Robust approachability and regret minimization in games with partial monitoring

Shie Mannor, Vianney Perchet, Gilles Stoltz

Robust approachability and regret minimization in games with partial monitoring

Shie Mannor
Israel Institute of Technology (Technion), Haifa, Israel
email: shie@ee.technion.ac.il http://webee.technion.ac.il/people/shie/

Vianney Perchet
Université Paris-Diderot, Paris, France
email: vianney.perchet@normalesup.org https://sites.google.com/site/vianneyperchet/home

Gilles Stoltz
Ecole Normale Supérieure – CNRS – INRIA, Paris, France & HEC Paris – CNRS, Jouy-en-Josas, France
email: gilles.stoltz@ens.fr http://www.math.ens.fr/~stoltz

Approachability has become a standard tool in analyzing learning algorithms in the adversarial online learning setup. We develop a variant of approachability for games where there is ambiguity in the obtained reward that belongs to a set, rather than being a single vector. Using this variant we tackle the problem of approachability in games with partial monitoring and develop simple and efficient algorithms (i.e., with constant per-step complexity) for this setup. We finally consider external regret and internal regret in repeated games with partial monitoring and derive regret-minimizing strategies based on approachability theory.

Key words: Approachability; repeated games; partial monitoring; regret; online learning

MSC2000 Subject Classification: Primary: 91A20, 91A26, 62L12; Secondary: 68Q32

OR/MS subject classification: Primary: decision analysis: sequential; games/group decisions: noncooperative; secondary: computer science: artificial intelligence

---

1. Introduction. Blackwell's approachability theory and its variants has become a standard and useful tool in analyzing online learning algorithms (Cesa-Bianchi and Lugosi [5]) and algorithms for learning in games (Hart and Mas-Colell [13, 14]). The first application of Blackwell's approachability to learning in the online setup is due to Blackwell [3] himself. Numerous other contributions are summarized in the monograph by Cesa-Bianchi and Lugosi [5]. Blackwell's approachability theory enjoys a clear geometric interpretation that allows it to be used in situations where online convex optimization or exponential weights do not seem to be easily applicable and, in some sense, to go beyond the minimization of the regret and/or to control quantities of a different flavor; e.g., in the article by Mannor et al. [20], to minimize the regret together with path constraints, and in the one by Mannor and Shimkin [18], to minimize the regret in games whose stage duration is not fixed. Recently, it has been shown by Abernethy et al. [1] that approachability and low regret learning are equivalent in the sense that efficient reductions exist from one to the other. Another recent paper by Rakhlin et al. [27] showed that approachability can be analyzed from the perspective of learnability using tools from learning theory.

In this paper we consider approachability and online learning with partial monitoring in games against Nature. In partial monitoring the decision maker does not know how much reward was obtained and only gets a (random) signal whose distribution depends on the action of the decision maker and the action of Nature. There are two extremes of this setup that are well studied. On the one extreme we have the case where the signal includes the reward itself (or a signal that can be used to unbiasedly estimate the reward), which is essentially the celebrated bandits setup. The other extreme is the case where the signal is not informative (i.e., it tells the decision maker nothing about the actual reward obtained); this setting then essentially consists of repeating the same situation over and over again, as no information is gained over time. We consider a setup encompassing these situations and more general ones, in which the signal is indicative of the actual reward, but is not necessarily a sufficient statistics thereof. The difficulty is that the decision maker cannot compute the actual reward he obtained nor the actions of Nature.

Regret minimization with partial monitoring has been studied in several papers in the learning theory community. Piccolboni and Schindelhauer [26], Mannor and Shimkin [17], Cesa-Bianchi et al. [6] study special cases where an accurate estimation of the rewards (or worst-case rewards) of the decision maker is possible thanks to some extra structure. A general policy with vanishing regret is presented by Lugosi et al. [16]. This policy is based on exponential weights and a specific estimation procedure

for the (worst-case) obtained rewards. In contrast, we provide approachability-based results for the problem of regret minimization. On route, we define a new type of approachability setup, with enables to re-derive the extension of approachability to the partial monitoring vector-valued setting proposed by Perchet [23]. More importantly, we provide concrete algorithms for this approachability problem that are more efficient in the sense that, unlike previous works in the domain, their complexity is constant over all steps. Moreover, their rates of convergence are independent of the game at hand, as in the seminal paper by Blackwell [3] but for the first time in this general framework. For example, the recent purely theoretical (and fairly technical) study of approachability Perchet and Quincampoix [25], which is based on somehow related arguments, does neither provide rates of convergence nor concrete algorithms for this matter.

Outline. The paper is organized as follows. In Section 2 we recall some basic facts from approachability theory in the standard vector-valued games setting where a decision maker is engaged in a repeated vector-valued game against an arbitrary opponent (or "Nature"). In Section 3 we propose a novel setup for approachability, termed "robust approachability," where instead of obtaining a vector-valued reward, the decision maker obtains a set, that represents the ambiguity concerning his reward. We provide a simple characterization of approachable convex sets and an algorithm for the set-valued reward setup under the assumption that the set-valued reward functions are linear. In Section 4 we extend the robust approachability setup to problems where the set-valued reward functions are not linear, but rather concave in the mixed action of the decision maker and convex in the mixed action of Nature. In Section 5 we show how to apply the robust approachability framework to the repeated vector-valued games with partial monitoring. In Section 6 we consider a special type of games where the signaling structure possesses a special property, called bi-piecewise linearity, that can be exploited to derive efficient strategies. This type of games is rich enough as it encompasses several useful special cases. In Section 6.1 we provide a simple and constructive algorithm for theses games. Previous results for approachability in this setup were either non-constructive (Rustichini [29]) or were highly inefficient as they relied on some sort of lifting to the space of probability measures on mixed actions (Perchet [23]) and typically required a grid that is progressively refined (leading to a step complexity that is exponential in the number T of past steps). In Section 6.2 we apply our results for both external-regret and internal-regret minimization in repeated games with partial monitoring. In both cases our proofs are simple, lead to algorithms with constant complexity at each step, and are accompanied with rates. Our results for external regret have rates similar to the ones obtained by Lugosi et al. [16], but our proof is direct and simpler. In Section 7 we mention the general signaling case and explain how it is possible to approach certain special sets such as polytopes efficiently and general convex sets although inefficiently.

2. Some basic facts from approachability theory. In this section we recall the most basic versions of Blackwell's approachability theorem for vector-valued payoff functions.

We consider a vector-valued game between two players, a decision maker (first player) and Nature (second player), with respective finite action sets $\mathcal{A}$ and $\mathcal{B}$, whose cardinalities are referred to as $N_{\mathcal{A}}$ and $N_{\mathcal{B}}$. We denote by d the dimension of the reward vector and equip $\mathbb{R}^d$ with the ℓ^2–norm $\|\cdot\|_2$. The payoff function of the first player is given by a mapping $m : \mathcal{A} \times \mathcal{B} \to \mathbb{R}^d$, which is multi-linearly extended to $\Delta(\mathcal{A}) \times \Delta(\mathcal{B})$, the set of product-distributions over $\mathcal{A} \times \mathcal{B}$.

We consider two frameworks, depending on whether pure or mixed actions are taken.

Pure actions taken and observed. We denote by A_1, A_2, ... and B_1, B_2, ... the actions in $\mathcal{A}$ and $\mathcal{B}$ sequentially taken by each player; they are possibly given by randomized strategies, i.e., the actions A_t and B_t were obtained by random draws according to respective probability distributions denoted by $\boldsymbol{x}_t \in \Delta(\mathcal{A})$ and $\boldsymbol{y}_t \in \Delta(\mathcal{B})$. For now, we assume that the first player has a full or bandit monitoring of the pure actions taken by the opponent player: at the end of round t, when receiving the payoff $m(A_t, B_t)$, either the pure action B_t (full monitoring) or only the indicated payoff (bandit monitoring) is revealed to him.

DEFINITION 2.1 *A set $\mathcal{C} \subseteq \mathbb{R}^d$ is* m–approachable with pure actions *if there exists a strategy of the first*

player such that, for all $\varepsilon > 0$, there exists an integer T_ε such that for all strategies of the second player,

$$\mathbb{P}\left\{\forall\, T \geqslant T_\varepsilon, \quad \inf_{c\in\mathcal{C}} \left\| c - \frac{1}{T}\sum_{t=1}^{T} m\big(A_t, B_t\big)\right\|_2 \leqslant \varepsilon\right\} \geqslant 1-\varepsilon\,.$$

In particular, the first player has a strategy that ensures that the average of his vector-valued payoffs converges almost surely to the set $\mathcal{C}$ (uniformly with respect to the strategies of the second player).

The above convergence will be achieved in the course of this paper under two forms. Most often we will exhibit strategies such that, for all strategies of the second player, for all $\delta > 0$, with probability at least $1-\delta$,

$$\inf_{c\in\mathcal{C}} \left\| c - \frac{1}{T}\sum_{t=1}^{T} m\big(A_t, B_t\big)\right\|_2 \leqslant \beta(T,\delta)\,.$$

A union bound shows that such strategies m–approach $\mathcal{C}$ as soon as there exists a positive sequence ε_T such that $\sum \varepsilon_t$ is finite and $\beta(T,\varepsilon_T) \to 0$. Sometimes we will also deal with strategies directly ensuring that, for all strategies of the second player, for all $\delta > 0$, with probability at least $1-\delta$,

$$\sup_{\tau\geqslant T}\ \inf_{c\in\mathcal{C}} \left\| c - \frac{1}{\tau}\sum_{t=1}^{\tau} m\big(A_t, B_t\big)\right\|_2 \leqslant \beta(T,\delta)\,.$$

Such strategies m–approach $\mathcal{C}$ as soon as $\beta(T,\delta) \to 0$ for all $\delta > 0$.

Mixed actions taken and observed. In this case, we denote by $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ and $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots$ the actions in $\Delta(\mathcal{A})$ and $\Delta(\mathcal{B})$ sequentially taken by each player. We also assume a full or bandit monitoring for the first player: at the end of round t, when receiving the payoff $m(\boldsymbol{x}_t, \boldsymbol{y}_t)$, either the mixed action $\boldsymbol{y}_t$ (full monitoring) or the indicated payoff (bandit monitoring) is revealed to him.

Definition 2.2 *A set $\mathcal{C} \subseteq \mathbb{R}^d$ is* m–approachable with mixed actions *if there exists a strategy of the first player such that, for all $\varepsilon > 0$, there exists an integer T_ε such that for all strategies of the second player,*

$$\mathbb{P}\left\{\forall\, T \geqslant T_\varepsilon, \quad \inf_{c\in\mathcal{C}} \left\| c - \frac{1}{T}\sum_{t=1}^{T} m\big(\boldsymbol{x}_t, \boldsymbol{y}_t\big)\right\|_2 \leqslant \varepsilon\right\} \geqslant 1-\varepsilon\,.$$

As indicated below, in this setting the first player may even have deterministic strategies such that, for all (deterministic or randomized) strategies of the second player,

$$\inf_{c\in\mathcal{C}} \left\| c - \frac{1}{T}\sum_{t=1}^{T} m\big(\boldsymbol{x}_t, \boldsymbol{y}_t\big)\right\|_2 \leqslant \beta(T)$$

with probability 1, where $\beta(T) \to 0$.

Necessary and sufficient condition for approachability. For closed convex sets there is a simple characterization of approachability that is a direct consequence of the minimax theorem; the condition is the same for the two settings, whether pure or mixed actions are taken and observed.

Theorem 2.1 (Theorem 3 of Blackwell [2]) *A closed convex set $\mathcal{C} \subseteq \mathbb{R}^d$ is approachable (with pure or mixed actions) if and only if*

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

An associated strategy (that is efficient depending on the geometry of $\mathcal{C}$). Blackwell suggested a simple strategy with a geometric flavor; it only requires a bandit monitoring.

Play an arbitrary $\boldsymbol{x}_1$. For $t \geqslant 1$, given the vector-valued quantities

$$\widehat{m}_t = \frac{1}{t}\sum_{s=1}^{t} m(A_s, B_s) \qquad \text{or} \qquad \widehat{m}_t = \frac{1}{t}\sum_{s=1}^{t} m(\boldsymbol{x}_s, \boldsymbol{y}_s)\,,$$

depending on whether pure or mixed actions are taken and observed, compute the projection c_t (in ℓ^2–norm) of $\widehat{m}_t$ on $\mathcal{C}$. Find a mixed action $\boldsymbol{x}_{t+1}$ that solves the minimax equation

$$\min_{\boldsymbol{x}\in\Delta(\mathcal{A})} \max_{\boldsymbol{y}\in\Delta(\mathcal{B})} \left\langle \widehat{m}_t - c_t,\, m(\boldsymbol{x},\boldsymbol{y}) \right\rangle, \tag{1}$$

where $\langle \cdot\,,\cdot \rangle$ is the Euclidian inner product in $\mathbb{R}^d$. In the case when pure actions are taken and observed, draw A_{t+1} at random according to $\boldsymbol{x}_{t+1}$.

The minimax problem used above to determine $\boldsymbol{x}_{t+1}$ is easily seen to be a (scalar) zero-sum game and is therefore efficiently solvable using, e.g., linear programming: the associated complexity is polynomial in $N_{\mathcal{A}}$ and $N_{\mathcal{B}}$. All in all, this strategy is efficient if the computations of the required projections onto $\mathcal{C}$ in ℓ^2–norm can be performed efficiently.

The strategy presented above enjoys the following rates of convergence for approachability.

THEOREM 2.2 (THEOREM 3 OF BLACKWELL [2]; THEOREM II.4.3 OF MERTENS ET AL. [21]) *We denote by M a bound in norm over m, i.e.,*

$$\max_{(a,b)\in\mathcal{A}\times\mathcal{B}} \big\| m(a,b) \big\|_2 \leqslant M\,.$$

With mixed actions taken and observed, the above strategy ensures that for all strategies of the second player, with probability 1,

$$\inf_{c\in\mathcal{C}} \left\| c - \frac{1}{T}\sum_{t=1}^{T} m\big(\boldsymbol{x}_t,\boldsymbol{y}_t\big) \right\|_2 \leqslant \frac{2M}{\sqrt{T}}\,;$$

while with pure actions taken and observed, for all $\delta\in(0,1)$ and for all strategies of the second player, with probability at least $1-\delta$,

$$\sup_{\tau\geqslant T} \inf_{c\in\mathcal{C}} \left\| c - \frac{1}{\tau}\sum_{t=1}^{\tau} m\big(A_t,B_t\big) \right\|_2 \leqslant 2M\sqrt{\frac{2}{\delta T}}\,.$$

An alternative strategy in the case where pure actions are taken and observed. Convergence rates of a slightly different flavor (but still implying approachability) can be proved, in the full monitoring case, by modifying the above procedure as follows. For $t\geqslant 1$, consider instead the vector-valued quantity

$$\widehat{m}_t = \frac{1}{t}\sum_{s=1}^{t} m(\boldsymbol{x}_s,B_s)\,,$$

compute its projection c_t (in ℓ^2–norm) on $\mathcal{C}$, and solve the associated minimax problem (1).

This modified strategy enjoys the following rates of convergence for approachability when pure actions are taken and observed.

THEOREM 2.3 (SECTION 7.7 AND EXERCISE 7.23 OF CESA-BIANCHI ET AL. [6]) *We denote by M a bound in norm over m, i.e.,*

$$\max_{(a,b)\in\mathcal{A}\times\mathcal{B}} \big\| m(a,b) \big\|_2 \leqslant M\,.$$

With pure actions taken and observed, the above strategy ensures that for all strategies of the second player, with probability at least $1-\delta$,

$$\inf_{c\in\mathcal{C}} \left\| c - \frac{1}{T}\sum_{t=1}^{T} m\big(A_t,B_t\big) \right\|_2 \leqslant \frac{2M}{\sqrt{T}}\Big(1+2\sqrt{\ln(2/\delta)}\Big)\,.$$

In the next section, we will rather resort to this slightly modified procedure as the form of the resulting bounds is closer to the one derived in the main section (Section 6) of this paper.

3. Robust approachability for finite set-valued games. In this section we extend the results from the previous section to set-valued payoff functions in the case of full monitoring. We denote by $\mathcal{S}\big(\mathbb{R}^d\big)$ the set of all subsets of $\mathbb{R}^d$ and consider a set-valued payoff function $\overline{m}:\mathcal{A}\times\mathcal{B}\to\mathcal{S}\big(\mathbb{R}^d\big)$.

Pure actions taken and observed. At each round t, the players choose simultaneously respective actions $A_t \in \mathcal{A}$ and $B_t \in \mathcal{B}$, possibly at random according to mixed distributions $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$. Full monitoring takes place for the first player: he observes B_t at the end of round t. However, as a result, the first player gets the *subset* $\overline{m}(A_t, B_t)$ as a payoff. This models the ambiguity or uncertainty associated with some true underlying payoff gained.

We extend $\overline{m}$ multi-linearly to $\Delta(\mathcal{A}) \times \Delta(\mathcal{B})$ and even to $\Delta(\mathcal{A} \times \mathcal{B})$, the set of joint probability distributions on $\mathcal{A} \times \mathcal{B}$, as follows. Let

$$\mu = \big(\mu_{a,b}\big)_{(a,b)\in\mathcal{A}\times\mathcal{B}}$$

be such a joint probability distribution; then $\overline{m}(\mu)$ is defined as a finite convex combination[1] of subsets of $\mathbb{R}^d$,

$$\overline{m}(\mu) = \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{B}} \mu_{a,b}\, \overline{m}(a,b)\,.$$

When μ is the product-distribution of some $\boldsymbol{x} \in \Delta(\mathcal{A})$ and $\boldsymbol{y} \in \Delta(\mathcal{B})$, we use the notation $\overline{m}(\mu) = \overline{m}(\boldsymbol{x}, \boldsymbol{y})$.

We denote by

$$\pi_T = \frac{1}{T} \sum_{t=1}^{T} \delta_{(A_t,B_t)}$$

the empirical distribution of the pairs (A_t, B_t) of actions taken during the first T rounds, and will be interested in the behavior of

$$\frac{1}{T} \sum_{t=1}^{T} \overline{m}(A_t, B_t)\,,$$

which can also be rewritten here in a compact way as $\overline{m}(\pi_T)$, by linearity of the extension of $\overline{m}$.

The distance of this set $\overline{m}(\pi_T)$ to the target set $\mathcal{C}$ will be measured in a worst-case sense: we denote by

$$\varepsilon_T = \sup_{d\in\overline{m}(\pi_T)} \inf_{c\in\mathcal{C}} \|c-d\|_2$$

the smallest value such that $\overline{m}(\pi_T)$ is included in an ε_T–neighborhood of $\mathcal{C}$. Robust approachability of a set $\mathcal{C}$ with the set-valued payoff function $\overline{m}$ then simply means that the sequence of ε_T tends almost-surely to 0, uniformly with respect to the strategies of the second player.

DEFINITION 3.1 *A set $\mathcal{C} \subseteq \mathbb{R}^d$ is* $\overline{m}$–robust approachable with pure actions *if there exists a strategy of the first player such that, for all $\varepsilon > 0$, there exists an integer T_ε such that for all strategies of the second player,*

$$\mathbb{P}\left\{ \forall T \geqslant T_\varepsilon, \quad \sup_{d\in\overline{m}(\pi_T)} \inf_{c\in\mathcal{C}} \|c-d\|_2 \leqslant \varepsilon \right\} \geqslant 1-\varepsilon\,.$$

Mixed actions taken and observed. At each round t, the players choose simultaneously respective mixed actions $\boldsymbol{x}_t \in \Delta(\mathcal{A})$ and $\boldsymbol{y}_t \in \Delta(\mathcal{B})$. Full monitoring still takes place for the first player: he observes $\boldsymbol{y}_t$ at the end of round t; he however gets the subset $\overline{m}(\boldsymbol{x}_t, \boldsymbol{y}_t)$ as a payoff (which, again, accounts for the uncertainty).

The product-distribution of two elements $\boldsymbol{x} = (x_a)_{a\in\mathcal{A}} \in \Delta(\mathcal{A})$ and $\boldsymbol{y} = (y_b)_{b\in\mathcal{B}} \in \Delta(\mathcal{B})$ will be denoted by $\boldsymbol{x} \otimes \boldsymbol{y}$; it gives a probability mass of $x_a y_b$ to each pair $(a,b) \in \mathcal{A} \times \mathcal{B}$. We consider the empirical joint distribution of mixed actions taken during the first T rounds,

$$\nu_T = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{x}_t \otimes \boldsymbol{y}_t\,,$$

and will be interested in the behavior of

$$\frac{1}{T} \sum_{t=1}^{T} \overline{m}(\boldsymbol{x}_t, \boldsymbol{y}_t)\,,$$

which can also be rewritten here in a compact way as $\overline{m}(\nu_T)$, by linearity of the extension of $\overline{m}$.

[1] For two sets S, T and $\alpha \in [0,1]$, the convex combination $\alpha S + (1-\alpha)T$ is defined as
$$\big\{\alpha s + (1-\alpha)t, \quad s \in S \text{ and } t \in T\big\}\,.$$

DEFINITION 3.2 *A set $\mathcal{C} \subseteq \mathbb{R}^d$ is* $\overline{m}$–robust approachable with mixed actions *if there exists a strategy of the first player such that, for all $\varepsilon > 0$, there exists an integer T_ε such that for all strategies of the second player,*

$$\mathbb{P}\left\{ \forall\, T \geqslant T_\varepsilon, \quad \sup_{d \in \overline{m}(\nu_T)} \inf_{c \in \mathcal{C}} \|c-d\|_2 \leqslant \varepsilon \right\} \geqslant 1-\varepsilon\,.$$

Actually, the bounds exhibited below in this setting will be of the form

$$\sup_{d \in \overline{m}(\nu_T)} \inf_{c \in \mathcal{C}} \|c-d\|_2 \leqslant \beta(T)$$

with probability 1 and uniformly over all (deterministic or randomized) strategies of the second player, where $\beta(T) \to 0$ and for deterministic strategies of the first player.

