JMLR: Workshop and Conference Proceedings vol 35 (2014) 1–17

Approachability in unknown games:
Online learning meets multi-objective optimization

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

Vianney Perchet VIANNEY.PERCHET@NORMALESUP.ORG
Université Paris Diderot, France

Gilles Stoltz STOLTZ@HEC.FR
GREGHEC: HEC Paris – CNRS, France

Abstract

In the standard setting of approachability there are two players and a target set. The players play a repeated vector-valued game where one of them wants to have the average vector-valued payoff converge to the target set which the other player tries to exclude. We revisit the classical setting and consider the setting where the player has a preference relation between target sets: she wishes to approach the smallest ("best") set possible given the observed average payoffs in hindsight. Moreover, as opposed to previous works on approachability, and in the spirit of online learning, we do not assume that there is a known game structure with actions for two players. Rather, the player receives an arbitrary vector-valued reward vector at every round. We show that it is impossible, in general, to approach the best target set in hindsight. We further propose a concrete strategy that approaches a non-trivial relaxation of the best-in-hindsight given the actual rewards. Our approach does not require projection onto a target set and amounts to switching between scalar regret minimization algorithms that are performed in episodes.

Keywords: Online learning, multi-objective optimization, approachability

1. Introduction

In online learning (or regret minimization) a decision maker is interested in obtaining as much reward as she would have obtained with perfect hindsight of the average rewards. The underlying assumption is that the decision maker can quantify the outcomes of her decision into a single value, e.g., money. However, the outcome of some sequential decision problems cannot be cast as a single dimensional optimization problem: different objectives that are possibly contradicting need to be considered. This arises in diverse fields such as finance, control, resource management, and many others. This is called multi-objective optimization.

Offline answers to multi-objective optimization. The fundamental solution concept used in offline multi-objective optimization is that of the Pareto front: given several criteria to be optimized this is the set of feasible points that are not (weakly) dominated by any other point. While every rationally optimal solution is on the Pareto front, it is not always clear which of the points on the front should be selected. One approach it to scalarize the different objectives and solve a single objective. However, scalarization leads to finding just a single point on the Pareto front. Other approaches include no-preference methods, a prior methods and a posteriori methods; see Hwang and Masud (1979); Miettinen (1999).

Online answers proposed so far. The approachability theory of Blackwell (1956) can be considered as the most general approach available so far for online multi-objective optimization. In the standard setting of approachability there are two players, a vector-valued payoff function, and a target set. The players play a repeated vector-valued game where one of them wants to have the average vector-valued payoff (representing the states in which the different objectives are) converge to the target set (representing the admissible values for the said states) which the other player tries to exclude. The target set is prescribed a priori before the game starts and the aim of the decision maker is that the average reward be asymptotically inside the target set.

We note that determining if a convex set is approachable may not be an easy task. In fact, Mannor and Tsitsiklis (2009) show that determining if a single point is approachable is NP-hard in the dimension of the reward vector.

Our view: approachability in "unknown games." The analysis in approachability has been limited to date to cases where the action of Nature, or a signal thereof, is revealed. We deviate from the standard setting by considering the decision problem to be an online problem where only (vector-valued) rewards are observed and there is no a priori assumption on what can and cannot be obtained. Moreover, we do not assume that there is some underlying game structure we can exploit. In our model for every action of the decision maker there is a reward that is only assumed to be arbitrary. This setting is referred to as the one of an "unknown game" and the minimization of regret could be extended to it (see, e.g., Cesa-Bianchi and Lugosi, 2006, Sections 7.5 and 7.10). One might wonder if it is possible to treat an unknown game as a known game with a very large class of actions and then use approachability. While such lifting is possible in principle, it would lead to unreasonable time and memory complexity as the dimensionality of the problem will explode.

In such unknown games, the decision maker does not try to approach a pre-specified target set, but rather tries to approach the best (smallest) target set given the observed (average) vector-valued rewards. Defining a goal in terms of the actual rewards is standard in online learning, but has not been pursued (with a few exceptions listed below) in the multi-objective optimization community.

