Journal of Machine Learning Research 1 (2014) 3247-3295 Submitted 6/13; Revised 4/14; Published 8/14

Set-Valued Approachability and Online Learning with Partial Monitoring

Shie Mannor SHIE@EE.TECHNION.AC.IL
Israel Institute of Technology (Technion)
Faculty of Electrical Engineering
32000 Haifa, Israel

Vianney Perchet VIANNEY.PERCHET@NORMALESUP.ORG
Université Paris Diderot, LPMA
8 place FM/13
75013 Paris, France

Gilles Stoltz STOLTZ@HEC.FR
GREGHEC: HEC Paris — CNRS
1 rue de la Libération
78351 Jouy-en-Josas, France

Editor: Alexander Rakhlin

Abstract

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: it 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 a simple and generally efficient strategy (i.e., with constant per-step complexity) for this setup. As an important example, we instantiate our general strategy to the case when external regret or internal regret is to be minimized under partial monitoring.

Keywords: online learning, approachability, regret, partial monitoring

1. Introduction

Blackwell's approachability theory and its variants have become a standard and useful tool in analyzing online learning algorithms (Cesa-Bianchi and Lugosi, 2006) and algorithms for learning in games (Hart and Mas-Colell, 2000, 2001). The first application of Blackwell's approachability to learning in the online setup is due to Blackwell (1956b) himself. Numerous other contributions are summarized in the monograph by Cesa-Bianchi and Lugosi (2006). Blackwell's approachability theory enjoys a natural geometric interpretation that allows it to be used in situations where other learning methods (online convex optimization or exponential weights) do not seem to be easily applicable. In some sense, it can be used to go beyond the minimization of the regret to control quantities of a different flavor. Examples of such uses can be found in Mannor et al. (2009), which minimizes the regret together with path constraints, and in Mannor and Shimkin (2008), which minimizes the regret in games whose stage duration is not fixed. Recently, it has been shown by Abernethy et al.

(2011) that approachability and low regret learning are equivalent in the sense that efficient reductions exist from one problem to the other. Another recent paper by Rakhlin et al. (2011) 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 an arbitrary opponent. That is, we will obtain worst-case performance guarantees: guarantees that are valid for *all* strategies of the opponent. 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 pair of actions taken by the decision maker and the opponent. 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 statistic thereof. The difficulty is that the decision maker cannot compute the actual reward obtained nor the actions of the opponent.

Regret minimization with partial monitoring (defined in the general sense of Rustichini, 1999) has been studied in several papers in the learning theory community. Piccolboni and Schindelhauer (2001), Mannor and Shimkin (2003), Cesa-Bianchi et al. (2006), Bartók et al. (2010, 2011), Foster and Rakhlin (2012) study games in which an accurate estimation of the rewards (or worst-case rewards) of the decision maker is possible thanks to some statistically sufficient monitoring; in this case, the notion of regret with partial monitoring reduces to the classical notion of regret with full monitoring. A general policy with vanishing external regret with partial monitoring is presented by Lugosi et al. (2008). This policy is based on exponential weights and a specific estimation procedure for the (worst-case) obtained rewards.

In contrast, we devise a general (efficient) algorithm for the problem of approachability under partial monitoring. We then apply it to the more restricted problem of regret minimization. More precisely, we first define a new type of approachability setup, for set-valued functions, which enables to re-derive the extension of approachability to the partial monitoring vector-valued setting proposed by Perchet (2011a). 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 (1956b) but for the first time in this general framework. For example, the recent theoretical study of approachability by Perchet and Quincampoix (2011), which is based on somehow related arguments, does neither provide rates of convergence nor concrete algorithms.

1.1 Outline and Comparison to Known Results

The paper is organized in three main parts. The first part consists of Section 2, where we recall basic facts from approachability theory, when a decision maker faces an arbitrary opponent in the standard vector-valued games setting.

The second part deals with our first contribution, a novel setup for approachability termed "set-valued approachability", where instead of obtaining a vector-valued reward, the decision maker obtains a set, that represents the ambiguity concerning his reward. In Section 3, 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 set-valued 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 the opponent. This new concept of set-valued approachability is interesting on its own, as it cannot be directly encompassed into classical vector-valued approachability; yet we retrieve several familiar results (characterization of approachable convex sets, rates of convergence that are independent of the dimension, and so on). More importantly, these results are the key tools for our second series of contributions, which we describe now.

The third part studies approachability in repeated games with partial monitoring. Previous general results in this setup suffered from at least one of the following drawbacks. They were either non-constructive (Rustichini, 1999) or were highly inefficient. The latter drawback refers to strategies that relied on some sort of lifting to the space of probability measures on mixed actions (see e.g., Lehrer and Solan, 2007 and Perchet, 2009, 2011a). They then typically required a fine grid of elements in this lifted space, which had to be progressively refined over time. This construction leads to two main issues: on the one hand, the step complexity continuously increases and becomes prohibitive in the number T of past steps. On the other hand, rates of convergence deteriorate and depend on the dimension. Our aim is therefore to devise algorithms that are efficient (as long as the projection onto some convex set can be done efficiently), with a constant step complexity (although it may depend on parameters of the problem at hand), and with rates of convergence independent of the ambient dimension. Our strategies are the first, to our knowledge, satisfying all of these properties in the general approachability framework. They do so because they do not rely on finer and finer grids; as a byproduct, they can also be considered more natural. Section 5 discusses in greater detail all the points mentioned in this paragraph.

More precisely, we state in Section 5.1 the necessary and sufficient condition for approachability in games with partial monitoring and show in Section 5.2 how to apply set-valued approachability framework to the repeated vector-valued games with partial monitoring. In Section 5.3 we then consider a specific type of games where the signaling structure possesses a special property, called bi-piecewise linearity, that can be exploited to derive simple, constructive and efficient strategies. This type of games is rich enough as it encompasses several useful special cases discussed in the later sections. In Section 5.4, we mention the general signaling case and explain how it is possible to approach certain special sets such as polytopes efficiently (thanks to a reduction to bi-piecewise linearity) with the same dimension-independent rates of convergence—and even general convex sets, although inefficiently in the latter case.

As an important other example of a setting where bi-piecewise linearity holds, we apply in Section 6 the results of Section 5.3 to both external-regret and internal-regret minimization in repeated games with partial monitoring. In this specific case, our algorithms have rates similar to the ones obtained by Lugosi et al. (2008) but slower than Perchet (2011b); however our proof is direct and simpler and the strategy is efficient.

1.2 Mixed Actions versus Pure Actions

Most of Sections 2–4 (classical approachability and set-valued approachability) is concerned with mixed actions, while Sections 5–6 (approachability in games with partial monitoring) are focused on pure actions. The explanation for this is as follows. Even though pure actions are inherent to the model of partial monitoring, the reduction from approachability in games with partial monitoring to set-valued approachability, as described in Section 5.2, is to set-valued approachability with mixed actions.

2. Some Basic Facts from Approachability Theory

In this section we recall the most basic version 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 an opponent (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 vectors 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 a framework in which mixed actions are taken. 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 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 only the indicated payoff (bandit monitoring) is revealed to him.

Strategies of the players are defined as mappings associating the information available at the beginning of each round $t \geqslant 1$ with a mixed action. In particular, strategies of the first player in the case of full monitoring associate with $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}$ and $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}$ a mixed action $\boldsymbol{x}_t \in \Delta(\mathcal{A})$, while in the case of bandit monitoring, they do this association based on $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}$ and $m(\boldsymbol{x}_1, \boldsymbol{y}_1), \ldots, m(\boldsymbol{x}_{t-1}, \boldsymbol{y}_{t-1})$. We do not restrict the opponent and assume a full monitoring for him: his strategies associate with $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}$ and $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}$ a mixed action $\boldsymbol{y}_t \in \Delta(\mathcal{B})$.

2.1 Necessary and Sufficient Condition for Approachability

Given a set $\mathcal{C}$, the aim of the first player is to ensure that his average payoff converges to $\mathcal{C}$, while the second player wants to prevent it. This gives rise to Blackwell's classical definition of approachability. (Here, we state it as m–approachability to remind the reader, in the notation, that the underlying payoff function is m.)

Definition 1 *Given a function $m : \mathcal{A} \times \mathcal{B} \to \mathbb{R}^d$, a set $\mathcal{C} \subseteq \mathbb{R}^d$ is m–*approachable *by the first player if he has a strategy such that, for all $\varepsilon > 0$, there exists an integer T_ε such that for all strategies of the second player,*

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

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.

As will be recalled below in Theorem 3, even stronger approachability guarantees can be achieved. Indeed, the first player has deterministic strategies such that, for all (deterministic or randomized) strategies of the second player, with probability 1, for all $T \geqslant 1$,

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

where $\beta(\cdot)$ is some decreasing mapping to 0 to be determined later.

For closed convex sets there is a simple characterization of approachability that is a direct consequence of von Neumann's minimax theorem.

Theorem 2 (see Blackwell, 1956a, Theorem 3) *A closed convex set $\mathcal{C} \subseteq \mathbb{R}^d$ is approachable if and only if*

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

2.2 An Associated Strategy (whose Efficiency Depends on the Geometry of $\mathcal{C}$)

Blackwell suggested a simple strategy with a geometric flavor; it only requires bandit monitoring.

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

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

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

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

where $\langle\,\cdot\,,\,\cdot\,\rangle$ is the Euclidean inner product in $\mathbb{R}^d$.

The stated minimax problem for determining $\boldsymbol{x}_{t+1}$ can be solved efficiently using, e.g., linear programming: the associated complexity is polynomial in $N_{\mathcal{A}}$ and $N_{\mathcal{B}}$. This strategy is efficient if computing 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, which can be derived as a special case of the results stated and proved in Theorem 25 later in this paper.

Theorem 3 (see Blackwell, 1956a, Theorems 1 and 3) *We consider an approachable closed convex set $\mathcal{C} \subseteq \mathbb{R}^d$ and we denote by M a bound in norm over m, i.e.,*

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

The above strategy ensures that for all strategies of the second player, with probability 1, for all $T \geqslant 1$,

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

3. Set-Valued Approachability for Finite 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 $\overline{m} : \mathcal{A}\times\mathcal{B} \to \mathcal{S}\big(\mathbb{R}^d\big)$. When the players choose respective actions $a \in \mathcal{A}$ and $b \in \mathcal{B}$, the first player gets the *subset* $\overline{m}(a,b)$ as a payoff. This models the ambiguity or uncertainty associated with some true underlying payoff.

3.1 Mixed Actions Taken and Observed

For the moment, we only consider the case of mixed actions taken and observed, keeping the same definition of a strategy as in the previous section. (The next subsection will briefly explain, for the sake of completeness, how to deal with the case of pure actions taken and observed.)

