contains it is approachable by Player 1*
4. *There exists a closed convex set that is neither weakly-approachable by Player 1 nor weakly-excludable by Player 2.*

As said in the introduction, the proof of Theorem 1.8 relies on the construction of a strategy that has no internal regret in an auxiliary game with partial monitoring.

2 Internal regret with partial monitoring

Consider the following two-person repeated game $\mathcal{G}$ with partial monitoring. At stage $n \in \mathbb{N}$, we denote by $x_n \in \Delta(I)$ and $y_n \in \Delta(J)$ the mixed actions chosen by Player 1 and Player 2 (*i.e.* the laws of i_n and j_n). As before, we denote by s_n the signal observed by Player 1, whose law is the i_n-th coordinate of $\mu_n = \mathbf{s}(j_n)$.

Although payoffs are unobserved, given a flag $\mu \in \Delta(S)^I$ and $x \in \Delta(I)$, Player 1 evaluates his payoff through $G(x,\mu)$ where G is a continuous map from $\Delta(I) \times \Delta(S)^I$ to $\mathbb{R}$, not necessarily linear.

In the full monitoring framework, Foster & Vohra [5] defined internally consistent strategies (or strategies that have no internal regret) as follows: Player 1 has asymptotically no internal regret if for every $i \in I$, either the action i is a best response to his opponent's empirical distribution of actions on the set of stages where he actually played i, or the density of this set (also called the frequency of the action i) converges to zero.

In our framework, G is not linear so every action $i \in I$ (or the Dirac mass on i) might never be a best response; best responses are indeed elements of $\Delta(I)$. Thus if we want to define internal regret, we cannot distinguish the stages as a function of the actions actually played (*i.e.* $i_n \in I$) but as a function of the laws of the actions (*i.e.* $x_n \in \Delta(I)$).

We consider strategies described as follows: at stage n Player 1 chooses (at random) a law $x(l_n)$ in a finite set $\{x(l) \in \Delta(I); l \in L\}$ and given that choice, i_n is drawn accordingly to $x(l_n)$; l_n is called the type of the stage n.

We denote by $N_n(l) = \{1 \leq m \leq n; l_m = l\}$ the set of stages (before the n-th) of type l and for any sequence $a = \{a_m \in \mathbb{R}^d\}_{m\in\mathbb{N}}$, $\overline{a}_n(l) = \sum_{m\in N_n(l)} a_m/|N_n(l)|$ is the average of a on $N_n(l)$.

Definition 2.1 *For every $n \in \mathbb{N}$ and every $l \in L$, the internal regret of type $l \in L$ at stage n is*

$$\mathcal{R}_n(l) = \sup_{x\in\Delta(I)} \left[G(x, \overline{\mu}_n(l)) - G(x(l), \overline{\mu}_n(l))\right],$$

where $\overline{\mu}_n(l)$ is the unobserved average flag on $N_n(l)$.

A strategy σ of Player 1 is (L,ε)-internally consistent if for every strategy τ of Player 2:

$$\limsup_{n\to+\infty} \frac{|N_n(l)|}{n}\left(\mathcal{R}_n(l) - \varepsilon\right) \leq 0, \quad \forall l \in L, \quad \mathbb{P}_{\sigma,\tau}\text{--as.}$$

The set L is assumed to be finite, otherwise there would exist trivial strategies such that the frequency of every $x(l)$ converges to zero. In words, if σ is an (L,ε)-internally consistent strategy then either $x(l)$ is an ε-best response to $\overline{\mu}_n(l)$, the unobserved average flag on $N_n(l)$, or this set has a very small density.

Theorem 2.2 (Lehrer & Solan [8]; Perchet [10]) *For every $\varepsilon > 0$, there exist a finite set L and a (L,ε)-internally consistent strategy σ such that for every strategy τ of Player 2:*

$$\mathbb{E}_{\sigma,\tau}\left[\sup_{l\in L} \frac{|N_n(l)|}{n}\left(\mathcal{R}_n(l) - \varepsilon\right)\right] = O\left(\frac{1}{\sqrt{n}}\right) \quad \text{and}$$

$$\forall \eta > 0, \mathbb{P}_{\sigma,\tau}\left(\exists n \geq N, l \in L, \frac{|N_n(l)|}{n}\left(\mathcal{R}_n(l) - \varepsilon\right) > \eta\right) \leq O\left(\frac{1}{\eta^2 N}\right).$$

3 Proofs of the main results

This section is devoted to the proofs of the theorems stated in the previous section.