Literature review. Our approach generalizes several existing works. Our proposed strategy can be used for standard approachability as it is computationally efficient. It can further be used for opportunistic approachability (when the decision maker tries to take advantage of suboptimal plays of Nature, see Bernstein et al., 2013). The proposed strategy further encompasses online learning with sample path constraints approachability Mannor et al. (2009) as a special case. The algorithm we present does not require projection which is the Achilles' heel of many approachability-based schemes (similarly to Bernstein and Shimkin, 2014). Our approach is also more general than one recently considered by Azar et al. (2014). An extensive comparison to the results by Bernstein and Shimkin (2014) and Azar et al. (2014) is offered in Section 4.2.

Contributions and outline. To summarize, we propose a strategy that works in the online setting where a game is not defined, but rather only reward vectors are obtained. This strategy can approach a good-in-hindsight set among a filtration of target sets. Furthermore, the convergence rate is independent of the dimension and the computational complexity is reasonable (i.e., polynomial).

We start the paper with defining the setting of approachability in unknown games in Section 2. In Section 3 we then move to discussing the issue of the target to be achieved. We review three different families of possible targets. The first is the best set based on average rewards in hindsight, which is not achievable. The second is the convexification of the former, which is achievable but not ambitious enough. The third goal is a sort of convexification of some individual-response-based target set; we show that the latter goal is never worse and often strictly better than the second one. In

Section 4 we devise a general strategy achieving this third goal. It amounts to playing a (standard) regret minimization in blocks and modifying the direction as needed. In Section 5 we finally work out the applications of our approach to the setting of classical approachability and to online learning with sample path constraints approachability.

2. Setup ("unknown games"), notation, and aim

The setting is the one of (classical) approachability, that is, vector payoffs are considered. The difference lies in the aim. In (classical) approachability theory, the average $\overline{r}_T$ of the obtained vector payoffs should converge asymptotically to some base approachable convex set $\mathcal{C}$. In our setting, we do not know whether $\mathcal{C}$ is approachable (because there is no underlying payoff function) and ask for convergence to some α–expansion of $\mathcal{C}$, where α should be as small as possible.

Setting: unknown game with vectors of vector payoffs. The following game is repeatedly played between two players, who will be called respectively the decision-maker (or first player) and the opponent (or second player). Vector payoffs in $\mathbb{R}^d$, where $d \geqslant 1$, will be considered. The first player has finitely many actions whose set we denote by $\mathcal{A} = \{1, \ldots, A\}$. The opponent chooses at each round $t \in \{1, 2, \ldots\}$ a vector $m_t = (m_{t,a})_{a \in \mathcal{A}}$ of vector payoffs $m_{t,a} \in \mathbb{R}^d$. We impose the restriction that these vectors m_t lie in a convex and bounded set K of $\mathbb{R}^{dA}$. The first player picks at each round t an action a_t, possibly at random according to some mixed action $x_t = (x_{t,a})_{a \in \mathcal{A}}$; we denote by $\Delta(\mathcal{A})$ the set of all such mixed actions. We consider a scenario where the player is informed of the whole vector m_t at the end of the round and we are interested in controlling the average of the payoffs m_{t,a_t}. Actually, because of martingale convergence results, this is equivalent to studying the averages $\overline{r}_T$ of the conditionally expected payoffs r_t, where

$$r_t = x_t \odot m_t = \sum_{a \in \mathcal{A}} x_{t,a} m_{t,a} \qquad \text{and} \qquad \overline{r}_T = \frac{1}{T} \sum_{t=1}^{T} r_t = \frac{1}{T} \sum_{t=1}^{T} x_t \odot m_t \,.$$

Remark 1 *We will not assume that the first player knows K (or any bound on the maximal norm of its elements); put differently, the scaling of the problem is unknown.*

Aim. This aim could be formulated in terms of a general filtration (see Remark 2 below); for the sake of concreteness we resort rather to expansions of a base set $\mathcal{C}$ in some ℓ_p–norm, which we denote by $\|\cdot\|$, for $0 < p < \infty$. Formally, we denote by $\mathcal{C}_\alpha$ the α–expansion in ℓ_p–norm of $\mathcal{C}$. The decision-maker wants that her average payoff $\overline{r}_T$ approaches an as small as possible set $\mathcal{C}_\alpha$. To get a formal definition of the latter aim, we consider the smallest set that would have been approachable in hindsight for a properly chosen target function $\varphi : K \to [0, +\infty)$. (Section 3 will indicate reasonable such choices of φ.) This function takes as argument the average of the past payoff vectors,

$$\overline{m}_T = \frac{1}{T} \sum_{t=1}^{T} m_t \,, \qquad \text{that is,} \qquad \forall\, a \in \mathcal{A}, \quad \overline{m}_{T,a} = \frac{1}{T} \sum_{t=1}^{T} m_{t,a} \,.$$

