Robust approachability and regret minimization in games with partial monitoring

Shie Mannor SHIE@EE.TECHNION.AC.IL
Technion, Haifa
Israel

Vianney Perchet VIANNEY.PERCHET@NORMALESUP.ORG
Ecole normale supérieure, Cachan
France

Gilles Stoltz GILLES.STOLTZ@ENS.FR
Ecole Normale Supérieure – CNRS – INRIA, Paris
France
&
HEC Paris – CNRS, Jouy-en-Josas
France

Editor: Sham Kakade, Ulrike von Luxburg

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 that belongs to a set, rather than being a single vector. Using this variant we tackle the problem of approachability in games with partial monitoring and develop simple and efficient algorithms (i.e., with constant per-step complexity) for this setup. We finally consider external and internal regret in repeated games with partial monitoring, for which we derive regret-minimizing strategies based on approachability theory.

Keywords: Approchability, partial monitoring, regret, adversarial learning

1. Introduction

Blackwell's approachability theory and its variants has become a standard and useful tool in analyzing online learning algorithms (Cesa-Bianchi and Lugosi, 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 himself in Blackwell (1956b). Numerous other contributions are summarized in Cesa-Bianchi and Lugosi (2006). Blackwell's approachability theory enjoys a clear geometric interpretation that allows it to be used in situations where online convex optimization or exponential weights do not seem to be easily applicable and, in some sense, to go beyond the minimization of the regret and/or to control quantities of a different flavor; e.g., in Mannor et al. (2009), to minimize the regret together with path constraints, and in Mannor and Shimkin (2008), to minimize the regret in games whose stage duration is not fixed. Recently, it has been shown that approachability and low regret learning are equivalent in the sense that efficient reductions exist from one to the other (Abernethy et al., 2011). Another recent paper (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 Nature. In partial monitoring the decision maker does not know how much reward was obtained and only gets a (random) signal whose distribution depends on the action of the decision maker and the action of Nature. There are two extremes of this setup that are well studied. On the one extreme we have the case where the signal includes the reward itself (or a signal that can be used to unbiasedly estimate the reward), which is essentially the celebrated bandits setup. The other extreme is the case where the signal is not informative (i.e., it tells the decision maker nothing about the actual reward obtained); this setting then essentially consists of repeating the same situation over and over again, as no information is gained over time. We consider a setup encompassing these situations and more general ones, in which the signal is indicative of the actual reward, but is not necessarily a sufficient statistics thereof. The difficulty is that the decision maker cannot compute the actual reward he obtained nor the actions of Nature.

Regret minimization with partial monitoring has been studied in several papers in the learning theory community. Piccolboni and Schindelhauer (2001); Mannor and Shimkin (2003); Cesa-Bianchi et al. (2006) study special cases where an accurate estimation of the rewards (or worst-case rewards) of the decision maker is possible thanks to some extra structure. A general policy with vanishing regret is presented in 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 provide approachability-based results for the problem of regret minimization. On route, we define a new type of approachability setup, with enables to re-derive the extension of approachability to the partial monitoring vector-valued setting proposed by Perchet (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, as in Blackwell (1956b) but for the first time in this general framework, independent of the game at hand. The paper is organized as follows. In Section 2 we recall some basic facts from approachability theory. In Section 3 we propose a novel setup for approachability, termed "robust approachability," where instead of obtaining a vector-valued reward, the decision maker obtains a set, that represents the ambiguity concerning his reward. We provide a simple characterization of approachable convex sets and an algorithm for the set-valued reward setup. In Section 4 we show how to apply the robust approachability framework to the repeated vector-valued games with partial monitoring. We start in Section 4.1 with the case where the signaling structure is bi-piecewise linear. For this important special case, we provide a simple and constructive algorithm. Previous results for approachability in this setup were either non-constructive (Rustichini, 1999) or were highly inefficient as they relied on some sort of lifting to the space of probability measures on mixed actions (Perchet, 2011a) and typically required a grid that is progressively refined (leading to a step complexity that is exponential in the number T of past steps). In Section 4.2 we apply our results for both external and internal regret minimization with partial monitoring.

In both cases our proofs are simple, lead to algorithms with constant complexity at each step, and are accompanied with rates. Our results for external regret have rates similar to Lugosi et al. (2008), but our proof is direct and simpler. For internal regret minimization we present the first algorithm not relying on a grid being refined over time and the first convergence rates. In Section 4.3 we mention the general signaling case and explain how it is possible to approach certain special sets such as polytopes efficiently and general convex sets inefficiently.