A useful continuity lemma. Before proceeding we provide a continuity lemma. It can be reformulated as indicating that for all joint distributions μ and ν over $\mathcal{A}\times\mathcal{B}$, the set $\overline{m}(\mu)$ is contained in a $M\,\|\mu-\nu\|_1$–neighborhood of $\overline{m}(\nu)$, where M is a bound in ℓ^2–norm on $\overline{m}$; this is a fact that we will use repeatedly below.

LEMMA 3.1 *Let μ and ν be two probability distributions over $\mathcal{A}\times\mathcal{B}$. We assume that the set-valued function $\overline{m}$ is bounded in norm by M, i.e., that there exists a real number $M>0$ such that*

$$\forall (a,b) \in \mathcal{A}\times\mathcal{B}, \qquad \sup_{d \in \overline{m}(a,b)} \|d\|_2 \leqslant M\,.$$

Then

$$\sup_{d \in \overline{m}(\mu)} \inf_{c \in \overline{m}(\nu)} \|d-c\|_2 \leqslant M\,\|\mu-\nu\|_1 \leqslant M\sqrt{N_{\mathcal{A}} N_{\mathcal{B}}}\,\|\mu-\nu\|_2\,,$$

where the norms in the right-hand side are respectively the ℓ^1 and ℓ^2–norms between probability distributions.

PROOF. Let d be an element of $\overline{m}(\mu)$; it can be written as

$$d = \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{B}} \mu_{a,b}\,\theta_{a,b}$$

for some elements $\theta_{a,b} \in \overline{m}(a,b)$. We consider

$$c = \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{B}} \nu_{a,b}\,\theta_{a,b}\,,$$

which is an element of $\overline{m}(\nu)$. Then by the triangle inequality,

$$\|d-c\|_2 = \left\| \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{B}} \big(\mu_{a,b}-\nu_{a,b}\big)\theta_{a,b} \right\|_2 \leqslant \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{B}} \big|\mu_{a,b}-\nu_{a,b}\big|\,\|\theta_{a,b}\|_2 \leqslant M \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{B}} \big|\mu_{a,b}-\nu_{a,b}\big|\,.$$

This entails the first claimed inequality. The second one follows from an application of the Cauchy-Schwarz inequality. □

COROLLARY 3.1 *When the set-valued function $\overline{m}$ is bounded in norm, for all $\boldsymbol{y} \in \Delta(\mathcal{B})$, the mapping $D_{\boldsymbol{y}} : \Delta(A) \to \mathbb{R}$ defined by*

$$\forall\, \boldsymbol{x} \in \Delta(\mathcal{A}), \qquad D_{\boldsymbol{y}}(\boldsymbol{x}) = \sup_{d \in \overline{m}(\boldsymbol{x},\boldsymbol{y})} \inf_{c\in\mathcal{C}} \|c-d\|_2$$

is continuous.

PROOF. We show that for all $\boldsymbol{x}$, $\boldsymbol{x}' \in \Delta(\mathcal{A})$, the condition $\|\boldsymbol{x}'-\boldsymbol{x}\|_1 \leqslant \varepsilon$ implies that $D_{\boldsymbol{y}}(\boldsymbol{x}) - D_{\boldsymbol{y}}(\boldsymbol{x}') \leqslant M\varepsilon$, where M is the bound in norm over $\overline{m}$. Indeed, fix $\delta>0$ and let $d_{\delta,\boldsymbol{x}} \in \overline{m}(\boldsymbol{x},\boldsymbol{y})$ be such that

$$D_{\boldsymbol{y}}(\boldsymbol{x}) \leqslant \inf_{c\in\mathcal{C}} \big\|\, c - d_{\delta,\boldsymbol{x}} \,\big\|_2 + \delta\,. \tag{2}$$

By Lemma 3.1 (with the choices $\mu = \boldsymbol{x} \otimes \boldsymbol{y}$ and $\nu = \boldsymbol{x}' \otimes \boldsymbol{y}$) there exists $d_{\delta,\boldsymbol{x}'} \in \overline{m}(\boldsymbol{x}', \boldsymbol{y})$ such that $\left\| d_{\delta,\boldsymbol{x}} - d_{\delta,\boldsymbol{x}'} \right\|_2 \leqslant M\varepsilon + \delta$. The triangle inequality entails that

$$\inf_{c\in\mathcal{C}} \left\| c - d_{\delta,\boldsymbol{x}} \right\|_2 \leqslant \inf_{c\in\mathcal{C}} \left\| c - d_{\delta,\boldsymbol{x}'} \right\|_2 + M\varepsilon + \delta \,.$$

Substituting in (2), we get that

$$D_{\boldsymbol{y}}(\boldsymbol{x}) \leqslant M\varepsilon + 2\delta + \inf_{c\in\mathcal{C}} \left\| c - d_{\delta,\boldsymbol{x}'} \right\|_2 \leqslant M\varepsilon + 2\delta + D_{\boldsymbol{y}}(\boldsymbol{x}') \,,$$

which, letting $\delta \to 0$, proves our continuity claim. □

Necessary and sufficient condition for robust approachability. This conditions reads as follows and will be referred to as (RAC), an acronym that stands for "robust approachability condition."

THEOREM 3.1 *Suppose that the set-valued function $\overline{m}$ is bounded in norm by M. A closed convex set $\mathcal{C} \subseteq \mathbb{R}^d$ is $\overline{m}$–approachable (with pure or mixed actions) if and only if the following robust approachability condition is satisfied,*

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

PROOF OF THE NECESSITY OF CONDITION (RAC). If the condition does not hold, then there exists $\boldsymbol{y}_0 \in \Delta(\mathcal{B})$ such that for every $\boldsymbol{x} \in \mathcal{A}$, the set $\overline{m}(\boldsymbol{x}, \boldsymbol{y}_0)$ is not included in $\mathcal{C}$, i.e., it contains at least one point not in $\mathcal{C}$. We consider the mapping $D_{\boldsymbol{y}_0}$ defined in the statement of Corollary 3.1. Since $\mathcal{C}$ is closed, distances of given individual points to $\mathcal{C}$ are achieved; therefore, by the choice of $\boldsymbol{y}_0$, we get that $D_{\boldsymbol{y}_0}(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in \Delta(\mathcal{A})$. Now, since $D_{\boldsymbol{y}_0}$ is continuous on the compact set $\Delta(\mathcal{A})$, as asserted by the indicated corollary, it attains its minimum, whose value we denote by $D_{\min} > 0$.

Assume now that the second player chooses at each round $\boldsymbol{y}_t = \boldsymbol{y}_0$ as his mixed action. In the case of mixed actions taken and observed, denoting

$$\overline{\boldsymbol{x}}_T = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{x}_t \,,$$

we get that $\nu_t = \overline{\boldsymbol{x}}_T \otimes \boldsymbol{y}_0$, and hence, for all strategies of the first player and for all $T \geqslant 1$,

$$\sup_{d\in\overline{m}(\nu_T)} \inf_{c\in\mathcal{C}} \|c - d\|_2 = D_{\boldsymbol{y}_0}(\overline{\boldsymbol{x}}_T) \geqslant D_{\min} > 0 \,,$$

which shows that $\mathcal{C}$ is not approachable.

The case of pure actions taken and observed is treated similarly, with the sole addition of a concentration argument. By martingale convergence (e.g., repeated uses of the Hoeffding-Azuma inequality together with an application of the Borel-Cantelli lemma), $\delta_T = \|\pi_T - \nu_T\|_1 \to 0$ almost surely as $T \to \infty$. By applying Lemma 3.1, we get

$$\sup_{d\in\overline{m}(\pi_T)} \inf_{c\in\mathcal{C}} \|c - d\|_2 \geqslant \sup_{d\in\overline{m}(\nu_T)} \inf_{c\in\mathcal{C}} \|c - d\|_2 - M\delta_T \geqslant D_{\min} - M\delta_T$$

and simply take the lim inf in the above inequalities to conclude the argument. □

That (RAC) is sufficient to get robust approachability is proved in a constructive way, by exhibiting suitable strategies. We identify probability distributions over $\mathcal{A} \times \mathcal{B}$ with vectors in $\mathbb{R}^{\mathcal{A}\times\mathcal{B}}$ and consider the vector-valued payoff function

$$m : (a, b) \in \mathcal{A} \times \mathcal{B} \longmapsto \delta_{(a,b)} \in \mathbb{R}^{\mathcal{A}\times\mathcal{B}} \,,$$

which we extend multi-linearly to $\Delta(\mathcal{A}) \times \Delta(\mathcal{B})$; the target set will be

$$\widetilde{\mathcal{C}} = \left\{ \mu \in \Delta(\mathcal{A} \times \mathcal{B}) : \ \ \overline{m}(\mu) \subseteq \mathcal{C} \right\}. \tag{3}$$

Since $\overline{m}$ is a linear function on $\Delta(\mathcal{A} \times \mathcal{B})$ and $\mathcal{C}$ is convex, the set $\widetilde{\mathcal{C}}$ is convex as well. In addition, since $\mathcal{C}$ is closed, $\widetilde{\mathcal{C}}$ is also closed.

LEMMA 3.2 *Condition* (RAC) *is equivalent to the m–approachability of $\widetilde{\mathcal{C}}$.*

PROOF. This equivalence is immediate via Theorem 2.1. The latter indeed states that the m–approachability of $\widetilde{\mathcal{C}}$ is equivalent to the fact that for all $\boldsymbol{y} \in \Delta(\mathcal{B})$, there exists some $\boldsymbol{x} \in \Delta(\mathcal{A})$ such that $\mu = m(\boldsymbol{x}, \boldsymbol{y})$, the product-distribution between $\boldsymbol{x}$ and $\boldsymbol{y}$, belongs to $\widetilde{\mathcal{C}}$, i.e., satisfies $\overline{m}(\mu) = \overline{m}(\boldsymbol{x}, \boldsymbol{y}) \subseteq \mathcal{C}$. □

The above definition of m entails the following rewriting,

$$\pi_T = \frac{1}{T} \sum_{t=1}^{T} m(A_t, B_t) \qquad \text{and} \qquad \nu_T = \frac{1}{T} \sum_{t=1}^{T} m(\boldsymbol{x}_t, \boldsymbol{y}_t) \,.$$

Let $P_{\widetilde{\mathcal{C}}}$ denote the projection operator onto $\widetilde{\mathcal{C}}$; the quantities at hand in the definition of m–approachability of $\widetilde{\mathcal{C}}$ are given by

$$\varepsilon_T = \Big\| \pi_T - P_{\widetilde{\mathcal{C}}}(\pi_T) \Big\|_2 = \inf_{\mu \in \widetilde{\mathcal{C}}} \|\pi_T - \mu\|_2 \qquad \text{and} \qquad \varepsilon'_T = \Big\| \nu_T - P_{\widetilde{\mathcal{C}}}(\nu_T) \Big\|_2 = \inf_{\mu \in \widetilde{\mathcal{C}}} \|\nu_T - \mu\|_2 \,.$$

We now relate the quantities of interest, i.e., the ones arising in the definition of $\overline{m}$–robust approachability of $\mathcal{C}$, to the former quantities.

LEMMA 3.3 *With pure actions taken and observed,*

$$\sup_{d \in \overline{m}(\pi_T)} \inf_{c \in \mathcal{C}} \|c - d\|_2 \leqslant M \sqrt{N_{\mathcal{A}} N_{\mathcal{B}}} \; \varepsilon_T \,.$$

With mixed actions taken and observed,

$$\sup_{d \in \overline{m}(\nu_T)} \inf_{c \in \mathcal{C}} \|c - d\|_2 \leqslant M \sqrt{N_{\mathcal{A}} N_{\mathcal{B}}} \; \varepsilon'_T \,.$$

PROOF. Lemma 3.1 entails that the sets $\overline{m}(\pi_T)$ are included in $M\sqrt{N_{\mathcal{A}} N_{\mathcal{B}}}\, \varepsilon_T$–neighborhoods of $\overline{m}\big(P_{\widetilde{\mathcal{C}}}(\pi_T)\big)$. Since by definition of $\widetilde{\mathcal{C}}$, one has $\overline{m}\big(P_{\widetilde{\mathcal{C}}}(\pi_T)\big) \subseteq \mathcal{C}$, we get in particular that the sets $\overline{m}(\pi_T)$ are included in $M\sqrt{N_{\mathcal{A}} N_{\mathcal{B}}}\, \varepsilon_T$–neighborhoods of $\mathcal{C}$, which is exactly what was stated. The argument can be repeated with the ν_T to get the second bound in the statement of the lemma. □

PROOF OF THE SUFFICIENCY OF CONDITION (RAC). First, Lemma 3.2 shows that Condition (RAC) (via Theorems 2.2 or 2.3) ensures the existence of strategies m–approaching $\widetilde{\mathcal{C}}$. Second, Lemma 3.3 indicates that these strategies also $\overline{m}$–robust approach $\mathcal{C}$. (It even translates the rates for the m–approachability of $\widetilde{\mathcal{C}}$ into rates for the $\overline{m}$–robust approachability of $\mathcal{C}$; for instance, in the case of mixed actions taken and observed, the $2/\sqrt{T}$ rate for the m–approachability of $\widetilde{\mathcal{C}}$ becomes a $2M\sqrt{N_{\mathcal{A}} N_{\mathcal{B}}/T}$ rate for the $\overline{m}$–robust approachability of $\mathcal{C}$, a fact that we will use in the proof of Theorem 6.1.) □

Two concluding remarks. Note that, as explained around Equation (1), the considered strategies for m–approaching $\widetilde{\mathcal{C}}$, or equivalently $\overline{m}$–robust approaching $\mathcal{C}$, are efficient as soon as projections in ℓ^2–norm onto the set $\widetilde{\mathcal{C}}$ defined in (3) can be computed efficiently. The latter fact depends on the respective geometries of $\overline{m}$ and $\mathcal{C}$. We will provide examples of favorable cases (see, e.g., Section 6.2.1 about minimization of external regret under partial monitoring).

A final remark is that the proposed strategies require full monitoring, as they rely on the observations of either the pair of played mixed actions $m(\boldsymbol{x}_t, \boldsymbol{y}_t)$ or of played pure actions $m(A_t, B_t)$. They enjoy no obvious extension to a case where only a bandit monitoring of the played sets $\overline{m}(\boldsymbol{x}_t, \boldsymbol{y}_t)$ or $\overline{m}(A_t, B_t)$ would be available.

4. Robust approachability for concave–convex set-valued games. We consider in this section the same setting of mixed actions taken and observed as in the previous section, that is, we deal with set-valued payoff functions $\overline{m} : \Delta(\mathcal{A}) \times \Delta(\mathcal{B}) \to \mathcal{S}\big(\mathbb{R}^d\big)$ under full monitoring. However, in the previous section $\overline{m}$ was linear on $\Delta(\mathcal{A}) \times \Delta(\mathcal{B})$, an assumption that we now weaken while still having that (RAC) is the necessary and sufficient condition for robust approachability. The price to pay for this is the loss of the possible efficiency of the approachability strategies exhibited and the worsening of the convergence rates.

Formally, the functions $\overline{m} : \Delta(\mathcal{A}) \times \Delta(\mathcal{B}) \to \mathcal{S}\big(\mathbb{R}^d\big)$ that we will consider will satisfy one or several of the following properties.

DEFINITION 4.1 *A function $\overline{m} : \Delta(\mathcal{A}) \times \Delta(\mathcal{B}) \to \mathcal{S}\big(\mathbb{R}^d\big)$ is uniformly continuous in its first argument if for all $\varepsilon > 0$, there exists $\eta > 0$ such that for all $\boldsymbol{x}, \boldsymbol{x}' \in \Delta(\mathcal{A})$ satisfying $\|\boldsymbol{x} - \boldsymbol{x}'\|_1 \leqslant \eta$ and for all $\boldsymbol{y} \in \Delta(\mathcal{B})$, the set $\overline{m}(\boldsymbol{x}', \boldsymbol{y})$ is included in an ε–neighborhood of $\overline{m}(\boldsymbol{x}, \boldsymbol{y})$ in the Euclidian norm. Put differently,*

$$\sup_{d \in \overline{m}(\boldsymbol{x}', \boldsymbol{y})} \inf_{c \in \overline{m}(\boldsymbol{x}, \boldsymbol{y})} \|d - c\|_2 \leqslant \varepsilon \qquad \text{or} \qquad \overline{m}(\boldsymbol{x}', \boldsymbol{y}) \subseteq \overline{m}(\boldsymbol{x}, \boldsymbol{y}) + \varepsilon \boldsymbol{B}\,,$$

where $\boldsymbol{B}$ is the unit Euclidian ball in $\mathbb{R}^d$.

Uniform continuity in the second argument is defined symmetrically.

DEFINITION 4.2 *A function $\overline{m} : \Delta(\mathcal{A}) \times \Delta(\mathcal{B}) \to \mathcal{S}\big(\mathbb{R}^d\big)$ is concave in its first argument if for all $\boldsymbol{x}, \boldsymbol{x}' \in \Delta(\mathcal{A})$, all $\boldsymbol{y} \in \Delta(\mathcal{B})$, and all $\alpha \in [0,1]$,*

$$\overline{m}\big(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{x}',\, \boldsymbol{y}\big) \subseteq \alpha\, \overline{m}(\boldsymbol{x}, \boldsymbol{y}) + (1-\alpha)\, \overline{m}(\boldsymbol{x}', \boldsymbol{y})\,.$$

A function $\overline{m} : \Delta(\mathcal{A}) \times \Delta(\mathcal{B}) \to \mathcal{S}\big(\mathbb{R}^d\big)$ is convex in its second argument if for all $\boldsymbol{x} \in \Delta(\mathcal{A})$, all $\boldsymbol{y}, \boldsymbol{y}' \in \Delta(\mathcal{B})$, and all $\alpha \in [0,1]$,

$$\alpha\, \overline{m}(\boldsymbol{x}, \boldsymbol{y}) + (1-\alpha)\, \overline{m}(\boldsymbol{x}, \boldsymbol{y}') \subseteq \overline{m}\big(\boldsymbol{x},\, \alpha \boldsymbol{y} + (1-\alpha)\boldsymbol{y}'\big)\,.$$

An example of such a function $\overline{m}$ is discussed in Lemma 5.1.

The following theorem indicates that (RAC) is the necessary and sufficient condition for the $\overline{m}$–robust approachability of a closed convex set $\mathcal{C}$ with mixed actions when the payoff function $\overline{m}$ satisfies all four properties stated above. (Boundedness of $\overline{m}$ indeed follows from the continuity of $\overline{m}$ in each variable.)

THEOREM 4.1 *If $\overline{m}$ is bounded, convex, and uniformly continuous in its second argument, then* (RAC) *entails that a closed convex set $\mathcal{C}$ is $\overline{m}$–robust approachable with mixed actions.*

On the contrary, if $\overline{m}$ is concave and uniformly continuous in its first argument, then a closed convex set $\mathcal{C}$ can be $\overline{m}$–robust approachable with mixed actions only if (RAC) *is satisfied.*

PROOF OF THE SECOND STATEMENT OF THEOREM 4.1. The proof of Corollary 3.1 extends to the case considered here and shows, thanks to the ad hoc consideration of the result stated in Lemma 3.1 as following from Definition 4.1, that for all $\boldsymbol{y} \in \Delta(\mathcal{B})$, the mapping $D_{\boldsymbol{y}}$ is still continuous over $\Delta(\mathcal{A})$. We now proceed by contradiction and assume that (RAC) is not satisfied; the first part of the proof of the necessity of (RAC) in Theorem 3.1 also applies to the present case: there exists $\boldsymbol{y}_0$ such that $D_{\boldsymbol{y}_0} \geqslant D_{\min} > 0$ over $\Delta(\mathcal{A})$. It then suffices to note that whenever the second player resorts to $\boldsymbol{y}_t = \boldsymbol{y}_0$ at all rounds $t \geqslant 1$, then for all strategies of the first player, the quantity of interest in robust approachability can be lower bounded as follows, thanks to the concavity in the first argument:

$$\begin{aligned}
&\sup \left\{ \inf_{c \in \mathcal{C}} \|d - c\|_2 : \quad d \in \frac{1}{T} \sum_{t=1}^{T} \overline{m}(\boldsymbol{x}_t, \boldsymbol{y}_0) \right\} \\
&\qquad \geqslant \sup \left\{ \inf_{c \in \mathcal{C}} \|d - c\|_2 : \quad d \in \overline{m}\left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{x}_t,\ \boldsymbol{y}_0 \right) \right\} = D_{\boldsymbol{y}_0}\left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{x}_t \right) \geqslant D_{\min} > 0\,.
\end{aligned}$$

Therefore, $\mathcal{C}$ is $\overline{m}$–robust approachable with mixed actions by no strategy of the first player. □

The proof of the first statement of Theorem 4.1 relies on the use of approximately calibrated strategies of the first player, as introduced and studied (among others) by Dawid [8], Foster and Vohra [9], Mannor and Stoltz [19]. Formally, given $\eta > 0$, an η–calibrated strategy of the first player considers some finite covering of $\Delta(\mathcal{B})$ by N_η balls of radius η and abides by the following constraints. Denoting by $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^{N_\eta}$ the centers of the balls in the covering (they form what will be referred to later on as an η–grid), such a strategy chooses only forecasts in $\big\{\boldsymbol{y}^1, \ldots, \boldsymbol{y}^{N_\eta}\big\}$. We thus denote by L_t the index chosen in $\big\{1, \ldots, N_\eta\big\}$ at round t and by

$$N_T(\ell) = \sum_{t=1}^{T} \mathbb{I}_{\{L_t = \ell\}}$$

the total number of rounds within the first T ones when the element ℓ of the grid was chosen. We denote by $(\,\cdot\,)_+$ the function that gives the nonnegative part of a real number. The final condition to be satisfied

is that for all $\delta > 0$, there exists an integer T_δ such that for all strategies of the second player, with probability at least $1-\delta$, for all $T \geqslant T_\delta$,

$$\sum_{\ell=1}^{N_\eta} \frac{N_T(\ell)}{T} \left(\left\| \boldsymbol{y}^\ell - \frac{1}{N_T(\ell)} \sum_{t=1}^{T} \boldsymbol{y}_t \mathbb{I}_{\{L_t=\ell\}} \right\|_1 - \eta \right)_+ \leqslant \delta \,. \tag{4}$$

This calibration criterion is slightly stronger than the classical η–calibration score usually considered in the literature, which consists of omitting nonnegative parts in the criterion above and ensuring that for all strategies of the second player, with probability at least $1-\delta$, for all $T \geqslant T_\delta$,

$$\sum_{\ell=1}^{N_\eta} \frac{N_T(\ell)}{T} \left\| \boldsymbol{y}^\ell - \frac{1}{N_T(\ell)} \sum_{t=1}^{T} \boldsymbol{y}_t \mathbb{I}_{\{L_t=\ell\}} \right\|_1 \leqslant \eta + \delta \,. \tag{5}$$

The existence of a calibrated strategy in the sense of (4) however follows from the same approachability-based construction studied in Mannor and Stoltz [19] to get (5) and is detailed in the appendix. In the sequel we will only use the following consequence of calibration: that for all strategies of the second player, with probability at least $1-\delta$, for all $T \geqslant T_\delta$,

$$\max_{\ell=1,\dots,N_\eta} \frac{N_T(\ell)}{T} \left(\left\| \boldsymbol{y}^\ell - \frac{1}{N_T(\ell)} \sum_{t=1}^{T} \boldsymbol{y}_t \mathbb{I}_{\{L_t=\ell\}} \right\|_1 - \eta \right)_+ \leqslant \delta \,. \tag{6}$$

Proof of the first statement of Theorem 4.1. The insight of this proof is similar to the one illustrated in Perchet [22]. We first note that it suffices to prove that for all $\varepsilon > 0$, the set $\mathcal{C}_\varepsilon$ defined as the ε–neighborhood of $\mathcal{C}$ is $\overline{m}$–robust approachable with mixed actions; this is so up to proceeding in regimes $r = 1, 2, \dots$ each corresponding to a dyadic value $\varepsilon_r = 2^{-r}$ and lasting for a number of rounds carefully chosen in terms of the length of the previous regimes.