3.1 Proof of Theorem 1.8

Let C be a convex set such that for every $\mu \in \Delta(S)^I$ there exists $x_\mu \in \Delta(I)$ such that $P(x_\mu, \mu) \subset C$. Given $\varepsilon > 0$, we are going to construct an ε-approachability strategy in Γ based on an (L, ϵ)-internally consistent strategy in some auxiliary game $\mathcal{G}$, where the evaluation function G is defined by:

$$G(x, \mu) = -\sup_{y \in \mathbf{s}^{-1}(\mu)} d_C\left(\rho(x, y)\right)$$

if $\mu \in \mathcal{S}$. If $\mu \notin \mathcal{S}$, then $G(x, \mu) = G\left(x, \Pi_{\mathcal{S}}(\mu)\right)$ where $\Pi_{\mathcal{S}}$ is the projection onto $\mathcal{S}$.

Sufficiency: Any strategy in the auxiliary game $\mathcal{G}$ naturally defines a strategy in the original game Γ. The main idea of the proof is quite simple: given $\varepsilon > 0$, consider the finite family $\{x(l); l \in L\}$ and the (L, ε)-internally consistent strategy σ of Player 1 given by Theorem 2.2. Then for every $l \in L$, either $|N_n(l)|/n$ is very small, or $\mathcal{R}_n(l) \leq \varepsilon$. In that last case, the definition of G implies that $\overline{\rho}_n(l)$ is ε-close to C. Since

$$\overline{\rho}_n = \sum_{l \in L} \frac{|N_n(l)|}{n} \overline{\rho}_n(l), \tag{5}$$

$\overline{\rho}_n$ is a convex combination of terms that are ε-close to C. Since C is convex, $\overline{\rho}_n$ is also close to C.

Formally, let σ be a (L, ε)-internally consistent strategy of Player 1 given by Theorem 2.2. For every $\theta > 0$, there exists $N^1 \in \mathbb{N}$ such that for any strategy τ of Player 2:

$$\mathbb{P}_{\sigma,\tau}\left(\forall n \geq N^1, \sup_{l \in L} \frac{|N_n(l)|}{n}\left(\mathcal{R}_n(l) - \varepsilon\right) \leq \theta\right) \geq 1 - \theta. \tag{6}$$

Recall that for any $\mu \in \Delta(S)^I$ there exists $x_\mu \in \Delta(I)$ such that $P(x_\mu, \mu) \subset C$, therefore $\sup_{z \in \Delta(I)} G(z, \mu) = G(x_\mu, \mu) = 0$ and

$$\mathcal{R}_n(l) = \sup_{y \in \mathbf{s}^{-1}(\overline{\mu}_n(l))} d_C\left(\rho(x(l), y)\right) \geq d_C\Big(\rho\big(x(l), \overline{\jmath}_n(l)\big)\Big),$$

because $\mathbf{s}(\overline{\jmath}_n(l)) = \mu_n(l)$ by linearity of $\mathbf{s}$.

The random variables l_n and j_n are independent (given the finite histories) and so are i_n and j_n given l_n. Thus Hoeffding-Azuma [3, 6]'s inequality for sums of bounded martingale differences implies that $\rho\left(x(l), \overline{\jmath}_n(l)\right)$ is asymptotically close to $\overline{\rho}_n(l)$. Explicitly, for every $\theta > 0$, there exists $N^2 \in \mathbb{N}$ (independent of σ and τ) such that:

$$\mathbb{P}_{\sigma,\tau}\left(\forall n \geq N^2, \exists l \in L, \frac{|N_n(l)|}{n}\left|\overline{\rho}_n(l) - \rho\left(x(l), \overline{\jmath}_n(l)\right)\right| \leq \theta\right) \geq 1 - \theta. \quad (7)$$

Equations (6) and (7) imply that for every $n \geq N = \max\{N^1, N^2\}$ and every $l \in L$, with probability at least $1 - 2\theta$:

$$\frac{|N_n(l)|}{n}\big(d_C\left(\overline{\rho}_n(l)\right) - \varepsilon\big) \leq 2\theta.$$

Since C is a convex set, $d_C(\cdot)$ is convex, thus for any strategy τ of Player 2, with $\mathbb{P}_{\sigma,\tau}$-probability at least $1 - 2\theta$, for every $n \geq N$:

$$d_C(\overline{\rho}_n) \leq \sum_{l \in L} \frac{|N_n(l)|}{n} d_C(\overline{\rho}_n(l)) \leq 2L\theta + \varepsilon,$$

and C is approachable by Player 1.