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 $\mathbb{R}^d$,

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

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 $(a,b) \in \mathcal{A}\times\mathcal{B}$. When μ is such a product-distribution, we use the notation $\overline{m}(\mu) = \overline{m}(\boldsymbol{x},\boldsymbol{y})$.

We can now describe how the game proceeds. 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 takes place for the first player: he observes $\boldsymbol{y}_t$ at the end of round t and he gets the subset $\overline{m}(\boldsymbol{x}_t,\boldsymbol{y}_t)$ as a payoff (which, again, accounts for the uncertainty).

1. For two sets S, T and $\alpha \in [0,1]$, the convex combination $\alpha S + (1-\alpha)T$ is defined as

$$\big\{\alpha s + (1-\alpha)t, \quad s \in S \text{ and } t \in T\big\}\,.$$

3.1.1 Definition of Set-Valued Approachability

We are interested in the behavior of

$$\frac{1}{T}\sum_{t=1}^{T}\overline{m}(\boldsymbol{x}_t,\boldsymbol{y}_t)=\overline{m}(\nu_T)\,,\qquad\text{where}\qquad \nu_T:=\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_t\otimes\boldsymbol{y}_t$$

is the empirical joint distribution of mixed actions taken during the first T rounds.

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

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

the smallest value such that $\overline{m}(\nu_T)$ is included in an ε_T–neighborhood of $\mathcal{C}$. 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. This is made formal in the following definition.

Definition 4 *Given a set-valued payoff function $\overline{m}:\mathcal{A}\times\mathcal{B}\to\mathcal{S}\big(\mathbb{R}^d\big)$, a set $\mathcal{C}\subseteq\mathbb{R}^d$ is $\overline{m}$–approachable by the first player if he has a strategy such that, for all $\varepsilon>0$, there exists an integer T_ε such that for all strategies of the second player,*

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

When the set-valued function $\overline{m}$ is clear from the context, we will simply say that $\mathcal{C}$ is set-valued approachable. Actually, just as in the classical case of approachability, the bounds exhibited below in Theorem 8 will be for deterministic strategies of the first player and will read as follows: for all (deterministic or randomized) strategies of the second player, with probability 1, for all $T\geqslant 1$,

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

where $\beta(\cdot)$ is a mapping decreasing to 0 to be determined.

3.1.2 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 an $M\,\|\mu-\nu\|_1$–neighborhood of $\overline{m}(\nu)$, where M is a bound in ℓ^2–norm on $\overline{m}$. This is a result that we will use repeatedly.

Definition 5 *The set-valued function $\overline{m}:\mathcal{A}\times\mathcal{B}\to\mathcal{S}\big(\mathbb{R}^d\big)$ is bounded in ℓ^2–norm by M if*

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

Lemma 6 *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 ℓ^2–norm by M. Then*

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

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

Proof Let ξ be an element of $\overline{m}(\mu)$; it can be written as

$$\xi = \sum_{a \in \mathcal{A}} \sum_{b \in \mathcal{B}} \mu_{a,b} \, \zeta_{a,b}$$

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

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

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

$$\|\xi - c\|_2 = \left\| \sum_{a \in \mathcal{A}} \sum_{b \in \mathcal{B}} \bigl(\mu_{a,b} - \nu_{a,b}\bigr) \zeta_{a,b} \right\|_2 \leqslant \sum_{a \in \mathcal{A}} \sum_{b \in \mathcal{B}} \bigl|\mu_{a,b} - \nu_{a,b}\bigr| \, \|\zeta_{a,b}\|_2 \leqslant M \sum_{a \in \mathcal{A}} \sum_{b \in \mathcal{B}} \bigl|\mu_{a,b} - \nu_{a,b}\bigr| .$$

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

Corollary 7 *If the set-valued function $\overline{m}$ is bounded in norm, then for all $\boldsymbol{y} \in \Delta(\mathcal{B})$, the mappings $D_{\boldsymbol{y}} : \Delta(A) \to \mathbb{R}$ defined, for all $\boldsymbol{x} \in \Delta(\mathcal{A})$, by*

$$D_{\boldsymbol{y}}(\boldsymbol{x}) = \sup_{\xi \in \overline{m}(\boldsymbol{x},\boldsymbol{y})} \inf_{c \in \mathcal{C}} \|c - \xi\|_2$$

are 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 ℓ^2–norm over $\overline{m}$. Indeed, fix $\delta > 0$ and let $\xi_{\delta,\boldsymbol{x}} \in \overline{m}(\boldsymbol{x},\boldsymbol{y})$ be such that

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

By Lemma 6 (with the choices $\mu = \boldsymbol{x} \otimes \boldsymbol{y}$ and $\nu = \boldsymbol{x}' \otimes \boldsymbol{y}$),

$$\inf_{\xi' \in \overline{m}(\boldsymbol{x}',\boldsymbol{y})} \|\xi_{\delta,\boldsymbol{x}} - \xi'\| \leqslant M\varepsilon ,$$

and therefore, there exists $\xi_{\delta,\boldsymbol{x}'} \in \overline{m}(\boldsymbol{x}',\boldsymbol{y})$ such that $\bigl\| \xi_{\delta,\boldsymbol{x}} - \xi_{\delta,\boldsymbol{x}'} \bigr\|_2 \leqslant M\varepsilon + \delta$. The triangle inequality entails that

$$\inf_{c \in \mathcal{C}} \bigl\| c - \xi_{\delta,\boldsymbol{x}} \bigr\|_2 \leqslant \inf_{c \in \mathcal{C}} \bigl\| c - \xi_{\delta,\boldsymbol{x}'} \bigr\|_2 + M\varepsilon + \delta .$$

Substituting in (2), we get that

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

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

3.1.3 Necessary and Sufficient Condition for Set-Valued Approachability

This condition will be referred to as (SVAC), an acronym that stands for "set-valued approachability condition."

Theorem 8 *Suppose that the set-valued function $\overline{m}$ is bounded in norm. A closed convex set $\mathcal{C} \subseteq \mathbb{R}^d$ is $\overline{m}$–approachable if and only if the following set-valued approachability condition is satisfied,*

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

In this case, an $\overline{m}$–approaching strategy for $\mathcal{C}$ is an m–approaching strategy of $\widetilde{\mathcal{C}}$ defined below at (3) *and* (4). *It satisfies, for all $T \geqslant 1$,*

$$\sup_{\xi \in \overline{m}(\nu_T)} \inf_{c \in \mathcal{C}} \|c - \xi\|_2 \leqslant 2M \sqrt{\frac{N_{\mathcal{A}} N_{\mathcal{B}}}{T}}\,,$$

where M is a bound in ℓ^2–norm on $\overline{m}$.

Proof [of the necessity of Condition (SVAC)] 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 7. 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. Then, 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 $T \geqslant 1$,

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

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

We now prove in a constructive way, by exhibiting a suitable strategy (the one alluded at in the statement of the theorem), that (SVAC) is sufficient for set-valued approachability.

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

$$m : (a,b) \in \mathcal{A} \times \mathcal{B} \longmapsto \delta_{(a,b)} \in \mathbb{R}^{\mathcal{A}\times\mathcal{B}}, \tag{3}$$

where $\delta_{(a,b)}$ is the point mass on (a,b). We extend m to $\Delta(\mathcal{A}) \times \Delta(\mathcal{B})$ in a multi-linear fashion. The target set will be

$$\widetilde{\mathcal{C}} = \bigl\{ \mu \in \Delta(\mathcal{A} \times \mathcal{B}) : \quad \overline{m}(\mu) \subseteq \mathcal{C} \bigr\} . \tag{4}$$

The linearity of the function $\overline{m}$ on $\Delta(\mathcal{A} \times \mathcal{B})$ entails that if $\mathcal{C}$ is a convex set (respectively, a closed set, or a polyhedron), then $\widetilde{\mathcal{C}}$ is a convex set as well (respectively, a closed set, or a polyhedron). In the case where $\widetilde{\mathcal{C}}$ is a polyhedron, it is actually a polytope (that is, a compact polyhedron).

We then consider the m–approaching strategy of $\widetilde{\mathcal{C}}$ described in (1) and now prove that it enjoys the convergence guarantees stated in Theorem 8.

Lemma 9 *Condition* (SVAC) *is equivalent to the m–approachability of $\widetilde{\mathcal{C}}$.*

Proof Since $\mathcal{C}$ and thus $\widetilde{\mathcal{C}}$ are closed and convex sets, we can resort to Theorem 2. The latter 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 $\mu = m(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{x} \otimes \boldsymbol{y}$, 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 definition (3) of m entails the rewriting

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

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

$$\varepsilon_T = \Bigl\| \nu_T - P_{\widetilde{\mathcal{C}}}(\nu_T) \Bigr\|_2 = \inf_{\mu \in \widetilde{\mathcal{C}}} \| \nu_T - \mu \|_2 .$$

We now relate them to the ones arising in the definition of $\overline{m}$–approachability of $\mathcal{C}$.

Lemma 10 *The following upper bound holds,*

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

Proof Lemma 6 entails that the sets $\overline{m}(\nu_T)$ are included in $M\sqrt{N_{\mathcal{A}} N_{\mathcal{B}}}\,\varepsilon_T$–neighborhoods of $\overline{m}\bigl(P_{\widetilde{\mathcal{C}}}(\nu_T)\bigr)$. Since by definition of $\widetilde{\mathcal{C}}$, one has $\overline{m}\bigl(P_{\widetilde{\mathcal{C}}}(\nu_T)\bigr) \subseteq \mathcal{C}$, we get in particular that the sets $\overline{m}(\nu_T)$ are included in $M\sqrt{N_{\mathcal{A}} N_{\mathcal{B}}}\,\varepsilon_T$–neighborhoods of $\mathcal{C}$, which is exactly the statement of the lemma. ■

Proof [of the sufficiency of Condition (SVAC)] First, Lemma 9 and Condition (SVAC) entail, via Theorem 3, that the considered strategy m–approaches $\widetilde{\mathcal{C}}$, at the following rate: $\varepsilon_T \leqslant 2/\sqrt{T}$, with probability 1. Second, Lemma 10 indicates that this strategy also $\overline{m}$–approaches $\mathcal{C}$, at the stated rate of $2M\sqrt{N_{\mathcal{A}} N_{\mathcal{B}}/T}$, with probability 1. ■

3.1.4 Remarks: on Efficiency; on Full versus Bandit Monitorings

Note that, as explained around (1), the considered strategy for m–approaching $\widetilde{\mathcal{C}}$, or equivalently $\overline{m}$–approaching $\mathcal{C}$, is efficient as long as projections in ℓ^2–norm onto the set $\widetilde{\mathcal{C}}$ defined in (4) can be computed efficiently. The latter depends on the respective geometries of $\overline{m}$ and $\mathcal{C}$. We will provide examples of favorable cases (see, e.g., Section 6.1 about minimization of external regret under partial monitoring). In the sequel the notion of "efficiency up to a projection oracle" will refer to this efficiency depending solely on the efficient computation of the needed projections.