Therefore, we fix $\varepsilon > 0$ and associate with it a modulus of continuity $\eta > 0$ given by the uniform continuity of $\overline{m}$ in its second argument. We consider an $\eta/2$–calibrated strategy of the first player, which we will use as an auxiliary strategy. Since (RAC) is satisfied, we may associate with each element $\boldsymbol{y}^\ell$ of the underlying $\eta/2$–grid a mixed action $\boldsymbol{x}^\ell \in \Delta(\mathcal{A})$ such that $\overline{m}(\boldsymbol{x}^\ell, \boldsymbol{y}^\ell) \subseteq \mathcal{C}$. The main strategy of the first player then prescribes the use of $\boldsymbol{x}_t = \boldsymbol{x}^{L_t}$ at each round $t \geqslant 1$. The intuition behind this definition is that if $\boldsymbol{y}^{L_t}$ is forecast by the auxiliary strategy, then since the latter is calibrated, one should play as good as possible against $\boldsymbol{y}^{L_t}$; in view of the aim at hand, which is approaching $\mathcal{C}$, such a good reply is given by $\boldsymbol{x}^{L_t}$.

To assess the constructed strategy, we group rounds according to the values ℓ taken by the L_t; to that end, we recall that $N_T(\ell)$ denotes the number of rounds in which $\boldsymbol{y}^\ell$ was forecast and $\boldsymbol{x}^\ell$ was played. The average payoff up to round T is then rewritten as

$$\frac{1}{T} \sum_{t=1}^{T} \overline{m}(\boldsymbol{x}_t, \boldsymbol{y}_t) = \sum_{\ell=1}^{N_{\eta/2}} \frac{N_T(\ell)}{T} \left(\frac{1}{N_T(\ell)} \sum_{t=1}^{T} \overline{m}(\boldsymbol{x}^\ell, \boldsymbol{y}_t) \mathbb{I}_{\{L_t=\ell\}} \right) .$$

We denote for all ℓ such that $N_T(\ell) > 0$ the average of their corresponding mixed actions $\boldsymbol{y}_t$ by

$$\overline{\boldsymbol{y}}_T^\ell = \frac{1}{N_T(\ell)} \sum_{t=1}^{T} \boldsymbol{y}_t \mathbb{I}_{\{L_t=\ell\}} \,.$$

The convexity of $\overline{m}$ in its second argument leads to the inclusion

$$\frac{1}{T} \sum_{t=1}^{T} \overline{m}(\boldsymbol{x}_t, \boldsymbol{y}_t) = \sum_{\ell=1}^{N_{\eta/2}} \frac{N_T(\ell)}{T} \left(\frac{1}{N_T(\ell)} \sum_{t=1}^{T} \overline{m}(\boldsymbol{x}^\ell, \boldsymbol{y}_t) \mathbb{I}_{\{L_t=\ell\}} \right) \subseteq \sum_{\ell=1}^{N_{\eta/2}} \frac{N_T(\ell)}{T} \, \overline{m}\big(\boldsymbol{x}^\ell, \overline{\boldsymbol{y}}_T^\ell\big) \,.$$

To show that the above-defined strategy $\overline{m}$–robust approaches $\mathcal{C}_\varepsilon = \mathcal{C} + \varepsilon \boldsymbol{B}$, it suffices to show that for all $\delta > 0$, there exists an integer T'_δ such that for all strategies of the second player,

$$\mathbb{P}\left\{ \forall \, T \geqslant T'_\delta, \quad \sum_{\ell=1}^{N_{\eta/2}} \frac{N_T(\ell)}{T} \, \overline{m}\big(\boldsymbol{x}^\ell, \overline{\boldsymbol{y}}_T^\ell\big) \subseteq \mathcal{C} + (\varepsilon + \delta) \boldsymbol{B} \right\} \geqslant 1 - \delta \,.$$

We denote by M a bound in ℓ^2–norm on $\overline{m}$, i.e., for all $\boldsymbol{x} \in \Delta(\mathcal{A})$ and $\boldsymbol{y} \in \Delta(\mathcal{B})$, the inclusion $\overline{m}(\boldsymbol{x},\boldsymbol{y}) \subseteq M\boldsymbol{B}$ holds. We let $\delta' = \delta(\eta/2)/\big(M\,N_{\eta/2}\big)$ and define T'_δ as the time $T_{\delta'}$ corresponding to (6). All statements that follow will be for all strategies of the second player and with probability at least $1-\delta' \geqslant 1-\delta$, for all $T \geqslant T'_\delta$, as required. For each index ℓ of the grid, either $\delta' T/N_T(\ell) \leqslant \eta/2$ or $\delta' T/N_T(\ell) > \eta/2$. In the first case, following (6), $\big\|\boldsymbol{y}^\ell - \overline{\boldsymbol{y}}_T^\ell\big\| \leqslant \eta/2 + \delta' T/N_T(\ell) \leqslant \eta$; since η is the modulus of continuity for ε, we get that

$$\frac{N_T(\ell)}{T}\,\overline{m}\big(\boldsymbol{x}^\ell, \overline{\boldsymbol{y}}_T^\ell\big) \subseteq \frac{N_T(\ell)}{T}\,\big(\overline{m}\big(\boldsymbol{x}^\ell, \boldsymbol{y}^\ell\big) + \varepsilon\boldsymbol{B}\big) \subseteq \frac{N_T(\ell)}{T}\big(\mathcal{C} + \varepsilon\boldsymbol{B}\big)\,,$$

where we used the definition of $\boldsymbol{x}^\ell$ to get the second inclusion. In the second case, using the boundedness of $\overline{m}$, we simply write

$$\frac{N_T(\ell)}{T}\,\overline{m}\big(\boldsymbol{x}^\ell, \overline{\boldsymbol{y}}_T^\ell\big) \subseteq \frac{N_T(\ell)}{T}\,M\boldsymbol{B} \subseteq \frac{\delta'}{\eta/2}\,M\boldsymbol{B}\,.$$

Summing these bounds over ℓ yields

$$\sum_{\ell=1}^{N_{\eta/2}} \frac{N_T(\ell)}{T}\,\overline{m}\big(\boldsymbol{x}^\ell, \overline{\boldsymbol{y}}_T^\ell\big) \subseteq \mathcal{C} + \varepsilon\boldsymbol{B} + \frac{N_{\eta/2}\delta'}{\eta/2}\,M\,\boldsymbol{B} = \mathcal{C} + (\varepsilon+\delta)\,\boldsymbol{B}\,,$$

where we used the definition of δ' in terms of δ. This concludes the proof. □

5. Approachability in games with partial monitoring: statement of the necessary and sufficient condition; links with robust approachability. A repeated vector-valued game with partial monitoring is described as follows (see, e.g., Mertens et al. [21], Rustichini [29], and the references therein). The players have respective finite action sets $\mathcal{I}$ and $\mathcal{J}$. We denote by $r : \mathcal{I}\times\mathcal{J} \to \mathbb{R}^d$ the vector-valued payoff function of the first player and extend it multi-linearly to $\Delta(\mathcal{I})\times\Delta(\mathcal{J})$. At each round, players simultaneously choose their actions $I_t \in \mathcal{I}$ and $J_t \in \mathcal{J}$, possibly at random according to probability distributions denoted by $\boldsymbol{p}_t \in \Delta(\mathcal{I})$ and $\boldsymbol{q}_t \in \Delta(\mathcal{J})$. At the end of a round, the first player does not observe J_t nor $r(I_t,J_t)$ but only a signal. There is a finite set $\mathcal{H}$ of possible signals; the feedback S_t that is given to the first player is drawn at random according to the distribution $H(I_t,J_t)$, where the mapping $H : \mathcal{I}\times\mathcal{J} \to \Delta(\mathcal{H})$ is known by the first player.

EXAMPLE 5.1 *Examples of such partial monitoring games are provided by, e.g., Cesa-Bianchi et al. [6], among which we can cite the apple tasting problem, the label-efficient prediction constraint, and the multi-armed bandit settings.*

Some additional notation will be useful. We denote by R the norm of (the linear extension of) r,

$$R = \max_{(i,j)\in\mathcal{I}\times\mathcal{J}} \big\|\,r(i,j)\,\big\|_2\,.$$

The cardinalities of the finite sets $\mathcal{I}$, $\mathcal{J}$, and $\mathcal{H}$ will be referred to as $N_\mathcal{I}$, $N_\mathcal{J}$, and $N_\mathcal{H}$.

Definition 2.1 can be extended as follows in this setting; the only new ingredient is the signaling structure, the aim is unchanged.

DEFINITION 5.1 *Let $\mathcal{C} \subseteq \mathbb{R}^d$ be some set; $\mathcal{C}$ is r–approachable for the signaling structure H if there exists a strategy of the first player such that, for all $\varepsilon > 0$, there exists an integer T_ε such that for all strategies of the second player,*

$$\mathbb{P}\left\{\forall\, T \geqslant T_\varepsilon, \quad \inf_{c\in\mathcal{C}} \left\| c - \frac{1}{T}\sum_{t=1}^T r(I_t,J_t)\right\|_2 \leqslant \varepsilon\right\} \geqslant 1-\varepsilon\,.$$

That is, the first player has a strategy that ensures that the sequence of his average vector-valued payoffs converges to the set $\mathcal{C}$ (uniformly with respect to the strategies of the second player), even if he only observes the random signals S_t as a feedback.

Our contributions. A necessary and sufficient condition for r–approachability with the signaling structure H was stated and proved by Perchet [23]; we therefore need to indicate where our contribution lies. First, both proofs are constructive but our strategy can be efficient (as soon as some projection operator can be computed efficiently, e.g., in the cases of external and internal regret minimization described below) whereas the one of Perchet [23] relies on auxiliary strategies that are calibrated and that require a grid that is progressively refined (leading to a step complexity that is exponential in the number T of past steps); the latter construction is in essence the one used in Section 4. Second, we are able to exhibit convergence rates. Third, as far as elegancy is concerned, our proof is short, compact, and more direct than the one of Perchet [23], which relied on several layers of notations (internal regret in games with partial monitoring, calibration of auxiliary strategies, etc.).

5.1 Statement of the necessary and sufficient condition for approachability in games with partial monitoring. To recall the mentioned approachability condition of Perchet [23] we need some additional notation: for all $\boldsymbol{q} \in \Delta(\mathcal{J})$, we denote by $\widetilde{H}(\boldsymbol{q})$ the element in $\Delta(\mathcal{H})^{\mathcal{I}}$ defined as follows. For all $i \in \mathcal{I}$, its i–th component is given by the convex combination of probability distributions over $\mathcal{H}$

$$\widetilde{H}(\boldsymbol{q})_i = H(i,\boldsymbol{q}) = \sum_{j\in\mathcal{J}} q_j H(i,j)\,.$$

Finally, we denote by $\mathcal{F}$ the convex set of feasible vectors of probability distributions over $\mathcal{H}$:

$$\mathcal{F} = \left\{\widetilde{H}(\boldsymbol{q}) : \quad \boldsymbol{q} \in \Delta(\mathcal{J})\right\}.$$

A generic element of $\mathcal{F}$ will be denoted by $\sigma \in \mathcal{F}$ and we define the set-valued function $\overline{m}$, for all $\boldsymbol{p} \in \Delta(\mathcal{I})$ and $\sigma \in \mathcal{F}$, by

$$\overline{m}(\boldsymbol{p},\sigma) = \left\{r(\boldsymbol{p},\boldsymbol{q}') : \quad \boldsymbol{q}' \in \Delta(\mathcal{J}) \text{ such that } \widetilde{H}(\boldsymbol{q}') = \sigma\right\}.$$

The necessary and sufficient condition exhibited by Perchet [23] for the r–approachability of $\mathcal{C}$ with the signaling structure H can now be recalled. In the sequel we will refer to this condition as Condition (APM), an acronym that stands for "approachability with partial monitoring."

Condition 1 (referred to as Condition (APM)) *The signaling structure H, the vector-payoff function r, and the set $\mathcal{C}$ satisfy*

$$\forall\, \boldsymbol{q} \in \Delta(\mathcal{J}), \quad \exists\, \boldsymbol{p} \in \Delta(\mathcal{I}), \quad \forall\, \boldsymbol{q}' \in \Delta(\mathcal{J}), \qquad \widetilde{H}(\boldsymbol{q}) = \widetilde{H}(\boldsymbol{q}') \quad \Rightarrow \quad r(\boldsymbol{p},\boldsymbol{q}') \in \mathcal{C}\,.$$

The condition can be equivalently reformulated as

$$\forall\, \sigma \in \mathcal{F}, \quad \exists\, \boldsymbol{p} \in \Delta(\mathcal{I}), \qquad \overline{m}(\boldsymbol{p},\sigma) \subseteq \mathcal{C}\,. \tag{APM}$$

This condition is necessary. The subsequent sections show (in a constructive way) that Condition (APM) is sufficient for r–approachability of closed convex sets $\mathcal{C}$ given the signaling structure H. That this condition is necessary was already proved in Section 3.1 of Perchet [23].

5.2 Links with robust approachability. As will become clear in the proof of Theorem 6.1, the key in our problem will be to ensure the robust approachability of $\mathcal{C}$ with the following non-linear set-valued payoff function, that is however concave–convex in the sense of Definition 4.2.

Lemma 5.1 *The function*

$$(\boldsymbol{p},\boldsymbol{q}) \in \Delta(\mathcal{I}) \times \Delta(\mathcal{J}) \;\longmapsto\; \overline{m}\big(\boldsymbol{p},\, H(\boldsymbol{q})\big)\,.$$

is concave in its first argument and convex in its second argument.

Unfortunately, efficient strategies for robust approachability were only proposed in the linear case, not in the concave–convex case. But we illustrate in the next example (and provide a general theory in the next section) how working in lifted spaces can lead to linearity and hence to efficiency.

Example 5.2 *We consider a game in which the second player (the column player) can force the first player (the row player) to play a game of matching pennies in the dark by choosing actions L or M; in the matrix below, the real numbers denote the payoff while ♣ and ♡ denote the two possible signals. The respective sets of actions are $\mathcal{I} = \{T,\, B\}$ and $\mathcal{J} = \{L,\, M,\, R\}$.*

	L	M	R
T	$1\,/\,\clubsuit$	$-1\,/\,\clubsuit$	$2\,/\,\heartsuit$
B	$-1\,/\,\clubsuit$	$1\,/\,\clubsuit$	$3\,/\,\heartsuit$

In this example we only study the mapping $\boldsymbol{p} \mapsto \overline{m}(\boldsymbol{p}, \clubsuit)$ and show that it is piecewise linear on $\Delta(\mathcal{I})$, thus, is induced by a linear mapping defined on a lifted space.

We introduce a set $\mathcal{A} = \{\boldsymbol{p}_T, \boldsymbol{p}_B, \boldsymbol{p}_{1/2}\}$ of possibly mixed actions extending the set $\mathcal{I} = \{T, B\}$ of pure actions; the set $\mathcal{A}$ is composed of

$$\boldsymbol{p}_T = \delta_T, \quad \boldsymbol{p}_B = \delta_B, \quad \text{and} \quad \boldsymbol{p}_{1/2} = \frac{1}{2}\delta_T + \frac{1}{2}\delta_B\,.$$

Each mixed action in $\Delta(\mathcal{I})$ can be uniquely written as $\boldsymbol{p}_\lambda = \lambda\,\delta_B + (1-\lambda)\,\delta_T$ for some $\lambda \in [0,1]$. Now, for $\lambda \geqslant 1/2$, first,

$$\boldsymbol{p}_\lambda = (2\lambda - 1)\,\delta_B + \big(1 - (2\lambda-1)\big)\,\boldsymbol{p}_{1/2}\,;$$

second, by definition of $\overline{m}$,

$$\overline{m}\big(\boldsymbol{p}_\lambda, \clubsuit\big) = [1-2\lambda,\, 2\lambda - 1]\,;$$

since in particular $\overline{m}\big(\boldsymbol{p}_{1/2}, \clubsuit\big) = \{0\}$ and $\overline{m}(\delta_B, \clubsuit) = [-1, 1]$, we have the convex decomposition

$$\overline{m}\big(\boldsymbol{p}_\lambda, \clubsuit\big) = (2\lambda-1)\,\overline{m}(\delta_B, \clubsuit) + \big(1-(2\lambda-1)\big)\,\overline{m}(\boldsymbol{p}_{1/2}, \clubsuit)\,,$$

which can be restated as

$$\overline{m}\big(\boldsymbol{p}_\lambda, \clubsuit\big) = \overline{m}\Big((2\lambda-1)\,\delta_B + \big(1-(2\lambda-1)\big)\,\boldsymbol{p}_{1/2}, \clubsuit\Big) = (2\lambda-1)\,\overline{m}(\delta_B, \clubsuit) + \big(1-(2\lambda-1)\big)\,\overline{m}(\boldsymbol{p}_{1/2}, \clubsuit)\,.$$

That is, $\overline{m}(\,\cdot\,, \clubsuit)$ is linear on the subset of $\Delta(\mathcal{I})$ corresponding to mixed actions $\boldsymbol{p}_\lambda$ with $\lambda \geqslant 1/2$.

A similar property holds the subset of distributions with $\lambda \leqslant 1/2$, so that we have proved that $\overline{m}(\,\cdot\,, \clubsuit)$ is piecewise linear on $\Delta(\mathcal{I})$.

The linearity on a lifted space comes from the following observation: $\overline{m}$ is induced by the linear extension to $\Delta(\mathcal{A})$ of the restriction of $\overline{m}$ to $\mathcal{A}$ (see Definition 6.1 for a more formal statement).

6. Application of robust approachability to games with partial monitoring: for a particular class of games encompassing regret minimization. In this section we consider the case where the signaling structure has some special properties described below (linked to linearity properties on lifted spaces) and that can be exploited to get efficient strategies. The case of general signaling structures is then considered in Section 7 but the particular class of games considered here is already rich enough to encompass the minimization of external and internal regret.

6.1 Approachability in bi-piecewise linear games. To define bi-piecewise linearity of a game, we start from a technical lemma that shows that $\overline{m}(\boldsymbol{p}, \sigma)$ can be written as a *finite* convex combination of sets of the form $\overline{m}(\boldsymbol{p}, b)$, where b belongs to some finite set $\mathcal{B} \subseteq \mathcal{F}$ that depends on the game. Under the additional assumption of piecewise linearity of the thus-defined mappings $\overline{m}(\,\cdot\,, b)$, we then describe a (possibly) efficient strategy for approachability followed by convergence rate guarantees.

6.1.1 Bi-piecewise linearity of a game – A preliminary technical result.

LEMMA 6.1 *For any game with partial monitoring, there exists a finite set $\mathcal{B} \subset \mathcal{F}$ and a piecewise-linear (injective) mapping $\Phi : \mathcal{F} \to \Delta(\mathcal{B})$ such that*

$$\forall\,\sigma \in \mathcal{F}, \quad \forall\,\boldsymbol{p} \in \Delta(\mathcal{I}), \qquad \overline{m}(\boldsymbol{p}, \sigma) = \sum_{b \in \mathcal{B}} \Phi_b(\sigma)\,\overline{m}(\boldsymbol{p}, b)\,,$$

where we denoted the convex weight vector $\Phi(\sigma) \in \Delta(\mathcal{B})$ by $\big(\Phi_b(\sigma)\big)_{b\in\mathcal{B}}$.