2. Some basic facts from approachability theory

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

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

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

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

Definition 1 *A set $\mathcal{C} \subseteq \mathbb{R}^d$ is m–*approachable with pure actions *if there exists a strategy*[1] *of the first player such that for all strategies of the second player,*

$$\limsup_{T\to\infty} \quad \inf_{c\in\mathcal{C}} \left\| c - \frac{1}{T}\sum_{t=1}^{T} m\bigl(A_t, B_t\bigr) \right\|_2 = 0 \qquad a.s.$$

That is, the first player has a strategy that ensures that the average of his vector-valued payoffs converges to the set $\mathcal{C}$.

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

[1] The original definition given by Blackwell requires uniformity w.r.t. the strategy set of the opponent. We ignore the uniformity to avoid excessive nomenclature.

Definition 2 *In this context, a set $\mathcal{C} \subseteq \mathbb{R}^d$ is m–*approachable with mixed actions *if there exists a strategy of the first player such that for all strategies of the second player,*

$$\limsup_{T\to\infty} \quad \inf_{c\in\mathcal{C}} \left\Vert c - \frac{1}{T}\sum_{t=1}^{T} m\big(\boldsymbol{x}_t, \boldsymbol{y}_t\big) \right\Vert_2 = 0 \qquad a.s.$$

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

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

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

In the latter case, an explicit strategy achieves the following convergence rates. We denote by M a bound in norm over m, i.e.,

$$\max_{(a,b)\in\mathcal{A}\times\mathcal{B}} \big\Vert\, m(a,b)\, \big\Vert_2 \leqslant M\,.$$

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

$$\inf_{c\in\mathcal{C}} \left\Vert c - \frac{1}{T}\sum_{t=1}^{T} m\big(\boldsymbol{x}_t, \boldsymbol{y}_t\big) \right\Vert_2 \leqslant \frac{2M}{\sqrt{T}}\,.$$

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

$$\inf_{c\in\mathcal{C}} \left\Vert c - \frac{1}{T}\sum_{t=1}^{T} m\big(A_t, B_t\big) \right\Vert_2 \leqslant \frac{2M}{\sqrt{T}}\Big(1 + 2\sqrt{\ln(2/\delta)}\Big).$$

The proof is standard and is omitted from this article; it is detailed in the extended version of this paper (Mannor et al., 2011a).

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

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

$$\widehat{m}_t = \frac{1}{t}\sum_{\tau=1}^{t} m(\boldsymbol{x}_\tau, B_\tau) \qquad \text{or} \qquad \widehat{m}_t = \frac{1}{t}\sum_{\tau=1}^{t} m(\boldsymbol{x}_\tau, \boldsymbol{y}_\tau)\,,$$

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

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

where $\langle \cdot\,, \cdot \rangle$ is the Euclidian inner product in $\mathbb{R}^d$. The minimax problem above is easily seen to be a (scalar) zero-sum game and is therefore efficiently solvable using, e.g., linear programming: the associated complexity is polynomial in $N_{\mathcal{A}}$ and $N_{\mathcal{B}}$. All in all, this strategy is efficient as soon as the computations of the required projections onto $\mathcal{C}$ in ℓ^2–norm can be performed efficiently.

In the case when pure actions are taken and observed, it only remains to draw A_{t+1} at random according to $\boldsymbol{x}_{t+1}$.

3. Robust approachability

In this section we extend the results of the previous section to set-valued payoff functions. To this end, we denote by $\mathcal{S}\bigl(\mathbb{R}^d\bigr)$ the set of all subsets of $\mathbb{R}^d$ and consider a set-valued payoff function $\overline{m} : \mathcal{A} \times \mathcal{B} \to \mathcal{S}\bigl(\mathbb{R}^d\bigr)$.

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

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

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

be such a joint probability distribution; then $\overline{m}(\mu)$ is defined as a finite convex combination[2] of subsets of $\mathbb{R}^d$,

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

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

We denote by

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

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

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

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

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

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

Definition 4 *Let $\mathcal{C} \subseteq \mathbb{R}^d$ be some set; $\mathcal{C}$ is $\overline{m}$–approachable with pure actions if there exists a strategy of the first player such that for all strategies of the second player,*

$$\limsup_{T\to\infty} \sup_{d\in\overline{m}(\pi_T)} \inf_{c\in\mathcal{C}} \|c-d\|_2 = 0 \qquad a.s.$$

That is, when $\mathcal{C}$ is $\overline{m}$–approachable with pure actions, the first player has a strategy that ensures that the average of the sets of payoffs converges to the set $\mathcal{C}$: the sets $\overline{m}(\pi_T)$ are included in ε_T–neighborhoods of $\mathcal{C}$, where the sequence of ε_T tends almost-surely to 0.