The proposed strategy does not require full monitoring, although it seems to rely on the observation of the pair of played mixed actions $m(\boldsymbol{x}_t, \boldsymbol{y}_t)$. With bandit monitoring, only the played sets $\overline{m}(\boldsymbol{x}_t, \boldsymbol{y}_t)$ would be available, not the $\boldsymbol{y}_t$ themselves; in that case, the player can act as if the other player chose any $\boldsymbol{y}'_t$ that generates this set, i.e., such that $\overline{m}(\boldsymbol{x}_t, \boldsymbol{y}'_t) = \overline{m}(\boldsymbol{x}_t, \boldsymbol{y}_t)$.

3.2 Pure Actions Taken and Observed

It is well-known that the basic results recalled in Section 2 extend to the case of pure actions. We briefly explain here how the developed theory of set-valued approachability for games with mixed actions extends as well to the case of pure actions taken (still under full monitoring).

The game goes as follows. At each round t, the players choose simultaneously respective pure actions $A_t \in \mathcal{A}$ and $B_t \in \mathcal{B}$, possibly at random according to distributions $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$. As a result, the first player gets the subset $\overline{m}(A_t, B_t)$ as a payoff and observes B_t. Strategies for the players now associate with $A_1, \ldots, A_{t-1}$ and $B_1, \ldots, B_{t-1}$ mixed actions $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$, according to which A_t and B_t are drawn independently.

We are interested in the behavior of

$$\frac{1}{T}\sum_{t=1}^{T} \overline{m}(A_t, B_t) = \overline{m}(\pi_T)\,, \qquad \text{where} \qquad \pi_T := \frac{1}{T}\sum_{t=1}^{T} \delta_{(A_t, B_t)}$$

is the empirical distribution of the pairs (A_t, B_t) of actions taken during the first T rounds. The definition of set-valued approachability extends as follows.

Definition 11 *A set $\mathcal{C} \subseteq \mathbb{R}^d$ is $\overline{m}$–approachable by the first player with pure actions if he has a strategy such that, for all $\varepsilon > 0$, there exists an integer T_ε such that for all strategies of the second player,*

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

The fact that Condition (SVAC) is still a necessary and sufficient condition for $\overline{m}$–approachability with pure actions of a closed convex set $\mathcal{C} \subseteq \mathbb{R}^d$ (where $\overline{m}$ is bounded) can be seen as follows.

Concerning the proof of the sufficiency of this condition, first recall that Lemma 9 indicates that Condition (SVAC) is equivalent to the m–approachability of $\widetilde{\mathcal{C}}$. In view of a version of Theorem 3 for pure actions (e.g., Theorem II.4.3 of Mertens et al., 1994) the strategy described around (3) and (4), with the replacement of the ν_T by the π_T and extra random draws of the pure actions A_t according to the mixed distributions $\boldsymbol{x}_t$ thus computed, is such that the quantities

$$\varepsilon'_T = \left\| \pi_T - P_{\widetilde{\mathcal{C}}}(\pi_T) \right\|_2 = \inf_{\mu \in \widetilde{\mathcal{C}}} \|\pi_T - \mu\|_2$$

satisfy the following convergence guarantees. For all $\delta \in (0,1)$, there exists an integer T_δ such that for all strategies of the second player,

$$\mathbb{P}\bigl\{ \forall\, T \geqslant T_\delta, \quad \varepsilon'_T \leqslant \delta \bigr\} \geqslant 1 - \delta \,.$$

This shows that this strategy also $\overline{m}$–approaches $\mathcal{C}$ since Lemma 10 is valid with the respective replacements of ν_T and ε_T by π_T and ε'_T.

The proof of the necessity of the condition is the same as for mixed actions taken, with the addition of a concentration argument. Indeed, by martingale convergence (e.g., repeated uses of the Hoeffding-Azuma inequality together with an application of the Borel-Cantelli lemma), $\delta_T = \|\pi_T - \nu_T\|_1$ converges to zero almost surely as T goes to infinity. By applying Lemma 6 and by using the notation of the proof of Theorem 8, we get

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

and we simply take the $\liminf$ in the above inequalities to conclude the argument.

4. Set-Valued Approachability for Concave–Convex Set-Valued Games

We consider in this section the same setting of mixed actions taken and observed as in Section 3.1, that is, we deal with set-valued payoff functions $\overline{m} : \Delta(\mathcal{A}) \times \Delta(\mathcal{B}) \to \mathcal{S}\bigl(\mathbb{R}^d\bigr)$ 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 (SVAC) is the necessary and sufficient condition for set-valued approachability. The price to pay for this is the loss of the exhibited efficiency (up to a projection oracle) of the approaching strategies and an inferior convergence rate.

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

Definition 12 *The set-valued function $\overline{m} : \Delta(\mathcal{A}) \times \Delta(\mathcal{B}) \to \mathcal{S}\bigl(\mathbb{R}^d\bigr)$ is bounded in ℓ^2–norm by M if*

$$\forall (\boldsymbol{x}, \boldsymbol{y}) \in \Delta(\mathcal{A}) \times \Delta(\mathcal{B}), \qquad \sup_{\xi \in \overline{m}(\boldsymbol{x}, \boldsymbol{y})} \|\xi\|_2 \leqslant M \,.$$

Definition 13 *A function $\overline{m} : \Delta(\mathcal{A}) \times \Delta(\mathcal{B}) \to \mathcal{S}\bigl(\mathbb{R}^d\bigr)$ is uniformly continuous in its first argument if for every $\varepsilon > 0$, 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 Euclidean norm. Put differently,*

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

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

Uniform continuity in the second argument is defined symmetrically.

Definition 14 *A function $\overline{m} : \Delta(\mathcal{A}) \times \Delta(\mathcal{B}) \to \mathcal{S}\bigl(\mathbb{R}^d\bigr)$ 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 $\alpha \in [0, 1]$,*

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

A function $\overline{m} : \Delta(\mathcal{A}) \times \Delta(\mathcal{B}) \to \mathcal{S}\bigl(\mathbb{R}^d\bigr)$ 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 $\alpha \in [0, 1]$,

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

An example of such a concave–convex function $\overline{m}$ is discussed in Lemma 17.

The following theorem indicates that (SVAC) is the necessary and sufficient condition for the $\overline{m}$–approachability of a closed convex set $\mathcal{C}$ when the payoff function $\overline{m}$ satisfies all four properties stated in Definitions 13 and 14. (Boundedness of $\overline{m}$ indeed follows from the continuity of $\overline{m}$ in each variable.)

Theorem 15 *If $\overline{m}$ is bounded, convex, and uniformly continuous in its second argument, then* (SVAC) *entails that a closed convex set $\mathcal{C}$ is $\overline{m}$–approachable.*

If $\overline{m}$ is concave and uniformly continuous in its first argument, then a closed convex set $\mathcal{C}$ is $\overline{m}$–approachable only if (SVAC) *is satisfied.*

The proof of the necessity statement follows closely the arguments used in the proof of Theorem 8. The sufficiency statement relies on the use of what is called a calibrated strategy, where we define calibration in a (slightly) stronger way than Foster and Vohra (1998) did. All the details, including the definition of the stronger notion of calibration and the construction of an algorithm controlling it, can be found in Appendix A.

5. Approachability in Games with Partial Monitoring

A repeated vector-valued game with partial monitoring is described as follows (see, e.g., Mertens et al., 1994, Rustichini, 1999, and 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 receives 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. We will refer to H as the signaling structure.

Formally, strategies of the first player now associate with $I_1, \ldots, I_{t-1}$ and $S_1, \ldots, S_{t-1}$ a mixed action $\boldsymbol{p}_t \in \Delta(\mathcal{I})$, according to which I_t is drawn independently. We do not impose any restriction on the opponent player, who enjoys a full monitoring: strategies of his associate with $I_1, \ldots, I_{t-1}$, with $J_1, \ldots, J_{t-1}$ and with $S_1, \ldots, S_{t-1}$ a mixed action $\boldsymbol{q}_t \in \Delta(\mathcal{I})$, according to which J_t is drawn independently.

Example 1 *Examples of such partial monitoring games are provided by, e.g., Cesa-Bianchi et al. (2006), 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 a bound on the norm of (the linear extension of) r,

$$R = \max_{(i,j) \in \mathcal{I} \times \mathcal{J}} \left\| r(i,j) \right\|_2 . \tag{5}$$

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 $N_{\mathcal{H}}$.

The definition of approachability can be extended from the setting of full information to the setting of partial monitoring as follows. The only new ingredient is the signaling structure H, the aim is unchanged.

Definition 16 *Let $\mathcal{C} \subseteq \mathbb{R}^d$ be some set; $\mathcal{C}$ is r–approachable by the first player for the signaling structure H if he has a strategy such that, for all $\varepsilon > 0$, there exists an integer T_ε such that for all strategies of the second player,*

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

In particular, the first player has a strategy that ensures that the sequence of his average vector-valued payoffs converges almost surely 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.

Here again, more precise approachability guarantees than the ones required by the definition will be obtained. Indeed, Corollary 27 exhibits bounds of the following form, for a suitable strategy of the first player. For all strategies of the second player and for all $T \geqslant 1$, with probability at least $1 - P_T$,

$$\sup_{\tau \geqslant T} \inf_{c \in \mathcal{C}} \left\| c - \frac{1}{\tau} \sum_{t=1}^{\tau} r(I_t, J_t) \right\|_2 \leqslant R_T ,$$

where $P_T = O(1/T)$ and $R_T = O\big(T^{-1/5} \ln T\big)$.

Our contributions to approachability in games with partial monitoring: A necessary and sufficient condition for r–approachability with the signaling structure H was already stated and proved by Perchet (2011a), together with an approaching strategy. We therefore need to detail where our contribution lies.

First, our strategy is efficient (as long as some projection operator can be computed efficiently, e.g., in the case when the target set is a polytope, see Sections 5.4.1–5.4.2 as well as in the cases of external and internal regret minimization described below in Section 6). In contrast, the one of Perchet (2011a) relies on auxiliary strategies that are calibrated and that require a grid that is progressively refined (leading to a step complexity that is prohibitive in the number T of past steps and to rates of convergence that become dependent on the dimension). The latter construction is in essence the one used in Section 4.

Second, we are able, for the first time, to exhibit convergence rates that are independent of the dimension (as is the case with full monitoring). A somehow related result appeared in Perchet (2011b), but only for the special case of regret minimization. The proof techniques used therein are involved and hold only for regret minimization, not for general approachability.