PROOF. Since $\widetilde{H}$ is linear on the polytope $\Delta(\mathcal{J})$, Proposition 2.4 in Rambau and Ziegler [28] implies that its inverse application $\widetilde{H}^{-1}$ is a piecewise linear mapping of $\mathcal{F}$ into the subsets of $\Delta(\mathcal{J})$. This means

that there exists a finite decomposition of $\mathcal{F}$ into polytopes $\{P_1, \dots, P_K\}$ each on which $\widetilde{H}^{-1}$ is linear. Up to a triangulation (see, e.g., Chapter 14 in [12]), we can assume that each P_k is a simplex. Denote by $\mathcal{B}_k \subseteq \mathcal{F}$ the set of vertices of P_k; then, the finite subset stated in the lemma is

$$\mathcal{B} = \bigcup_{k=1}^{K} \mathcal{B}_k \,,$$

the set of all vertices of all the simplices.

Fix any $\sigma \in \mathcal{F}$. It belongs to some simplex P_k, so that there exists a convex decomposition $\sigma = \sum_{b \in \mathcal{B}_k} \lambda_b \, b$; this decomposition is unique within the simplex P_k. If σ belongs to two different simplices, then it actually belongs to their common face and the two possible decompositions coincide (some coefficients λ_b in the above decomposition are null). All in all, with each $\sigma \in \mathcal{F}$, we can associate a unique decomposition in $\mathcal{B}$,

$$\sigma = \sum_{b \in \mathcal{B}} \Phi_b(\sigma) \, b \,,$$

where the coefficients $\big(\Phi_b(\sigma)\big)_{b \in \mathcal{B}}$ form a convex weight vector over $\mathcal{B}$, i.e., belong to $\Delta(\mathcal{B})$; in addition, $\Phi_b(\sigma) > 0$ only if $b \in \mathcal{B}_k$, where k is such that $\sigma \in P_k$.

Since $\widetilde{H}^{-1}$ is linear on each simplex $P_1, \dots, P_K$, we therefore get

$$\widetilde{H}^{-1}(\sigma) = \sum_{b \in \mathcal{B}} \Phi_b(\sigma) \, \widetilde{H}^{-1}(b) \,.$$

Finally, the result is a consequence of the fact that

$$\overline{m}(\boldsymbol{p}, \sigma) = r\Big(\boldsymbol{p}, \, \widetilde{H}^{-1}(\sigma)\Big) = r\left(\boldsymbol{p}, \sum_{b \in \mathcal{B}} \Phi_b(\sigma) \, \widetilde{H}^{-1}(b)\right) ,$$

which implies, by linearity of r, that

$$\overline{m}(\boldsymbol{p}, \sigma) = \sum_{b \in \mathcal{B}} \Phi_b(\sigma) \, r\Big(\boldsymbol{p}, \widetilde{H}^{-1}(b)\Big) = \sum_{b \in \mathcal{B}} \Phi_b(\sigma) \, \overline{m}(\boldsymbol{p}, b) \,,$$

which concludes the proof. □

REMARK 6.1 *The proof shows that Φ is piecewise linear on a finite decomposition of $\mathcal{F}$; it is therefore Lipschitz on $\mathcal{F}$. We denote by κ_Φ its Lipschitz constant with respect to the ℓ^2–norms.*

The main contribution of this subsection (Definition 6.1) relies on the following additional assumption.

ASSUMPTION 6.1 *A game is bi-piecewise linear if $\overline{m}(\,\cdot\,, b)$ is piecewise linear on $\Delta(\mathcal{I})$ for every $b \in \mathcal{B}$.*

Assumption 6.1 means that for all $b \in \mathcal{B}$ there exists a decomposition of $\Delta(\mathcal{I})$ into polytopes each on which $\overline{m}(\,\cdot\,, b)$ is linear. Since $\mathcal{B}$ is finite, there exists a finite number of such decompositions, and thus there exists a decomposition to polytopes that refines all of them. (The latter is generated by the intersection of all considered polytopes as b varies.) By construction, every $\overline{m}(\,\cdot\,, b)$ is linear on any of the polytopes of this common decomposition. We denote by $\mathcal{A} \subset \Delta(\mathcal{I})$ the finite subset of all their vertices: a construction similar to the one used in the proof of Lemma 6.1 (provided below) then leads to a piecewise linear (injective) mapping $\Theta : \Delta(\mathcal{I}) \to \Delta(\mathcal{A})$, where $\Theta(\boldsymbol{p})$ is the decomposition of $\boldsymbol{p}$ on the vertices of the polytope(s) of the decomposition to which it belongs, satisfying

$$\forall \, b \in \mathcal{B}, \quad \forall \, \boldsymbol{p} \in \Delta(\mathcal{I}), \qquad \overline{m}(\boldsymbol{p}, b) = \sum_{a \in \mathcal{A}} \Theta_a(\boldsymbol{p}) \, \overline{m}(a, b) \,,$$

where we denoted the convex weight vector $\Theta(\boldsymbol{p}) \in \Delta(\mathcal{A})$ by $\big(\Theta_a(\boldsymbol{p})\big)_{a \in \mathcal{A}}$. This, Lemma 6.1, and Assumption 6.1 show that on a lifted space, $\overline{m}$ coincides with a bi-linear mapping $\overline{\overline{m}}$, as is made formal in the next definition.

DEFINITION 6.1 *We denote by $\overline{\overline{m}}$ the linear extension to $\Delta(\mathcal{A} \times \mathcal{B})$ of the restriction of $\overline{m}$ to $\mathcal{A} \times \mathcal{B}$, so that for all $\boldsymbol{p} \in \Delta(\mathcal{I})$ and $\sigma \in \mathcal{F}$,*

$$\overline{m}(\boldsymbol{p}, \sigma) = \overline{\overline{m}}\big(\Theta(\boldsymbol{p}), \, \Phi(\sigma)\big) \,.$$

Approaching Strategy in Games with Partial Monitoring

Parameters: an integer block length $L \geqslant 1$, an exploration parameter $\gamma \in [0,1]$, a strategy Ψ for $\overline{\overline{m}}$–robust approachability of $\mathcal{C}$

Notation: $\boldsymbol{u} \in \Delta(\mathcal{I})$ is the uniform distribution over $\mathcal{I}$, $P_{\mathcal{F}}$ denotes the projection operator in ℓ^2–norm of $\mathbb{R}^{\mathcal{H}\times\mathcal{I}}$ onto $\mathcal{F}$

Initialization: compute the finite set $\mathcal{B}$ and the mapping $\Phi : \mathcal{F} \to \Delta(\mathcal{B})$ of Lemma 6.1, compute the finite set $\mathcal{A}$ and the mapping $\Theta : \Delta(\mathcal{I}) \to \Delta(\mathcal{A})$ defined based on Assumption 6.1, pick an arbitrary $\boldsymbol{\theta}_1 \in \Delta(\mathcal{A})$

For all blocks $n = 1, 2, \ldots,$

(i) define $\boldsymbol{x}_n = \sum_{a\in\mathcal{A}} \theta_{n,a}\, a$ and $\boldsymbol{p}_n = (1-\gamma)\,\boldsymbol{x}_n + \gamma\,\boldsymbol{u}$;

(ii) for rounds $t = (n-1)L+1, \ldots, nL$,

2.1 draw an action $I_t \in \mathcal{I}$ at random according to $\boldsymbol{p}_n$;

2.2 get the signal S_t;

(iii) form the estimated vector of probability distributions over signals,

$$\widetilde{\sigma}_n = \left(\frac{1}{L} \sum_{t=(n-1)L+1}^{nL} \frac{\mathbb{I}_{\{S_t=s\}}\mathbb{I}_{\{I_t=i\}}}{p_{I_t,n}} \right)_{(i,s)\in\mathcal{I}\times\mathcal{H}} ;$$

(iv) compute the projection $\widehat{\sigma}_n = P_{\mathcal{F}}\big(\widetilde{\sigma}_n\big)$;

(v) choose $\boldsymbol{\theta}_{n+1} = \Psi\Big(\boldsymbol{\theta}_1, \Phi\big(\widehat{\sigma}_1\big), \ldots, \boldsymbol{\theta}_n, \Phi\big(\widehat{\sigma}_n\big)\Big)$.

Figure 1: The proposed strategy, which plays in blocks.

6.1.2 Construction of a strategy to approach $\mathcal{C}$. The approaching strategy for the original problem is based on a strategy Ψ for $\overline{\overline{m}}$–approachability of $\mathcal{C}$, provided by Theorem 3.1; we therefore first need to prove the existence of such a Ψ.

LEMMA 6.2 *Under Condition* (APM)*, the closed convex set $\mathcal{C}$ is $\overline{\overline{m}}$–robust approachable.*

PROOF. We show that Condition (RAC) in Theorem 3.1 is satisfied, that is, that for all $\boldsymbol{y} \in \Delta(\mathcal{B})$, there exists some $\boldsymbol{x} \in \Delta(\mathcal{A})$ such that $\overline{\overline{m}}(\boldsymbol{x},\boldsymbol{y}) \subseteq \mathcal{C}$. With such a given $\boldsymbol{y} \in \Delta(\mathcal{B})$, we associate[2] the feasible vector of signals $\sigma = \sum_{b\in\mathcal{B}} y_b\, b \in \mathcal{F}$ and let $\boldsymbol{p}$ be given by Condition (APM), so that $\overline{m}(\boldsymbol{p},\sigma) \subseteq \mathcal{C}$. By linearity of $\overline{\overline{m}}$ (for the first equality), by convexity of $\overline{m}$ in its second argument (for the first inclusion), by Lemma 6.1 (for the second and fourth equalities), by construction of $\mathcal{A}$ (for the third equality),

$$\begin{aligned}\overline{\overline{m}}\big(\Theta(\boldsymbol{p}),\boldsymbol{y}\big) = \sum_{a\in\mathcal{A}} \Theta_a(\boldsymbol{p}) \sum_{b\in\mathcal{B}} y_b\, \overline{m}(a,b) \quad &\subseteq \quad \sum_{a\in\mathcal{A}} \Theta_a(\boldsymbol{p})\, \overline{m}(a,\sigma) = \sum_{a\in\mathcal{A}} \Theta_a(\boldsymbol{p}) \sum_{b\in\mathcal{B}} \Phi_b(\sigma)\, \overline{m}(a,b) \\ &= \quad \sum_{b\in\mathcal{B}} \Phi_b(\sigma)\, \overline{m}(\boldsymbol{p},b) = \overline{m}(\boldsymbol{p},\sigma) \subseteq \mathcal{C}\,,\end{aligned}$$

which concludes the proof. □

We consider the strategy described in Figure 1. It forces exploration at a γ rate, as is usual in situations with partial monitoring. One of its key ingredient, that conditionally unbiased estimators are available, is extracted from Section 6 in the article by Lugosi et al. [16]: in block n we consider sums of elements of the form

$$\widehat{H}_t = \left(\frac{\mathbb{I}_{\{S_t=s\}}\mathbb{I}_{\{I_t=i\}}}{p_{I_t,n}} \right)_{(i,s)\in\mathcal{I}\times\mathcal{H}} \in \mathbb{R}^{\mathcal{H}\times\mathcal{I}};$$

averaging over the respective random draws of I_t and S_t according to $\boldsymbol{p}_n$ and $H(I_t,J_t)$, i.e., taking the conditional expectation $\mathbb{E}_t$ with respect to $\boldsymbol{p}_n$ and J_t, we get

$$\mathbb{E}_t\big[\widehat{H}_t\big] = \widetilde{H}\big(\delta_{J_t}\big). \tag{7}$$

[2]Note however that we do not necessarily have that $\Phi(\sigma)$ and $\boldsymbol{y}$ are equal, as Φ is not a one-to-one mapping (it is injective but not surjective).

Indeed, the conditional expectation of the component i of $\widehat{H}_t$ equals

$$\mathbb{E}_t\left[\left(\frac{\mathbb{I}_{\{S_t=s\}}\mathbb{I}_{\{I_t=i\}}}{p_{I_t,n}}\right)_{s\in\mathcal{H}}\right] = \mathbb{E}_t\left[\frac{H(I_t,J_t)\,\mathbb{I}_{\{I_t=i\}}}{p_{I_t,n}}\right] = \frac{H(i,J_t)}{p_{i,n}}\,\mathbb{E}_t\big[\mathbb{I}_{\{I_t=i\}}\big] = H(i,J_t)\,,$$

where we first took the expectation over the random draw of S_t (conditionally to $\boldsymbol{p}_n$, J_t, and I_t) and then over the one of I_t. Consequently, concentration-of-the-measure arguments can show that for L large enough,

$$\widetilde{\sigma}_n = \frac{1}{L}\sum_{t=(n-1)L+1}^{nL}\widehat{H}_t \qquad \text{is close to} \qquad \widetilde{H}\big(\widehat{\boldsymbol{q}}_n\big)\,, \qquad \text{where} \quad \widehat{\boldsymbol{q}}_n = \frac{1}{L}\sum_{t=(n-1)L+1}^{nL}\delta_{J_t}\,.$$

Actually, since $\mathcal{F}\subseteq\Delta(\mathcal{H})^{\mathcal{I}}$, we have a natural embedding of $\mathcal{F}$ into $\mathbb{R}^{\mathcal{H}\times\mathcal{I}}$ and we can define $P_{\mathcal{F}}$, the convex projection operator onto $\mathcal{F}$ (in ℓ^2–norm). Instead of using directly $\widetilde{\sigma}_n$, we consider in our strategy $\widehat{\sigma}_n = P_{\mathcal{F}}\big(\widetilde{\sigma}_n\big)$, which is even closer to $\widetilde{H}\big(\widehat{\boldsymbol{q}}_n\big)$.

More precisely, the following result can be extracted from the proof of Theorem 6.1 in Lugosi et al. [16]. The proof is provided in Appendix B.

LEMMA 6.3 *With probability* $1-\delta$,

$$\left\|\widehat{\sigma}_n - \widetilde{H}\big(\widehat{\boldsymbol{q}}_n\big)\right\|_2 \leqslant \sqrt{N_{\mathcal{I}}N_{\mathcal{H}}}\left(\sqrt{\frac{2N_{\mathcal{I}}}{\gamma L}\ln\frac{2N_{\mathcal{I}}N_{\mathcal{H}}}{\delta}} + \frac{1}{3}\frac{N_{\mathcal{I}}}{\gamma L}\ln\frac{2N_{\mathcal{I}}N_{\mathcal{H}}}{\delta}\right).$$

6.1.3 A performance guarantee for the strategy of Figure 1. For the sake of simplicity, we provide first a performance bound for fixed parameters γ and L tuned as functions of T. Adaptation to $T\to\infty$ is then described in the next section; note that it cannot be performed by simply proceeding in regimes, as the approachability guarantees offered by the second part of the theorem are only at time round T. (This is so because the considered strategy depends on T via the parameters γ and L.)

THEOREM 6.1 *Consider a closed convex set* $\mathcal{C}$ *and a game* (r,H) *for which Condition* (APM) *is satisfied and that is bi-piecewise linear in the sense of Assumption 6.1. Then, for all* $T\geqslant 1$, *the strategy of Figure 1, run with parameters* $\gamma\in[0,1]$ *and* $L\geqslant 1$ *and fed with a strategy* Ψ *for* $\overline{\overline{m}}$*–approachability of* $\mathcal{C}$ *(provided by Lemma 6.2) is such that, with probability at least* $1-\delta$,

$$\begin{aligned}\inf_{c\in\mathcal{C}}\left\|c-\frac{1}{T}\sum_{t=1}^{T}r(I_t,J_t)\right\|_2 \;\leqslant\; &\frac{2L}{T}R + 4R\sqrt{\frac{\ln\big((2T)/(L\delta)\big)}{T}} + 2\gamma R + \frac{2R}{\sqrt{T/L-1}}\sqrt{N_{\mathcal{A}}N_{\mathcal{B}}}\\ &+ R\kappa_\Phi\sqrt{N_{\mathcal{I}}N_{\mathcal{H}}N_{\mathcal{A}}}\left(\sqrt{\frac{2N_{\mathcal{I}}}{\gamma L}\ln\frac{2N_{\mathcal{I}}N_{\mathcal{H}}T}{L\delta}} + \frac{1}{3}\frac{N_{\mathcal{I}}}{\gamma L}\ln\frac{2N_{\mathcal{I}}N_{\mathcal{H}}T}{L\delta}\right).\end{aligned}$$

In particular, for all $T\geqslant 1$, *the choices of* $L=\lceil T^{3/5}\rceil$ *and* $\gamma = T^{-1/5}$ *imply that with probability at least* $1-\delta$,

$$\inf_{c\in\mathcal{C}}\left\|c-\frac{1}{T}\sum_{t=1}^{T}r(I_t,J_t)\right\|_2 \;\leqslant\; \square\left(T^{-1/5}\sqrt{\ln\frac{T}{\delta}} + T^{-2/5}\ln\frac{T}{\delta}\right)$$

for some constant $\square$ *depending only on* $\mathcal{C}$ *and on the game* $(r,\,H)$ *at hand.*

The efficiency of the strategy of Figure 1 depends on whether it can be fed with an efficient approachability strategy Ψ, which in turn depends on the respective geometries of $\overline{m}$ and $\mathcal{C}$, as was indicated before the statement of Theorem 3.1. (Note that the projection onto $\mathcal{F}$ can be performed in polynomial time, as the latter closed convex set is defined by finitely many linear constraints, and that the computation of $\mathcal{A}$, $\mathcal{B}$, and $\overline{\overline{m}}$ can be performed beforehand.) In any case, the per-round complexity is constant (though possibly large).

PROOF. We write T as $T=NL+k$ where N is an integer and $0\leqslant k\leqslant L-1$ and will show successively that (possibly with overwhelming probability only) the following statements hold.

$$\frac{1}{T}\sum_{t=1}^{T}r(I_t,J_t) \qquad \text{is close to} \qquad \frac{1}{NL}\sum_{t=1}^{NL}r(I_t,J_t)\,; \tag{8}$$

$$\frac{1}{NL}\sum_{t=1}^{NL} r(I_t,J_t) \qquad \text{is close to} \qquad \frac{1}{N}\sum_{n=1}^{N} r\bigl(\boldsymbol{p}_n,\widehat{\boldsymbol{q}}_n\bigr)\,; \tag{9}$$

$$\frac{1}{N}\sum_{n=1}^{N} r\bigl(\boldsymbol{p}_n,\widehat{\boldsymbol{q}}_n\bigr) \qquad \text{is close to} \qquad \frac{1}{N}\sum_{n=1}^{N} r\bigl(\boldsymbol{x}_n,\widehat{\boldsymbol{q}}_n\bigr)\,; \tag{10}$$

$$\frac{1}{N}\sum_{n=1}^{N} r\bigl(\boldsymbol{x}_n,\widehat{\boldsymbol{q}}_n\bigr) = \frac{1}{N}\sum_{n=1}^{N}\sum_{a\in\mathcal{A}} \theta_{n,a}\, r\bigl(a,\widehat{\boldsymbol{q}}_n\bigr) \qquad \text{belongs to the set} \qquad \frac{1}{N}\sum_{n=1}^{N}\sum_{a\in\mathcal{A}} \theta_{n,a}\,\overline{m}\Bigl(a,\,\widetilde{H}\bigl(\widehat{\boldsymbol{q}}_n\bigr)\Bigr)\,;$$

$$\frac{1}{N}\sum_{n=1}^{N}\sum_{a\in\mathcal{A}} \theta_{n,a}\,\overline{m}\Bigl(a,\,\widetilde{H}\bigl(\widehat{\boldsymbol{q}}_n\bigr)\Bigr) \qquad \text{is equal to the set} \qquad \frac{1}{N}\sum_{n=1}^{N}\overline{\overline{m}}\Bigl(\boldsymbol{\theta}_n,\,\Phi\Bigl(\widetilde{H}\bigl(\widehat{\boldsymbol{q}}_n\bigr)\Bigr)\Bigr)\,;$$

$$\frac{1}{N}\sum_{n=1}^{N}\overline{\overline{m}}\Bigl(\boldsymbol{\theta}_n,\,\Phi\Bigl(\widetilde{H}\bigl(\widehat{\boldsymbol{q}}_n\bigr)\Bigr)\Bigr) \qquad \text{is close to the set} \qquad \frac{1}{N}\sum_{n=1}^{N}\overline{\overline{m}}\Bigl(\boldsymbol{\theta}_n,\,\Phi\bigl(\widehat{\sigma}_n\bigr)\Bigr)\,; \tag{11}$$

$$\frac{1}{N}\sum_{n=1}^{N}\overline{\overline{m}}\Bigl(\boldsymbol{\theta}_n,\,\Phi\bigl(\widehat{\sigma}_n\bigr)\Bigr) \qquad \text{is close to the set} \qquad \mathcal{C}\,; \tag{12}$$

where we recall that the notation $\widehat{\boldsymbol{q}}_n$ was defined above and is referring to the empirical distribution of the J_t in the n–th block. Actually, we will show below the numbered statements only. The first unnumbered statement is immediate by the definition of $\boldsymbol{x}_n$, the linearity of r, and the very definition of $\overline{m}$; while the second one follows from Definition 6.1:

$$\frac{1}{N}\sum_{n=1}^{N}\sum_{a\in\mathcal{A}} \theta_{n,a}\,\overline{m}\Bigl(a,\,\widetilde{H}\bigl(\widehat{\boldsymbol{q}}_n\bigr)\Bigr) = \frac{1}{N}\sum_{n=1}^{N}\sum_{(a,b)\in\mathcal{A}\times\mathcal{B}} \theta_{n,a}\,\Phi_b\Bigl(\widetilde{H}\bigl(\widehat{\boldsymbol{q}}_n\bigr)\Bigr)\,\overline{m}(a,b) = \frac{1}{N}\sum_{n=1}^{N}\overline{\overline{m}}\Bigl(\boldsymbol{\theta}_n,\,\Phi\Bigl(\widetilde{H}\bigl(\widehat{\boldsymbol{q}}_n\bigr)\Bigr)\Bigr)\,.$$