Necessity: Conversely, assume that there exists $\mu_0 \in \Delta(S)^I$ such that for all $x \in \Delta(I)$, there is some $y(= y(x)) \in \mathbf{s}^{-1}(\mu_0)$ such that $d_C\left(\rho(x, y)\right) > 0$. Since $\Delta(I)$ is compact, we can assume that there exists $\delta > 0$ such that $d_C\left(\rho(x, y(x))\right) \geq \delta$.

Let $\mathcal{T}_0$ be the subset of strategies of Player 2 that generate at any stage the same flag μ_0 (explicitly, a strategy τ belongs to $\mathcal{T}_0$ if for every finite history h_n^2, $\tau(h_n^2) \in \mathbf{s}^{-1}(\mu_0)$). Recall that a strategy σ of Player 1 depends only on his past actions and on the signals he received. Since at any stage, two strategies τ and τ' in $\mathcal{T}_0$ induce the same laws of signals, the couples (σ, τ) and (σ, τ') generate the same probability on the infinite sequences of moves of Player 1. Therefore $\mathbb{E}_{\sigma,\tau}\left[\overline{\imath}_n\right] = \mathbb{E}_{\sigma,\tau'}\left[\overline{\imath}_n\right] := \overline{x}_n$ is independent of τ.

For every $n \in \mathbb{N}$, define the strategy τ_n in $\mathcal{T}_0$ by $\tau_n(h) = y(\overline{x}_n)$, for all finite history h. Since $d_C(\cdot)$ is convex, by Jensen's inequality

$$\mathbb{E}_{\sigma,\tau_n}\left[d_C\left(\overline{\rho}_n\right)\right] \geq d_C\left(\mathbb{E}_{\sigma,\tau_n}\left[\overline{\rho}_n\right]\right).$$

Since j_m is independent of the history h_{m-1}:

$$\mathbb{E}_{\sigma,\tau_n}\left[\rho(i_m, j_m)\middle|h_{m-1}\right] = \mathbb{E}_{\sigma,\tau_n}\left[\rho(i_m, y(\overline{x}_n))\middle|h_{m-1}\right]$$

hence by linearity of $\rho(\cdot, y(\overline{x}_n))$,

$$\mathbb{E}_{\sigma,\tau_n}\left[\rho(i_m, j_m)\big|h_{m-1}\right] = \rho\left(\mathbb{E}_{\sigma,\tau_n}\left[i_m\big|h_{m-1}\right], y(\overline{x}_n)\right).$$

Therefore $\mathbb{E}_{\sigma,\tau_n}\left[\overline{\rho}_n\right] = \rho(\overline{x}_n, y(\overline{x}_n))$. Consequently

$$\mathbb{E}_{\sigma,\tau_n}\left[d_C\left(\overline{\rho}_n\right)\right] \geq d_C\left(\mathbb{E}_{\sigma,\tau_n}\left[\overline{\rho}_n\right]\right) = d_C\left(\rho(\overline{x}_n, y(\overline{x}_n))\right) \geq \delta$$

and for any strategy σ of Player 1 and any stage $n \in \mathbb{N}$, Player 2 has a strategy such that the expected average payoff is at a distance greater than $\delta > 0$ from C. Thus C is not approachable by Player 1.

Remark 3.1 *The fact that C is a convex set is crucial in both parts of the proof. In the sufficient part, it would otherwise be possible that $\overline{\rho}_n(l) \in C$ for every $l \in L$, while $\overline{\rho}_n \notin C$. In the necessary part, the counterpart could happen: $d_C\left(\mathbb{E}\left[\overline{\rho}_n\right]\right) \geq \delta$ while $\mathbb{E}\left[d_C(\overline{\rho}_n)\right] = 0$.*

Remark 3.2 *The ε-approachability strategy constructed relies on a (L, ε)-internally consistent strategy, so one can easily show that:*

$$\mathbb{E}_{\sigma,\tau}\left[d_C\left(\overline{\rho}_n\right)\right] = \varepsilon + O\left(\frac{1}{\sqrt{n}}\right) \quad \text{and}$$

$$\mathbb{P}_{\sigma,\tau}\left(\exists n \geq N, d_C\left(\overline{\rho}_n\right) - \varepsilon > \eta\right) \leq O\left(\frac{1}{\eta^2 N}\right).$$