Third, as far as elegance is concerned, our proof of the sufficiency of the condition for r–approachability with the signaling structure H is short, compact, and more direct than the one of Perchet (2011a) or even of Perchet (2011b), which relied on several layers of concepts (for example, calibration or internal regret in games with partial monitoring).

5.1 Statement of the Necessary and Sufficient Condition for Approachability

To recall the mentioned approachability condition of Perchet (2011a) 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 $\mathcal{H}$

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

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

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

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

$$\overline{m}(\boldsymbol{p}, \sigma) = \big\{ r(\boldsymbol{p}, \boldsymbol{q}_{\mathrm{eqv}}) : \quad \boldsymbol{q}_{\mathrm{eqv}} \in \Delta(\mathcal{J}) \text{ such that } \widetilde{H}(\boldsymbol{q}_{\mathrm{eqv}}) = \sigma \big\}\,.$$

We use in $\boldsymbol{q}_{\mathrm{eqv}}$ the subscript "eqv" (standing for "equivalent") as all considered $\boldsymbol{q}_{\mathrm{eqv}}$ vectors induce the same distributions of signals σ and are thus equivalent from the monitoring perspective.

The necessary and sufficient condition exhibited by Perchet (2011a) 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

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

The subsequent sections show (in a constructive way, with a strategy efficient up to a projection oracle) 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 (2011a).

5.2 Links with Set-Valued Approachability

As will become clear in the proof of Theorem 24, the key in our problem will be to ensure the set-valued approachability of $\mathcal{C}$ with the following non-linear set-valued payoff function, that is however concave–convex in the sense of Definition 14.

Lemma 17 *The function*

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

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

Proof For the concavity part, we consider some pair $\boldsymbol{p}, \boldsymbol{p}' \in \Delta(\mathcal{I})$, some $\boldsymbol{q} \in \Delta(\mathcal{J})$ and some $\alpha \in [0,1]$. By the linearity of r, the elements of the set of interest can be written as

$$\begin{aligned} &\overline{m}\Big(\alpha \boldsymbol{p} + (1-\alpha)\boldsymbol{p}',\, \widetilde{H}(\boldsymbol{q})\Big) \\ = &\Big\{\alpha\, r(\boldsymbol{p}, \boldsymbol{q}_{\mathrm{eqv}}) + (1-\alpha) r(\boldsymbol{p}', \boldsymbol{q}_{\mathrm{eqv}}) : \quad \boldsymbol{q}_{\mathrm{eqv}} \in \Delta(\mathcal{J}) \text{ such that } \widetilde{H}(\boldsymbol{q}_{\mathrm{eqv}}) = \widetilde{H}(\boldsymbol{q})\Big\}\,. \end{aligned}$$

This set is therefore indeed included in (but in general, not equal to)

$$\begin{aligned} &\alpha\, \overline{m}\Big(\boldsymbol{p},\, \widetilde{H}(\boldsymbol{q})\Big) + (1-\alpha)\, \alpha\, \overline{m}\Big(\boldsymbol{p},\, \widetilde{H}(\boldsymbol{q})\Big) \\ = \quad &\alpha \Big\{r(\boldsymbol{p}, \boldsymbol{q}_{\mathrm{eqv}}) : \quad \boldsymbol{q}_{\mathrm{eqv}} \in \Delta(\mathcal{J}) \text{ such that } \widetilde{H}(\boldsymbol{q}_{\mathrm{eqv}}) = \widetilde{H}(\boldsymbol{q})\Big\} \\ &+ (1-\alpha) \Big\{r(\boldsymbol{p}, \boldsymbol{q}'_{\mathrm{eqv}}) : \quad \boldsymbol{q}'_{\mathrm{eqv}} \in \Delta(\mathcal{J}) \text{ such that } \widetilde{H}(\boldsymbol{q}'_{\mathrm{eqv}}) = \widetilde{H}(\boldsymbol{q})\Big\}\,. \end{aligned}$$

Similarly, for the convexity part, we consider some pair $\boldsymbol{q}, \boldsymbol{q}' \in \Delta(\mathcal{J})$, some $\boldsymbol{p} \in \Delta(\mathcal{I})$ and some $\alpha \in [0,1]$. Elements of the convex combination of sets

$$\alpha\, \overline{m}\Big(\boldsymbol{p},\, \widetilde{H}(\boldsymbol{q})\Big) + (1-\alpha)\, \overline{m}\Big(\boldsymbol{p},\, \widetilde{H}(\boldsymbol{q}')\Big)$$

are of the form

$$\alpha\, r(\boldsymbol{p}, \boldsymbol{q}_{\mathrm{eqv}}) + (1-\alpha)\, r(\boldsymbol{p}, \boldsymbol{q}'_{\mathrm{eqv}}) = r\big(\boldsymbol{p},\, \alpha \boldsymbol{q}_{\mathrm{eqv}} + (1-\alpha)\boldsymbol{q}'_{\mathrm{eqv}}\big)\,,$$

where $\boldsymbol{q}_{\mathrm{eqv}}$ and $\boldsymbol{q}'_{\mathrm{eqv}}$ are such that

$$\widetilde{H}(\boldsymbol{q}_{\mathrm{eqv}}) = \widetilde{H}(\boldsymbol{q}) \qquad \text{and} \qquad \widetilde{H}(\boldsymbol{q}'_{\mathrm{eqv}}) = \widetilde{H}(\boldsymbol{q}')\,. \tag{6}$$

In particular, by linearity of $\widetilde{H}$, we have

$$\widetilde{H}\big(\alpha \boldsymbol{q}_{\text{eqv}} + (1-\alpha)\boldsymbol{q}'_{\text{eqv}}\big) = \widetilde{H}\big(\alpha \boldsymbol{q} + (1-\alpha)\boldsymbol{q}'\big)\,,$$

which shows that

$$r\big(\boldsymbol{p},\, \alpha \boldsymbol{q}_{\text{eqv}} + (1-\alpha)\boldsymbol{q}'_{\text{eqv}}\big) \in \overline{m}\Big(\boldsymbol{p},\, \widetilde{H}\big(\alpha \boldsymbol{q} + (1-\alpha)\boldsymbol{q}'\big)\Big)\,.$$

The desired inclusion

$$\alpha\,\overline{m}\Big(\boldsymbol{p},\, \widetilde{H}\big(\boldsymbol{q}\big)\Big) + (1-\alpha)\,\overline{m}\Big(\boldsymbol{p},\, \widetilde{H}\big(\boldsymbol{q}'\big)\Big) \subseteq \overline{m}\Big(\boldsymbol{p},\, \widetilde{H}\big(\alpha \boldsymbol{q} + (1-\alpha)\boldsymbol{q}'\big)\Big)$$

follows. Note that this inclusion is not an equality in general, as it cannot be guaranteed that any $\boldsymbol{q}''_{\text{eqv}}$ such that

$$\widetilde{H}\big(\alpha \boldsymbol{q} + (1-\alpha)\boldsymbol{q}'\big) = \widetilde{H}(\boldsymbol{q}''_{\text{eqv}})$$

can be decomposed under the form $\alpha \boldsymbol{q}_{\text{eqv}} + (1-\alpha)\boldsymbol{q}'_{\text{eqv}}$, where $\boldsymbol{q}_{\text{eqv}}$ and $\boldsymbol{q}'_{\text{eqv}}$ satisfy (6). ■

Unfortunately, efficient strategies for set-valued approachability were only proposed in the linear case (Section 3), not in the concave–convex case (Section 4), and the proof of Lemma 17 shows that linearity cannot be guaranteed per se. However, 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 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. More formally, 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 / ♣	-1 / ♣	2 / ♡
B	-1 / ♣	1 / ♣	3 / ♡

In this example we only study the mapping $\boldsymbol{p} \mapsto \overline{m}(\boldsymbol{p}, ♣)$ 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

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

that can be restated as

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

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 on the subset of distributions with $\lambda \leqslant 1/2$, so that we 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 21 for a more formal statement).

5.3 A Particular Class of Games, Encompassing Regret Minimization

In this section we consider the case where the payoff function and the signaling structure have some special properties described below (linked to linearity properties on lifted spaces and called “bi-piecewise linearity”) and that can be exploited to get efficient strategies. The case of general games with partial monitoring is then considered in Section 5.4 but the particular class of games considered here is already rich enough to encompass the minimization of external and internal regret, as will be seen in Section 6.

To define bi-piecewise linearity of a game with partial monitoring, 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 the so-called piecewise linearity of the thus-defined mappings $\overline{m}(\,\cdot\,, b)$, we then describe an efficient strategy for approachability (up to a projection oracle) followed by convergence rate guarantees.

Definition 18 *Let P be a polytope and let $\mathcal{X}$ be a convex set. A mapping $f : P \to \mathcal{X}$ is piecewise linear if f is continuous and*

- *there exist finitely many sub-polytopes $P_1, \ldots, P_K$ covering P and such that two different sub-polytopes P_k, $P_{k'}$ have an intersection with empty interior; we call these sub-polytopes a decomposition of P;*

- *f is linear on each sub-polytope P_k.*

5.3.1 Bi-Piecewise Linearity of a Game — a Preliminary Technical Result

Lemma 19 *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 $\bigl(\Phi_b(\sigma)\bigr)_{b \in \mathcal{B}}$.

Proof $\widetilde{H}$ is linear on the polytope $\Delta(\mathcal{J})$; Proposition 2.4 in Rambau and Ziegler (1996) thus implies that its inverse mapping $\widetilde{H}^{-1}$ is a piecewise linear mapping of $\mathcal{F}$ into the set of the subsets of $\Delta(\mathcal{J})$. (Note that the latter set has a structure of a convex set, see Footnote 1.) This means by definition that there exists a finite decomposition of $\mathcal{F}$ into polytopes $P_1, \ldots, P_K$ each on which $\widetilde{H}^{-1}$ is linear. Up to a triangulation (see, e.g., Goodman and O'Rourke, 2004, Chapter 14), 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 $\bigl(\Phi_b(\sigma)\bigr)_{b \in \mathcal{B}}$ 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 $P_1, \ldots, P_K$, 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

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

that implies, by the linearity of r, that

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

which concludes the proof. ■

Remark 20 *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 21) relies on the following additional assumption.

Assumption 1 *We assume that $\overline{m}(\,\cdot\,,b)$ is piecewise linear on $\Delta(\mathcal{I})$ for every $b \in \mathcal{B}$. We then call the corresponding game (r,H) a bi-piecewise linear game.*

Assumption 1 means that for each $b \in \mathcal{B}$ 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 exist finitely many such decompositions to consider, 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 19 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 19, and Assumption 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 21 *For a bi-piecewise linear game, 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 $\sigma \in \mathcal{F}$,*

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

5.3.2 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 8; we therefore first need to prove the existence of such a Ψ.

