

Robust approachability

with applications to regret minimization in games with partial monitoring

Gilles Stoltz

CNRS — École normale supérieure — INRIA, project-team CLASSIC
& HEC Paris



Joint work with **Vianney Perchet** (Université Paris-Diderot)
and **Shie Mannor** (Technion)

Blackwell's approachability

For games with full or bandit monitoring

A vector-valued base game

- Finite action sets \mathcal{A} and \mathcal{B}
- Sets of distributions over these action sets $\Delta(\mathcal{A})$ and $\Delta(\mathcal{B})$
- Payoff function $r : \mathcal{A} \times \mathcal{B} \rightarrow \mathbb{R}^d$, linearly extended to $\Delta(\mathcal{A}) \times \Delta(\mathcal{B})$
- Target (often closed convex) set $\mathcal{C} \subset \mathbb{R}^d$

Which is repeated

At each round $t \geq 1$,

- The decision maker chooses $\mathbf{x}_t \in \Delta(\mathcal{A})$ and possibly draws $I_t \sim \mathbf{x}_t$ at random
- Nature chooses $\mathbf{y}_t \in \Delta(\mathcal{B})$ and possibly draws $J_t \sim \mathbf{y}_t$ at random
- The payoffs $r(\mathbf{x}_t, \mathbf{y}_t)$ or $r(I_t, J_t)$ are obtained and observed

Summary

In the non-randomized version, players

- sequentially choose $\mathbf{x}_t \in \Delta(\mathcal{A})$ and $\mathbf{y}_t \in \Delta(\mathcal{B})$
- obtain the payoff $r(\mathbf{x}_t, \mathbf{y}_t) \in \mathbb{R}^d$

We will be interested in the average payoff

$$\bar{\mathbf{r}}_T = \frac{1}{T} \sum_{t=1}^T r(\mathbf{x}_t, \mathbf{y}_t)$$

Note that by concentration inequalities, it has the same behavior as its randomized counterpart

$$\tilde{\mathbf{r}}_T = \frac{1}{T} \sum_{t=1}^T r(I_t, J_t)$$

where $I_t \sim \mathbf{x}_t$ and $J_t \sim \mathbf{y}_t$.

We will thus focus on $\bar{\mathbf{r}}_T$, that is, on the game with mixed actions taken.

Summary

In the non-randomized version, players

- sequentially choose $\mathbf{x}_t \in \Delta(\mathcal{A})$ and $\mathbf{y}_t \in \Delta(\mathcal{B})$
- obtain the payoff $r(\mathbf{x}_t, \mathbf{y}_t) \in \mathbb{R}^d$

We will be interested in the average payoff

$$\bar{\mathbf{r}}_T = \frac{1}{T} \sum_{t=1}^T r(\mathbf{x}_t, \mathbf{y}_t)$$

Goals: Approachability of \mathcal{C}

A set $\mathcal{C} \subset \mathbb{R}^d$ is r -approachable if there exists a strategy for the decision maker such that for **all strategies of Nature**,

$$d(\bar{\mathbf{r}}_T, \mathcal{C}) = \inf_{c \in \mathcal{C}} \left\| c - \frac{1}{T} \sum_{t=1}^T r(\mathbf{x}_t, \mathbf{y}_t) \right\|_2 \rightarrow 0$$

Theorem (Blackwell '56)

A closed convex set \mathcal{C} is approachable if and only

$$\forall \mathbf{y} \in \Delta(\mathcal{B}), \exists \mathbf{x} \in \Delta(\mathcal{A}), \quad \mathbf{r}(\mathbf{x}, \mathbf{y}) \in \mathcal{C}.$$

The **sufficiency** of the condition was proved in a constructive way.

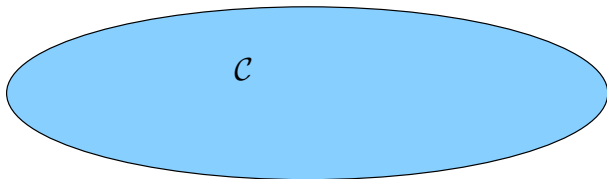
Denoting by M a bound in ℓ^2 -norm over \mathbf{r} , i.e.,

$$\max_{(a,b) \in \mathcal{A} \times \mathcal{B}} \|\mathbf{r}(a,b)\|_2 \leq M,$$

Blackwell's approachability strategy (described on the next slide) ensures that for all strategies of Nature,

$$d(\bar{\mathbf{r}}_T, \mathcal{C}) = \inf_{\mathbf{c} \in \mathcal{C}} \left\| \mathbf{c} - \frac{1}{T} \sum_{t=1}^T \mathbf{r}(\mathbf{x}_t, \mathbf{y}_t) \right\|_2 \leq \frac{2M}{\sqrt{T}}.$$

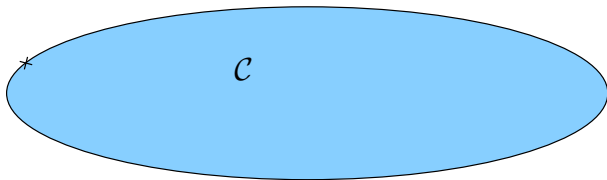
Blackwell's approachability strategy for a closed convex set

 $\times \bar{r}_t$ 

We indicate how to choose \mathbf{x}_{t+1} based on the past.