Corollary 3.3 *There exists σ a strategy of Player 1 such that for every $\eta > 0$, there exists $N \in \mathbb{N}$ such that for every strategy τ of Player 2 and $n \geq N$, $\mathbb{E}_{\sigma,\tau}\left[d_C\left(\overline{\rho}_n\right)\right] \leq \eta$.*

The proof is rather classical and relies on a careful concatenation of ε_k-approachability strategies (where the sequence $(\varepsilon_k)_{k\in\mathbb{N}}$ decreases towards 0) called *doubling trick* (see e.g. Sorin [11], Proposition 3.2). It is therefore omitted.

3.2 Proof of Proposition 1.9

In the proof of Theorem 1.8, we have shown that if a convex set is not approachable by Player 1 then for any of his strategy and any $n \in \mathbb{N}$, Player 2 has a strategy τ_n such that $\overline{\rho}_n$ is at, at least, δ from C. It does not imply that C is excludable by Player 2; indeed this would require that τ_n does not depend on σ nor n. The proof of Proposition 1.9 relies mainly on the study of the following example.

Proof of Proposition 1.9. Consider the following matrix two-person repeated game where Player 1 (the row player) receives no signal and his one-dimensional payoffs are defined by :

	L	R
T	0	1
B	-1	0

$\mathbf{C := [0; 1/2]}$ is neither approachable nor excludable: The closed convex set $C := [0; 1/2]$ is obviously not approachable by Player 1 (otherwise Theorem 1.8 implies that there exists $x \in \Delta(I)$ such that $\rho(x, y) \in [0, 1/2]$ for every $y \in \Delta(J)$). More precisely, given a strategy σ of Player 1, we define τ_n as follows: if $\overline{x}_n$ (the expected frequency of T up to stage $n \in \mathbb{N}$ — it does not depend on Player 2's strategy) is smaller than $1/4$, then τ_n is the strategy that always plays L, otherwise that always plays R. Then the law of large numbers implies that, for n big enough, $\mathbb{E}_{\sigma,\tau_n}\left[d_C\left(\overline{\rho}_n\right)\right]$ is arbitrarily close to $1/4$.

It remains to show that Player 2 cannot exclude C. We prove this by constructing a strategy σ of Player 1 such that the average payoff is infinitely often close to 0: σ is played in blocks and the length of the p-th block is p^{2p+1}. On odd blocks, Player 1 plays T while on even blocks he plays B. At the end of the block p, the average payoff is at most $1/p$ if it is an odd block and at least $-1/p$ otherwise. Hence on two consecutive blocks (the p-th and the $p+1$-th) there is at least one stage such that the average payoff is at a distance smaller than $1/p$ to $\{0\}$. Therefore $\{0\}$ and C (since it contains $\{0\}$) is not excludable by Player 2.

An half-space is either approachable by Player 1 or excludable by Player 2: Let E be an half-space not approachable by Player 1. Then there exists $\mu_0 \in \Delta(S)^I$ such that, for every $x \in \Delta(I)$, $P(x, \mu_0) \not\subset E$. This implies that there exists $\delta > 0$ such that $\inf_{x \in \Delta(I)} \sup_{y \in \mathbf{s}^{-1}(\mu_0)} d_E\left(\rho(x, y)\right) \geq \delta > 0$ and therefore for every $x \in \Delta(I)$, there exists $y \in \Delta(J)$ such that $\rho(x, y)$ is in the complement of E^δ which is convex, since E is an half-space. Blackwell's result applies for Player 2 (since we assumed he has full monitoring), so he can approach the complement of E^δ and exclude E.

C is not approachable by Player 1 while every half-space that contains it is: An half-space that contains C contains either $(-\infty, 0]$ or $[0, +\infty)$ which are approachable by, respectively, always playing T or always playing B.

C is neither weakly-approachable by Player 1 nor weakly excludable by Player 2 : we proved that for every strategy σ of Player 1 and every $n \in \mathbb{N}$ big enough, Player 2 has a strategy τ_n such that $\mathbb{E}_{\sigma,\tau_n}[d_C(\overline{\rho}_n)] = 1/2$. Hence C is not weakly approachable.