Lemma 22 *Under Condition* (APM)*, the closed convex set $\mathcal{C}$ is $\overline{\overline{m}}$–approachable.*

Proof We show that Condition (SVAC) in Theorem 8 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

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).

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 $\overline{\overline{m}}$–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 19, compute the finite set $\mathcal{A}$ and the mapping $\Theta : \Delta(\mathcal{I}) \to \Delta(\mathcal{A})$ defined based on Assumption 1, pick an arbitrary $\boldsymbol{\theta}_1 \in \Delta(\mathcal{A})$

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

1. define $\boldsymbol{x}_n = \sum_{a\in\mathcal{A}} \theta_{n,a}\, a$ and $\boldsymbol{p}_n = (1-\gamma)\,\boldsymbol{x}_n + \gamma\,\boldsymbol{u}$; refer to the components of $\boldsymbol{p}_n$ as $(p_{i,n})_{i\in\mathcal{I}}$;
2. 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;
3. form the estimated vector of probability distributions over signals,
$$\widetilde{\sigma}_n = \left(\frac{1}{L} \sum_{t=(n-1)L+1}^{nL} \frac{\mathbb{I}_{\{S_t=s\}}\mathbb{I}_{\{I_t=i\}}}{p_{I_t,n}} \right)_{(i,s)\in\mathcal{I}\times\mathcal{H}} ;$$
4. compute the projection $\widehat{\sigma}_n = P_{\mathcal{F}}\big(\widetilde{\sigma}_n\big)$;
5. 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.

convexity of $\overline{m}$ in its second argument (for the first inclusion), by Lemma 19 (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 (and the notation introduced therein). It forces exploration at a γ rate, as is usual in situations with partial monitoring. One of its key ingredients, that conditionally unbiased estimators are available, is extracted from Lugosi et al. (2008, Section 6): 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

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

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 on $\boldsymbol{p}_n$, J_t, and I_t) and then over the one of I_t. Consequently, concentration arguments show that for L large enough,

$$\widetilde{\sigma}_n = \frac{1}{L}\sum_{t=(n-1)L+1}^{nL} \widehat{H}_t \qquad \text{is close to} \qquad \widetilde{H}\big(\widehat{\boldsymbol{q}}_n\big)\,, \qquad \text{where} \quad \widehat{\boldsymbol{q}}_n = \frac{1}{L}\sum_{t=(n-1)L+1}^{nL} \delta_{J_t}\,. \tag{8}$$

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. (2008). For the convenience of the reader, a self-contained proof is provided in Appendix C.

Lemma 23 *For all $\delta \in (0,1)$, for all blocks $n \geqslant 1$, with probability at least $1-\delta$,*

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

5.3.3 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 a known horizon T. We then obtain a bound holding only for the specific round T. Adaptation to $T \to \infty$ (and the obtention of bounds for all $T \geqslant 1$) are then described in the next section. We recall that R was defined in (5) as a bound in norm on r.

Theorem 24 *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 1. Then, for all strategies of the second player, 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 22) is such that, for all $T \geqslant L+1$, for all $\delta \in (0,1)$, with probability at least $1-\delta$,*

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

In particular, for all $T \geqslant 1$, the choices of $L = \lceil T^{3/5} \rceil$ and $\gamma = T^{-1/5}$ imply that for all strategies of the second player, for all $\delta \in (0,1)$, with probability at least $1-\delta$,

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

for some constant Ξ 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 approaching strategy Ψ, which in turn depends on the respective geometries of $\overline{m}$ and $\mathcal{C}$, as was indicated at the end of Section 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.

$$\frac{1}{T} \sum_{t=1}^{T} r(I_t, J_t) \qquad \text{is close to} \qquad \frac{1}{NL} \sum_{t=1}^{NL} r(I_t, J_t)\,; \tag{9}$$

$$\frac{1}{NL} \sum_{t=1}^{NL} r(I_t, J_t) \qquad \text{is close to} \qquad \frac{1}{N} \sum_{n=1}^{N} r\big(\boldsymbol{p}_n, \widehat{\boldsymbol{q}}_n\big)\,; \tag{10}$$

$$\frac{1}{N} \sum_{n=1}^{N} r\big(\boldsymbol{p}_n, \widehat{\boldsymbol{q}}_n\big) \qquad \text{is close to} \qquad \frac{1}{N} \sum_{n=1}^{N} r\big(\boldsymbol{x}_n, \widehat{\boldsymbol{q}}_n\big)\,; \tag{11}$$

$$\frac{1}{N} \sum_{n=1}^{N} r\big(\boldsymbol{x}_n, \widehat{\boldsymbol{q}}_n\big)$$

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

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

$$\frac{1}{N} \sum_{n=1}^{N} \overline{\overline{m}}\Big(\boldsymbol{\theta}_n, \Phi\Big(\widetilde{H}\big(\widehat{\boldsymbol{q}}_n\big)\Big)\Big) \qquad \text{is close to the set} \qquad \frac{1}{N} \sum_{n=1}^{N} \overline{\overline{m}}\Big(\boldsymbol{\theta}_n, \Phi\big(\widehat{\sigma}_n\big)\Big)\,; \tag{12}$$

$$\frac{1}{N} \sum_{n=1}^{N} \overline{\overline{m}}\Big(\boldsymbol{\theta}_n, \Phi\big(\widehat{\sigma}_n\big)\Big) \qquad \text{is close to the set} \qquad \mathcal{C}\,; \tag{13}$$

where we recall that the notation $\widehat{\boldsymbol{q}}_n$ was defined in (8) 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 21:

$$\begin{aligned}\frac{1}{N}\sum_{n=1}^{N}\sum_{a\in\mathcal{A}}\theta_{n,a}\,\overline{m}\Big(a,\,\widetilde{H}\big(\widehat{\boldsymbol{q}}_n\big)\Big) &= \frac{1}{N}\sum_{n=1}^{N}\sum_{(a,b)\in\mathcal{A}\times\mathcal{B}}\theta_{n,a}\,\Phi_b\Big(\widetilde{H}\big(\widehat{\boldsymbol{q}}_n\big)\Big)\,\overline{m}(a,b)\\ &= \frac{1}{N}\sum_{n=1}^{N}\overline{\overline{m}}\Big(\boldsymbol{\theta}_n,\,\Phi\Big(\widetilde{H}\big(\widehat{\boldsymbol{q}}_n\big)\Big)\Big)\,.\end{aligned}$$

Step 1: Assertion (9)**.** 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 (10)**.** We note that by defining $\mathbb{E}_t$ as the conditional expectation with respect to (I_1,S_1,J_1), $\ldots$, $(I_{t-1},S_{t-1},J_{t-1})$ and J_t, which fixes the values of the distribution $\boldsymbol{p}'_t$ of I_t and the value of J_t, we have

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

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 (1996), we therefore get that with probability at least $1-\delta$,

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

The second inequality comes from $NL=T-k\geqslant T-L$.

Step 3: Assertion (11)**.** Since by definition $\boldsymbol{p}_n=(1-\gamma)\,\boldsymbol{x}_n+\gamma\,\boldsymbol{u}$, we get

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

Step 4: Assertion (12)**.** We fix a given block n. Lemma 23 indicates that with probability $1-\delta$,

$$\left\|\widehat{\sigma}_n-\widetilde{H}\big(\widehat{\boldsymbol{q}}_n\big)\right\|_2\leqslant\sqrt{N_{\mathcal{I}}N_{\mathcal{H}}}\left(\sqrt{\frac{2N_{\mathcal{I}}}{\gamma L}\ln\frac{2N_{\mathcal{I}}N_{\mathcal{H}}}{\delta}}+\frac{1}{3}\frac{N_{\mathcal{I}}}{\gamma L}\ln\frac{2N_{\mathcal{I}}N_{\mathcal{H}}}{\delta}\right).\tag{14}$$

Since Φ is Lipschitz (see Remark 20), with a Lipschitz constant in ℓ^2–norms denoted by κ_Φ, we get that with probability $1-\delta$,

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

3. Note that the martingale increments are bounded in norm by $2R$ in our framework and 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 $n = 1, \ldots, N$ with probability at least $1 - N\delta$. Finally, an application of Lemma 6 shows that

$$\frac{1}{N} \sum_{n=1}^{N} \overline{\overline{m}}\Big(\boldsymbol{\theta}_n,\, \Phi\Big(\widetilde{H}\big(\widehat{\boldsymbol{q}}_n\big)\Big)\Big)$$

is in a ε_T–neighborhood (in ℓ^2–norm) of

$$\frac{1}{N} \sum_{n=1}^{N} \overline{\overline{m}}\Big(\boldsymbol{\theta}_n,\, \Phi\big(\widehat{\sigma}_n\big)\Big)\,,$$

where

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

Step 5: Assertion (13)**.** Since $\mathcal{C}$ is $\overline{\overline{m}}$–approachable and by definition of the choices of the $\boldsymbol{\theta}_n$ in Figure 1, we get by Theorem 8, 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, as used already at the end of step 2, $N \geqslant T/L - 1$.

Conclusion of the proof. The proof is concluded by putting the pieces together, thanks to a triangle inequality. By a union bound, the obtained bound holds however only with probability at least $1 - (N+1)\delta \geqslant 1 - \big((T+L)/L\big)\delta$, where we used $N \leqslant T/L$. The stated bound follows by replacing all occurrences of δ in the previous steps by $\delta L/(T+L)$. ■

5.3.4 Uniform Guarantees over Time

We present here a variant of the strategy of Figure 1 that r–approaches $\mathcal{C}$ for the signaling structure H. This is achieved by making the strategy independent of the horizon T. (The strategy of the previous section depended on the knowledge of T, via suitable choices for L and γ.) Two options could have been worked out: resorting to some "doubling trick" or having the parameters L and γ vary over time. In the latter option, the lengths L_n of blocks n and the exploration rates γ_n used therein are no longer constant but of lengths polynomial in n. We chose the latter option for the sake of elegance and because it relies on a result of independent interest, namely a generalization of Theorem 3 to polynomial averages. We only state this generalization for mixed actions taken and observed, but the adaptation for pure actions follows easily.

Consider the setting of Theorem 3. 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

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

Theorem 25 *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 $T \geqslant 1$,

$$\inf_{c\in\mathcal{C}} \left\| c - \frac{1}{\sum_{t=1}^T t^\alpha} \sum_{t=1}^T t^\alpha\, m(\boldsymbol{x}_t,\boldsymbol{y}_t) \right\|_2 \leqslant 2M \frac{\sqrt{\sum_{t=1}^T t^{2\alpha}}}{\sum_{t=1}^T t^\alpha} \leqslant \frac{2M(\alpha+1)}{\sqrt{T}} . \tag{15}$$

It is interesting to note that the convergence rate is independent of α and is the same as standard approachability ($1/\sqrt{T}$). The proof of this theorem is a slight modification of the proof of Theorem 3 and is hence deferred to Appendix D.