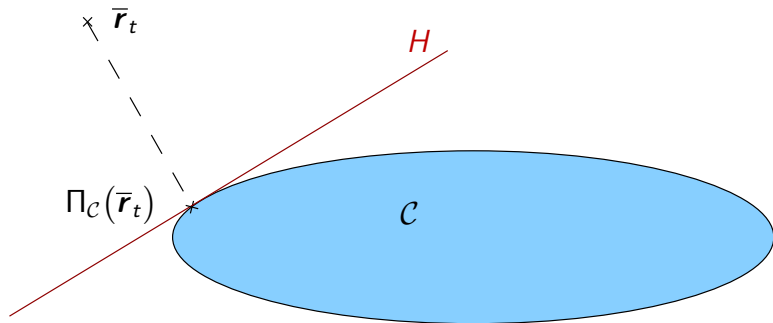
At round t the average payoff is $\bar{r}_t = \frac{1}{t} \sum_{s=1}^t r(\mathbf{x}_s, \mathbf{y}_s)$.

Blackwell's approachability strategy for a closed convex set

 $\times \bar{r}_t$ $\Pi_C(\bar{r}_t)$ 

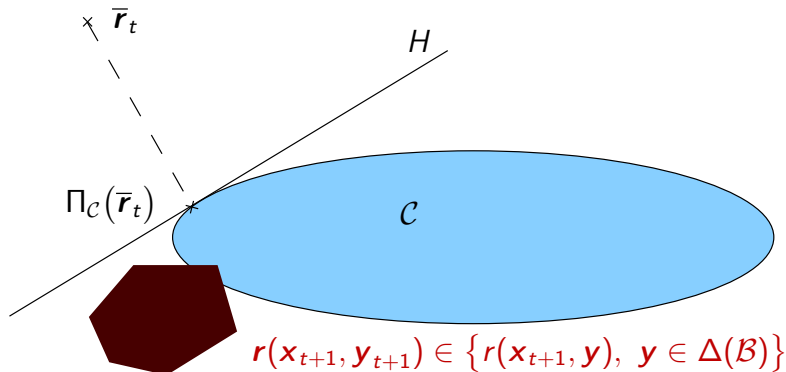
Aim: That \bar{r}_{t+1} gets closer to $\Pi_C(\bar{r}_t)$, the projection of \bar{r}_t onto C .

Blackwell's approachability strategy for a closed convex set



This is true as soon as $r(x_{t+1}, y_{t+1})$ is on the other side of H .

Blackwell's approachability strategy for a closed convex set



Given H , the existence of a $x_{t+1} \in \Delta(\mathcal{A})$ such that the property illustrated above takes place is guaranteed by the approachability condition (and the **minmax theorem**); it can be found by solving a **minmax program**.

An **application** is formed by the **minimization of regret**.

A (scalar) payoff function $r : \mathcal{A} \times \mathcal{B} \rightarrow \mathbb{R}$ is given and the goal of the decision maker is to ensure that

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \left(\sum_{t=1}^T r(\mathbf{x}_t, \mathbf{y}_t) - \max_{a \in \mathcal{A}} \sum_{t=1}^T r(a, \mathbf{y}_t) \right) \geq 0$$

That is, his payoff should on average be almost as large as what he would have obtained by playing a constant action $a \in \mathcal{A}$, all things being equal.

To do so, it suffices to r -approach \mathcal{C} , where $r(\mathbf{x}, \mathbf{y}) = \begin{bmatrix} r(\mathbf{x}, \mathbf{y}) \\ \mathbf{y} \end{bmatrix}$

and $\mathcal{C} = \left\{ (z, \mathbf{y}) \in \mathbb{R} \times \Delta(\mathcal{B}) : z \geq \max_{a \in \mathcal{A}} r(a, \mathbf{y}) \right\}$

Approachability in games with partial monitoring

Our initial motivation and starting point

In games **with partial monitoring**, the decision maker gets less information at the end of a round.

Description (mixed extension)

A finite set of signals \mathcal{S} is available and a matrix $H : \mathcal{A} \times \mathcal{B} \rightarrow \Delta(\mathcal{S})$ indicates the feedback received by the decision maker:

He only gets to see $H(\mathbf{x}_t, \mathbf{y}_t)$ instead of \mathbf{y}_t or of $r(\mathbf{x}_t, \mathbf{y}_t)$, observation of which would be enough to use Blackwell's strategy.

Of course, the most natural formulation would be for the **randomized version** of the game, in which only a signal $s_t \in \mathcal{S}$ drawn independently at random according to $H(I_t, J_t)$ is observed.

But for simplicity we only deal with the mixed extension of the game: $\mathbf{x}_t \in \Delta(\mathcal{A})$ and $\mathbf{y}_t \in \Delta(\mathcal{B})$ lead to $H(\mathbf{x}_t, \mathbf{y}_t) \in \Delta(\mathcal{S})$.