Conversely, let τ be a strategy of Player 2 in the game repeated $2n$ times (where n is large enough) and $M \in \mathbb{N}$ be any integer. Consider the strategy σ of Player 1 that consists in playing T during the first n stages. Since $\overline{\rho}_n$, the average payoff after those n stages, belongs to $[0;1]$, there exists an integer $k_1 \in \{1,\ldots,M\}$ such that $\overline{\rho}_n$ belongs to $[\frac{k_1-1}{M};\frac{k_1}{M}]$ with $\mathbb{P}_{\sigma,\tau}$-probability at least $\frac{1}{M}$. Note that, given τ , Player 1 can compute this k.

Assume that, from stage $n+1$ on, the strategy σ dictates to play i.i.d action B with probability $\frac{k_1}{M}$ and action T with probability $1-\frac{k_1}{M}$. If n is large enough, the probability that the average payoff between stages $n+1$ and $2n$ belongs to $[-\frac{k_1}{M}-\frac{1}{M};1-\frac{k_1}{M}+\frac{1}{M}]$ is close to one (say bigger than 1/2, this is again a direct consequence of the law of large number). Therefore, this strategy σ ensures that with $\mathbb{P}_{\sigma,\tau}$-probability at least $\frac{1}{2M}$ the average payoff over the $2n$ stages belongs to $[-\frac{1}{M};\frac{1}{2}+\frac{1}{2M}]$.

Denote by $\left(C^{2/M}\right)^c$ the complement of the $\frac{2}{M}$-neighborhood of C. Given a strategy τ of Player 2 and an integer n big enough, the strategy σ we described ensures that $\mathbb{E}_{\sigma,\tau}\left[d_{\left(C^{2/M}\right)^c}(\overline{\rho}_{2n})\right] \geq \frac{1}{2M^2}$. Therefore, for every $M \in \mathbb{N}$, $\left(C^{2/M}\right)^c$ is not weakly-approachable by Player 2 hence C is not weakly-excludable.

The strategy σ we described can be easily made independent of τ by, for example, choosing $k_1 \in \{1,\ldots,M\}$ at random; indeed, this would imply that $\mathbb{E}_{\sigma,\tau}\left[d_{\left(C^{2/M}\right)^c}(\overline{\rho}_{2n})\right] \geq \frac{1}{2M^3}$. □

These results hold if one chooses $C_3 := [0;1/3]$ instead of $[0;1/2]$. In fact, it only remains to prove that C_3 is not weakly-excludable by Player 2. Consider the game repeated $3n$ times and the strategy σ, defined by block of size n, that plays on the first block always T, on the second block i.i.d. action B with probability $\frac{k_1}{M}$. The average payoff on those two block belongs to a small neighborhood of $[0;1/2]$, hence to some $[\frac{k_2-1}{M},\frac{k_2}{M}]$ (where $k_2 \leq \frac{M}{2}$) with probability at least $\frac{1}{M}$. Assume that on the third block Player 1 plays i.i.d action B with probability $\frac{2k_2}{M}$ then the average payoff over the three blocks belongs to a small neighborhood of $[0;1/3]$ with probability at least $\frac{1}{(2M)^2}$. Therefore C_3 is not weakly excludable.

Since this proof can be generalized to any set $C_k = [0;\frac{1}{k}]$, even the singleton $\{0\}$ is neither weakly-approachable nor weakly-excludable; we recall that in the full monitoring framework all those convex sets are approachable by Player 1.

3.3 Remarks on the counterexample

Following Mertens, Sorin & Zamir's notations [9] (see Definition 1.2 p. 149), Player 1 can guarantee $\underline{v}$ in a zero-sum repeated game Γ_∞ if

$$\forall \varepsilon > 0, \exists \sigma_\varepsilon, \exists N \in \mathbb{N}, \mathbb{E}_{\sigma_\varepsilon,\tau}\left[\overline{\rho}_n\right] \geq \underline{v} - \varepsilon, \forall \tau, \forall n \geq N,$$

where σ_ε is a strategy of Player 1, and τ any strategy of Player 2. Player 2 can defend $\underline{v}$ if:

$$\forall \varepsilon > 0, \forall \sigma_\varepsilon, \exists \tau, \exists N \in \mathbb{N}, \mathbb{E}_{\sigma_\varepsilon,\tau}\left[\overline{\rho}_n\right] \leq \underline{v} + \varepsilon, \forall n \geq N.$$

If Player 1 can guarantee $\underline{v}$ and Player 2 defend $\underline{v}$, then $\underline{v}$ is the maxmin of Γ_∞. The minmax $\overline{v}$ is defined in a dual way and Γ_∞ has a value if $\underline{v} = \overline{v}$.