The extension to polynomially weighted averages can also be obtained in the context of set-valued approachability. This is because the key to Theorem 8 is Lemma 10, which indicates that to get set-valued 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 $\overline{\overline{m}}$–approaching strategy of $\mathcal{C}$ with respect to polynomially weighted averages with parameter $\alpha = 3/2$. We call it a time-adaptive version of the strategy of Figure 1; indeed, this choice of α ensures that there are roughly $T^{2/5}$ blocks and that the length of the last one is of the order of $T^{3/5}$. Note that it does not depend anymore on any time horizon T, hence guarantees can be obtained for all T.

Theorem 26 *In the same setting and under the same assumptions as in Theorem 24, 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 strategies of the second player, for all $T \geqslant 1$ and all $\delta \in (0,1)$, with probability at least $1-\delta$,*

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

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

The proof follows closely the one of Theorem 24 and is presented in Appendix E.

Corollary 27 *In the same setting and under the same assumptions as in Theorem 24, the time-adaptive version of the strategy described in Figure 1 (with $L_n = n^\alpha$ and $\gamma_n = n^{-\alpha/3}$ for $\alpha = 3/2$) indeed r–approaches $\mathcal{C}$ for the signalling structure H.*

Proof The strategy at hand is such that for all $T \geqslant 1$, with probability at least $1 - 1/T^2$,

$$\inf_{c\in\mathcal{C}} \left\| c - \frac{1}{T}\sum_{t=1}^{T} r(I_t, J_t) \right\|_2 \leqslant \Xi\big(T^{-1/5}\ln T^3 + T^{-2/5}\ln T^3\big)\,.$$

In particular, a union bound shows that for all $T \geqslant 2$,

$$\sup_{\tau\geqslant T}\,\inf_{c\in\mathcal{C}} \left\| c - \frac{1}{\tau}\sum_{t=1}^{\tau} r(I_t, J_t) \right\|_2 \leqslant R_T \stackrel{\text{def}}{=} \sup_{\tau\geqslant T}\, \Xi\big(\tau^{-1/5}\ln \tau^3 + \tau^{-2/5}\ln \tau^3\big)\,,$$

with probability at least $1-P_T$, where $P_T = \sum_{\tau\geqslant T} 1/\tau^2$. We note that $P_T \to 0$ and $R_T \to 0$ as $T \to \infty$. To see that the definition of approachability is satisfied, given $\varepsilon > 0$, it suffices to define T_ε as the minimal T such that $R_T \leqslant \varepsilon$ and $P_T \leqslant \varepsilon$. ■

5.4 Approachability in the Case of General Games with Partial Monitoring

Unfortunately, as is illustrated in the example below, there exist games with partial monitoring that are not bi-piecewise linear. However, we will show that if Condition (APM) holds there exist strategies with a constant per-round complexity that 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 main part of the paper by re-proving, using a doubling trick, that 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 of inefficiency.

Example 3 *The following game (with the same action and signal sets as in Example 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 $\overline{m}\big((p, 1-p), \clubsuit\big)$ 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 19 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

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

the inclusion $\subseteq$ holds by concavity of $\overline{m}$ in its first argument (Lemma 17) 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. ■

5.4.1 Approachability of the Negative Orthant in General Games

For the sake of simplicity, we start with the case of the negative orthant $\mathbb{R}^d_-$ and prove the following result. Note that for the first time in this general framework, the rates for approachability of polytopes are independent of the dimension (as is the case with full monitoring).

Theorem 28 *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.*

Our argument is based on Lemma 19; 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 $k \in \{1, \ldots, d\}$, 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 $b \in \mathcal{B}$,

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

We then linearly extend this mapping to 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 19 (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 29 *The set-valued mappings $\breve{m}$ and $\overline{m}$ satisfy that for all $p \in \Delta(\mathcal{I})$ and $\sigma \in \mathcal{F}$,*

1. *the inclusion $\overline{m}(\boldsymbol{p}, \sigma) \subseteq \breve{m}(\boldsymbol{p}, \sigma)$ holds;*
2. *if $\overline{m}(\boldsymbol{p}, \sigma) \subseteq \mathbb{R}^d_-$, then one also has $\breve{m}(\boldsymbol{p}, \sigma) \subseteq \mathbb{R}^d_-$.*

The interpretation of these two properties is: 1. $\breve{m}$–approaching a set $\mathcal{C}$ is more difficult than $\overline{m}$–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 the first property, 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 19, 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 the second property, it suffices to work component-wise. Note that (by Lemma 19 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 $k \in \{1,\ldots,d\}$; 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.

Proof [of Theorem 28] The assumption of the theorem and the second property of Lemma 29 imply that Condition (APM) holds for $\mathbb{R}_-^d$ and $\breve{m}$. Furthermore, the latter corresponds to a bi-piecewise linear game, i.e., Assumption 1 is satisfied. Indeed, we show below that each $\widetilde{m}_k(\,\cdot\,,b)$ is a piecewise linear function. As a consequence, each $\breve{m}(\,\cdot\,,b)$ is also a piecewise linear function.

Each $\widetilde{m}_k(\,\cdot\,,b)$ is piecewise linear since $\widetilde{m}_k$ is based on the scalar payoff function r_k. Indeed, since $\widetilde{H}$ is linear, the set

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

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

$$\begin{aligned}\widetilde{m}_k(\boldsymbol{p},b) &= \Big\{r_k(\boldsymbol{p},\boldsymbol{q}): \quad \boldsymbol{q} \in \Delta(\mathcal{J}) \text{ such that } \widetilde{H}(\boldsymbol{q}) = b\Big\} \\ &= \mathrm{co}\Big\{r_k(\boldsymbol{p},\boldsymbol{q}_{b,1}),\,\ldots,\,r(\boldsymbol{p},\boldsymbol{q}_{b,M})\Big\} = \left[\min_{k\in\{1,\ldots,M\}} r_k(\boldsymbol{p},\boldsymbol{q}_{b,k})\,,\quad \max_{k'\in\{1,\ldots,M\}} r_k(\boldsymbol{p},\boldsymbol{q}_{b,k'})\right], \end{aligned} \tag{16}$$

where co stands for the convex hull. Since all mappings $r_k(\,\cdot\,,\boldsymbol{q}_{b,k})$ are linear, their minimum and their maximum are piecewise linear functions, therefore $\widetilde{m}_k(\,\cdot\,,b)$ is also piecewise linear.

Therefore, the steps between Equations (11)–(13) of the proof of Theorem 24 (or the corresponding statements in the proof of Theorem 26) can be adapted by replacing $\overline{m}$ and $\overline{\overline{m}}$ by, respectively, $\widetilde{m}$, $\breve{m}$, and its extension corresponding to Definition 21. The result about approachability rates follows.

We now prove that the strategy constructed here is efficient. Indeed, recall that this is always the case as long as the projection onto the associated convex set $\widetilde{\mathcal{C}}$ defined by (4) with the linear function $\overline{\overline{m}}$ is also efficient. But this follows from the fact that as $\mathbb{R}_-^d$ is a polyhedron, the set $\widetilde{\mathcal{C}}$ is a polytope. ■

We now generalize the above ideas to more complex sets.

5.4.2 Approachability of Polytopes for General Games

If 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 $\{(e_k,\ f_k) \in \mathbb{R}^d \times \mathbb{R},\ k \in \mathcal{K}\}$ 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 is a mere exercise of rewriting.

Lemma 30 *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 solving one problem translates to a strategy for solving 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 28 indicates that Condition (APM) for (r,H) and $\mathcal{C}$ is a sufficient condition for the (r,H)–approachability of $\mathcal{C}$ and provides an efficient strategy to do so. (The per-round complexity of this strategy depends in particular at least linearly on the cardinality of $\mathcal{K}$.) Again, rates of convergence are also, for the first time, independent of the dimensions (yet the question of their optimality remains open).

5.4.3 Approachability of General Convex Sets in the Case of General Games

In the above, we provided efficient strategies in the following cases:

– Up to projection oracles, when the games (r,H) are bi-piecewise linear, with no assumption on the target set $\mathcal{C}$; see Section 5.3. This includes at least the minimization of external and internal regret, for which the projections can indeed be performed efficiently; see the upcoming Section 6.

– When the target set $\mathcal{C}$ is a polytope, with no assumption on the game (r,H); see Sections 5.4.1 and 5.4.2.

We only mention the case of general games (r,H) and general closed convex target sets $\mathcal{C}$ in this section to have a complete, self-contained, and constructive proof of the sufficiency of Condition (APM) for (r,H)–approachability. (Perchet, 2011a already proved the latter.)

Theorem 15 and Lemma 17 show that Condition (APM) is indeed sufficient to (r,H)–approach any general closed convex set $\mathcal{C}$. However, the computational complexity of the resulting strategy is much larger as the per-round complexity increases over time. Another way to deal with a general closed convex set is based on the fact that it can be approximated arbitrarily well by a polytope (where the number of vertices of the latter increases as the quality of the approximation does). Playing in regimes approachability strategies of such a sequence of approximations also gives an approachability strategy of the original set $\mathcal{C}$. However, the per-round complexity increases over time (as the numbers of vertices of the approximating polytopes do).

6. Application to Regret Minimization

In this section we analyze external and internal regret minimization in repeated games with partial monitoring from the perspective of approachability. We show how to 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 1) is satisfied.

The results instantiated below are not necessarily new in terms of efficiency or convergence rates, and some are even slightly suboptimal. However, our point is that all previous good strategies were specifically designed for the problem of regret minimization, while we introduced above a general strategy for all approachability problems, including regret minimization. And what we gained in generality (the wider range of problems that we can deal with) has no impact (or little impact only) on the efficiency or on the rates, which we think is an important contribution.

6.1 External Regret

We consider in this section the framework and aim introduced by Rustichini (1999) and studied, sometimes for restricted classes of games, by Piccolboni and Schindelhauer (2001), Mannor and Shimkin (2003), Cesa-Bianchi et al. (2006), Lugosi et al. (2008), Bartók et al. (2010, 2011), Foster and Rakhlin (2012). 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 slight contrast with the notation $\widehat{\boldsymbol{q}}_n$ used in Section 5.3 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 $\sigma \in \mathcal{F}$,

$$\rho(\boldsymbol{p}, \sigma) = \min \Big\{ r(\boldsymbol{p}, \boldsymbol{q}) \,:\, \boldsymbol{q} \text{ such that } \widetilde{H}(\boldsymbol{q}) = \sigma \Big\}\,. \tag{17}$$

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 (more formal arguments are given in Rustichini, 1999). 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.