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

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

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

and will be interested in the behavior of

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

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

Definition 5 *Let $\mathcal{C} \subseteq \mathbb{R}^d$ be some set; $\mathcal{C}$ is $\overline{m}$–approachable with mixed actions if there exists a strategy of the first player such that for all strategies of the second player,*

$$\limsup_{T\to\infty} \sup_{d\in\overline{m}(\nu_T)} \inf_{c\in\mathcal{C}} \|c-d\|_2 = 0 \qquad a.s.$$

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

Lemma 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 norm by M, i.e., that there exists a real number $M > 0$ such that*

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

Then

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

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

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

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

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

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

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

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

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

Necessary and sufficient condition for approachability. We state the condition in the theorem below, as well as the associated convergence rates. Explicit strategies can be deduced from the proof, which is based on Theorem 3; these strategies are efficient as soon as projections in ℓ^2–norm onto the set $\widetilde{\mathcal{C}}$ defined in (3) can be computed efficiently. The latter fact depends on the respective geometries of $\overline{m}$ and $\mathcal{C}$.

Theorem 7 *Suppose that the set-valued function $\overline{m}$ is bounded in norm by M. A closed convex set $\mathcal{C} \subseteq \mathbb{R}^d$ is approachable (with pure or mixed actions) if and only if the following robust approachability condition is satisfied,*

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

In the latter case, the following convergence rates are achieved by a strategy constructed in the proof. With mixed actions taken and observed, for all strategies of the second player, with probability 1,

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

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

$$\sup_{d \in \overline{m}(\pi_T)} \inf_{c \in \mathcal{C}} \|c - d\|_2 \leqslant \frac{2M}{\sqrt{T}} \sqrt{N_{\mathcal{A}} N_{\mathcal{B}}} \Bigl(1 + 2\sqrt{\ln(2/\delta)}\Bigr).$$

Proof that Condition (RAC) **is necessary.** 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 then define a mapping $D : \Delta(A) \to \mathbb{R}$ by

$$\forall \, \boldsymbol{x} \in \Delta(\mathcal{A}), \qquad D(\boldsymbol{x}) = \sup_{d \in \overline{m}(\boldsymbol{x}, \boldsymbol{y}_0)} \inf_{c \in \mathcal{C}} \|c - d\|_2 \,.$$

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 $D(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in \Delta(\mathcal{A})$.

We now show that D is continuous on the compact set $\Delta(\mathcal{A})$; it thus attains its minimum, whose value we denote by $D_{\min} > 0$. More precisely, it suffices to show that for all $\boldsymbol{x}$, $\boldsymbol{x}' \in \Delta(\mathcal{A})$, the condition $\|\boldsymbol{x}' - \boldsymbol{x}\|_1 \leqslant \varepsilon$ implies that $D(\boldsymbol{x}) - D(\boldsymbol{x}') \leqslant M\varepsilon$. Indeed, fix $\delta > 0$ and let $d_{\delta,\boldsymbol{x}} \in \overline{m}(\boldsymbol{x}, \boldsymbol{y}_0)$ be such that

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

By Lemma 6 (with the choices $\mu = \boldsymbol{x} \otimes \boldsymbol{y}_0$ and $\nu = \boldsymbol{x}' \otimes \boldsymbol{y}_0$) there exists $d_{\delta,\boldsymbol{x}'} \in \overline{m}(\boldsymbol{x}', \boldsymbol{y}_0)$ such that $\left\| d_{\delta,\boldsymbol{x}} - d_{\delta,\boldsymbol{x}'} \right\|_2 \leqslant M\varepsilon + \delta$. The triangle inequality entails that

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

Substituting in (2), we get that

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

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

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

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

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

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

which shows that $\mathcal{C}$ is not approachable. The case of pure actions taken and observed is treated similarly, with the sole addition of a concentration argument. By repeated uses of the Hoeffding-Azuma inequality together with an application of the Borel-Cantelli lemma, $\delta_T = \|\pi_T - \nu_T\|_1 \to 0$ almost surely as $T \to \infty$. By applying Lemma 6 as above, we get

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

we simply take the lim inf in the above inequalities to conclude the argument. ■

Proof that Condition (RAC) **is sufficient.** We first show that there exists a strategy of the first player such that, for all strategies of the opponent player, the sequences (π_T) or (ν_T) of the empirical distributions of actions converge to the set

$$\widetilde{\mathcal{C}} = \bigl\{ \mu \in \Delta(\mathcal{A} \times \mathcal{B}) : \;\; \overline{m}(\mu) \subseteq \mathcal{C} \bigr\} \tag{3}$$

in ℓ^2–norm, at the rates prescribed by Theorem 3.