These definitions can be extended to the vector payoff framework: we say that Player 1 can guarantee a set E if he can approach E:

$$\forall \varepsilon > 0, \exists \sigma_\varepsilon, \exists N \in \mathbb{N}, \mathbb{E}_{\sigma_\varepsilon,\tau}\left[d_E\left(\overline{\rho}_n\right)\right] \leq \varepsilon, \forall \tau, \forall n \geq N.$$

In the counterexample of the proof of Proposition 1.9, Player 1 cannot guarantee the convex set $C = \{0\}$ and Player 2 cannot defend it since:

$$\exists \sigma, \forall \varepsilon > 0, \forall \tau, \forall N \in \mathbb{N}, \exists n \geq N, \mathbb{E}_{\sigma,\tau}\left[d_C\left(\overline{\rho}_n\right)\right] \leq \varepsilon.$$

To keep the notations of zero-sum repeated game, one could say that the game we constructed *has no maxmin.*

Blackwell [4] also gave an example of a game (with vector payoff) without maxmin in the full monitoring case. The main differences between the two examples are:

i) in the partial monitoring case this set can be convex (which cannot occur in the full monitoring framework);

ii) the strategy of Player 1 is such that the average payoff is infinitely often close to C. However, unlike Blackwell's example, he does not know at which stages.

4 Repeated game with incomplete information on one side, with partial monitoring

Aumann & Maschler [2] introduced the class of two-person zero-sum games with incomplete information on one side. Those games are described as follows: Nature chooses k_0 from a finite set of states K according to some known probability $p \in \Delta(K)$. Player 1 (the maximizer) is informed about k_0 but not Player 2. At stage $m \in \mathbb{N}$, Player 1 (resp. Player 2) chooses $i_m \in I$ (resp. $j_m \in J$) and the payoff is $\rho_m^{k_0} = \rho^{k_0}(i_m, j_m)$. Player 1 observes j_m and Player 2 does not observe i_m nor ρ_m but receives a signal s_m whose law is $s^{k_0}(i_m, j_m) \in \Delta(S)$. As in the previous sections, we define $\mathbf{s}^k(x) = \left(s^k(x,j)\right)_{j\in K}$, for every $x \in \Delta(I)$.

A strategy σ (resp. τ) of Player 1 (resp. Player 2) is a mapping from $K \times \bigcup_{\in\mathbb{N}} (I \times J \times S)^m$ to $\Delta(I)$ (resp. from $\bigcup_{m\in\mathbb{N}} (J \times S)^m$ to $\Delta(J)$). At stage $m+1$, $\sigma(k, h_m^1)$ is the law of i_{m+1} after the history h_m^1 if the chosen state is k.

We define Γ_1 the one-shot game with expected payoff $\sum_{k\in K} p^k \rho^k(x^k, y)$ and $\Gamma_\infty(p)$ the infinitely repeated game. We denote by $v_\infty(p)$ its value, if it exists (*i.e.* if both Player 1 and Player 2 can guarantee it). Aumann & Maschler [2] (Theorem C, p. 191) proved that $\Gamma_\infty(p)$ has a value and characterized it.

Let us first introduce the operator **Cav** and the non-revealing game $D(p)$: for any function f from $\Delta(I) \times \Delta(J)$ to $\mathbb{R}$, $\mathbf{Cav}(f)(\cdot)$ is the smallest (pointwise) concave function greater than f.

A profile of mixed actions $x = (x^k)_{k\in K} \in \Delta(I)^K$ is non-revealing at $p \in \Delta(K)$ (and induces the flag $\mu \in \Delta(S)^J$) if the flag induced by x is independent of the state:

$$NR(p,\mu) = \left\{x = (x^1, \dots, x^K) \in \Delta(I)^K \,|\, \mathbf{s}^k(x^k) = \mu, \forall k \text{ st } p^k > 0\right\}.$$

We denote by $NR(p) = \bigcup_{\mu\in\Delta(S)^J} NR(p,\mu)$ the set of non-revealing strategies. For every $\mu \in \Delta(S)^J$, $D(p,\mu)$ (resp. $D(p)$) is the one-stage game Γ_1 where Player 1 is restricted to $NR(p,\mu)$ (resp. $NR(p)$) and its value is denoted by $u(p,\mu)$ (resp. $u(p)$), with $u(p,\mu) = -\infty$ if $NR(p,\mu) = \emptyset$ (resp. $u(p) = -\infty$ if $NR(p) = \emptyset$).