Rustichini (1999) calls the partial monitoring in the game (r, H) statistically sufficient when

$$\max_{\boldsymbol{p}\in\Delta(\mathcal{I})} r\big(\boldsymbol{p}, \widehat{\boldsymbol{q}}_T\big) = \max_{\boldsymbol{p}\in\Delta(\mathcal{I})} \rho\Big(\boldsymbol{p}, \widetilde{H}\big(\widehat{\boldsymbol{q}}_T\big)\Big).$$

In general, only an inequality $\geqslant$ holds between the two quantities. A line of research initiated by Piccolboni and Schindelhauer (2001) first studied efficient strategies to minimize the regret in the said case of a statistically sufficient monitoring.

6.1.1 A Strategy Minimizing External Regret

The following result is a consequence of Theorem 26, as its proof shows; it corresponds to the main result of Lugosi et al. (2008), with the same convergence rate but with a different strategy. (However, Perchet, 2011b, Section 2.3 exhibited an efficient strategy achieving a convergence rate of order $T^{-1/3}$, which is optimal; this strategy was an ad hoc strategy for regret minimization. Nonetheless, a question that remains open is thus whether the rates exhibited in Theorem 26 could be improved.)

Corollary 31 *The first player can apply the strategy of Theorem 25 such that for all strategies of the second player, for all T and all $\delta \in (0,1)$, with probability at least $1-\delta$,*

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

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

The discussion about the efficiency of the strategy is postponed to the end of this section, as it relies on some objects that will be introduced in the proof of the corollary. The latter proof is an extension to the setting of partial monitoring of the original proof and strategy of Blackwell (1956b) 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 1 is satisfied. To do so, we will use the same arguments as around (16) and 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. But this is true because r takes scalar values, as indicated around (16).

Theorem 26 can then 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 in the last paragraph of this proof.

Now, the regret is non-positive if $\sum_{t=1}^{T} \underline{r}(I_t,J_t)/T$ 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 26.

It only remains to prove the indicated Lipschitz continuity. (All Lipschitz continuity statements that follow will be with respect to the ℓ^2–norms.) We have by Definition 21 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 $\overline{\overline{m}}_1$ indifferently either is 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 20 the mapping $\sigma \in \mathcal{F} \mapsto \Phi(\sigma)$ is κ_Φ–Lipschitz; this entails, by Lemma 6, 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. ∎

6.1.2 Discussion about Efficiency

An argument similar to the one in Perchet (2011b) shows that the convex set $\mathcal{C}$ is defined by a finite number of piecewise linear equations, it is therefore a polyhedron; so that the projection onto it, as well as the computation of the strategy, can be done efficiently. We only sketch here the argument. The argument used when referring to (16) indicates a priori that for each $\sigma \in \mathcal{F}$, there exist a finite number K_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 $\rho(\boldsymbol{p}, \sigma) = \min\bigl\{r(\boldsymbol{p}, \boldsymbol{q}_{\sigma,1}), \ldots, r(\boldsymbol{p}, \boldsymbol{q}_{\sigma,M_\sigma})\bigr\}$. But by an argument stated in Perchet (2011b),

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

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

$$\bigl\{\boldsymbol{q}_{\sigma,1}, \ldots, \boldsymbol{q}_{\sigma,K_\sigma}\bigr\} = \bigl\{\boldsymbol{q}'_{\sigma,1}, \ldots, \boldsymbol{q}'_{\sigma,K}\bigr\}\,.$$

(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 $\sigma \in \mathcal{F}$,

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

Each function $\sigma \mapsto \boldsymbol{q}'_{\sigma,k}$ being piecewise linear, one can construct a finite set $\{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{\tilde{K}}\} \subset \Delta(\mathcal{I})$ 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 a polyhedron. Its lifted image $\widetilde{\mathcal{C}}$ is a then a polytope: thus, the projection onto it, hence the computation of the proposed strategy, can be done efficiently.

6.2 Internal / Swap Regret

Foster and Vohra (1999) 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 (if there are enough such stages). In other words, i is the best response to the empirical distribution of actions 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 in (17). 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 (2007) 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 grid of $\Delta(\mathcal{I})$, which we denote by $\{\boldsymbol{p}_g,\ g \in \mathcal{G}\}$, 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. A discussion about the choice of $\mathcal{G}$ is provided below. For now, we define formally $\mathcal{G}$–internal regret as internal regret with respect to the set of mixed actions $\mathcal{G}$.

Up to a standard concentration-of-the-measure argument, we will measure the payoff at round t with $r(\boldsymbol{p}_{G_t}, J_t)$ 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 $\{1, \dots, T\}$ 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 $\widehat{\boldsymbol{q}}_{T,g}$ in an arbitrary way when $N_T(g) = 0$. The $\mathcal{G}$–internal regret of the first player at round T is measured as

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

Actually, our proof technique rather leads to the minimization of some $\mathcal{G}$–swap regret (see Blum and Mansour, 2007, for the definition of swap regret in full monitoring):

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

At first sight, to handle all possible alternatives one should take $\mathcal{G}$ as a thin grid in $\Delta(\mathcal{I})$, i.e., some ε–discretization of the latter. This is what Lehrer and Solan (2007) do. However, Perchet (2011b) showed that there exists a finite subset $\mathcal{G}_0$ of $\Delta(\mathcal{I})$ such that $\mathcal{G}_0$ contains a best response to any mixed action of the second player: for all $\boldsymbol{q} \in \Delta(\mathcal{J})$,

$$\left(\operatorname*{argmin}_{\boldsymbol{p} \in \Delta(\mathcal{I})} \ \rho\big(\boldsymbol{p},\, \widetilde{H}(\boldsymbol{q})\big) \right) \cap \mathcal{G}_0 \neq \emptyset \,.$$

The strategy we discuss below will have a complexity polynomial in the size of $\mathcal{G}$. We thus advise to take $\mathcal{G} = \mathcal{G}_0$ for the sake of efficiency.

Again, the following bound on the swap regret easily follows from Theorem 24. The latter constructs a simple and direct strategy to control the swap regret, thus also the internal regret. It therefore improves on the results of Lehrer and Solan (2007) and Perchet (2009, 2011b), three papers that presented more involved and less efficient strategies to do so. These strategies were indeed based on auxiliary strategies using thin grids that need to be refined over time; this resulted in complexities that were at least exponential in the number of rounds. (The ideas used therein bear some resemblance with what is done in calibration, see the references provided in Section 4.) In contrast, our strategy can have a constant per-round complexity (when used with the grid $\mathcal{G}_0$). This is a major improvement in efficiency. However, as far as convergence rates are concerned, we must note that again, as in the case of external regret, Perchet (2011b) obtained rates of the faster order $T^{-1/3}$, for an ad hoc (inefficient) strategy. We thus sacrifice efficiency for rates.

Corollary 32 *The first player has an explicit strategy such that for all strategies of the second player, for all T and all $\delta \in (0,1)$, with probability at least $1-\delta$,*

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

for some constant Ξ 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 31; $\mathcal{G}$ 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 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 possesses some smoothness properties by providing a uniform Lipschitz bound on certain functions.

Step 1: We denote by

$$\mathcal{F}_{\mathrm{cone}} = \bigl\{ \lambda\sigma, \quad \sigma \in \mathcal{F}, \ \lambda \in \mathbb{R}_+ \bigr\}$$

the cone generated by $\mathcal{F}$ and extend linearly $\rho : \Delta(\mathcal{I}) \times \mathcal{F} \to \mathbb{R}$ into a mapping $\rho : \Delta(\mathcal{I}) \times \mathcal{F}_{\mathrm{cone}} \to R$ 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 $\mathcal{F}_{\mathrm{cone}}$ 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

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

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

$$\underline{r}(g, j) = \begin{bmatrix} r\bigl(\boldsymbol{p}_g, j\bigr)\, \mathbb{I}_{\{g'=g\}} \\ \widetilde{H}(\delta_j)\, \mathbb{I}_{\{g'=g\}} \end{bmatrix}_{g' \in \mathcal{G}} .$$

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

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

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 $\underline{r}\bigl(g^\star(\boldsymbol{q}), \boldsymbol{q}'\bigr)$ are all null but the ones corresponding to $g^\star(\boldsymbol{q})$, for which we have

$$\begin{aligned} & r\bigl(\boldsymbol{p}_{g^\star(\boldsymbol{q})}, \boldsymbol{q}'\bigr) \\ & \qquad \geqslant \rho\Bigl(\boldsymbol{p}_{g^\star(\boldsymbol{q})}, \widetilde{H}(\boldsymbol{q}')\Bigr) = \rho\Bigl(\boldsymbol{p}_{g^\star(\boldsymbol{q})}, \widetilde{H}(\boldsymbol{q})\Bigr) = \max_{g' \in \mathcal{G}} \rho\Bigl(\boldsymbol{p}_{g'}, \widetilde{H}(\boldsymbol{q})\Bigr) = \max_{g' \in \mathcal{G}} \rho\Bigl(\boldsymbol{p}_{g'}, \widetilde{H}(\boldsymbol{q}')\Bigr) , \end{aligned}$$

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 22 and Theorem 24 is thus satisfied, so that we have approachability.

Step 2: We then show that Assumption 1 is satisfied. It suffices to show that for all $\sigma \in \mathcal{F}$, the mapping

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

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 mapping

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

is seen to be piecewise linear, by using the same one-dimensional argument as the one stated around (16) and also used in the proof of Corollary 31.

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 subset of coordinates. The subset of coordinates of $\sum_{t=1}^T \underline{r}(G_t, J_t)/T$ corresponding to a given g 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

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

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,

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

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,

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

where we used a fact proved in step 4, that the mapping

$$\boldsymbol{v} \in \mathcal{F}_{\text{cone}} \longmapsto \max_{g' \in \mathcal{G}} \rho\big(\boldsymbol{p}_{g'}, \boldsymbol{v}\big) \tag{18}$$

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,

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

where we denoted by $N_\mathcal{G}$ the cardinality of $\mathcal{G}$. Resorting to the convergence rate stated in Theorem 26 concludes the proof.

Step 4: It only remains to prove the claimed Lipschitzness of the mapping (18). (All Lipschitzness statements that follow will be with respect to the ℓ^2–norms.) 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 31 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,

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

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 $\big|\lambda - \lambda'\big|$ can be bounded by $\big\|\, \lambda\sigma - \lambda'\sigma' \,\big\|_2$ 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

$$\big|\lambda - \lambda'\big| = \frac{1}{N_\mathcal{I}}\Big(\lambda\,\|\sigma\|_1 - \lambda'\,\big\|\,\sigma'\,\big\|_1\Big) \leqslant \frac{1}{N_\mathcal{I}}\,\big\|\,\lambda\sigma - \lambda'\sigma'\,\big\|_1 \leqslant \frac{\sqrt{N_\mathcal{H} N_\mathcal{I}}}{N_\mathcal{I}}\,\big\|\,\lambda\sigma - \lambda'\sigma'\,\big\|_2 ,$$

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

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

This concludes the proof. ■

7. Summary of the Results

This paper extended Blackwell's classical approachability theory to the case where set-valued functions are considered, which models ambiguity in the obtained reward. In the case of mixed actions taken, this extension was provided in the case of linear (Section 3) and concave–convex (Section 4) set-valued functions; only in the former case efficient strategies (up to a projection oracle) could be constructed.