Step 1: Assertion (8)**.** A direct calculation decomposing the sum over T elements into a sum over the NL first elements and the k remaining ones shows that

$$\left\|\frac{1}{T}\sum_{t=1}^{T} r(I_t,J_t) - \frac{1}{NL}\sum_{t=1}^{NL} r(I_t,J_t)\right\|_2 \leqslant R\left(\frac{k}{T} + \left(\frac{1}{NL}-\frac{1}{T}\right)NL\right) = \frac{2k}{T}R \leqslant \frac{2L}{T}R\,.$$

Step 2: Assertion (9)**.** We note that by defining $\mathbb{E}_t$ the conditional expectation with respect to (I_1,S_1,J_1), …, $(I_{t-1},S_{t-1},J_{t-1})$ and J_t, which fixes the values of the distribution $\boldsymbol{p}'_t$ of I_t and the value of J_t, we have

$$\mathbb{E}_t\bigl[r(I_t,J_t)\bigr] = r(\boldsymbol{p}'_t,J_t)\,.$$

We note that by definition of the forecaster, $\boldsymbol{p}'_t = \boldsymbol{p}_n$ if t belongs to the n–th block. By a version of the Hoeffding-Azuma inequality for sums of Hilbert space-valued martingale differences stated as[3] Lemma 3.2 in Chen and White [7], we therefore get that with probability at least $1-\delta$,

$$\left\|\frac{1}{NL}\sum_{t=1}^{NL} r(I_t,J_t) - \frac{1}{N}\sum_{n=1}^{N} r\bigl(\boldsymbol{p}_n,\widehat{\boldsymbol{q}}_n\bigr)\right\|_2 \leqslant 4R\sqrt{\frac{\ln(2/\delta)}{T}}\,.$$

Step 3: Assertion (10)**.** Since by definition $\boldsymbol{p}_n = (1-\gamma)\,\boldsymbol{x}_n + \gamma\,\boldsymbol{u}$, we get

$$\left\|\frac{1}{N}\sum_{n=1}^{N} r\bigl(\boldsymbol{p}_n,\widehat{\boldsymbol{q}}_n\bigr) - \frac{1}{N}\sum_{n=1}^{N} r\bigl(\boldsymbol{x}_n,\widehat{\boldsymbol{q}}_n\bigr)\right\|_2 \leqslant 2\gamma R\,.$$

Step 4: Assertion (11)**.** We fix a given block n. Lemma 6.3 indicates that with probability $1-\delta$,

$$\left\|\widehat{\sigma}_n - \widetilde{H}\bigl(\widehat{\boldsymbol{q}}_n\bigr)\right\|_2 \leqslant \sqrt{N_{\mathcal{I}}N_{\mathcal{H}}}\left(\sqrt{\frac{2N_{\mathcal{I}}}{\gamma L}\ln\frac{2N_{\mathcal{I}}N_{\mathcal{H}}}{\delta}} + \frac{1}{3}\frac{N_{\mathcal{I}}}{\gamma L}\ln\frac{2N_{\mathcal{I}}N_{\mathcal{H}}}{\delta}\right). \tag{13}$$

Since Φ is Lipschitz (see Remark 6.1), with a Lipschitz constant in ℓ^2–norms denoted by κ_Φ, we get that with probability $1-\delta$,

$$\left\|\Phi\bigl(\widehat{\sigma}_n\bigr) - \Phi\Bigl(\widetilde{H}\bigl(\widehat{\boldsymbol{q}}_n\bigr)\Bigr)\right\|_2 \leqslant \kappa_\Phi\sqrt{N_{\mathcal{I}}N_{\mathcal{H}}}\left(\sqrt{\frac{2N_{\mathcal{I}}}{\gamma L}\ln\frac{2N_{\mathcal{I}}N_{\mathcal{H}}}{\delta}} + \frac{1}{3}\frac{N_{\mathcal{I}}}{\gamma L}\ln\frac{2N_{\mathcal{I}}N_{\mathcal{H}}}{\delta}\right).$$

[3] Together with the fact that $\sqrt{u}\,e^{-u} \leqslant e^{-u/2}$ for all $u \geqslant 0$.

By a union bound, the above bound holds for all blocks $n = 1, \dots, N$ with probability at least $1 - N\delta$. Finally, an application of Lemma 3.1 shows that

$$\frac{1}{N}\sum_{n=1}^{N} \overline{\overline{m}}\Big(\boldsymbol{\theta}_n, \Phi\Big(\widetilde{H}\big(\widehat{\boldsymbol{q}}_n\big)\Big)\Big) \qquad \text{is in a } \varepsilon_T\text{–neighborhood (in } \ell^2\text{–norm) of} \qquad \frac{1}{N}\sum_{n=1}^{N} \overline{\overline{m}}\Big(\boldsymbol{\theta}_n, \Phi\big(\widehat{\sigma}_n\big)\Big)\,,$$

where

$$\varepsilon_T = R\sqrt{N_{\mathcal{B}}} \;\times\; \kappa_\Phi \sqrt{N_{\mathcal{I}} N_{\mathcal{H}}} \left(\sqrt{\frac{2N_{\mathcal{I}}}{\gamma L} \ln \frac{2N_{\mathcal{I}} N_{\mathcal{H}}}{\delta}} + \frac{1}{3}\frac{N_{\mathcal{I}}}{\gamma L} \ln \frac{2N_{\mathcal{I}} N_{\mathcal{H}}}{\delta} \right).$$

Step 5: Assertion (12)**.** Since $\mathcal{C}$ is $\overline{\overline{m}}$–robust approachable and by definition of the choices of the $\boldsymbol{\theta}_n$ in Figure 1, we get by (the proof of the sufficiency part of) Theorem 3.1, with probability 1,

$$\inf_{c\in\mathcal{C}} \left\| c - \frac{1}{N}\sum_{n=1}^{N} \overline{\overline{m}}\Big(\boldsymbol{\theta}_n, \Phi\big(\widehat{\sigma}_n\big)\Big) \right\|_2 \leqslant \frac{2R}{\sqrt{N}}\sqrt{N_{\mathcal{A}} N_{\mathcal{B}}} \leqslant \frac{2R}{\sqrt{T/L - 1}}\sqrt{N_{\mathcal{A}} N_{\mathcal{B}}}\,,$$

since $T/L \leqslant N + k/L \leqslant N + 1$.

Conclusion of the proof. The proof is concluded by putting the pieces together, thanks to a triangle inequality and by considering $L\delta/T \leqslant \delta/(N+1)$ instead of δ. □

6.1.4 Uniform guarantees over time for a time-adaptive version of the strategy of Figure 1. We present here a variant of the strategy of Figure 1 for which the lengths L_n of blocks n and the exploration rates γ_n are no longer constant. To do so, we need the following generalization of Theorem 2.2 to polynomial averages; this result is of independent interest. We only state the result for mixed actions taken and observed, but the generalization for pure actions follows easily.

Consider the setting of Theorem 2.2. The studied strategy relies on a parameter $\alpha \geqslant 0$. It plays an arbitrary $\boldsymbol{x}_1$. For $t \geqslant 1$, it forms at stage $t+1$ the vector-valued polynomial average

$$\widehat{m}_t^\alpha = \frac{1}{T_t^\alpha} \sum_{s=1}^{t} s^\alpha\, m(\boldsymbol{x}_s, \boldsymbol{y}_s) \qquad \text{where} \qquad T_t^\alpha = \sum_{s=1}^{t} s^\alpha\,,$$

computes its projection c_t^α onto $\mathcal{C}$, and resorts to a mixed action $\boldsymbol{x}_{t+1}$ solving the minimax equation

$$\min_{\boldsymbol{x}\in\Delta(\mathcal{A})} \max_{\boldsymbol{y}\in\Delta(\mathcal{B})} \big\langle \widehat{m}_t^\alpha - c_t^\alpha,\, m(\boldsymbol{x},\boldsymbol{y}) \big\rangle\,.$$

THEOREM 6.2 *We denote by M a bound in norm over m, i.e.,*

$$\max_{(a,b)\in\mathcal{A}\times\mathcal{B}} \big\| m(a,b) \big\|_2 \leqslant M\,.$$

For all $\alpha \geqslant 0$, when $\mathcal{C}$ is an approachable closed convex set, the above strategy ensures that for all strategies of the second player, with probability 1, for all $T \geqslant 1$,

$$\inf_{c\in\mathcal{C}} \left\| c - \frac{1}{\sum_{t=1}^{T} t^\alpha} \sum_{t=1}^{T} t^\alpha\, m(\boldsymbol{x}_t, \boldsymbol{y}_t) \right\|_2 \leqslant 2M \frac{\sqrt{\sum_{t=1}^{T} t^{2\alpha}}}{\sum_{t=1}^{T} t^\alpha} \leqslant \frac{2MK_\alpha}{\sqrt{T}}\,, \tag{14}$$

where K_α is a constant depending only α.

It is interesting to note that the convergence rate are independent of α and are the same as standard approachability ($1/\sqrt{T}$).

PROOF. The proof is a slight modification of the one of Theorem 2.2. We denote by d_t^α the squared distance of $\widehat{m}_t^\alpha$ to $\mathcal{C}$,

$$d_t^\alpha = \inf_{c\in\mathcal{C}} \big\| c - \widehat{m}_t^\alpha \big\|^2 = \big\| c_t^\alpha - \widehat{m}_t^\alpha \big\|^2$$

and use the shortcut notation $m_t = m(\boldsymbol{x}_t, \boldsymbol{y}_t)$ for all $t \geqslant 1$. Then,

$$\begin{aligned}
d_{t+1}^\alpha &\leqslant \left\| \widehat{m}_{t+1}^\alpha - c_t^\alpha \right\|^2 = \left\| \widehat{m}_t^\alpha - c_t^\alpha + \frac{(t+1)^\alpha}{T_{t+1}^\alpha} \left(m_{t+1} - \widehat{m}_t^\alpha \right) \right\|^2 \\
&\leqslant \left\| \widehat{m}_t^\alpha - c_t^\alpha \right\|^2 + \frac{2(t+1)^\alpha}{T_{t+1}^\alpha} \left\langle \widehat{m}_t^\alpha - c_t^\alpha,\, m_{t+1} - m_t^\alpha \right\rangle + \left(\frac{(t+1)^\alpha}{T_{t+1}^\alpha} \right)^2 \left\| m_{t+1} - \widehat{m}_t^\alpha \right\|^2 \\
&\leqslant d_t^\alpha + \frac{2(t+1)^\alpha}{T_{t+1}^\alpha} \underbrace{\left\langle \widehat{m}_t^\alpha - c_t^\alpha,\, m_{t+1} - c_t^\alpha \right\rangle}_{\leqslant 0} + \frac{2(t+1)^\alpha}{T_{t+1}^\alpha} \left\langle \widehat{m}_t^\alpha - c_t^\alpha,\, c_t^\alpha - m_t^\alpha \right\rangle + \left(\frac{(t+1)^\alpha}{T_{t+1}^\alpha} \right)^2 4M^2 \\
&\leqslant d_t^\alpha \left(1 - \frac{2(t+1)^\alpha}{T_{t+1}^\alpha} \right) + \left(\frac{(t+1)^\alpha}{T_{t+1}^\alpha} \right)^2 4M^2,
\end{aligned}$$

where we used in the third inequality the same convex projection inequality as in the proof of Theorem 2.2.

The first inequality in (14) then follows by induction: the bound $2M$ for $t = 1$ is by boundedness of m. If the stated bound holds for d_t^α, then

$$d_{t+1}^\alpha \leqslant \left(2M \frac{\sqrt{\sum_{s=1}^t s^{2\alpha}}}{\sum_{s=1}^t s^\alpha} \right)^2 \left(1 - \frac{2(t+1)^\alpha}{T_{t+1}^\alpha} \right) + \left(\frac{(t+1)^\alpha}{T_{t+1}^\alpha} \right)^2 4M^2 \leqslant 4M^2 \, \frac{\sum_{s=1}^{t+1} s^{2\alpha}}{\left(T_{t+1}^\alpha \right)^2},$$

as desired, since

$$\frac{1}{\left(T_t^\alpha \right)^2} \left(1 - \frac{2(t+1)^\alpha}{T_{t+1}^\alpha} \right) = \frac{1}{T_{t+1}^\alpha \left(T_t^\alpha \right)^2} \left(T_t^\alpha - (t+1)^\alpha \right) \leqslant \frac{1}{T_{t+1}^\alpha \left(T_t^\alpha \right)^2} \frac{\left(T_t^\alpha \right)^2 - (t+1)^{2\alpha}}{T_t^\alpha + (t+1)^\alpha} \leqslant \frac{1}{\left(T_{t+1}^\alpha \right)^2}.$$

The second inequality in (14) can be proved as follows. First, for all $\alpha \geqslant 0$, by comparing sums and integrals, we get that for all $t \geqslant 1$,

$$\frac{t^{\alpha+1}}{\alpha+1} = \int_0^t s^\alpha \, \mathrm{d}s \leqslant \sum_{s=1}^t s^\alpha \leqslant \int_1^{t+1} s^\alpha \, \mathrm{d}s \leqslant \frac{(t+1)^{\alpha+1}}{\alpha+1} \leqslant \frac{(2t)^{\alpha+1}}{\alpha+1}.$$

Therefore,

$$\frac{\sqrt{\sum_{s=1}^t s^{2\alpha}}}{\sum_{s=1}^t s^\alpha} \leqslant \frac{\alpha+1}{\sqrt{2\alpha+1}} \, \frac{\sqrt{(2t)^{\alpha+1}}}{t^{\alpha+1}} = K_\alpha \frac{1}{\sqrt{t}}$$

for

$$K_\alpha = \frac{\alpha+1}{\sqrt{2\alpha+1}} \sqrt{2^{\alpha+1}}.$$

This concludes the proof. □

The extension to polynomially weighted averages can also be obtained in the context of robust approachability as the key to Theorem 3.1 is Lemma 3.3, which indicates that to get robust approachability, it suffices to approach, in the usual sense, $\widetilde{\mathcal{C}}$; both can thus be performed with polynomially weighted averages.

Consider now the variant of the strategy of Figure 1 for which the length of the n-th block, denoted by L_n, is equal to n^α, the exploration rate on this block comes at a rate $\gamma_n = n^{-\alpha/3}$ and Ψ is an $\overline{\overline{m}}$-robust approachability strategy of $\mathcal{C}$ with respect to polynomially weighted averages with parameter $\alpha = 3/2$. We call it a time-adaptive version of this strategy; note that it does not depend anymore on any time horizon T, hence guarantees can be obtained for all T.

THEOREM 6.3 *The time-adaptive version of the strategy described in Figure 1 (with $L_n = n^\alpha$ and $\gamma_n = n^{-\alpha/3}$ for $\alpha = 3/2$) ensures that, for all $T \geqslant 1$, with probability at least $1 - \delta$,*

$$\inf_{c \in \mathcal{C}} \left\| c - \frac{1}{T} \sum_{t=1}^T r(I_t, J_t) \right\|_2 \leqslant \square \left(T^{-1/5} \sqrt{\ln \frac{T}{\delta}} + T^{-2/5} \ln \frac{T}{\delta} \right)$$

for some constant $\square$ depending only on $\mathcal{C}$ and the game (r, H) at hand.

PROOF. The proof follows closely the one of Theorem 6.1. We choose N so as to write $T = T_N^\alpha + k$ where $0 \leqslant k \leqslant L_{N+1} - 1$. We adapt step 1 as follows,

$$\left\| \frac{1}{T} \sum_{t=1}^{T} r(I_t, J_t) - \frac{1}{T_N^\alpha} \sum_{t=1}^{T_N^\alpha} r(I_t, J_t) \right\|_2 \leqslant R \left(\frac{k}{T} + \left(\frac{1}{T_N^\alpha} - \frac{1}{T} \right) T_N^\alpha \right) = \frac{2k}{T} R \leqslant \frac{2L_{N+1}}{T} R \,.$$

Second, as in step 2, we resort again to the Hoeffding-Azuma inequality for sums of Hilbert space-valued martingale differences; with probability at least $1 - \delta$,

$$\left\| \frac{1}{T_N^\alpha} \sum_{t=1}^{T_N^\alpha} r(I_t, J_t) - \frac{1}{T_N^\alpha} \sum_{n=1}^{N} n^\alpha \, r\big(\boldsymbol{p}_n, \widehat{\boldsymbol{q}}_n\big) \right\|_2 \leqslant 4R \sqrt{\frac{\ln(2/\delta)}{T_N^\alpha}} \leqslant 4R \sqrt{\frac{\ln(2/\delta)}{T}} \,.$$

In view of the choice $\gamma_n = n^{-\alpha/3}$, step 3 translates here to

$$\left\| \frac{1}{T_N^\alpha} \sum_{n=1}^{N} n^\alpha \, r\big(\boldsymbol{p}_n, \widehat{\boldsymbol{q}}_n\big) - \frac{1}{T_N^\alpha} \sum_{n=1}^{N} n^\alpha \, r\big(\boldsymbol{x}_n, \widehat{\boldsymbol{q}}_n\big) \right\|_2 \leqslant 2R \, \frac{\sum_{n=1}^N n^\alpha \gamma_n}{T_N^\alpha} = 2R \, \frac{\sum_{n=1}^N n^{2\alpha/3}}{T_N^\alpha} = 2R \frac{T_N^{(2\alpha/3)}}{T_N^\alpha} \,.$$

The same argument as the one at the beginning of the proof of Theorem 6.1 shows that

$$\frac{1}{T_N^\alpha} \sum_{n=1}^{N} n^\alpha \, r\big(\boldsymbol{x}_n, \widehat{\boldsymbol{q}}_n\big) \;\in\; \frac{1}{T_N^\alpha} \sum_{n=1}^{N} n^\alpha \, \overline{\overline{m}}\Big(\boldsymbol{\theta}_n, \, \Phi\Big(\widetilde{H}\big(\widehat{\boldsymbol{q}}_n\big)\Big)\Big).$$

Step 4 starts also by an application of Lemma 6.3 together with the Lipschitzness of Φ to get that for all regimes $n = 1, \dots, N$, with probability at least $1 - \delta$,

$$\left\| \Phi\big(\widehat{\sigma}_n\big) - \Phi\Big(\widetilde{H}\big(\widehat{\boldsymbol{q}}_n\big)\Big) \right\|_2 \leqslant \kappa_\Phi \sqrt{N_{\mathcal{I}} N_{\mathcal{H}}} \left(\sqrt{\frac{2N_{\mathcal{I}}}{\gamma_n L_n} \ln \frac{2N_{\mathcal{I}} N_{\mathcal{H}}}{\delta}} + \frac{1}{3} \frac{N_{\mathcal{I}}}{\gamma_n L_n} \ln \frac{2N_{\mathcal{I}} N_{\mathcal{H}}}{\delta} \right).$$

By a union bound, the above bound holds for all regimes $n = 1, \dots, N$ with probability at least $1 - N\delta$. Then, an application of Lemma 3.1 shows that

$$\frac{1}{T_N^\alpha} \sum_{n=1}^{N} n^\alpha \, \overline{\overline{m}}\Big(\boldsymbol{\theta}_n, \, \Phi\Big(\widetilde{H}\big(\widehat{\boldsymbol{q}}_n\big)\Big)\Big) \qquad \text{is in a } \varepsilon_N\text{–neighborhood of} \qquad \frac{1}{T_N^\alpha} \sum_{n=1}^{N} n^\alpha \, \overline{\overline{m}}\Big(\boldsymbol{\theta}_n, \, \Phi\big(\widehat{\sigma}_n\big)\Big) \,,$$

where, substituting the values of $L_n = n^\alpha$ and $\gamma_n = n^{-\alpha/3}$,

$$\begin{aligned} \varepsilon_N &= R\sqrt{N_{\mathcal{B}}} \;\times\; \kappa_\Phi \sqrt{N_{\mathcal{I}} N_{\mathcal{H}}} \; \frac{1}{T_N^\alpha} \sum_{n=1}^{N} n^\alpha \left(\sqrt{\frac{2N_{\mathcal{I}}}{\gamma_n L_n} \ln \frac{2N_{\mathcal{I}} N_{\mathcal{H}}}{\delta}} + \frac{1}{3} \frac{N_{\mathcal{I}}}{\gamma_n L_n} \ln \frac{2N_{\mathcal{I}} N_{\mathcal{H}}}{\delta} \right) \\ &= R\sqrt{N_{\mathcal{B}}} \;\times\; \kappa_\Phi \sqrt{N_{\mathcal{I}} N_{\mathcal{H}}} \left(\frac{T_N^{(2\alpha/3)}}{T_N^\alpha} \sqrt{2N_{\mathcal{I}} \ln \frac{2N_{\mathcal{I}} N_{\mathcal{H}}}{\delta}} + \frac{T_N^{(\alpha/3)}}{T_N^\alpha} \, \frac{N_{\mathcal{I}}}{3} \ln \frac{2N_{\mathcal{I}} N_{\mathcal{H}}}{\delta} \right). \end{aligned}$$

It then suffices, as in step 5 of the original proof, to write the convergence rates for robust approachability guaranteed by the strategy Ψ. By combining the result of Lemma 3.3 with Theorem 6.2 and Lemma 3.1, we get

$$\inf_{c \in \mathcal{C}} \left\| c - \frac{1}{T_n^\alpha} \sum_{n=1}^{N} n^\alpha \, \overline{\overline{m}}\Big(\boldsymbol{\theta}_n, \, \Phi\big(\widehat{\sigma}_n\big)\Big) \right\|_2 \leqslant \frac{2R \, K_\alpha}{\sqrt{N}} \sqrt{N_{\mathcal{A}} N_{\mathcal{B}}} \,.$$

Putting all things together and applying a union bound, we obtain that with probability at least $1 - \delta$,

$$\begin{aligned} &\inf_{c \in \mathcal{C}} \left\| c - \frac{1}{T} \sum_{t=1}^{T} r(I_t, J_t) \right\|_2 \\ &\qquad = O\left(\frac{(N+1)^\alpha}{T} + \sqrt{\frac{\ln(N/\delta)}{T_n^\alpha}} + \frac{T_N^{(2\alpha/3)}}{T_N^\alpha} + \frac{T_N^{(2\alpha/3)}}{T_N^\alpha} \sqrt{\ln \frac{N}{\delta}} + \frac{T_N^{(\alpha/3)}}{T_N^\alpha} \ln \frac{N}{\delta} + \frac{1}{\sqrt{N}} \right). \end{aligned}$$

Since (as proved at the end of Theorem 6.2) $T_N^\beta \sim N^{\beta+1}/(\beta+1)$ for all $\beta \geqslant 0$, we get that

$$N \sim \big((\alpha+1)T\big)^{1/(\alpha+1)} \qquad \text{and} \qquad T_N^\beta \sim \frac{N^{\beta+1}}{\beta+1} \sim \kappa_{\alpha,\beta} \, T^{(\beta+1)/(\alpha+1)} \,,$$

where $\kappa_{\alpha,\beta}$ is a constant that only depends on α and β. Choosing $\alpha = 3/2$ and substituting these equivalences ensures the result. □

6.2 Application to regret minimization. In this section we analyze external and internal regret minimization in repeated games with partial monitoring from the approachability perspective. We show how to—in particular—efficiently minimize regret in both setups using the results developed for vector-valued games with partial monitoring; to do so, we indicate why the assumption of bi-piecewise linearity (Assumption 6.1) is satisfied.