Theorem 4.1 (Aumann & Maschler [2]) *The game* Γ_∞ *has a value defined by* $v_\infty(p) = \mathbf{Cav}(u)(p)$.

Proof. Player 1 can guarantee $u(p)$: indeed if $NR(p) \neq \emptyset$, he just has to play i.i.d. an optimal strategy in $NR(p)$ and otherwise $u(p) = -\infty$. Therefore, using the splitting procedure (see Lemma 5.2 p. 25 in [2]), Player 1 can guarantee $\mathbf{Cav}(u)(p)$.

It remains to show that Player 2 can also guarantee $\mathbf{Cav}(u)(p)$. The function $\mathbf{Cav}(u)(\cdot)$ is concave and continuous, therefore there exists $\mathbf{m} = (\mathbf{m}^1, \ldots, \mathbf{m}^k) \in \mathbb{R}^K$ such that $\mathbf{Cav}(u)(p) = \langle \mathbf{m}, p \rangle$ and $u(q) \leq \mathbf{Cav}(u)(q) \leq \langle \mathbf{m}, q \rangle$. Instead of constructing a strategy of Player 2 that minimizes the expected payoff $\sum_{k \in K} p^k \overline{\rho}_n^k$, it is enough to construct a strategy such that each $\overline{\rho}_n^k$ is smaller than $\mathbf{m}^k$, for every state k that has a positive probability accordingly to Player 2's posterior.

Therefore, we consider an auxiliary two-person repeated game with vector payoff where at stage $n \in \mathbb{N}$, Player 2 (resp. Player 1) chooses j_n accordingly to $y_n \in \Delta(J)$ (resp. $(i_n^1, \ldots, i_n^K)$ accordingly to $(x_n^1, \ldots, x_n^K) \in \Delta(I)^K$). Player 2 receives a signal s_n whose law is $s^{k_0}(i_n^{k_0}, j_n)$ where k_0 is the true state. We denote by $\mu_n = \mathbf{s}^{k_0}(x_n^{k_0})$ the expected flag of stage n. The k-th component of the vector payoff ρ_n is defined by $\rho^k(i_n^k, j_n)$ if μ_n belongs to $\mathcal{S}^k$, the range of $\mathbf{s}^k$ and $-A := -\max_{k \in K} \|\rho^k\|_\infty$ otherwise[1]. Conversely, the set of compatible payoffs given a flag $\mu \in \Delta(S)^J$, $y \in \Delta(J)$ and a state k, is defined by:

$$P^k(\mu, y) = \left\{ \rho^k(x^k, y) \,\middle|\, \mathbf{s}^k(x^k) = \mu \right\} \text{ if } \mu \in \mathcal{S}^k, \text{ otherwise } P^k(\mu, y) = \{-A\},$$

and the set of compatible vector payoffs is $P(\mu, y) = \Pi_{k \in K} P^k(\mu, y) \subset \mathbb{R}^K$.

If Player 2 can approach $M = \{m \in \mathbb{R}^K; m^k \leq \mathbf{m}^k, \forall k \in K\} = \mathbf{m} + \mathbb{R}^K_-$, then he can guarantee $\mathbf{Cav}(u)(p)$. Theorem 1.8 implies that the convex set M is approachable if and only if, for every $\mu \in \Delta(S)^I$ there exists $y \in \Delta(J)$ such that $P(\mu, y) \subset M$.

Hence it is enough to prove that this property holds. Assume the converse: there exists $\mu_0 \in \Delta(S)^I$ such that for every $y \in \Delta(J)$, $P(\mu_0, y)$ is not included in M.