In games **with partial monitoring**, the decision maker gets less information at the end of a round.

Description (mixed extension)

A finite set of signals \mathcal{S} is available and a matrix $H : \mathcal{A} \times \mathcal{B} \rightarrow \Delta(\mathcal{S})$ indicates the feedback received by the decision maker:

He only gets to see $H(\mathbf{x}_t, \mathbf{y}_t)$ instead of \mathbf{y}_t or of $r(\mathbf{x}_t, \mathbf{y}_t)$, observation of which would be enough to use Blackwell's strategy.

Yet, he still aims at controlling the average payoff

$$\bar{r}_T = \frac{1}{T} \sum_{t=1}^T r(\mathbf{x}_t, \mathbf{y}_t)$$

Approachability is defined in the same manner as before.

Two mixed actions \mathbf{y}' , $\mathbf{y} \in \Delta(\mathcal{B})$ of Nature are indistinguishable to the decision maker if

$$H(\cdot, \mathbf{y}) = [H(a, \mathbf{y})]_{a \in \mathcal{A}} = [H(a, \mathbf{y}')]_{a \in \mathcal{A}} = H(\cdot, \mathbf{y}')$$

This is why we introduced the set-valued mapping

$$m : (\mathbf{x}, \mathbf{y}) \mapsto \left\{ \mathbf{r}(\mathbf{x}, \mathbf{y}'), \quad \mathbf{y}' \text{ s.t. } H(\cdot, \mathbf{y}') = H(\cdot, \mathbf{y}) \right\}$$

There are **uncertainties** in the obtained payoffs: at best, the decision maker knows that

$$\mathbf{r}(\mathbf{x}_t, \mathbf{y}_t) \in m(\mathbf{x}_t, \mathbf{y}_t)$$

[Note: “at best” as he only sees $H(\mathbf{x}_t, \mathbf{y}_t)$ and not $H(\cdot, \mathbf{y}_t)$, but this is a detail...]

Perchet provided a constructive proof of sufficiency, but for a strategy with an exponentially increasing computational cost (and relying on several layers of notions –internal regret, calibration– all known to be directly related to Blackwell's approachability).

Theorem (Perchet '11)

A closed convex set \mathcal{C} is r -approachable with the feedback matrix H if and only

$$\forall \mathbf{y} \in \Delta(\mathcal{B}), \quad \exists \mathbf{x} \in \Delta(\mathcal{A}), \quad m(\mathbf{x}, \mathbf{y}) \subseteq \mathcal{C}.$$

Our motivation was mostly to **purify the argument** (and also to exhibit a strategy with **constant per-round complexity**).

Our hope was that this had already been possible for external regret in this setting (compare Rustichini '99 to Lugosi, Mannor and Stoltz '08)...

The proposed strategy is such that

$$\frac{1}{T} \sum_{t=1}^T m(x_t, y_t)$$

converges to \mathcal{C} , in the sense that it is eventually included in any ε -neighborhood of \mathcal{C} , for $\varepsilon > 0$.

(Actually, the convergence is at a $T^{-1/5}$ rate.)

This is of course enough to guarantee that the true average of payoffs \bar{r}_T converges to \mathcal{C} .

We therefore introduced the notion of **robust approachability**, i.e., approachability for set-valued mappings, which, for simplicity, we will assume to be linear in a first time.

(It is **almost** the case for the m considered here.)

Robust approachability

The key concept to approachability in games with partial monitoring

A payoff function m associates with each $(a, b) \in \mathcal{A} \times \mathcal{B}$ a subset $m(a, b) \subset \mathbb{R}^d$.

It is linearly extended into a mapping m defined on $\Delta(\mathcal{A} \times \mathcal{B})$.

Definition

A set $\mathcal{C} \subset \mathbb{R}^d$ is m -approachable if there exists a strategy for the decision maker such that for all strategies of Nature,

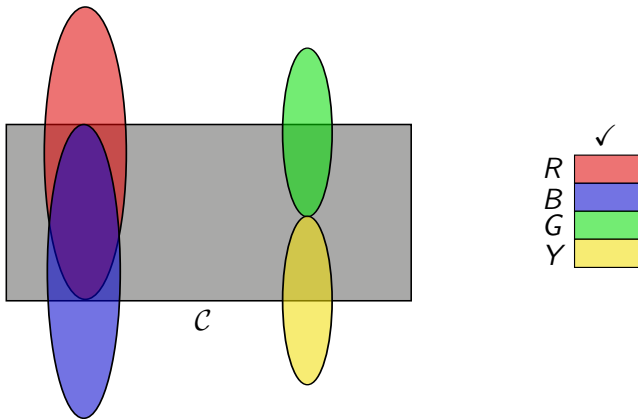
$$\bar{m}_T = \frac{1}{T} \sum_{t=1}^T m(x_t, y_t) \subseteq \mathcal{C}_{\varepsilon_T}$$

where $\mathcal{C}_{\varepsilon_T}$ is the ε_T -neighborhood of \mathcal{C} , with $\varepsilon_T \rightarrow 0$.