6.2.1 External regret. We consider in this section the framework and aim introduced by Rustichini [29] and studied, sometimes in special cases, by Piccolboni and Schindelhauer [26], Mannor and Shimkin [17], Cesa-Bianchi et al. [6], Lugosi et al. [16]. We show that our general strategy can be used for regret minimization.

Scalar payoffs are obtained (but not observed) by the first player, i.e., $d = 1$: the payoff function r is a mapping $\mathcal{I} \times \mathcal{J} \to \mathbb{R}$; we still denote by R a bound on $|r|$. We define in this section

$$\widehat{\boldsymbol{q}}_T = \frac{1}{T} \sum_{t=1}^{T} \delta_{J_T}$$

as the empirical distribution of the actions taken by the second player during the first T rounds. (This is in contrast with the notation $\widehat{\boldsymbol{q}}_T$ used in the previous section to denote such an empirical distribution, but only taken within regime n.)

The external regret of the first player at round T equals by definition

$$R_T^{\mathrm{ext}} = \max_{\boldsymbol{p} \in \Delta(\mathcal{I})} \rho\Big(\boldsymbol{p}, \widetilde{H}\big(\widehat{\boldsymbol{q}}_T\big)\Big) - \frac{1}{T} \sum_{t=1}^{T} r(I_t, J_t)\,,$$

where $\rho : \Delta(\mathcal{I}) \times \mathcal{F}$ is defined as follows: for all $\boldsymbol{p} \in \Delta(\mathcal{I})$ and $\sigma \in \mathcal{F}$,

$$\rho(\boldsymbol{p}, \sigma) = \min \Big\{ r(\boldsymbol{p}, \boldsymbol{q}) \,:\, \boldsymbol{q} \text{ such that } \widetilde{H}(\boldsymbol{q}) = \sigma \Big\}\,.$$

The function ρ is continuous in its first argument and therefore the supremum in the defining expression of R_T^{ext} is a maximum.

We recall briefly why, intuitively, this is the natural notion of external regret to consider in this case. Indeed, the first term in the definition of R_T^{ext} is (close to) the worst-case average payoff obtained by the first player when playing consistently a mixed action $\boldsymbol{p}$ against a sequence of mixed actions inducing on average the same laws on the signals as the sequence of actions actually played.

The following result is an easy consequence of Theorem 6.3, as is explained below; it corresponds to the main result of Lugosi et al. [16], with the same convergence rate but with a different strategy. (However, Section 2.3 of Perchet [24] exhibited an efficient strategy achieving a convergence rate of order $T^{-1/3}$, which is optimal; a question that remains open is thus whether the rates exhibited in Theorem 6.3 could be improved.)

COROLLARY 6.1 *The first player has a strategy such that for all T and all strategies of the second player, with probability at least $1 - \delta$,*

$$R_T^{\mathrm{ext}} \;\leqslant\; \square \left(T^{-1/5} \sqrt{\ln \frac{T}{\delta}} + T^{-2/5} \ln \frac{T}{\delta} \right)$$

for some constant $\square$ depending only on the game (r, H) at hand.

The proof below is an extension to the setting of partial monitoring of the original proof and strategy of Blackwell [3] for the case of external regret under full monitoring: in the latter case the vector-payoff function $\underline{r}$ and the set $\mathcal{C}$ considered in our proof are equal to the ones considered by Blackwell.

PROOF. We embed $\mathcal{F}$ into $\mathbb{R}^{\mathcal{I} \times \mathcal{H}}$ so that in this proof we will be working in the vector space $\mathbb{R}^d = \mathbb{R} \times \mathbb{R}^{\mathcal{I} \times \mathcal{H}}$. We consider the closed convex set $\mathcal{C}$ and the vector-valued payoff function $\underline{r}$ respectively defined by

$$\mathcal{C} = \left\{ (z, \sigma) \in \mathbb{R} \times \mathcal{F} : \quad z \geqslant \max_{\boldsymbol{p} \in \Delta(\mathcal{I})} \rho(\boldsymbol{p},\, \sigma) \right\} \qquad \text{and} \qquad \underline{r}(i,j) = \begin{bmatrix} r(i,j) \\ \widetilde{H}(\delta_j) \end{bmatrix},$$

for all $(i,j) \in \mathcal{I} \times \mathcal{J}$.

We first show that Condition (APM) is satisfied for the considered convex set $\mathcal{C}$ and game $(\underline{r}, H)$. To do so, by continuity of ρ in its first argument, we associate with each $\boldsymbol{q} \in \Delta(\mathcal{J})$ an element $\phi(\boldsymbol{q}) \in \Delta(\mathcal{I})$ such that

$$\phi(\boldsymbol{q}) \in \operatorname*{argmax}_{\boldsymbol{p} \in \Delta(\mathcal{I})} \rho\big(\boldsymbol{p}, \widetilde{H}(\boldsymbol{q})\big) \,.$$

Then, given any $\boldsymbol{q} \in \Delta(\mathcal{J})$, we note that for all $\boldsymbol{q}'$ satisfying $\widetilde{H}(\boldsymbol{q}') = \widetilde{H}(\boldsymbol{q})$, we have by definition of ρ,

$$r\big(\phi(\boldsymbol{q}), \boldsymbol{q}'\big) \geqslant \rho\big(\phi(\boldsymbol{q}), \widetilde{H}(\boldsymbol{q}')\big) = \max_{\boldsymbol{p} \in \Delta(\mathcal{I})} \rho\big(\boldsymbol{p}, \widetilde{H}(\boldsymbol{q}')\big) \,,$$

which shows that $\underline{r}\big(\phi(\boldsymbol{q}), \boldsymbol{q}'\big) \in \mathcal{C}$. The required condition is thus satisfied.

We then show that Assumption 6.1 is satisfied. To do so, we will actually prove the stronger property that the mappings $\overline{m}(\,\cdot\,, \sigma)$ are piecewise linear for all $\sigma \in \mathcal{F}$; we fix such a σ in the sequel. Only the first coordinate r of $\underline{r}$ depends on $\boldsymbol{p}$, so the desired property is true if and only if the mapping $\overline{m}_1(\,\cdot\,, \sigma)$ defined by

$$\boldsymbol{p} \in \Delta(\mathcal{I}) \;\longmapsto\; \overline{m}_1(\boldsymbol{p}, \sigma) = \Big\{ r(\boldsymbol{p}, \boldsymbol{q}) : \quad \boldsymbol{q} \in \Delta(\mathcal{J}) \text{ such that } \widetilde{H}(\boldsymbol{q}) = \sigma \Big\}$$

is piecewise linear. Since $\widetilde{H}$ is linear, the set

$$\Big\{ \boldsymbol{q} \in \Delta(\mathcal{J}) \text{ such that } \widetilde{H}(\boldsymbol{q}) = \sigma \Big\}$$

is a polytope, thus, the convex hull of some finite set $\{\boldsymbol{q}_{\sigma,1}, \ldots, \boldsymbol{q}_{\sigma,M}\} \subset \Delta(\mathcal{J})$. Therefore, for every $\boldsymbol{p} \in \Delta(\mathcal{I})$, by linearity of r (and by the fact that it takes one-dimensional values),

$$\overline{m}_1(\boldsymbol{p}, \sigma) = \operatorname{co} \Big\{ r(\boldsymbol{p}, \boldsymbol{q}_{\sigma,1}), \ldots, r(\boldsymbol{p}, \boldsymbol{q}_{\sigma,M}) \Big\} = \Big[\min_{k \in \{1,..,M\}} r(\boldsymbol{p}, \boldsymbol{q}_{\sigma,k}) \,, \;\; \max_{k' \in \{1,..,M\}} r(\boldsymbol{p}, \boldsymbol{q}_{\sigma,k'}) \Big] \,, \tag{15}$$

where co stands for the convex hull. Since all applications $r(\,\cdot\,, \boldsymbol{q}_{\sigma,k})$ are linear, their minimum and their maximum are piecewise linear functions, thus $\overline{m}_1(\,\cdot\,, \sigma)$ is also piecewise linear. Assumption 6.1 is thus satisfied, as claimed.

Theorem 6.1 can therefore be applied to exhibit the convergence rates; we simply need to relate the quantity of interest here to the one considered therein. To that end we use the fact that the mapping

$$\sigma \in \mathcal{F} \longmapsto \max_{\boldsymbol{p} \in \Delta(\mathcal{I})} \rho(\boldsymbol{p}, \sigma)$$

is Lipschitz, with Lipschitz constant in ℓ^2–norm denoted by L_ρ; the proof of this fact is detailed below.

Now, the regret is non positive as soon as $\sum_{t=1}^T \underline{r}(I_t, J_t)/T$ belongs to $\mathcal{C}$; we therefore only need to consider the case when this average is not in $\mathcal{C}$. In the latter case, we denote by $(\widetilde{r}_T, \widetilde{\sigma}_T)$ its projection in ℓ^2–norm onto $\mathcal{C}$. We have first that the defining inequality of $\mathcal{C}$ is an equality on its border, so that

$$\widetilde{r}_T = \max_{\boldsymbol{p} \in \Delta(\mathcal{I})} \; \rho\big(\boldsymbol{p}, \widetilde{\sigma}_T\big) \,;$$

and second, that

$$\begin{aligned}
R_T^{\mathrm{ext}} &= \max_{\boldsymbol{p} \in \Delta(\mathcal{I})} \; \rho\Big(\boldsymbol{p}, \widetilde{H}\big(\widehat{\boldsymbol{q}}_T\big)\Big) - \frac{1}{T} \sum_{t=1}^T r(I_t, J_t) \\
&\leqslant \left| \max_{\boldsymbol{p} \in \Delta(\mathcal{I})} \; \rho\Big(\boldsymbol{p}, \widetilde{H}\big(\widehat{\boldsymbol{q}}_T\big)\Big) - \max_{\boldsymbol{p} \in \Delta(\mathcal{I})} \; \rho\big(\boldsymbol{p}, \widetilde{\sigma}_T\big) \right| + \left| \widetilde{r}_T - \frac{1}{T} \sum_{t=1}^T r(I_t, J_t) \right| \\
&\leqslant L_\rho \left\| \widetilde{\sigma}_T - \widetilde{H}\big(\widehat{\boldsymbol{q}}_T\big) \right\|_2 + \left| \widetilde{r}_T - \frac{1}{T} \sum_{t=1}^T r(I_t, J_t) \right| \\
&\leqslant \sqrt{2} \max\big\{L_\rho, 1\big\} \left\| \begin{bmatrix} \widetilde{r}_T \\ \widetilde{\sigma}_T \end{bmatrix} - \frac{1}{T} \sum_{t=1}^T \underline{r}(I_t, J_t) \right\|_2 \\
&= \sqrt{2} \max\big\{L_\rho, 1\big\} \inf_{c \in \mathcal{C}} \left\| c - \frac{1}{T} \sum_{t=1}^T \underline{r}(I_t, J_t) \right\|_2 .
\end{aligned}$$

The claimed rates are now seen to follow from the ones indicated in Theorem 6.3.

It only remains to prove the indicated Lipschitzness. (All Lipschitzness statements that follow will be with respect to the ℓ^2–norms.) We have by Definition 6.1 that for all $\boldsymbol{p} \in \Delta(\mathcal{I})$ and $\sigma \in \mathcal{F}$,

$$\rho(\boldsymbol{p}, \sigma) = \min \, \overline{\overline{m}}_1\big(\boldsymbol{p}, \Phi(\sigma)\big) \,,$$

where the linear $\overline{\overline{m}}_1$ is indifferently either relative to $\overline{m}_1$ or is the projection onto the first component of the function $\overline{\overline{m}}$ relative to $\overline{m}$. By Remark 6.1 the mapping $\sigma \in \mathcal{F} \mapsto \Phi(\sigma)$ is κ_Φ–Lipschitz; this entails, by Lemma 3.1, that for all $\boldsymbol{p} \in \Delta(\mathcal{I})$, the mapping $\sigma \in \mathcal{F} \mapsto \rho(\boldsymbol{p}, \sigma)$ is $R\sqrt{N_{\mathcal{B}}}\ \kappa_\Phi$–Lipschitz. In particular, since the latter Lipschitz constant is independent of $\boldsymbol{p}$, the mapping

$$\sigma \in \mathcal{F} \ \longmapsto \ \max_{\boldsymbol{p} \in \Delta(\mathcal{I})} \rho(\boldsymbol{p}, \sigma)$$

is $R\sqrt{N_{\mathcal{B}}}\ \kappa_\Phi$–Lipschitz as well, which concludes the proof. □

A similar argument to the one in Perchet [24] shows that the convex set $\mathcal{C}$ is defined by a finite number of piecewise linear equations, it is therefore a polyhedron so the projection onto it, and as well the computation of the strategy, can be done efficiently. We sketch the argument below, and refer the reader to Perchet [24] for details. Equation (15) indicates a priori that for each $\sigma \in \mathcal{F}$, there exist a finite number M_σ (depending on σ) of mixed actions $\boldsymbol{q}_{\sigma,1}, \ldots, \boldsymbol{q}_{\sigma,M_\sigma}$ such that for all $\boldsymbol{p} \in \Delta(\mathcal{I})$, we have $\rho(\boldsymbol{p}, \sigma) = \min\big\{r(\boldsymbol{p}, \boldsymbol{q}_{\sigma,1}), \ldots, r(\boldsymbol{p}, \boldsymbol{q}_{\sigma,M_\sigma})\big\}$. But by an argument stated in Perchet [24],

$$\sigma \ \longmapsto \ \Big\{ \boldsymbol{q} \in \Delta(\mathcal{J}) \text{ such that } \widetilde{H}(\boldsymbol{q}) = \sigma \Big\}$$

evolves in a piecewise linear way and thus there exist a finite number M of piecewise linear functions $\sigma \mapsto q'_{\sigma,k}$, with $k = 1, \ldots, M$, such that, for all $\sigma \in \mathcal{F}$,

$$\big\{\boldsymbol{q}_{\sigma,1}, \ldots, \boldsymbol{q}_{\sigma,M_\sigma}\big\} = \big\{\boldsymbol{q}'_{\sigma,1}, \ldots, \boldsymbol{q}'_{\sigma,M}\big\} \,.$$

(There can be some redundancies between the $\boldsymbol{q}'_{\sigma,k}$.) Because of this, we have that for all $\boldsymbol{p} \in \Delta(\mathcal{I})$ and $\sigma \in \mathcal{F}$,

$$\rho(\boldsymbol{p}, \sigma) = \min\big\{r(\boldsymbol{p}, \boldsymbol{q}'_{\sigma,1}), \ldots, r(\boldsymbol{p}, \boldsymbol{q}'_{\sigma,M})\big\} \,.$$

Each function $\sigma \mapsto \boldsymbol{q}'_{\sigma,k}$ being piecewise linear, one can construct a finite set $\{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_K\} \subset \Delta(\mathcal{I})$ such that, for any $\sigma \in \mathcal{F}$, the mapping $\boldsymbol{p} \mapsto \rho(\boldsymbol{p}, \sigma)$ is maximized at one of these $\boldsymbol{p}_k$. The convex set $\mathcal{C}$ is therefore defined by a finite number of piecewise linear equations, it is therefore a polyhedron; therefore the projection onto it, hence the computation of the proposed strategy, can be done efficiently.

6.2.2 Internal / swap regret. Foster and Vohra [10] defined internal regret with full monitoring as follows. A player has no internal regret if, for every action $i \in \mathcal{I}$, he has no external regret on the stages when this specific action i was played. In other words, i is the best response to the empirical distribution of action of the other player on these stages.

With partial monitoring, the first player evaluates his payoffs in a pessimistic way through the function ρ defined above. This function is not linear over $\Delta(\mathcal{I})$ in general (it is concave), so that the best responses are not necessarily pure actions $i \in \mathcal{I}$ but mixed actions, i.e., elements of $\Delta(\mathcal{I})$. Following Lehrer and Solan [15] one therefore can partition the stages not depending on the pure actions actually played but on the mixed actions $\boldsymbol{p}_t \in \Delta(\mathcal{I})$ used to draw them. To this end, it is convenient to assume that the strategies of the first player need to pick these mixed actions in a finite (but possibly thin) grid of $\Delta(\mathcal{I})$, which we denote by $\big\{\boldsymbol{p}_g,\ g \in \mathcal{G}\big\}$, where $\mathcal{G}$ is a finite set. At each round t, the first player picks an index $G_t \in \mathcal{G}$ and uses the distribution $\boldsymbol{p}_{G_t}$ to draw his action I_t. Up to a standard concentration-of-the-measure argument, we will measure the payoff at round t with $r\big(\boldsymbol{p}_{G_t}, J_t\big)$ rather than with $r(I_t, J_t)$.

For each $g \in \mathcal{G}$, we denote by $N_T(g)$ the number of stages in $\{1, \ldots, T\}$ for which we had $G_t = g$ and, whenever $N_T(g) > 0$,

$$\widehat{\boldsymbol{q}}_{T,g} = \frac{1}{N_T(g)} \sum_{t : G_t = g} \delta_{J_t} \,.$$

We define $\widehat{\boldsymbol{q}}_{T,g}$ is an arbitrary way when $N_T(g) = 0$. The internal regret of the first player at round T is measured as

$$R_T^{\text{int}} = \max_{g, g' \in \mathcal{G}} \frac{N_T(g)}{T} \Big(\rho\Big(\boldsymbol{p}_{g'}, \widetilde{H}\big(\widehat{\boldsymbol{q}}_{T,g}\big)\Big) - r\big(\boldsymbol{p}_g, \widehat{\boldsymbol{q}}_{T,g}\big) \Big) \,.$$

Actually, our proof technique rather leads to the minimization of some swap regret (see Blum and Mansour [4] for the definition of swap regret in full monitoring):

$$R_T^{\text{swap}} = \sum_{g\in\mathcal{G}} \frac{N_T(g)}{T} \left(\max_{g'\in\mathcal{G}} \rho\Big(\boldsymbol{p}_{g'}, \widetilde{H}\big(\widehat{\boldsymbol{q}}_{T,g}\big)\Big) - r\big(\boldsymbol{p}_g, \widehat{\boldsymbol{q}}_{T,g}\big) \right)_+ .$$

Again, the following bound on the swap regret easily follows from Theorem 6.1; the latter constructs a simple and direct strategy to control the swap regret, thus also the internal regret. It therefore improves the results of Lehrer and Solan [15] and Perchet [22], two articles that presented more involved and less efficient strategies to do so (strategies based on auxiliary strategies using grids that need to be refined over time and whose complexities is exponential in the size of these grids; ideas all in all similar to what is done in calibration, see the references provided in Section 4). Moreover, we provide convergence rates.

COROLLARY 6.2 *The first player has an explicit strategy such that for all T and all strategies of the second player, with probability at least $1-\delta$,*

$$R_T^{\text{swap}} \leqslant \square \left(T^{-1/5} \sqrt{\ln \frac{T}{\delta}} + T^{-2/5} \ln \frac{T}{\delta} \right)$$

for some constant $\square$ depending only on the game (r, H) at hand and on the size of the finite grid $\mathcal{G}$.

PROOF. The proof of this corollary is based on ideas similar to the ones used in the proof of Corollary 6.1; $\mathcal{G}$ will play the role of the action set of the first player. The proof proceeds in four steps. In the first step, we construct an approachability setup and show that Condition (APM) applies. In the second step, we show that Assumption 6.1 is satisfied. In the third step we analyze the convergence rates of the swap regret. In the fourth and final step, we show that the set we are approaching possess some smoothness properties by providing a uniform Lipschitz bound on certain functions.