To do so, we identify probability distributions over $\mathcal{A} \times \mathcal{B}$ with vectors in $\mathbb{R}^{\mathcal{A}\times\mathcal{B}}$ and consider the vector-valued payoff function

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

which we extend multi-linearly to $\Delta(\mathcal{A}) \times \Delta(\mathcal{B})$. We have that

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

and we therefore only need to show that $\widetilde{C}$ is m–approachable (with pure or mixed actions).

Since $\overline{m}$ is a linear function on $\Delta(\mathcal{A} \times \mathcal{B})$ and $\mathcal{C}$ is convex, the set $\widetilde{\mathcal{C}}$ is convex as well. In addition, since $\mathcal{C}$ is closed, $\widetilde{\mathcal{C}}$ is also closed. We can therefore apply the original version of the approachability theorem (stated in Theorem 3). The desired existence result follows therefore from the fact that by assumption, for all $\boldsymbol{y} \in \Delta(\mathcal{B})$, there exists some $\boldsymbol{x} \in \Delta(\mathcal{A})$ such that $\mu = m(\boldsymbol{x}, \boldsymbol{y})$, the product-distribution between $\boldsymbol{x}$ and $\boldsymbol{y}$, belongs to $\widetilde{\mathcal{C}}$, as it satisfies $\overline{m}(\mu) = \overline{m}(\boldsymbol{x}, \boldsymbol{y}) \subseteq \mathcal{C}$.

Let $P_{\widetilde{\mathcal{C}}}$ denote the projection operator onto $\widetilde{\mathcal{C}}$. We therefore have proved the existence of explicit (and possibly efficient) strategies—along the lines of the ones presented around (1)—such that, for all strategies of the second player, with probability $1-\delta$,

$$\varepsilon_T := \Big\| \pi_T - P_{\widetilde{\mathcal{C}}}(\pi_T) \Big\|_2 = \inf_{\mu \in \widetilde{\mathcal{C}}} \|\pi_T - \mu\|_2 \leqslant \frac{2}{\sqrt{T}}\Big(1 + \sqrt{2\ln(2/\delta)}\Big),$$

and with probability 1, $\varepsilon'_T := \Big\| \nu_T - P_{\widetilde{\mathcal{C}}}(\nu_T) \Big\|_2 = \inf_{\mu \in \widetilde{\mathcal{C}}} \|\nu_T - \mu\|_2 \leqslant \dfrac{2}{\sqrt{T}}$.

Lemma 6 entails that the sets $\overline{m}(\pi_T)$ are included in $M\sqrt{N_{\mathcal{A}}N_{\mathcal{B}}}\,\varepsilon_T$–neighborhoods of $\overline{m}\big(P_{\widetilde{\mathcal{C}}}(\pi_T)\big)$, and thus, by definition of $\widetilde{\mathcal{C}}$, in $M\sqrt{N_{\mathcal{A}}N_{\mathcal{B}}}\,\varepsilon_T$–neighborhoods of $\mathcal{C}$. A similar statement holds for the sets the sets $\overline{m}(\nu_T)$ and this completes the proof. ■

4. Application to 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 the references therein). The players have respective finite action sets $\mathcal{I}$ and $\mathcal{J}$. We denote by $r : \mathcal{I} \times \mathcal{J} \to \mathbb{R}^d$ the vector-valued payoff function of the first player and extend it multi-linearly to $\Delta(\mathcal{I}) \times \Delta(\mathcal{J})$. At each round, players simultaneously choose their actions $I_t \in \mathcal{I}$ and $J_t \in \mathcal{J}$, possibly at random according to probability distributions denoted by $\boldsymbol{p}_t \in \Delta(\mathcal{I})$ and $\boldsymbol{q}_t \in \Delta(\mathcal{J})$. At the end of a round, the first player does not observe J_t or $r(I_t, J_t)$ but only a signal. There is a finite set $\mathcal{H}$ of possible signals; the feedback S_t that is given to the first player is drawn at random according to the distribution $H(I_t, J_t)$, where the mapping $H : \mathcal{I} \times \mathcal{J} \to \Delta(\mathcal{H})$ is known by the first player.

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

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

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

Definition 1 can be extended as follows in this setting; the only new ingredient is the signaling structure, the aim is unchanged.

Definition 8 *Let $\mathcal{C} \subseteq \mathbb{R}^d$ be some set; $\mathcal{C}$ is r–approachable for the signaling structure H if there exists a strategy of the first player such that for all strategies of the second player,*

$$\limsup_{T\to\infty} \quad \inf_{c\in\mathcal{C}} \left\| c - \frac{1}{T}\sum_{t=1}^{T} r(I_t, J_t) \right\|_2 = 0 \qquad a.s.$$

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

A necessary and sufficient condition for r–approachability with the signaling structure H was stated and proved by Perchet (2011a); we therefore need to indicate where our contribution lies. First, both proofs are constructive but our strategy can be efficient (as soon as some projection operator can be efficiently implemented) whereas the one of Perchet (2011a) relies on auxiliary strategies that are calibrated and that require a grid that is progressively refined to be so (leading to a step complexity that is exponential in the number T of past steps). Second, we are able to exhibit convergence rates. Third, as far as elegancy is concerned, our proof is short, compact, and more direct than the one of Perchet (2011a), which relied on several layers of complicated notions (internal regret in games with partial monitoring, calibration of auxiliary strategies, etc.).