We will see that the necessary and sufficient condition to do so for a closed convex set \mathcal{C} will read:

$$\forall y \in \Delta(\mathcal{B}), \exists x \in \Delta(\mathcal{A}), \quad m(x, y) \subseteq \mathcal{C}.$$

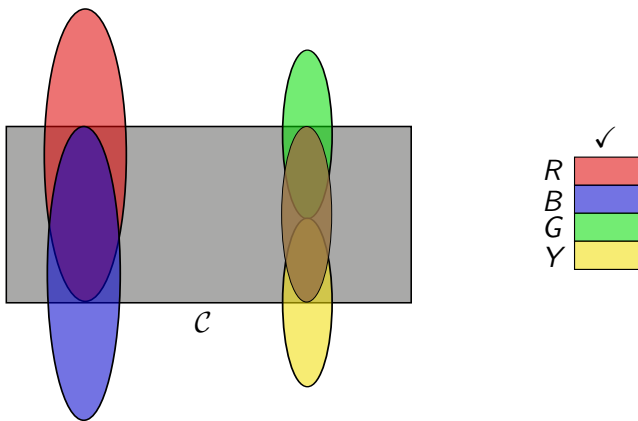
Blackwell's strategy to the farthest point does not work



Actions sets are $\mathcal{A} = \{R, B, G, Y\}$ and $\mathcal{B} = \{\checkmark\}$

$m(R, \checkmark)$ is the red set, $m(B, \checkmark)$ is the blue set, and so on

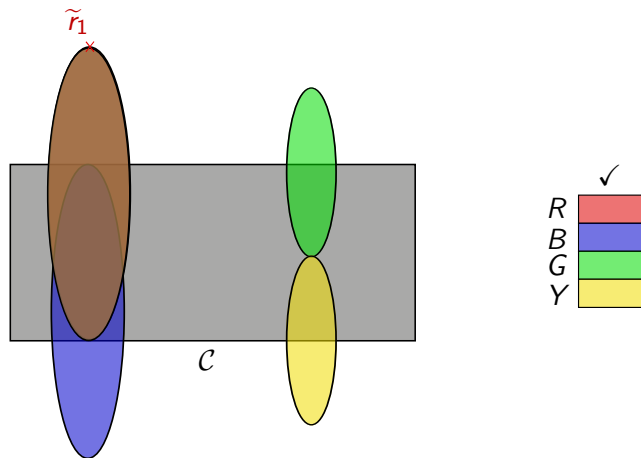
Blackwell's strategy to the farthest point does not work



In this case, there exists a mixed action \mathbf{x} such that $m(\mathbf{x}, \checkmark) \subseteq \mathcal{C}$,
with

$$\mathbf{x} = \frac{1}{2}\delta_G + \frac{1}{2}\delta_Y$$

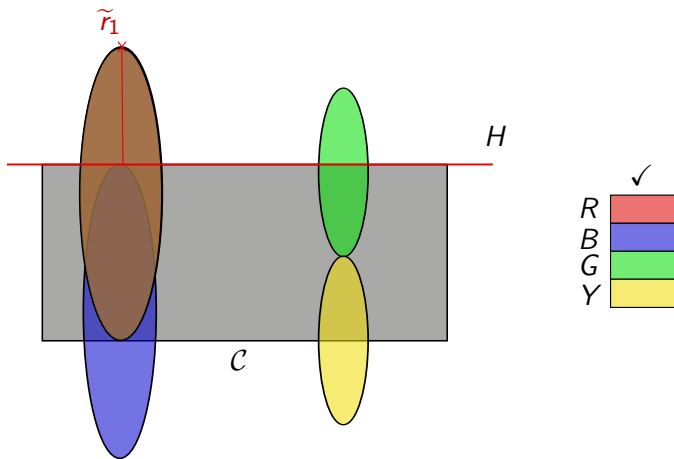
Blackwell's strategy to the farthest point does not work



Assume that R was played; the set \bar{m}_1 is in brown

The farthest point in \bar{m}_1 to C is denoted by \tilde{r}_1

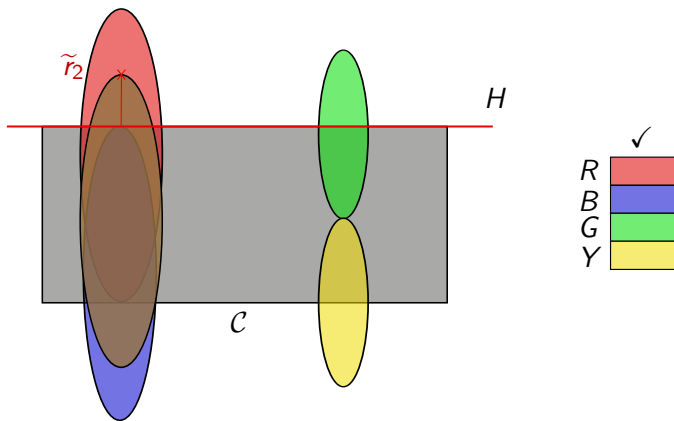
Blackwell's strategy to the farthest point does not work