It associates with it the $\varphi(\overline{m}_T)$–expansion of $\mathcal{C}$. Therefore, our aim is that

$$\mathrm{d}_p\big(\overline{r}_T,\, \mathcal{C}_{\varphi(\overline{m}_T)}\big) \longrightarrow 0 \qquad \text{as } T \to \infty \,, \tag{1}$$

where $\mathrm{d}_p(\,\cdot\,, S)$ denotes the distance in ℓ_p–norm to a set S.

Concrete example. Consider a decision problem where a decision maker has to decide how to transmit bits on a wireless channel in a cognitive network (Simon, 2005; Beibei and Liu, 2011). The objectives of the decision maker are to have minimum power and maximum throughput. The decision maker decides at every stage how to transmit: which channels to use, what code to select and how much power to use. The transmissions of multiple other players, modeled as Nature, dictate the success of each transmission. The ideal working point is where throughput is maximal and power is zero. This working point is untenable and the decision maker will be looking for a better balance between the objectives. The model presented here fits the application naturally with $d = 2$ where the two axes are power and throughput. The set $\mathcal{C}$ is the point in the power-throughput plane with values 0 for power and maximal throughput for throughput.

Remark 2 *More general filtrations $\alpha \in [0, +\infty) \mapsto \mathcal{C}_\alpha$ could be considered than expansions in some norm, as long as this mapping is Lipschitz for the Hausdorff distance between sets. (By "filtration" we mean that $\mathcal{C}_\alpha \subseteq \mathcal{C}_{\alpha'}$ for all $a \leqslant \alpha'$.) For instance, if $0 \in \mathcal{C}$, one can consider shrinkages and blow-ups, $\mathcal{C}_0 = \{0\}$ and $\mathcal{C}_\alpha = \alpha\,\mathcal{C}$ for $\alpha > 0$. Or, given some compact set $\mathcal{B}$ with non-empty interior, $\mathcal{C}_\alpha = \mathcal{C} + \alpha\mathcal{B}$ for $\alpha \geqslant 0$.*

2.1. Link with approachability in known finite games

We link here our general setting above with the classical setting considered by Blackwell. Therein the opponent also has a finite set of actions $\mathcal{B}$ and chooses at each round t an action $b_t \in \mathcal{B}$, possibly at random according to some mixed action $y_t = (y_{t,b})_{b\in\mathcal{B}}$. A payoff function $r : \mathcal{A} \times \mathcal{B} \to \mathbb{R}^d$ is given and is linearly extended to $\Delta(\mathcal{A}) \times \Delta(\mathcal{B})$, where $\Delta(\mathcal{A})$ and $\Delta(\mathcal{B})$ are the sets of probability distributions over $\mathcal{A}$ and $\mathcal{B}$, respectively. The conditional expectation of the payoff obtained at round t is $r_t = r(x_t, y_t)$. Therefore, the present setting can be encompassed in the more general one described above by thinking of the opponent as choosing the vector payoff $m_t = r(\,\cdot\,, b_t)$. A target set $\mathcal{C}$ is to be approached, that is, the convergence $\overline{r}_T = (1/T)\sum_{t\leqslant T} r(x_t, y_t) \longrightarrow \mathcal{C}$ should hold uniformly over the opponent's strategies. A necessary and sufficient condition for this when $\mathcal{C}$ is non-empty, closed, and convex is that for all $y \in \Delta(\mathcal{B})$, there exists some $x \in \Delta(\mathcal{A})$ such that $r(x, y) \in \mathcal{C}$. Of course, this condition, called the dual condition for approachability, is not always met. However, in view of the dual condition, the least approachable α–Euclidian expansion of such a non-empty, closed, and convex set $\mathcal{C}$ is given by

$$\alpha_{\text{unif}} = \max_{y\in\Delta(\mathcal{B})} \min_{x\in\Delta(\mathcal{A})} \mathrm{d}_2\big(r(x, y), \mathcal{C}\big)\,. \tag{2}$$

Approaching $\mathcal{C}_{\alpha_{\text{unif}}}$ corresponds to considering the constant target function $\varphi \equiv \alpha_{\text{unif}}$. Better (uniformly smaller) choices of target functions exist, as will be discussed in Section 5.1. This will be put in correspondance therein with what is called "opportunistic approachability."