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 following convex combination of probability distributions over $\mathcal{H}$,

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

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

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

A generic element of $\mathcal{F}$ will be denoted by $\sigma \in \mathcal{F}$. The necessary and sufficient condition exhibited by Perchet (2011a) for the r–approachability of $\mathcal{C}$ for the signaling structure H can now be recalled.

Condition 1 *The signaling structure H, the vector-payoff function r, and the set $\mathcal{C}$ satisfy*

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

Defining the set-valued function $\overline{m}$, for all $\boldsymbol{p} \in \Delta(\mathcal{I})$ and $\sigma \in \mathcal{F}$, by

$$\overline{m}(\boldsymbol{p},\sigma) = \bigl\{ r(\boldsymbol{p},\boldsymbol{q}') : \quad \boldsymbol{q}' \in \Delta(\mathcal{J}) \ \textit{such that}\ \widetilde{H}(\boldsymbol{q}') = \sigma \bigr\}\,,$$

the condition can be equivalently reformulated as

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

This condition is necessary. The subsequent sections show (in a constructive way and by constructing strategies) that Condition 1 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 Perchet (2011a); a slightly simpler argument can however be found in the extended version of this paper (Mannor et al., 2011a).

4.1. Approachability in bi-piecewise linear games

In this section we consider the case where the signaling structure has some special property described below; the case of general signaling structures is considered in Section 4.3.

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

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

With general signaling structures, $\overline{m}$ is not linear, it only satisfies that for all $\boldsymbol{p} \in \Delta(\mathcal{I})$, all pairs σ, $\sigma' \in \mathcal{F}$, and all $\alpha \in [0, 1]$,

$$\alpha\,\overline{m}(\boldsymbol{p}, \sigma) + (1-\alpha)\,\overline{m}(\boldsymbol{p}, \sigma') \;\subseteq\; \overline{m}\big(\boldsymbol{p},\, \alpha\sigma + (1-\alpha)\sigma'\big)\,,$$

with a strict inclusion in general. (Specific examples can be provided.) Therefore, a direct appeal to Theorem 7 is not possible.

However, a suitable linearity property on a lifted finite set is almost given by the geometric lemma stated below. It follows from an application of Rambau and Ziegler (1996, Proposition 2.4), which entails that since $\widetilde{H}$ is linear on the polytope $\Delta(\mathcal{J})$, its inverse application $\widetilde{H}^{-1}$ is a piecewise linear mapping of $\mathcal{F}$ into the subsets of $\Delta(\mathcal{J})$; the detailed proof can be found in the extended version of this paper (Mannor et al., 2011a,b).

Lemma 9 *For any game of 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*

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

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

The results of this subsection will rely on the following assumption.

Assumption 1 *A game is bi-piecewise linear if $\overline{m}(\,\cdot\,, b)$ is piecewise linear on $\Delta(\mathcal{I})$ for every $b \in \mathcal{B}$.*

Assumption 1 means that for all $b \in \mathcal{B}$ there exists a decomposition of $\Delta(\mathcal{I})$ into polytopes each on which $\overline{m}(\,\cdot\,, b)$ is linear. Since $\mathcal{B}$ is finite, there exists a finite number of such decompositions, and thus there exists a polytopial decomposition 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 9 then leads to a piecewise linear (injective) mapping $\Theta : \Delta(\mathcal{I}) \to \Delta(\mathcal{A})$, where $\Theta(\boldsymbol{p})$ is the decomposition of $\boldsymbol{p}$ on the vertices of the polytope(s) of the decomposition to which it belongs, satisfying

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

where we denoted the convex weight vector $\Theta(\boldsymbol{p}) \in \Delta(\mathcal{B})$ by $\bigl(\Theta_a(\boldsymbol{p})\bigr)_{a \in \mathcal{A}}$. Therefore, on a lifted space, $\overline{m}$ is seen to coincide with a bi-linear mapping.

Definition 10 *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 $\sigma \in \mathcal{F}$,*

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

4.1.2. Construction of a strategy to approach $\mathcal{C}$

The approaching strategy for the original problem is based on a strategy Ψ for $\overline{\overline{m}}$–approachability of $\mathcal{C}$, provided by Theorem 7 and thus solving repeatedly minimax problems of the form (1). We therefore first need to prove the existence of such a Ψ.