We denote by $K(\mu_0) = \left\{k \in K; \mu_0 \in \mathcal{S}^k\right\}$ the set of states that are compatible with μ_0: if Player 2 observes μ_0, then he knows that the true state is in $K(\mu_0)$. For every $y \in \Delta(J)$ and $k \in K(\mu_0)$, $\omega_0^k(y) = \sup_{\mathbf{s}^k(x^k) = \mu_0} \rho^k(x^k, y)$ is the worst payoff for Player 2 in state k. The fact that $P(\mu_0, y)$ is not included in M implies that $\omega_0(y) = (\omega_0^k(y))_{k \in K(\mu_0)}$ does not belong to

[1]We use this notation, because if μ_n is not in the range of $\mathbf{s}^k$, then Player 2 knows that the true state is not k, and therefore does not need to minimize the k-th component of the payoff vector

$M_0 = \{m \in \mathbb{R}^{K(\mu_0)}; m^k \le \mathbf{m}^k, \forall k \in K(\mu_0)\}$. Define the convex set:

$$W_0 = \{\omega_0(y); y \in \Delta(J)\} + \mathbb{R}_+^{K(\mu_0)} \bigcap B(0, A),$$

with $B(0, A)$ the closed ball of radius A. Obviously $W_0 \cap M_0 = \emptyset$ and, by linearity of each ρ^k, W_0 is a compact convex set. So there exists a strongly separating hyperplane $H_0 = \{\omega \in \mathbb{R}^{K(\mu_0)}; \langle \omega, q_0 \rangle = b\}$ such that $\sup_{m \in M_0} \langle m, q_0 \rangle < \inf_{\omega \in W_0} \langle \omega, q_0 \rangle$. Every component of q_0 must be non-negative (since M_0 is negatively comprehensive), therefore up to a normalization, we can assume that q_0 belongs to $\Delta(K(\mu_0))$.

Define $W = W_0 \times \mathbb{R}^{K \setminus K(\mu_0)}$ and $q \in \Delta(K)$ by $q(k) = q_0(k)$ if $k \in K(\mu_0)$ and 0 otherwise. Then, $H = \{\omega \in \mathbb{R}^K; \langle \omega, q \rangle = b\}$ strongly separates M and W, therefore:

$$\langle \mathbf{m}, q \rangle < \min_{\omega \in W_0} \langle \omega, q \rangle = \min_{y \in \Delta(J)} \max_{x \in NR(q, \mu_0)} \sum_{k \in K} q^k \rho^k(x^k, y) = u(q, \mu_0) \le u(q)$$

and by definition of $\mathbf{m}$, $u(q) \le \langle \mathbf{m}, q \rangle$ which is impossible.

So M is approachable by Player 2, he can guarantee $\mathbf{Cav}(u)(p)$ in $\Gamma_\infty(p)$ and $v_\infty(p) = \mathbf{Cav}(u)(p)$. □

Acknowledgments: I deeply thank my advisor Sylvain Sorin for its great help and numerous comments. I also acknowledge useful comments from Jérôme Renault and Gilles Stoltz and, of course, thank them for pointing out the counter example to me.

References

[1] S. As Soulaimani, M. Quincampoix, S. Sorin (2009) *Repeated games and qualitative differential games: approachability and comparison of strategies*. SIAM J. Control Optim., Vol. 48, 2461–2479

[2] R.J. Aumann and M. B. Maschler (1955) Repeated Games with Incomplete Information *MIT Press*,

[3] K. Azuma (1967) *Weighted sums of certain dependent random variables*, Tôhoku Math. J. (2) V. 19, 357–367

[4] D. Blackwell (1956) *An analog of the minimax theorem for vector payoffs*. Pacific J. Math., Vol. 6, 1–8.

[5] D. P. Foster and R. V. Vohra (1997) *Calibrated learning and correlated equilibrium* Games Econom. Behav., Vol. 21, 40–55

[6] W. Hoeffding (1963) *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. Vol. 58, 13–30

[7] E. Kohlberg (1975) *Optimal strategies in repeated games with incomplete information* Internat. J. Game Theory, Vol. 4, 7–24

[8] E. Lehrer, and E. Solan (2007) *Learning to play partially-specified equilibrium*, manuscript.

[9] J.-F. Mertens, S. Sorin and S. Zamir. (1994) Repeated Games *CORE discussion paper 9420-9422*

[10] V. Perchet (2009) *Calibration and internal no-regret with random signals*, Proceedings of the 20th International Conference on Algorithmic Learning Theory, 68–82

[11] S. Sorin (1987) *Supergames*, in: Game theory and applications, Academic Press, San Diego, CA, 46–82

[12] N. Vieille (1992) *Weak Approachability*, Math. Oper. Res., Vol. 17, 781–791