The second part of this paper (Section 5) applies this theory of set-valued approachability to approachability with partial monitoring. The necessary and sufficient Condition (APM) for this was exhibited by Perchet (2011a) and was recalled in Section 5.1; its link with the necessary and sufficient condition for set-valued approachability was discussed in Section 5.2. Then, under a so-called assumption of bi-piecewise linearity of the game (r, H) at hand, an efficient strategy (up to a projection oracle) was constructed and studied in Section 5.3, for the approachability of any closed convex set $\mathcal{C}$. Alternatively, Section 5.4 showed that for any game (r, H) at hand but under the constraint that the target set $\mathcal{C}$ is a polytope, the above efficient construction could still be used. In both cases, the novelty also relies not only the gained efficiency with respect to the construction by Perchet (2011a) but also on getting for the first time rates of convergence that are independent of the ambient dimension. The case of any game (r, H) and any closed target set $\mathcal{C}$ was discussed, for the sake of completeness, at then end of Section 5.4, so that the present article contains a complete and self-contained constructive proof of the sufficiency of Condition (APM).

Finally, Section 6 showed that the well-studied case of regret minimization, a special case of approachability, could fall under the umbrella of bi-piecewise linearity, and hence be performed efficiently, as was already known.

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 grants ANR-10-BLAN 0112 and ANR-13-JS01-0004-01. Gilles Stoltz acknowledges support from Investissements d'Avenir (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047).

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.

Appendix A. Proof of Theorem 15

Proof [of the second statement of Theorem 15] The proof of Corollary 7 extends to the case considered here and shows, thanks to the ad hoc consideration of the result stated in Lemma 6 as following from Definition 13, 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 (SVAC) is not satisfied. The first part of the proof of the necessity of (SVAC) in Theorem 8 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 the set-valued approachability can be lower bounded as follows. Thanks to the concavity in the first argument,

$$\sup\left\{\inf_{c\in\mathcal{C}}\ \|\xi-c\|_2:\quad \xi\in\frac{1}{T}\sum_{t=1}^{T}\overline{m}(\boldsymbol{x}_t,\boldsymbol{y}_0)\right\}$$

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

Therefore, $\mathcal{C}$ is $\overline{m}$–approachable by no strategy of the first player. ■

The proof of the first statement of Theorem 15 relies on the use of approximately calibrated strategies of the first player, as introduced and studied (among others) by Dawid (1982), Foster and Vohra (1998), Mannor and Stoltz (2010). 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 $\{\boldsymbol{y}^1,\ldots,\boldsymbol{y}^{N_\eta}\}$. We thus denote by L_t the index chosen in $\{1,\ldots,N_\eta\}$ 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 $\delta\in(0,1)$, there exists an integer T_δ such that for all strategies of the second player, with probability at least $1-\delta$, for all $T\geqslant T_\delta$,

$$\sum_{\ell=1}^{N_\eta}\frac{N_T(\ell)}{T}\left(\left\|\boldsymbol{y}^\ell-\frac{1}{N_T(\ell)}\sum_{t=1}^{T}\boldsymbol{y}_t\mathbb{I}_{\{L_t=\ell\}}\right\|_1-\eta\right)_+\leqslant\delta\,. \tag{19}$$

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 $T\geqslant T_\delta$,

$$\sum_{\ell=1}^{N_\eta}\frac{N_T(\ell)}{T}\left\|\boldsymbol{y}^\ell-\frac{1}{N_T(\ell)}\sum_{t=1}^{T}\boldsymbol{y}_t\mathbb{I}_{\{L_t=\ell\}}\right\|_1\leqslant\eta+\delta\,. \tag{20}$$

The existence of a calibrated strategy in the sense of (19) however follows from the same approachability-based construction studied in Mannor and Stoltz (2010) to get (20) and is detailed below in Section B. 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 $T\geqslant T_\delta$,

$$\max_{\ell=1,\ldots,N_\eta}\frac{N_T(\ell)}{T}\left(\left\|\boldsymbol{y}^\ell-\frac{1}{N_T(\ell)}\sum_{t=1}^{T}\boldsymbol{y}_t\mathbb{I}_{\{L_t=\ell\}}\right\|_1-\eta\right)_+\leqslant\delta\,. \tag{21}$$

Proof [of the first statement of Theorem 15] The insight of this proof is similar to the one illustrated in Perchet (2009). 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 $\overline{m}$–approachable. This is so up to proceeding in regimes $r = 1, 2, \ldots$ 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 (SVAC) 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 $\overline{m}\big(\boldsymbol{x}^\ell, \boldsymbol{y}^\ell\big) \subseteq \mathcal{C}$. 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 forecasted 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 forecasted 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

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

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

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

We recall that $\boldsymbol{B}$ denotes the unit Euclidean ball in $\mathbb{R}^d$. To show that the above-defined strategy $\overline{m}$–approaches $\mathcal{C}_\varepsilon = \mathcal{C} + \varepsilon \boldsymbol{B}$, it suffices to show that for all $\delta \in (0,1)$, there exists an integer T'_δ such that for all strategies of the second player,

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

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 $\delta' = \delta(\eta/2)\big/\big(M\, N_{\eta/2}\big)$ and define T'_δ as the time $T_{\delta'}$ corresponding to (21). 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 (21), $\left\| \boldsymbol{y}^\ell - \overline{\boldsymbol{y}}^\ell_T \right\| \leqslant \eta/2 + \delta' T / N_T(\ell) \leqslant \eta$; since η is the modulus of continuity for ε, we get that

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

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

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

Summing these bounds over ℓ yields

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

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

Appendix B. An Auxiliary Result of Calibration

We prove here (19) for a given $\eta > 0$ and do so by following closely the methodology of Mannor and Stoltz (2010). (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 (19) 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 $\left\| \boldsymbol{y} - \boldsymbol{y}^\ell \right\|_1 \leqslant \eta/N_\mathcal{B}$.

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. 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 $\left\| \boldsymbol{y} - \boldsymbol{y}^\ell \right\|_1 \leqslant \eta/N_{\mathcal{B}}$; such an index always exists as noted at the beginning of this proof. Indeed, one then has in particular that for each component $b \in \mathcal{B}$,

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

A straightforward adaptation of the proof of Theorem 3 (see, e.g., Mertens et al., 1994) then yields a strategy such that for all $\delta \in (0,1)$ and for all strategies of the second player, with probability at least $1-\delta$,

$$\sup_{\tau \geqslant T} \inf_{c \in (\mathbb{R}_-)^{2N_\eta N_{\mathcal{B}}}} \left\| c - \frac{1}{\tau}\sum_{t=1}^{\tau} C(L_t, \boldsymbol{y}_t) \right\|_2 \leqslant 2M\sqrt{\frac{2}{\delta T}}\,, \tag{22}$$

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

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

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

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

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

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

where the last inequality holds, by (22), 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 (19).

Appendix C. Proof of Lemma 23

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

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

The Bernstein-Freedman inequality (see Freedman, 1975 or Cesa-Bianchi et al., 2006, 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. ■

Appendix D. Proof of Theorem 25

Proof We denote by d_t^α the squared distance of $\widehat{m}_t^\alpha$ to $\mathcal{C}$,

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

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

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

where we used in the third inequality the same convex projection inequality as in the proof of Theorem 3.

The first inequality in (15) 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

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

The second inequality in (15) is straightforward for $\alpha = 0$ and is proved for $\alpha > 0$ as follows. First, by comparing sums and integrals, we get that for all $t \geqslant 1$,

$$\frac{t^{\alpha+1}}{\alpha+1} = \int_0^t s^\alpha \, \mathrm{d}s \leqslant \sum_{s=1}^t s^\alpha \leqslant t \times t^\alpha = t^{\alpha+1}.$$

Therefore,

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

This concludes the proof. Note for later purposes that upper bounding above the sum of the s^α as

$$\sum_{s=1}^t s^\alpha \leqslant t^\alpha + \int_1^t s^\alpha \, \mathrm{d}s \leqslant t^\alpha + \frac{t^{\alpha+1}}{\alpha+1}$$

shows that

$$\sum_{s=1}^{t} s^{\alpha} \sim \frac{t^{\alpha+1}}{\alpha+1}\,. \qquad \blacksquare$$

Appendix E. Proof of Theorem 26

Proof The proof follows closely the proof of Theorem 24. 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 $1-\delta$,

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

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

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

The same argument as the one at the beginning of the proof of Theorem 24 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 23 together with the Lipschitzness of Φ to get that for all regimes $n = 1, \ldots, N$, with probability at least $1-\delta$,

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

By a union bound, the above bound holds for all regimes $n = 1, \ldots, N$ with probability at least $1 - N\delta$. Then, an application of Lemma 6 shows that

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

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

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

It then suffices, as in step 5 of the original proof, to write the convergence rates for set-valued approachability guaranteed by the strategy Ψ. By combining the result of Lemma 10 with Theorem 25 and Lemma 6, we get

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

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

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

Since (as proved at the end of the proof of Theorem 25) $T_N^\beta \sim N^{\beta+1}/(\beta+1)$ 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. Replacing δ by $\delta/(N+1)$ as we did in step 5 of the proof of Theorem 24, choosing $\alpha = 3/2$ and substituting the equivalences above ensures the result. ■

References

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)*. JMLR Workshop and Conference Proceedings, 2011.

G. Bartók, D. Pál, and C. Szepesvári. Toward a classification of finite partial-monitoring games. In *Proceedings of the Twenty-First International Conference on Algorithmic Learning Theory (ALT'10)*, pages 224–238. Springer, 2010.

G. Bartók, D. Pál, and C. Szepesvári. Minimax regret of finite partial-monitoring games in stochastic environments. In *Proceedings of the Twenty-Fourth Annual Conference on Learning Theory (COLT'11)*. JMLR Workshop and Conference Proceedings, 2011.

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

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

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

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

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

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

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

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

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

D.P. Foster and A. Rakhlin. No internal regret via neighborhood watch. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics (AISTATS'12)*, 2012.

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

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.

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

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

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

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.

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.

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

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

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

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

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.

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

V. Perchet. Internal regret with partial monitoring calibration-based optimal algorithms. *Journal of Machine Learning Research*, 12:1893–1921, 2011b.

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

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.

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

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

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