The blue set is on the other side of the hyperplane

Hence, playing B at stage 2 is a choice compatible with Blackwell's strategy

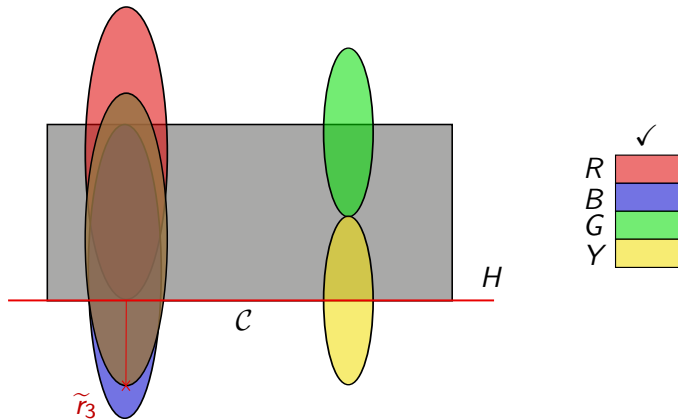
Blackwell's strategy to the farthest point does not work



The blue set is still on the other side of the hyperplane

Hence, playing B at stage 3 is still a choice compatible with Blackwell's strategy

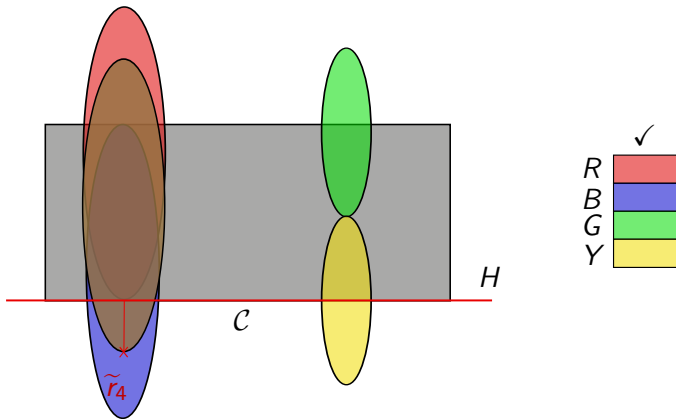
Blackwell's strategy to the farthest point does not work



Now, the red set is on the other side of the hyperplane

Hence, playing R at stage 4 is a choice compatible with Blackwell's strategy

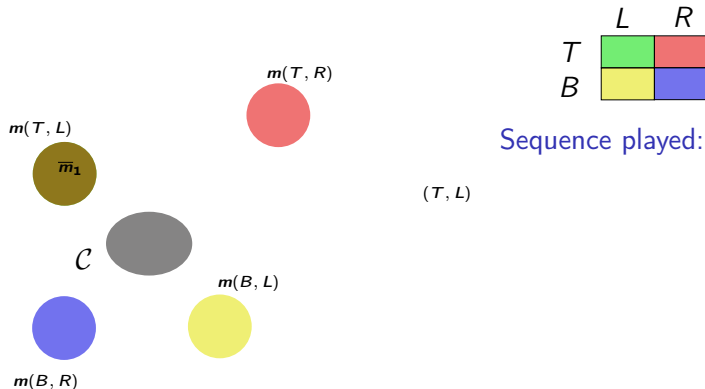
Blackwell's strategy to the farthest point does not work



The algorithm can oscillate indefinitely between R and B without \bar{m}_T converging to C

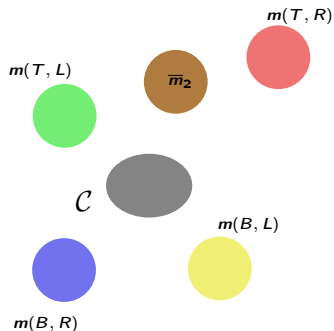
The payoff sets must be considered in their entirety!

An example of a good path of actions



Action sets $\mathcal{A} = \{T, B\}$ and $\mathcal{B} = \{L, R\}$; the set C is m -approachable if T and L are chosen at the first round: $\bar{m}_1 = m(T, L)$ is in brown

An example of a good path of actions



	L	R
T		
B		

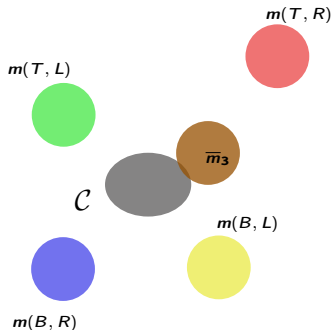
Sequence played:

$(T, L); (T, R)$

Action sets $\mathcal{A} = \{T, B\}$ and $\mathcal{B} = \{L, R\}$; the set C is m -approachable

Second round: (T, R) is played, $\bar{m}_2 = \frac{1}{2}m(T, L) + \frac{1}{2}m(T, R)$