Lemma 11 *Under Condition 1, the closed convex set $\mathcal{C}$ is $\overline{\overline{m}}$–robust approachable.*

Proof We show that Condition (RAC) in Theorem 7 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 a given such $\boldsymbol{y} \in \Delta(\mathcal{B})$, we associate the feasible vector of signals $\sigma = \sum_{b \in \mathcal{B}} y_b\, b$ and let $\boldsymbol{p}$ be given by Condition 1, so[3] that $\overline{m}(\boldsymbol{p}, \sigma) \subset \mathcal{C}$. By linearity of $\overline{\overline{m}}$ (for the first equality), by definition of $\overline{m}$ (for the first inclusion), by Lemma 9 (for the second and fourth equalities), by construction of $\mathcal{A}$ (for the third equality),

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

which concludes the proof. ■

We consider the strategy described in Figure 1. It forces exploration at a γ rate, as is usual in situations with partial monitoring. One of its key ingredient, that conditionally unbiased estimators are available, is extracted from Lugosi et al. (2008, Section 6): in block n we consider

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

[3] 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.

```
Parameters: an integer block length $L \geqslant 1$, an exploration parameter $\gamma \in [0,1]$, a strategy $\Psi$ 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 $\ell^2$–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 9, 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  $\boldsymbol{p}_n = (1-\gamma)\,\boldsymbol{x}_n + \gamma\,\boldsymbol{u}$;
    2. for rounds $t = (n-1)L+1, \ldots, nL$,
        2.1 drawn 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.

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

$$\mathbb{E}_t\big[\widehat{H}_t\big] = \widetilde{H}\big(\delta_{J_t}\big). \tag{4}$$

This is why, by concentration-of-the-measure argument, we will be able to show that for L large enough, $\widetilde{\sigma}_n$ is close to $\widetilde{H}\big(\widehat{\boldsymbol{q}}_n\big)$, where

$$\widehat{\boldsymbol{q}}_n = \frac{1}{L} \sum_{t=(n-1)L+1}^{nL} \delta_{J_t}. \tag{5}$$

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 $H\big(\widehat{\boldsymbol{q}}_n\big)$.

4.1.3. Performance guarantee

We provide a performance bound for fixed parameters γ and L tuned as functions of T. The proof is provided in the extended version of this paper (Mannor et al., 2011a,b). Adaptation to $T \to \infty$ can be performed either by resorting to a standard doubling trick (see, e.g., Cesa-Bianchi and Lugosi 2006, page 17) or by taking time-varying parameters γ_t and L_t.

Theorem 12 *Consider a closed convex set $\mathcal{C}$ and a game (r,H) for which Condition 1 is satisfied and that is bi-piecewise linear in the sense of Assumption 1. Then, for all $T \geqslant 1$, the strategy of Figure 1, run with parameters $L = \lceil T^{3/5} \rceil$ and $\gamma = T^{-1/5}$ and fed with a strategy Ψ for $\overline{\overline{m}}$–approachability of $\mathcal{C}$ (provided by Lemma 11) is such that, with probability at least $1-\delta$,*

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

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

The efficiency of the strategy of Figure 1 depends on whether it can be fed with an efficient approachability strategy Ψ, which in turn depends on the respective geometries of $\overline{m}$ and $\mathcal{C}$, as was indicated before the statement of Theorem 7. (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 $\overline{\overline{m}}$ can be performed beforehand.)

4.2. Application to regret minimization

In this section we analyze external and internal regret minimization in repeated games with partial monitoring from the approachability perspective. Using the results developed for vector-valued games with partial monitoring, we show how to—in particular—minimize regret in both setups.

4.2.1. External regret

We consider in this section the framework and aim introduced by Rustichini (1999) and studied, sometimes in special cases, by Piccolboni and Schindelhauer (2001); Mannor and Shimkin (2003); Cesa-Bianchi et al. (2006); Lugosi et al. (2008). We show that our general strategy can be used for regret minimization.

Scalar payoffs are obtained (but not observed) by the first player: 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. The external regret of the first player at round T equals by definition

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

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

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

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

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

The following result is an easy consequence of Theorem 12, as is explained below; 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; a question is thus whether the rates exhibited in Theorem 12 could be improved.)

Corollary 13 *For all T, the first player has a strategy such that, for all strategies of the second player and with probability at least $1-\delta$,*

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

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

The proof below is an extension to the setting of partial monitoring of the original proof and strategy of Blackwell (1956b) for the case of external regret under full monitoring: in the case of full monitoring 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 $\mathbb{R}^{\mathcal{H}\times\mathcal{I}}$ so that in this proof we will be working in the vector space $\mathbb{R}\times\mathbb{R}^{\mathcal{H}\times\mathcal{I}}$. We consider the closed convex set $\mathcal{C}$ and the vector-valued payoff function $\underline{r}$ respectively defined by