Step 1: We denote by

$$\mathcal{F}_{\text{cone}} = \big\{ \lambda\sigma, \quad \sigma\in\mathcal{F}, \ \lambda\in\mathbb{R}_+ \big\}$$

the cone generated by $\mathcal{F}$ and extend linearly $\rho : \Delta(\mathcal{I}) \times \mathcal{F} \to \mathbb{R}$ into a mapping $\rho : \Delta(\mathcal{I}) \times \mathcal{F}_{\text{cone}} \to R$ as follows: for all $\boldsymbol{p}\in\Delta(\mathcal{I})$, for all $\lambda \geqslant 0$ with $\lambda \neq 1$, and all $\sigma\in\mathcal{F}$,

$$\rho(\boldsymbol{p}, \lambda\sigma) = \begin{cases} 0 & \text{if } \lambda = 0, \\ \lambda\, \rho(\boldsymbol{p}, \sigma) & \text{if } \lambda > 0. \end{cases}$$

In the sequel, we embed $\mathcal{F}_{\text{cone}}$ into $\mathbb{R}^{\mathcal{I}\times\mathcal{H}}$.

The closed convex set $\mathcal{C}$ and the vector-valued payoff function $\underline{r}$ are then respectively defined by

$$\mathcal{C} = \left\{ (z_g, \boldsymbol{v}_g)_{g\in\mathcal{G}} \in \big(\mathbb{R}\times\mathcal{F}_{\text{cone}}\big)^{\mathcal{G}} : \quad \forall\, g\in\mathcal{G}, \quad z_g \geqslant \max_{g'\in\mathcal{G}} \rho\big(\boldsymbol{p}_{g'}, \boldsymbol{v}_g\big) \right\}$$

and, for all $(g,j)\in\mathcal{G}\times\mathcal{J}$,

$$\underline{r}(g,j) = \left[\begin{array}{c} r\big(\boldsymbol{p}_g, j\big)\, \mathbb{I}_{\{g'=g\}} \\ \widetilde{H}(\delta_j)\, \mathbb{I}_{\{g'=g\}} \end{array} \right]_{g'\in\mathcal{G}} .$$

To show that $\mathcal{C}$ is $\underline{r}$–approachable, we associate with each $\boldsymbol{q}\in\Delta(\mathcal{J})$ an element $g^\star(\boldsymbol{q})\in\mathcal{G}$ such that

$$g^\star(\boldsymbol{q}) \in \operatorname*{argmax}_{g\in\mathcal{G}} \rho\big(\boldsymbol{p}_g, \widetilde{H}(\boldsymbol{q})\big) .$$

Then, given any $\boldsymbol{q}\in\Delta(\mathcal{J})$, we note that for all $\boldsymbol{q}'$ satisfying $\widetilde{H}(\boldsymbol{q}') = \widetilde{H}(\boldsymbol{q})$, the components of the vector $\underline{r}\big(g^\star(\boldsymbol{q}), \boldsymbol{q}'\big)$ are all null but the ones corresponding to $g^\star(\boldsymbol{q})$, for which we have

$$r\big(\boldsymbol{p}_{g^\star(\boldsymbol{q})}, \boldsymbol{q}'\big) \geqslant \rho\Big(\boldsymbol{p}_{g^\star(\boldsymbol{q})}, \widetilde{H}\big(\boldsymbol{q}'\big)\Big) = \rho\Big(\boldsymbol{p}_{g^\star(\boldsymbol{q})}, \widetilde{H}\big(\boldsymbol{q}\big)\Big) = \max_{g'\in\mathcal{G}} \rho\Big(\boldsymbol{p}_{g'}, \widetilde{H}\big(\boldsymbol{q}\big)\Big) = \max_{g'\in\mathcal{G}} \rho\Big(\boldsymbol{p}_{g'}, \widetilde{H}\big(\boldsymbol{q}'\big)\Big) ,$$

where the first inequality is by definition of ρ. Therefore, $\underline{r}\big(g^\star(\boldsymbol{q}), \boldsymbol{q}'\big) \in \mathcal{C}$. Condition (APM) in Lemma 6.2 and Theorem 6.1 is thus satisfied, so that we have approachability.

Step 2: We then show that Assumption 6.1 is satisfied. It suffices to show that for all $\sigma \in \mathcal{F}$, the application

$$\pi = (\pi_g)_{g\in\mathcal{G}} \in \Delta(\mathcal{G}) \;\longmapsto\; \overline{m}_1(\pi,\sigma) = \Big\{ \big(\pi_g\, r(\boldsymbol{p}_g,\boldsymbol{q})\big)_{g\in\mathcal{G}} : \quad \boldsymbol{q} \in \Delta(\mathcal{J}) \text{ such that } \widetilde{H}(\boldsymbol{q}) = \sigma \Big\}$$

is piecewise linear (as the other components in the definition of $\overline{m}$ are linear in π). This is the case since for each g, the application

$$\pi \in \Delta(\mathcal{G}) \;\longmapsto\; \Big\{ \pi_g\, r(\boldsymbol{p}_g,\boldsymbol{q}) : \quad \boldsymbol{q} \in \Delta(\mathcal{J}) \text{ such that } \widetilde{H}(\boldsymbol{q}) = \sigma \Big\}$$

is seen to be piecewise linear, by using the same one-dimensional argument as in the proof of Corollary 6.1.

Step 3: We now exhibit the convergence rates. In view of the form of the defining set of constraints for $\mathcal{C}$, the coordinates of the elements in $\mathcal{C}$ can be grouped according to each $g \in \mathcal{G}$ and projections onto $\mathcal{C}$ can therefore be done separately for each such group. The group g of coordinates of $\sum_{t=1}^T \underline{r}(G_t,J_t)/T$ is formed by

$$\frac{N_T(g)}{T}\, r\big(\boldsymbol{p}_g, \widehat{\boldsymbol{q}}_{T,g}\big) \qquad \text{and} \qquad \frac{N_T(g)}{T}\, \widetilde{H}\big(\widehat{\boldsymbol{q}}_{T,g}\big)\,;$$

when

$$\frac{N_T(g)}{T}\, r\big(\boldsymbol{p}_g, \widehat{\boldsymbol{q}}_{T,g}\big) \geqslant \max_{g'\in\mathcal{G}}\; \rho\bigg(\boldsymbol{p}_{g'},\, \frac{N_T(g)}{T}\, \widetilde{H}\big(\widehat{\boldsymbol{q}}_{T,g}\big)\bigg)\,,$$

we denote these quantities by $\widetilde{r}_{T,g}$ and $\widetilde{\boldsymbol{v}}_{T,g}$. Otherwise, we project this pair on the set

$$\mathcal{C}_g = \Big\{ (z_g,\boldsymbol{v}_g) \in \mathbb{R}\times\mathcal{F}_{\mathrm{cone}} : \quad z_g \geqslant \max_{g'\in\mathcal{G}} \rho\big(\boldsymbol{p}_{g'},\boldsymbol{v}_g\big) \Big\}$$

and denote by $\widetilde{r}_{T,g}$ and $\widetilde{\boldsymbol{v}}_{T,g}$ the coordinates of the projection; they satisfy the defining inequality of $\mathcal{C}_g$ with equality,

$$\widetilde{r}_{T,g} = \max_{g'\in\mathcal{G}} \rho\big(\boldsymbol{p}_{g'},\widetilde{\boldsymbol{v}}_g\big)\,.$$

By distinguishing for each g according to which of the two cases above arose (for the first inequality), we may decompose and upper bound the swap regret as follows,

$$\begin{aligned}
R_T^{\mathrm{swap}} &= \sum_{g\in\mathcal{G}} \frac{N_T(g)}{T} \bigg(\max_{g'\in\mathcal{G}} \rho\Big(\boldsymbol{p}_{g'}, \widetilde{H}\big(\widehat{\boldsymbol{q}}_{T,g}\big)\Big) - r\big(\boldsymbol{p}_g,\widehat{\boldsymbol{q}}_{T,g}\big) \bigg)_+ \\
&= \sum_{g\in\mathcal{G}} \bigg(\max_{g'\in\mathcal{G}} \rho\bigg(\boldsymbol{p}_{g'},\, \frac{N_T(g)}{T}\, \widetilde{H}\big(\widehat{\boldsymbol{q}}_{T,g}\big)\bigg) - \frac{N_T(g)}{T}\, r\big(\boldsymbol{p}_g,\widehat{\boldsymbol{q}}_{T,g}\big) \bigg)_+ \\
&\leqslant \sum_{g\in\mathcal{G}} \bigg| \max_{g'\in\mathcal{G}} \rho\bigg(\boldsymbol{p}_{g'},\, \frac{N_T(g)}{T}\, \widetilde{H}\big(\widehat{\boldsymbol{q}}_{T,g}\big)\bigg) - \max_{g'\in\mathcal{G}} \rho\big(\boldsymbol{p}_{g'},\widetilde{\boldsymbol{v}}_{g,T}\big) \bigg| + \sum_{g\in\mathcal{G}} \bigg| \widetilde{r}_{T,g} - \frac{N_T(g)}{T}\, r\big(\boldsymbol{p}_g,\widehat{\boldsymbol{q}}_{T,g}\big) \bigg| \\
&\leqslant \sum_{g\in\mathcal{G}} \overline{L}_\rho \bigg\| \frac{N_T(g)}{T}\, \widetilde{H}\big(\widehat{\boldsymbol{q}}_{T,g}\big) - \widetilde{\boldsymbol{v}}_{g,T} \bigg\|_2 + \sum_{g\in\mathcal{G}} \bigg| \widetilde{r}_{T,g} - \frac{N_T(g)}{T}\, r\big(\boldsymbol{p}_g,\widehat{\boldsymbol{q}}_{T,g}\big) \bigg|\,,
\end{aligned}$$

where we used a fact proved below, that the application

$$\boldsymbol{v} \in \mathcal{F}_{\mathrm{cone}} \longmapsto \max_{g'\in\mathcal{G}} \rho\big(\boldsymbol{p}_{g'},\boldsymbol{v}\big)$$

is $\overline{L}_\rho$–Lipschitz. In the last inequality we had a sum of ℓ^2–norms, which can be bounded by a single ℓ^2–norm,

$$\begin{aligned}
R_T^{\mathrm{swap}} &\leqslant \max\big\{\overline{L}_\rho, 1\big\} \sqrt{2N_{\mathcal{G}}} \left\| \begin{bmatrix} \widetilde{r}_{T,g} \\ \widetilde{\boldsymbol{v}}_{T,g} \end{bmatrix}_{g\in\mathcal{G}} - \frac{1}{T}\sum_{t=1}^T \underline{r}(I_t,J_t) \right\|_2 \\
&\leqslant \max\big\{\overline{L}_\rho, 1\big\} \sqrt{2N_{\mathcal{G}}} \inf_{c\in\mathcal{C}} \left\| c - \frac{1}{T}\sum_{t=1}^T \underline{r}(I_t,J_t) \right\|_2\,,
\end{aligned}$$

where we denoted by $N_{\mathcal{G}}$ the cardinality of $\mathcal{G}$. Resorting to the convergence rate stated in Theorem 6.3 concludes the proof, up to the claimed Lipschitzness, which we now prove. (All Lipschitzness statements that follow will be with respect to the ℓ^2–norms.)

Step 4: To do so, it suffices to show that for all fixed elements $\boldsymbol{p} \in \Delta(\mathcal{I})$, the functions $\boldsymbol{v} \in \mathcal{F}_{\text{cone}} \mapsto \rho(\boldsymbol{p}, \boldsymbol{v})$ are Lipschitz, with a Lipschitz constant $\overline{L}_\rho$ that is independent of $\boldsymbol{p}$. Note that we already proved at the end of the proof of Corollary 6.1 that $\sigma \in \mathcal{F} \mapsto \rho(\boldsymbol{p}, \sigma)$ is Lipschitz, with a Lipschitz constant L_ρ independent of $\boldsymbol{p}$. Consider now two elements $\boldsymbol{v}$, $\boldsymbol{v}' \in \mathcal{F}_{\text{cone}}$, which we write as $\boldsymbol{v} = \lambda\sigma$ and $\boldsymbol{v}' = \lambda'\sigma'$, with σ, $\sigma' \in \mathcal{F}$ and λ, $\lambda' \in \mathbb{R}_+$. Using triangle inequalities, the Lipschitzness of ρ on $\mathcal{F}$, and the fact that r thus ρ are bounded by R,

$$\begin{aligned}
\bigl|\rho(\boldsymbol{p}, \lambda\sigma) - \rho(\boldsymbol{p}, \lambda'\sigma')\bigr| &\leqslant \bigl|\lambda\bigl(\rho(\boldsymbol{p}, \sigma) - \rho(\boldsymbol{p}, \sigma')\bigr)\bigr| + \bigl|(\lambda - \lambda')\rho(\boldsymbol{p}, \sigma')\bigr| \\
&\leqslant \lambda\, L_\rho \bigl\|\, \sigma - \sigma' \,\bigr\|_2 + R\,\bigl|\lambda - \lambda'\bigr| \\
&\leqslant L_\rho \bigl\|\, \lambda\sigma - \lambda'\sigma' + (\lambda' - \lambda)\sigma' \,\bigr\|_2 + R\,\bigl|\lambda - \lambda'\bigr| \\
&\leqslant L_\rho \bigl\|\, \lambda\sigma - \lambda'\sigma' \,\bigr\|_2 + \bigl(R + L_\rho N_{\mathcal{I}}\bigr)\,\bigl|\lambda - \lambda'\bigr|\,,
\end{aligned}$$

where we used also for the last inequality that since σ is a vector of $N_{\mathcal{I}}$ probability distributions over the signals, $\|\sigma\|_2 \leqslant \|\sigma\|_1 = N_{\mathcal{I}}$. To conclude the argument, we simply need to show that $\bigl|\lambda - \lambda'\bigr|$ can be bounded by $\bigl\|\, \lambda\sigma - \lambda'\sigma' \,\bigr\|_2$ up to some universal constant, which we do now. We resort again to the fact that $\|\sigma\|_1 = \|\sigma'\|_1 = N_{\mathcal{I}}$ and can thus write, thanks to a triangle inequality and assuming with no loss of generality that $\lambda' < \lambda$, that

$$\bigl|\lambda - \lambda'\bigr| = \frac{1}{N_{\mathcal{I}}}\Bigl(\lambda \|\sigma\|_1 - \lambda' \|\sigma'\|_1\Bigr) \leqslant \frac{1}{N_{\mathcal{I}}} \|\lambda\sigma - \lambda'\sigma'\|_1 \leqslant \frac{\sqrt{N_{\mathcal{H}} N_{\mathcal{I}}}}{N_{\mathcal{I}}} \|\lambda\sigma - \lambda'\sigma'\|_2 \;,$$

where we used the Cauchy-Schwarz inequality for the final step. One can thus take, for instance,

$$\overline{L}_\rho = L_\rho + \bigl(R + L_\rho N_{\mathcal{I}}\bigr)\sqrt{\frac{N_{\mathcal{H}}}{N_{\mathcal{I}}}}\,.$$

This concludes the proof. □

7. Approachability in the case of general games with partial monitoring. Unfortunately, as is illustrated in the following example, there exist games with partial monitoring that are not bi-piecewise linear.

EXAMPLE 7.1 *The following game (with the same action and signal sets as in Example 5.2) is not bi-piecewise linear.*

	L	M	R
T	$(1,0,0,0)$ / ♣	$(0,0,1,0)$ / ♣	$(2,0,4,0)$ / ♡
B	$(0,1,0,0)$ / ♣	$(0,0,0,1)$ / ♣	$(0,3,0,5)$ / ♡

PROOF. We denote mixed actions of the first player by $(p, 1-p)$, where $p \in [0,1]$ denotes the probability of playing T and $1-p$ is the probability of playing B. It is immediate that $\overline{m}\bigl((p,1-p), \clubsuit\bigr)$ can be identified with the set of all product distributions on 2×2 elements with first marginal distribution $(p, 1-p)$. The proof of Lemma 6.1 shows that the set $\mathcal{B}$ associated with any game always contains the Dirac masses on each signal; that is, $\delta_{\clubsuit} \in \mathcal{B}$. But for $p \neq p'$ and $\lambda \in (0,1)$, denoting $\overline{p} = \lambda\, p + (1-\lambda)p'$, one necessarily has that

$$\overline{m}\bigl((\overline{p}, 1-\overline{p}), \clubsuit\bigr) \;\subsetneq\; \lambda\, \overline{m}\bigl((p,1-p), \clubsuit\bigr) + (1-\lambda)\, \overline{m}\bigl((p',1-p'), \clubsuit\bigr)\,;$$

the inclusion $\subseteq$ holds by concavity of $\overline{m}$ in its first argument (Lemma 5.1) but this inclusion is always strict here since the left-hand side is formed by product distributions while the right-hand side also contains distributions with correlations. Hence, bi-piecewise linearity cannot hold for this game. □

However, we will show that if Condition (APM) holds there exist strategies with a constant per-round complexity to approach polytopes even when the game is not bi-piecewise linear. That is, by considering simpler closed convex sets $\mathcal{C}$, no assumption is needed on the pair (r, H).

We will conclude this section by indicating that thanks to a doubling trick, Condition (APM) is still sufficient for approachability in the most general case when no assumption is made neither on (r, H) nor on $\mathcal{C}$—at the cost, however, of inefficiency.

7.1 Approachability of the negative orthant in the case of general games. For the sake of simplicity, we start with the case of the negative orthant $\mathbb{R}^d_-$. Our argument will be based on Lemma 6.1; we use in the sequel the objects and notation introduced therein. We denote by $r = (r_k)_{1 \leqslant k \leqslant d}$ the components of the d–dimensional payoff function r and introduce, for all $k \in \{1, \dots, d\}$, the set-valued mapping $\widetilde{m}_k$ defined by

$$\widetilde{m}_k : \quad (\boldsymbol{p},\, b) \in \Delta(\mathcal{I}) \times \mathcal{B} \;\longmapsto\; \widetilde{m}_k(\boldsymbol{p}, b) = \Big\{ r_k(\boldsymbol{p}, \boldsymbol{q}) : \quad \boldsymbol{q} \in \Delta(\mathcal{J}) \text{ such that } \widetilde{H}(\boldsymbol{q}) = b \Big\} .$$

The mapping $\widetilde{m}$ is then defined as the Cartesian product of the $\widetilde{m}_k$; formally, for all $\boldsymbol{p} \in \Delta(\mathcal{I})$ and $b \in \mathcal{B}$,

$$\widetilde{m}(\boldsymbol{p}, b) = \Big\{ (z_1, \dots, z_d) : \qquad \forall k \in \{1, \dots, d\}, \quad z_k \in \widetilde{m}_k(\boldsymbol{p}, b) \Big\} .$$

We then linearly extend this mapping into a set-valued mapping $\widetilde{m}$ defined on $\Delta(\mathcal{I}) \times \Delta(\mathcal{B})$ and finally consider the set-valued mapping $\breve{m}$ defined on $\Delta(\mathcal{I}) \times \mathcal{F}$ by

$$\forall\, \sigma \in \mathcal{F}, \quad \forall\, \boldsymbol{p} \in \Delta(\mathcal{I}), \qquad \breve{m}(\boldsymbol{p}, \sigma) = \widetilde{m}\big(\boldsymbol{p}, \Phi(\sigma)\big) = \sum_{b \in \mathcal{B}} \Phi_b(\sigma)\, \widetilde{m}(\boldsymbol{p}, b) \,,$$

where Φ refers to the mapping defined in Lemma 6.1 (based on $\overline{m}$). The lemma below indicates why $\breve{m}$ is an excellent substitute to $\overline{m}$ in the case of the approachability of the orthant $\mathbb{R}^d_-$.

LEMMA 7.1 *The set-valued mappings $\breve{m}$ and $\overline{m}$ satisfy that for all $p \in \Delta(\mathcal{I})$ and $\sigma \in \mathcal{F}$,*

(i) the inclusion $\overline{m}(\boldsymbol{p}, \sigma) \subseteq \breve{m}(\boldsymbol{p}, \sigma)$ holds;

(ii) if $\overline{m}(\boldsymbol{p}, \sigma) \subseteq \mathbb{R}^d_-$, then one also has $\breve{m}(\boldsymbol{p}, \sigma) \subseteq \mathbb{R}^d_-$.

The interpretations of these two properties are that: 1. $\breve{m}$–robust approaching a set $\mathcal{C}$ is more difficult than $\overline{m}$–robust approaching it; and 2. that if Condition (APM) holds for $\overline{m}$ and $\mathbb{R}^d_-$, it also holds for $\breve{m}$ and $\mathbb{R}^d_-$.

PROOF. For property 1., note that by the component-wise construction of $\widetilde{m}$,

$$\forall\, b \in \mathcal{B}, \quad \forall\, \boldsymbol{p} \in \Delta(\mathcal{I}), \qquad \overline{m}(\boldsymbol{p}, b) \subseteq \widetilde{m}(\boldsymbol{p}, b) \,;$$

Lemma 6.1, the linear extension of $\widetilde{m}$, and the definition of $\breve{m}$ then show that

$$\forall\, \sigma \in \mathcal{F}, \quad \forall\, \boldsymbol{p} \in \Delta(\mathcal{I}), \qquad \overline{m}(\boldsymbol{p}, \sigma) = \sum_{b \in \mathcal{B}} \Phi_b(\sigma)\, \overline{m}(\boldsymbol{p}, b) \subseteq \widetilde{m}\big(\boldsymbol{p},\, \Phi(\sigma)\big) = \breve{m}(\boldsymbol{p}, \sigma) \,.$$

As for property 2., it suffices to work component-wise. Note that (by Lemma 6.1 again) the stated assumption exactly means that $\sum_{b \in \mathcal{B}} \Phi_b(\sigma)\, \overline{m}(\boldsymbol{p}, b) \subset \mathbb{R}^d_-$. In particular, rewriting the non-positivity constraint for each of the d components of the payoff vectors, we get

$$\sum_{b \in \mathcal{B}} \Phi_b(\sigma)\, \widetilde{m}_k(\boldsymbol{p}, b) \subseteq \mathbb{R}_- \,,$$

for all $k \in \{1, \dots, d\}$; thus, in particular, $\sum_{b \in \mathcal{B}} \Phi_b(\sigma)\, \widetilde{m}(\boldsymbol{p}, b) = \breve{m}(\boldsymbol{p}, \sigma) \subseteq \mathbb{R}^d_-$. □

We can then extend the result of the previous section without the bi-piecewise linearity assumption.