An example of a good path of actions



	L	R
T		
B		

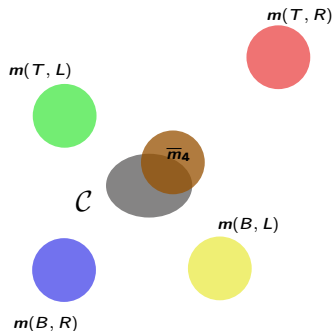
Sequence played:

$(T, L); (T, R); (B, L)$

Action sets $\mathcal{A} = \{T, B\}$ and $\mathcal{B} = \{L, R\}$; the set C is m -approachable

Third round: $\bar{m}_3 = \frac{1}{3}m(T, L) + \frac{1}{3}m(T, R) + \frac{1}{3}m(B, L)$

An example of a good path of actions



	L	R
T		
B		

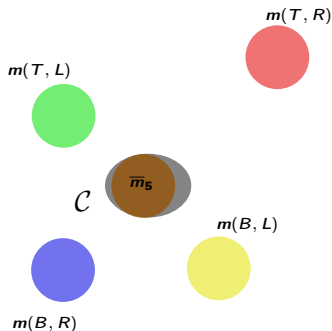
Sequence played:

$(T, L); (T, R); (B, L); (B, R)$

Action sets $\mathcal{A} = \{T, B\}$ and $\mathcal{B} = \{L, R\}$; the set \mathcal{C} is m -approachable

$$\bar{m}_4 = \frac{1}{4}m(T, L) + \frac{1}{4}m(T, R) + \frac{1}{4}m(B, L) + \frac{1}{4}m(B, R)$$

An example of a good path of actions



	L	R
T		
B		

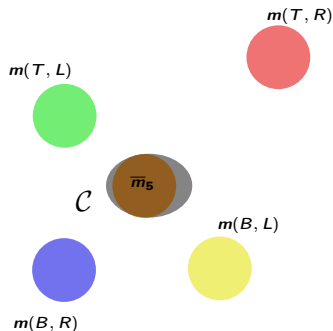
Sequence played:

$(T, L); (T, R); (B, L); (B, R); (B, R)$

Action sets $\mathcal{A} = \{T, B\}$ and $\mathcal{B} = \{L, R\}$; the set \mathcal{C} is m -approachable

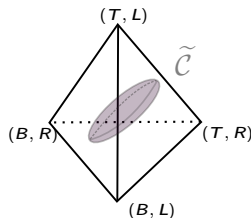
$$\bar{m}_5 = \frac{1}{5}m(T, L) + \frac{1}{5}m(T, R) + \frac{1}{5}m(B, L) + \frac{2}{5}m(B, R)$$

An example of a good path of actions



	L	R
T		
B		

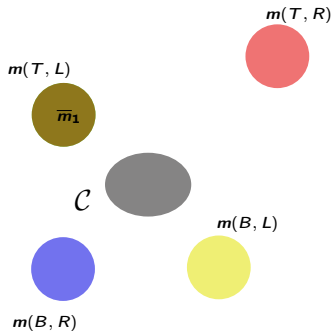
Sequence played:

 $(T, L); (T, R); (B, L); (B, R); (B, R)$ 

$$\frac{1}{5}m(T, L) + \frac{1}{5}m(T, R) + \frac{1}{5}m(B, L) + \frac{2}{5}m(B, R) \subseteq \mathcal{C}, \text{ that is,}$$

$$\left(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{2}{5}\right) \in \tilde{\mathcal{C}} = \left\{ \mu \in \Delta(\mathcal{A} \times \mathcal{B}) : \mathbb{E}_\mu[m(\mathcal{A}, \mathcal{B})] \subseteq \mathcal{C} \right\}$$

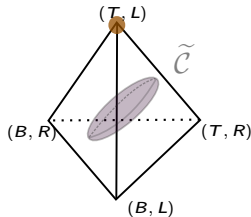
An example of a good path of actions



	L	R
T		
B		

Sequence played:

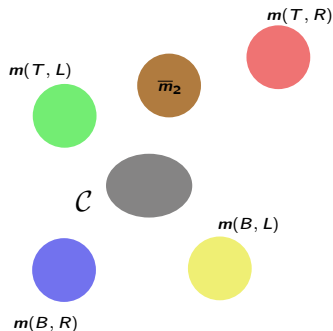
(T, L)



Approaching $\mathcal{C} \subset \mathbb{R}^d$ is equivalent to approaching $\tilde{\mathcal{C}} \subset \Delta(\mathcal{A} \times \mathcal{B})$

Abstract payoff of first round: $\mathbf{a}_1 = \delta_{(T,L)}$

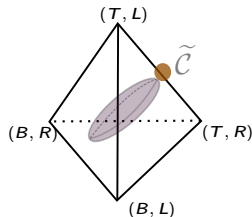
An example of a good path of actions



	L	R
T		
B		

Sequence played:

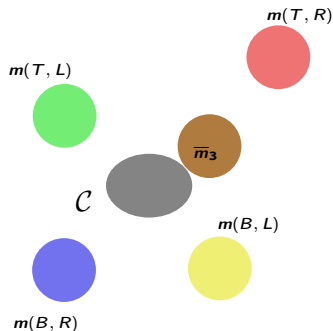
$(T, L); (T, R)$



Approaching $\mathcal{C} \subset \mathbb{R}^d$ is equivalent to approaching $\tilde{\mathcal{C}} \subset \Delta(\mathcal{A} \times \mathcal{B})$

Average payoff after second round: $\bar{a}_2 = \frac{1}{2}\delta_{(T,L)} + \frac{1}{2}\delta_{(T,R)}$

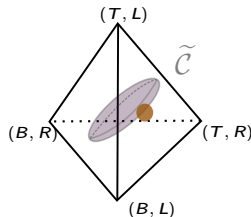
An example of a good path of actions



	L	R
T		
B		

Sequence played:

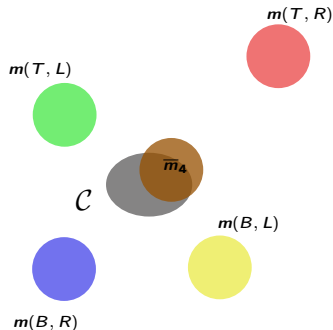
$(T, L); (T, R); (B, L)$



Approaching $\mathcal{C} \subset \mathbb{R}^d$ is equivalent to approaching $\tilde{\mathcal{C}} \subset \Delta(\mathcal{A} \times \mathcal{B})$

After third round: $\bar{a}_3 = \frac{1}{3}\delta_{(T,L)} + \frac{1}{3}\delta_{(T,R)} + \frac{1}{3}\delta_{(B,L)}$

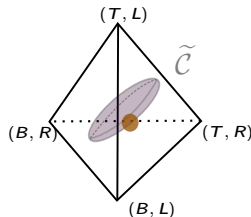
An example of a good path of actions



	L	R
T		
B		

Sequence played:

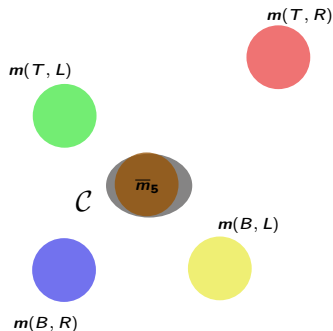
$(T, L); (T, R); (B, L); (B, R)$



Approaching $\mathcal{C} \subset \mathbb{R}^d$ is equivalent to approaching $\tilde{\mathcal{C}} \subset \Delta(\mathcal{A} \times \mathcal{B})$

$$\bar{a}_4 = \frac{1}{4}\delta_{(T,L)} + \frac{1}{4}\delta_{(T,R)} + \frac{1}{4}\delta_{(B,L)} + \frac{1}{4}\delta_{(B,R)}$$

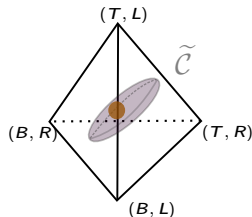
An example of a good path of actions



	L	R
T		
B		

Sequence played:

$(T, L); (T, R); (B, L); (B, R); (B, R)$



Approaching $\mathcal{C} \subset \mathbb{R}^d$ is equivalent to approaching $\tilde{\mathcal{C}} \subset \Delta(\mathcal{A} \times \mathcal{B})$

$$\bar{\mathbf{a}}_5 = \frac{1}{5}\delta_{(T,L)} + \frac{1}{5}\delta_{(T,R)} + \frac{1}{5}\delta_{(B,L)} + \frac{2}{5}\delta_{(B,R)}$$

There is an equivalence between the following two settings.

Payoffs with uncertainties

- Actions taken in $\Delta(\mathcal{A})$ and $\Delta(\mathcal{B})$
- Payoff given by the subset $\mathbf{m}(\mathbf{x}, \mathbf{y}) \subset \mathbb{R}^d$
- Target closed convex set \mathcal{C}

Payoffs without uncertainties (classical setting)

- Actions taken in $\Delta(\mathcal{A})$ and $\Delta(\mathcal{B})$
- Payoff given by a single point: the product-distribution $\mathbf{x} \otimes \mathbf{y} \in \mathbb{R}^{\mathcal{A} \times \mathcal{B}}$
- Target closed convex set $\tilde{\mathcal{C}}$, where

$$\tilde{\mathcal{C}} = \left\{ \mu \in \Delta(\mathcal{A} \times \mathcal{B}) : \mathbb{E}_\mu[\mathbf{m}(\mathcal{A}, \mathcal{B})] \subset \mathcal{C} \right\}$$

In the sense that
if and only if

\mathcal{C} is \mathbf{m} -approachable
 $\tilde{\mathcal{C}}$ is \otimes -approachable

Approachability in games with partial monitoring

As a consequence of the stated theory of robust approachability

Summary: A finite set of signals \mathcal{S} is available and a matrix $H : \mathcal{A} \times \mathcal{B} \rightarrow \Delta(\mathcal{S})$ indicates the feedback received by the decision maker:

He only gets to see $H(\mathbf{x}_t, \mathbf{y}_t) \in \Delta(\mathcal{S})$ at the end of round t .