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

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

We now first show that Assumption 1 is satisfied. To do so, we will actually prove the stronger property that the mappings $\overline{m}(\cdot,\sigma)$ are piecewise linear for all $\sigma\in\mathcal{F}$; we fix such a σ in the sequel. Only the first coordinate 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

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

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

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

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

$$\overline{m}_1(\boldsymbol{p},\sigma) = \mathrm{co}\left\{ r(\boldsymbol{p},\boldsymbol{q}_{\sigma,1}),\,\ldots,\,r(\boldsymbol{p},\boldsymbol{q}_{\sigma,M}) \right\} = \left[\min_{k\in\{1,..,M\}} r(\boldsymbol{p},\boldsymbol{q}_{\sigma,k})\,, \quad \max_{k'\in\{1,..,M\}} r(\boldsymbol{p},\boldsymbol{q}_{\sigma,k'}) \right],$$

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

We then show that Condition 1 is satisfied for the considered convex set $\mathcal{C}$ and game $(\underline{r}, H)$. To do so, we associate with each $\boldsymbol{q} \in \Delta(\mathcal{J})$ an element $\phi(\boldsymbol{q}) \in \Delta(\mathcal{I})$ such that

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

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

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

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

Theorem 12 can therefore be applied to exhibit the convergence rates; we simply need to relate the quantity of interest here to the one considered therein. To that end we use the fact that the mapping

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

is Lipschitz, with Lipschitz constant in ℓ^2–norm denoted by L_ρ; the proof of this fact is detailed in the extended version of this paper (Mannor et al., 2011a,b). Now, the regret is non positive as soon as $\sum_{t=1}^{T} \underline{r}(I_t, J_t)/T$ belongs to $\mathcal{C}$; we therefore only need to consider the case when this average is not in $\mathcal{C}$. In the latter case, we denote by $(\widetilde{r}_T, \widetilde{\sigma}_T)$ its projection in ℓ^2–norm onto $\mathcal{C}$. We have first that the defining inequality of $\mathcal{C}$ is an equality on its border, so that

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

and second, that

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

As claimed, the rates are now seen to follow from the ones indicated in Theorem 12. ■

4.2.2. 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. In other words, i is the best response to the empirical distribution of action of the other player on these stages.

With partial monitoring, the first player evaluates his payoffs in some pessimistic way through the function ρ defined above. This function is not linear over $\Delta(\mathcal{I})$ in general (it is concave), so that the best responses are not necessarily pure actions $i \in \mathcal{I}$ but mixed actions, i.e., elements of $\Delta(\mathcal{I})$. Following Lehrer and Solan (2007) we therefore should partition the stages not depending on the pure actions actually played but on the mixed actions $\boldsymbol{p}_t \in \Delta(\mathcal{I})$ used to draw them. To this end, it is convenient to assume that the strategies of the first player need to pick these mixed actions in a finite (but possibly thin) grid of $\Delta(\mathcal{I})$, which we denote by $\{\boldsymbol{p}_g,\ g \in \mathcal{G}\}$, where $\mathcal{G}$ is a finite set. At each round, the first player picks an index $G_t \in \mathcal{G}$ and uses the distribution $\boldsymbol{p}_{G_t}$ to draw his action I_t. Up to a standard concentration-of-the-measure argument, we will measure the payoff at round t with $r\big(\boldsymbol{p}_{G_t}, J_t\big)$ rather than with $r(I_t, J_t)$.

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

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

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

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

Actually, our proof technique rather leads to the minimization of some swap regret (see Blum and Mansour, 2007 for the definition of swap regret in full monitoring):

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

Again, the following bound on the swap regret easily follows from Theorem 12; 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); Perchet (2009), two articles which presented complicated strategies to do so (strategies based on auxiliary strategies using a grid that needs to be refined over time and whose complexities is exponential in the size of these grids). Moreover, we provide convergence rates.

Corollary 14 *For all T, the first player has an explicit strategy such that, for all strategies of the second player and with probability at least $1-\delta$,*

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

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

The proof of this corollary is based on ideas similar to the ones used in the proof of Corollary 13; it can be found in the extended version of this paper (Mannor et al., 2011a,b).

4.3. Approachability in the case of general games

Unfortunately, as is illustrated in the extended version of this paper (Mannor et al., 2011b), there exist games with partial monitoring that are not bi-piecewise linear.