THEOREM 7.1 *If Condition* (APM) *is satisfied for $\overline{m}$ and $\mathbb{R}^d_-$, then there exists a strategy for (r, H)–approaching $\mathbb{R}^d_-$ at a rate of the order of $T^{-1/5}$, with a constant per-round complexity.*

PROOF. The assumption of the theorem and Property 2. of Lemma 7.1 imply that Condition (APM) holds for $\mathbb{R}^d_-$ and $\breve{m}$; furthermore, the latter corresponds to a bi-piecewise linear game as can be seen by noting, similarly to what was done in the section devoted to regret minimization (Section 6.2), that each $\widetilde{m}_k$, being based on the scalar payoff function r_k, is a piecewise linear function. Thus, $\breve{m}$ is also a piecewise linear function.

Therefore, the steps between Equations (10)–(12) of the proof of Theorem 6.1 (or the corresponding statements in the proof of Theorem 6.3) can be adapted by replacing $\overline{m}$ and $\overline{\overline{m}}$ by, respectively, $\widetilde{m}$, $\breve{m}$, and its extension corresponding to Definition 6.1. The result follows. □

7.2 Approachability of polytopes in the case of general games. If that the target set $\mathcal{C}$ is a polytope, then $\mathcal{C}$ can be written as the intersection of a finite number of half-planes, i.e., there exists a finite family $\big\{(e_k,\, f_k) \in \mathbb{R}^d \times \mathbb{R},\ k \in \mathcal{K}\big\}$ such that

$$\mathcal{C} = \big\{z \in \mathbb{R}^d : \quad \langle z, e_k\rangle \leqslant f_k, \quad \forall\, k \in \mathcal{K}\big\}.$$

Given the original (not necessarily bi-piecewise linear) game (r, H), we introduce another game $(r_{\mathcal{C}}, H)$, whose payoff function $r_{\mathcal{C}} : \mathcal{I} \times \mathcal{J} \to \mathbb{R}^{\mathcal{K}}$ is defined as

$$\forall\, i \in \mathcal{I}, \quad \forall\, j \in \mathcal{J}, \qquad r_{\mathcal{C}}(i,j) = \Big[\langle r(i,j), e_k\rangle - f_k\Big]_{k\in\mathcal{K}}.$$

The following lemma follows by rewriting the above.

LEMMA 7.2 *Given a polytope $\mathcal{C}$, the (r,H)–approachability of $\mathcal{C}$ and the $\big(r_{\mathcal{C}}, H\big)$–approachability of $\mathbb{R}^d_-$ are equivalent in the sense that every strategy for one problem translates to a strategy for the other problem. In addition, Condition* (APM) *holds for (r,H) and $\mathcal{C}$ if and only if it holds for $\big(r_{\mathcal{C}}, H\big)$ and $\mathbb{R}^d_-$.*

Via the lemma above, Theorem 7.1 indicates that Condition (APM) for (r,H) and $\mathcal{C}$ is a sufficient condition for the (r,H)–approachability of $\mathcal{C}$ and provides a strategy to do so. (The per-round complexity of this strategy depends in particular at least linearly on the cardinality of $\mathcal{K}$.)

7.3 Approachability of general convex sets in the case of general games. A general closed convex set can always be approximated arbitrarily well by a polytope (where the number of vertices of the latter however increases as the quality of the approximation does). Therefore, via playing in regimes, Condition (APM) is also seen to be sufficient to (r,H)–approach any general closed convex set $\mathcal{C}$. However, the computational complexity of the resulting strategy is much larger: the per-round complexity increases over time (as the numbers of vertices of the approximating polytopes do).

Appendix A. An auxiliary result of calibration.

We prove here (4) for a given $\eta > 0$ and do so by following the methodology of Mannor and Stoltz [19]. (Note that this result is of independent interest.)

We actually assume that the covering $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^{N_\eta}$ is slightly finer than what was required around (4) and that it forms an $\eta/N_{\mathcal{B}}$–grid of $\Delta(\mathcal{B})$, i.e., that for all $\boldsymbol{y} \in \Delta(\mathcal{B})$, there exists $\ell \in \{1, \ldots, N_\eta\}$ such that $\big\|\boldsymbol{y} - \boldsymbol{y}^\ell\big\|_1 \leqslant \eta/N_{\mathcal{B}}$.

We recall that elements $\boldsymbol{y} \in \mathcal{B}$ are denoted by $\boldsymbol{y} = (y_b)_{b\in\mathcal{B}}$ and we identify $\Delta(\mathcal{B})$ with a subset of $\mathbb{R}^{N_{\mathcal{B}}}$. In particular, $\mathbb{I}_b$, the Dirac mass on a given $b \in \mathcal{B}$, is a binary vector whose only non-null component is the one indexed by b. Finally, we denote by

$$\underline{0} = (0, \ldots, 0) \qquad \text{and} \qquad \underline{1} = (1, \ldots, 1)$$

the elements of $\mathbb{R}^{\mathcal{B}}$ respectively formed by zeros and ones only.

We consider a vector-valued payoff function $C : \{1, \ldots, N_\eta\} \times \mathcal{B} \to \mathbb{R}^{2N_\eta N_{\mathcal{B}}}$ defined as follows; for all $\ell \in \{1, \ldots, N_\eta\}$ and for all $b \in \mathcal{B}$,

$$C(\ell, b) = \left(\underline{0},\ \ldots,\ \underline{0},\ \ \boldsymbol{y}^\ell - \mathbb{I}_b - \frac{\eta}{N_{\mathcal{B}}}\underline{1},\ \ \mathbb{I}_b - \boldsymbol{y}^\ell - \frac{\eta}{N_{\mathcal{B}}}\underline{1},\ \ \underline{0},\ \ldots,\ \underline{0}\right),$$

which is a vector of $2N_\eta$ elements of $\mathbb{R}^{\mathcal{B}}$ composed by $2(N_\eta - 1)$ occurrences of the zero element $\underline{0} \in \mathbb{R}^{\mathcal{B}}$ and two non-zero elements, located in the positions indexed by $2\ell - 1$ and 2ℓ.

We now show that the closed convex set $(\mathbb{R}_-)^{2N_\eta N_{\mathcal{B}}}$ is C–approachable; to do so, we resort to the characterization stated in Theorem 2.1. To each $\boldsymbol{y} \in \Delta(\mathcal{B})$ we will associate a pure action $\ell_{\boldsymbol{y}}$ in $\{1, \ldots, N_\eta\}$ so that $C\big(\ell_{\boldsymbol{y}}, \boldsymbol{y}\big) \in (\mathbb{R}_-)^{2N_\eta N_{\mathcal{B}}}$; note that to satisfy the necessary and sufficient condition, it is not necessary here to resort to mixed actions of the first player. The index $\ell_{\boldsymbol{y}}$ is any index ℓ such that $\big\|\boldsymbol{y} - \boldsymbol{y}^\ell\big\|_1 \leqslant \eta/N_{\mathcal{B}}$; such an index always exists as noted at the beginning of this proof. Indeed, one then has in particular that for each component $b \in \mathcal{B}$,

$$\big|y_b^{\ell_{\boldsymbol{y}}} - y_b\big| \leqslant \big\|\boldsymbol{y}^{\ell_{\boldsymbol{y}}} - \boldsymbol{y}\big\|_1 \leqslant \eta/N_{\mathcal{B}}\,.$$

A straightforward adaptation of the proof of Theorem 2.2 then yields a strategy such that for all $\delta \in (0,1)$ and for all strategies of the second player, with probability at least $1-\delta$,

$$\sup_{\tau \geqslant T} \inf_{c \in (\mathbb{R}_-)^{2N_\eta N_\mathcal{B}}} \left\| c - \frac{1}{\tau} \sum_{t=1}^{\tau} C(L_t, \boldsymbol{y}_t) \right\|_2 \leqslant 2M \sqrt{\frac{2}{\delta T}}\,, \tag{16}$$

where M is a bound in Euclidian norm over C, e.g., $M = 4 + 2\eta$. The quantities of interest can be rewritten as

$$\frac{1}{\tau} \sum_{t=1}^{\tau} C(L_t, \boldsymbol{y}_t) = \left(\frac{N_\tau(\ell)}{\tau} \big(\boldsymbol{y}^\ell - \overline{\boldsymbol{y}}_\tau^\ell \big) - \frac{N_\tau(\ell)}{\tau} \frac{\eta}{N_\mathcal{B}} \underline{1},\ \frac{N_\tau(\ell)}{\tau} \big(\overline{\boldsymbol{y}}_\tau^\ell - \boldsymbol{y}^\ell \big) - \frac{N_\tau(\ell)}{\tau} \frac{\eta}{N_\mathcal{B}} \underline{1} \right)_{\ell \in \{1,\dots,N_\eta\}},$$

where we recall that we denoted for all ℓ such that $N_\tau(\ell) > 0$ the average of their corresponding mixed actions $\boldsymbol{y}_t$ by

$$\overline{\boldsymbol{y}}_\tau^\ell = \frac{1}{N_\tau(\ell)} \sum_{t=1}^{\tau} \boldsymbol{y}_t \mathbb{I}_{\{L_t = \ell\}}\,.$$

The projection in ℓ^2–norm of quantity of interest onto $(\mathbb{R}_-)^{2N_\eta N_\mathcal{B}}$ is formed by its non-positive components, so that its square distance to $(\mathbb{R}_-)^{2N_\eta N_\mathcal{B}}$ equals

$$\inf_{c \in (\mathbb{R}_-)^{2N_\eta N_\mathcal{B}}} \left\| c - \frac{1}{\tau} \sum_{t=1}^{\tau} C(L_t, \boldsymbol{y}_t) \right\|_2^2 = \sum_{\ell=1}^{N_\eta} \left(\frac{N_\tau(\ell)}{\tau} \right)^2 \sum_{b \in \mathcal{B}} \underbrace{\left(\left(y_b^\ell - \overline{y}_{\tau,b}^\ell - \frac{\eta}{N_\mathcal{B}} \right)_+^2 + \left(\overline{y}_{\tau,b}^\ell - y_b^\ell - \frac{\eta}{N_\mathcal{B}} \right)_+^2 \right)}_{= \left(|\overline{y}_{\tau,b}^\ell - y_b^\ell| - \eta/N_\mathcal{B} \right)_+^2}.$$

Therefore, our target is achieved: using first that $(\,\cdot\,)_+$ is subadditive, then applying the Cauchy-Schwarz inequality,

$$\begin{aligned}
\sum_{\ell=1}^{N_\eta} \frac{N_\tau(\ell)}{\tau} \Big(\big\| \boldsymbol{y}^\ell - \overline{\boldsymbol{y}}_\tau \big\|_1 - \eta \Big)_+ &\leqslant \sum_{\ell=1}^{N_\eta} \frac{N_\tau(\ell)}{\tau} \sum_{b \in \mathcal{B}} \left(\big| y_b^\ell - \overline{y}_{\tau,b}^\ell \big| - \frac{\eta}{N_\mathcal{B}} \right)_+ \\
&\leqslant \sqrt{N_\eta N_\mathcal{B}} \sqrt{\sum_{\ell=1}^{N_\eta} \left(\frac{N_\tau(\ell)}{\tau} \right)^2 \sum_{b \in \mathcal{B}} \left(\big| y_b^\ell - \overline{y}_{\tau,b}^\ell \big| - \frac{\eta}{N_\mathcal{B}} \right)_+^2} \\
&\leqslant 2M \sqrt{N_\eta N_\mathcal{B}} \sqrt{\frac{2}{\delta T}}\,,
\end{aligned}$$

where the last inequality holds, by (16), for all $\tau \geqslant T$ with probability at least $1-\delta$. Choosing an integer T_δ sufficiently large so that

$$2M \sqrt{N_\eta N_\mathcal{B}} \sqrt{\frac{2}{\delta\, T_\delta}} \leqslant \delta$$

concludes the proof of the property stated in (4).

Appendix B. Proof of Lemma 6.3.

Proof. For all $(i,j) \in \mathcal{I} \times \mathcal{J}$, the quantity $H(i,j)$ is a probability distribution over the set of signals $\mathcal{H}$; we denote by $H_s(i,j)$ the probability mass that it puts on some signal $s \in \mathcal{H}$.

Equation (7) indicates that for each pair $(i,s) \in \mathcal{I} \times \mathcal{H}$,

$$\sum_{t=(n-1)L+1}^{nL} \left(\frac{\mathbb{I}_{\{S_t = s\}} \mathbb{I}_{\{I_t = i\}}}{p_{I_t,n}} - H_s(i, J_t) \right)$$

is a sum of L elements of a martingale difference sequence, with respect to the filtration whose t-th element is generated by $\boldsymbol{p}_n$, the pairs (I_s, S_s) for $s \leqslant t$, and J_s for $s \leqslant t+1$. The conditional variances of the increments are bounded by

$$\mathbb{E}_t \left[\left(\frac{\mathbb{I}_{\{S_t = s\}} \mathbb{I}_{\{I_t = i\}}}{p_{I_t,n}} \right)^2 \right] \leqslant \frac{1}{p_{i,n}^2} \mathbb{E}_t \big[\mathbb{I}_{\{I_t = i\}} \big] = \frac{1}{p_{i,n}}\,;$$

since by definition of the strategy, $\boldsymbol{p}_n = (1-\gamma)\,\boldsymbol{x}_n + \gamma\,\boldsymbol{u}$, we have that $p_{i,n} \geqslant \gamma/N_{\mathcal{I}}$, which shows that the sum of the conditional variances is bounded by

$$\sum_{t=(n-1)L+1}^{nL} \mathrm{Var}_t\left(\frac{\mathbb{I}_{\{S_t=s\}}\mathbb{I}_{\{I_t=i\}}}{p_{I_t,n}}\right) \leqslant \frac{LN_{\mathcal{I}}}{\gamma}\,.$$

The Bernstein-Freedman inequality (see Freedman [11] or Cesa-Bianchi et al. [6], Lemma A.1) therefore indicates that with probability at least $1-\delta$,

$$\left|\frac{1}{L}\sum_{t=(n-1)L+1}^{nL} \frac{\mathbb{I}_{\{S_t=s\}}\mathbb{I}_{\{I_t=i\}}}{p_{I_t,n}} - \underbrace{\frac{1}{L}\sum_{t=(n-1)L+1}^{nL} H_s(i,J_t)}_{=\,H_s(i,\widehat{\boldsymbol{q}}_n)}\right| \leqslant \sqrt{2\frac{N_{\mathcal{I}}}{\gamma L}\ln\frac{2}{\delta}} + \frac{1}{3}\frac{N_{\mathcal{I}}}{\gamma L}\ln\frac{2}{\delta}\,.$$

Therefore, by summing the above inequalities over $i \in \mathcal{I}$ and $s \in \mathcal{H}$, we get (after a union bound) that with probability at least $1 - N_{\mathcal{I}}N_{\mathcal{H}}\delta$,

$$\left\|\widetilde{\sigma}_n - \widetilde{H}\big(\widehat{\boldsymbol{q}}_n\big)\right\|_2 \leqslant \sqrt{N_{\mathcal{I}}N_{\mathcal{H}}}\left(\sqrt{\frac{2N_{\mathcal{I}}}{\gamma L}\ln\frac{2}{\delta}} + \frac{1}{3}\frac{N_{\mathcal{I}}}{\gamma L}\ln\frac{2}{\delta}\right).$$

Finally, since $\widehat{\sigma}_n$ is the projection in the ℓ^2–norm of $\widetilde{\sigma}_n$ onto the convex set $\mathcal{F}$, to which $\widetilde{H}\big(\widehat{\boldsymbol{q}}_n\big)$ belongs, we have that

$$\left\|\widehat{\sigma}_n - \widetilde{H}\big(\widehat{\boldsymbol{q}}_n\big)\right\|_2 \leqslant \left\|\widetilde{\sigma}_n - \widetilde{H}\big(\widehat{\boldsymbol{q}}_n\big)\right\|_2\,,$$

and this concludes the proof. □

References

[1] J. Abernethy, P. L. Bartlett, and E. Hazan. Blackwell approachability and low-regret learning are equivalent. In *Proceedings of the Twenty-Fourth Annual Conference on Learning Theory (COLT'11)*. Omnipress, 2011.

[2] D. Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6:1–8, 1956.

[3] D. Blackwell. Controlled random walks. In *Proceedings of the International Congress of Mathematicians, 1954, Amsterdam, vol. III*, pages 336–338, 1956.

[4] A. Blum and Y. Mansour. From external to internal regret. *Journal of Machine Learning Research*, 8:1307–1324, 2007.

[5] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[6] N. Cesa-Bianchi, G. Lugosi, and G. Stoltz. Regret minimization under partial monitoring. *Mathematics of Operations Research*, 31:562–580, 2006.

[7] X. Chen and H. White. Laws of large numbers for Hilbert space-valued mixingales with applications. *Econometric Theory*, 12:284–304, 1996.

[8] A.P. Dawid. The well-calibrated Bayesian. *Journal of the American Statistical Association*, 77: 605–613, 1982.

[9] D. Foster and R. Vohra. Asymptotic calibration. *Biometrika*, 85:379–390, 1998.

[10] D. Foster and R. Vohra. Regret in the on-line decision problem. *Games and Economic Behavior*, 29:7–36, 1999.

[11] D.A. Freedman. On tail probabilities for martingales. *Annals of Probability*, 3:100–118, 1975.

[12] J.E. Goodman and J. O'Rourke, editors. *Handbook of Discrete and Computational Geometry*. Discrete Mathematics and its Applications. Chapman & Hall/CRC, Boca Raton, FL, second edition, 2004.

[13] S. Hart and A. Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68:1127–1150, 2000.

[14] S. Hart and A. Mas-Colell. A general class of adaptive strategies. *Journal of Economic Theory*, 98: 26–54, 2001.

[15] E. Lehrer and E. Solan. Learning to play partially-specified equilibrium. Mimeo, 2007.

[16] G. Lugosi, S. Mannor, and G. Stoltz. Strategies for prediction under imperfect monitoring. *Mathematics of Operations Research*, 33:513–528, 2008. An extended abstract was presented at COLT'07.

[17] S. Mannor and N. Shimkin. On-line learning with imperfect monitoring. In *Proceedings of the Sixteenth Annual Conference on Learning Theory (COLT'03)*, pages 552–567. Springer, 2003.

[18] S. Mannor and N. Shimkin. Regret minimization in repeated matrix games with variable stage duration. *Games and Economic Behavior*, 63(1):227–258, 2008.

[19] S. Mannor and G. Stoltz. A geometric proof of calibration. *Mathematics of Operations Research*, 35:721–727, 2010.

[20] S. Mannor, J. Tsitsiklis, and J. Y. Yu. Online learning with sample path constraints. *Journal of Machine Learning Research*, 10(Mar):569–590, 2009.

[21] J.-F. Mertens, S. Sorin, and S. Zamir. Repeated games. Technical Report no. 9420, 9421, 9422, Université de Louvain-la-Neuve, 1994.

[22] V. Perchet. Calibration and internal no-regret with random signals. In *Proceedings of the Twentieth International Conference on Algorithmic Learning Theory (ALT'09)*, pages 68–82, 2009.

[23] V. Perchet. Approachability of convex sets in games with partial monitoring. *Journal of Optimization Theory and Applications*, 149:665–677, 2011.

[24] V. Perchet. Internal regret with partial monitoring calibration-based optimal algorithms. *Journal of Machine Learning Research*, 2011. In press.

[25] V. Perchet and M. Quincampoix. On an unified framework for approachability in games with or without signals. Mimeo, 2011.

[26] A. Piccolboni and C. Schindelhauer. Discrete prediction games with arbitrary feedback and loss. In *Proceedings of the Fourteenth Annual Conference on Computational Learning Theory (COLT'01)*, pages 208–223, 2001.

[27] A. Rakhlin, K. Sridharan, and A. Tewari. Online learning: Beyond regret. In *Proceedings of the Twenty-Fourth Annual Conference on Learning Theory (COLT'11)*. Omnipress, 2011.

[28] J. Rambau and G. Ziegler. Projections of polytopes and the generalized Baues conjecture. *Discrete and Computational Geometry*, 16:215–237, 1996.

[29] A. Rustichini. Minimizing regret: The general case. *Games and Economic Behavior*, 29:224–243, 1999.

Acknowledgments. Shie Mannor was partially supported by the ISF under contract 890015 and the Google Inter-university center for Electronic Markets and Auctions. Vianney Perchet benefited from the support of the ANR under grant ANR-10-BLAN 0112. Gilles Stoltz acknowledges support from the French National Research Agency (ANR) under grant EXPLO/RA ("Exploration–exploitation for efficient resource allocation") and by the PASCAL2 Network of Excellence under EC grant no. 506778.

An extended abstract of this paper appeared in the *Proceedings of the 24th Annual Conference on Learning Theory* (COLT'11), JMLR Workshop and Conference Proceedings, Volume 19, pages 515–536, 2011.