Yet, he still aims at having the average payoff

$$\bar{\mathbf{r}}_T = \frac{1}{T} \sum_{t=1}^T \mathbf{r}(\mathbf{x}_t, \mathbf{y}_t)$$

converge to a given closed convex set \mathcal{C} .

We denote by \mathcal{F} the image of $\Delta(\mathcal{B})$ by $\mathbf{y} \mapsto H(\cdot, \mathbf{y})$.

Key was the set-valued mapping

$$\bar{\mathbf{m}} : (\mathbf{x}, \sigma) \in \Delta(\mathcal{A}) \times \mathcal{F} \mapsto \left\{ \mathbf{r}(\mathbf{x}, \mathbf{y}), \mathbf{y} \text{ s.t. } H(\cdot, \mathbf{y}) = \sigma \right\}$$

Necessary and sufficient condition: A closed convex set \mathcal{C} is \mathbf{r} -approachable with the feedback matrix H if and only

$$\forall \sigma \in \mathcal{F}, \exists \mathbf{x} \in \Delta(\mathcal{A}), \quad \bar{\mathbf{m}}(\mathbf{x}, \sigma) \subseteq \mathcal{C}.$$

The mapping \bar{m} is not linear, where we recall that

$$\bar{m} : (\mathbf{x}, \sigma) \in \Delta(\mathcal{A}) \times \mathcal{F} \mapsto \left\{ \mathbf{r}(\mathbf{x}, \mathbf{y}), \quad \mathbf{y} \text{ s.t. } H(\cdot, \mathbf{y}) = \sigma \right\}$$

But there exists a finite set $\mathcal{B}^{\text{ext}} \subset \mathcal{F}$ and a **piecewise-linear** mapping $\Phi : \mathcal{F} \rightarrow \Delta(\mathcal{B}^{\text{ext}})$ such that

$$\forall \sigma \in \mathcal{F}, \quad \forall \mathbf{x} \in \Delta(\mathcal{A}), \quad \bar{m}(\mathbf{x}, \sigma) = \sum_{b \in \mathcal{B}^{\text{ext}}} \Phi_b(\sigma) \bar{m}(\mathbf{x}, b)$$

For some problems (e.g., the minimization of **external regret**, see Rustichini '99, or **internal regret**, see Lehrer and Solan '07), the mappings

$$\mathbf{x} \in \Delta(\mathcal{A}) \mapsto \bar{m}(\mathbf{x}, \sigma)$$

are **piecewise linear** as well, for all $\sigma \in \mathcal{F}$.

Thus, there exists a finite set $\mathcal{A}^{\text{ext}} \subset \Delta(\mathcal{A})$ and another **piecewise linear** mapping $\Theta : \Delta(\mathcal{A}) \rightarrow \Delta(\mathcal{A}^{\text{ext}})$ such that

\bar{m} is induced by the **linear** extension $\overline{\bar{m}}$ of the restriction of \bar{m} to $\mathcal{A}^{\text{ext}} \times \mathcal{B}^{\text{ext}}$, in the sense that

$$\forall \mathbf{x} \in \Delta(\mathcal{A}), \quad \forall \sigma \in \mathcal{F}, \quad \bar{m}(\mathbf{x}, \sigma) = \overline{\bar{m}}(\Theta(\mathbf{x}), \Phi(\sigma))$$

We therefore can use a **linear** mapping \bar{m} .

In addition, the stated condition,

$$\forall \mathbf{y} \in \Delta(\mathcal{B}), \exists \mathbf{x} \in \Delta(\mathcal{A}), \quad \bar{m}(\mathbf{x}, H(\cdot, \mathbf{y})) \subseteq \mathcal{C},$$

can be seen to imply that \mathcal{C} is \bar{m} -robust approachable, i.e., that

$$\forall \mathbf{y}^{\text{ext}} \in \Delta(\mathcal{B}^{\text{ext}}), \exists \mathbf{x}^{\text{ext}} \in \Delta(\mathcal{A}^{\text{ext}}) \quad \bar{m}(\mathbf{x}^{\text{ext}}, \mathbf{y}^{\text{ext}}) \subseteq \mathcal{C}.$$

The **associated strategy** for \bar{m} -robust approachability of \mathcal{C} is then the key ingredient for our r -approachability strategy under the partial feedback given by H .

We resort to some additional classical ingredients:

- we need to **play in blocks** as we do not observe the $H(\cdot, \mathbf{y}_t)$ but only the $H(\mathbf{x}_t, \mathbf{y}_t)$;
- some **exploration–exploitation tradeoff** is useful as well...

Convergence rates (of the suboptimal order of $T^{-1/5}$) can be obtained.

But putting all these arguments and techniques together is somewhat technical.

We refer you to our article ([arXiv:1105.4995](https://arxiv.org/abs/1105.4995)), published at COLT'11, with a journal version in progress.

Thanks for your attention!