However, we will show that if Condition 1 holds there exist strategies with a constant per-round complexity to approach polytopes even when the game is not bi-piecewise linear. That is, by considering simpler closed convex sets $\mathcal{C}$, no assumption is needed on the pair (r, H). We will conclude this subsection by indicating that thanks to a doubling trick, Condition 1 is still seen to be sufficient for approachability in the most general case when no assumption is made neither on (r, H) nor on $\mathcal{C}$, at the cost however of inefficiency.

4.3.1. Approachability of the negative orthant in the case of general games

For the sake of simplicity, we start with the case of the negative orthant $\mathbb{R}^d_-$. Our argument will be based on Lemma 9; we use in the sequel the objects and notation introduced therein. We denote by $r = (r_k)_{1 \leqslant k \leqslant d}$ the components of the d–dimensional payoff function r and introduce, for all $k \in \{1, \dots, d\}$, the set-valued mapping $\widetilde{m}_k$ defined by

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

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

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

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

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

where Φ refers to the mapping defined in Lemma 9. 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 15 *The set-valued mappings $\breve{m}$ and $\overline{m}$ are linked by the following two properties: 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 interpretations of these two properties are that 1. $\breve{m}$–robust approaching a set $\mathcal{C}$ is more difficult than $\overline{m}$–robust approaching it; and 2. that if Condition 1 holds for $\overline{m}$ and

$\mathbb{R}^d_-$, it also holds for $\breve{m}$ and $\mathbb{R}^d_-$.

Proof For property 1., note that by construction of $\breve{m}$,

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

Lemma 9 and the linear extension of $\widetilde{m}$ then show that

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

As for property 2., it suffices to note that (by Lemma 9 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 as announced; note that no bi-piecewise linearity assumption is needed on the game.

Theorem 16 *If Condition 1 is satisfied for $\overline{m}$ and $\mathbb{R}^d_-$, then there exists a strategy for (r, H)–approaching $\mathbb{R}^d_-$ at a rate of the order of $T^{-1/5}$, with a constant per-round complexity.*

Proof (sketched) The assumption of the theorem and Property 2. of Lemma 15 imply that Condition 1 holds for $\mathbb{R}^d_-$ and $\breve{m}$; furthermore, the latter corresponds to a bi-piecewise linear game as can be seen by noting, similarly to what was done in the section devoted to regret minimization, that each $\widetilde{m}_k$, thus also $\breve{m}$, is a piecewise linear function. Thus, (the proof of) Theorem 12 guarantees that $\mathcal{C}$ is $\breve{m}$–robust approachable. Now, Property 1. of Lemma 15 implies that any $\breve{m}$–robust approachability strategy of $\mathcal{C} = \mathbb{R}^d_-$ is also a $\overline{m}$–robust approachability strategy. Therefore, $\mathcal{C}$ is $\overline{m}$–robust approachable, hence, following again the methodology used in the proof of Theorem 12, is also (r, H)–approachable. ■

4.3.2. Approachability of polytopes in the case of general games

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

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

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 an exercise of mere rewriting.

Lemma 17 *Given a polytope $\mathcal{C}$, the (r,H)–approachability of $\mathcal{C}$ and the $\bigl(r_{\mathcal{C}},H\bigr)$–approachability of $\mathbb{R}^d_-$ are equivalent in the sense that all strategies for one problem translates to a strategy for the other problem.*

In addition, Condition 1 holds for (r,H) and $\mathcal{C}$ if and only if it holds for $\bigl(r_{\mathcal{C}},H\bigr)$ and $\mathbb{R}^d_-$.

Via the lemma above, Theorem 16 indicates that Condition 1 for (r,H) and $\mathcal{C}$ is a sufficient condition for the (r,H)–approachability of $\mathcal{C}$ and provides a strategy to do so.

4.3.3. Approachability of general convex sets in the case of general games

A general closed convex set can always be approximated arbitrarily well by a polytope (where the number of vertices of the latter however increases as the quality of the approximation does). There, via a doubling trick, Condition 1 is also seen to be sufficient to (r,H)–approach any general closed convex set $\mathcal{C}$, However, the computational complexity of the resulting strategy is much larger: the per-round complexity increases over time (as the numbers of vertices of the approximating polytopes do).

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)*. Omnipress, 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.

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

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

S. Mannor, V. Perchet, and G. Stoltz. Robust approachability and regret minimization in games with partial monitoring. 2011a. URL `http://hal.archives-ouvertes.fr/hal-00595695`.

S. Mannor, V. Perchet, and G. Stoltz. Corrigendum to "Robust approachability and regret minimization in games with partial monitoring". 2011b. URL `http://hal.archives-ouvertes.fr/hal-00617554`.

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*, 2011b. In press.

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)*. Omnipress, 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.

Acknowledgments

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