2.2. Applications

We describe in this section two related mathematical applications we have in mind.

Regret minimization under sample path constraints. We rephrase (and slightly generalize) here the setting of Mannor et al. (2009). A vector $m_a \in \mathbb{R}^d$ now not only represents some payoff but also some cost. The aim of the player here is to control the average payoff vector (to have it converge to the smallest expansion of a given target set $\mathcal{C}$) while abiding by some cost constraints (ensuring that the average cost vector converges to a prescribed set).

Formally, two matrices P and G associate with a vector $m_a \in \mathbb{R}^d$ a payoff vector $Pm_a \in \mathbb{R}^p$ and a cost vector $Gm_a \in \mathbb{R}^g$. By an abuse of notation, we extend P and G to work with vectors $m = (m_a)_{a\in\mathcal{A}}$ of vectors $m_a \in \mathbb{R}^d$ by defining $Pm = (Pm_a)_{a\in\mathcal{A}}$ and $Gm = (Gm_a)_{a\in\mathcal{A}}$. In our setting, the opponent player and the decision-maker thus choose simultaneously and respectively a vector $(m_a)_{a\in\mathcal{A}} \in K \subseteq \mathbb{R}^{dA}$ and a mixed action $x_t \in \Delta(\mathcal{A})$; the decision-maker then gets as a payoff and cost vectors $x_t \odot Pm_t = P\bigl(x_t \odot m_t\bigr)$ and $x_t \odot Gm_t = G\bigl(x_t \odot m_t\bigr)$. The admissible costs are represented by a set $\Gamma \subseteq \mathbb{R}^g$, while some set $\mathcal{P} \subseteq \mathbb{R}^p$ is to be approached.

We adapt here slightly the exposition above. We define some base set $\mathcal{C} = \mathcal{C}_0 \subseteq \mathbb{R}^d$ and its α–expansions $\mathcal{C}_\alpha$ in ℓ_p–norm by forcing the constraints stated by Γ: for all $\alpha \geqslant 0$,

$$\mathcal{C}_\alpha = \Bigl\{ m' \in \mathbb{R}^d : \;\; Gm' \in \Gamma \;\text{ and }\; \mathrm{d}_p\bigl(Pm', \,\mathcal{P}\bigr) \leqslant \alpha \Bigr\}\,.$$

We also denote by $\mathcal{P}_\alpha$ the (unconstrained) α–expansions of $\mathcal{P}$ in ℓ_p–norm. For all $m' \in \mathbb{R}^d$ with $Gm' \in \Gamma$, one has $\mathrm{d}_p(m', \mathcal{C}) \leqslant \mathrm{d}_p(Pm', \mathcal{P})$. Therefore, the general aim (1) is now satisfied as soon as the following convergences are realized: as $T \to \infty$,

$$\mathrm{d}_p\bigl(P\overline{r}_T, \,\mathcal{P}_{\varphi(\overline{m}_T)}\bigr) \longrightarrow 0 \qquad \text{and} \qquad \mathrm{d}_p\bigl(G\overline{r}_T, \,\Gamma\bigr) \longrightarrow 0\,, \tag{3}$$

for some target function φ to be defined (taking into account the cost constraints); see Section 5.2.

Approachability of an approachable set at a minimal cost. This is the dual problem of the previous problem: have the vector-valued payoffs approach an approachable convex set while suffering some costs and trying to control the overall cost. In this case, the set $\mathcal{P}$ is fixed and the α–expansions are in terms of Γ. Actually, this is a problem symmetric to the previous one, when the roles of P and $\mathcal{P}$ are exchanged with G and Γ. This is why we will not study it for itself in Section 5.2.

3. Choices of target functions

We discuss in this section what a reasonable choice of a target function φ can be. To do so, we start with an unachievable target function $\varphi^\star$. We then provide a relaxation given by its concavification $\mathrm{cav}[\varphi^\star]$, which can be aimed for but is not ambitious enough. Based on the intuition given by the formula for concavification, we finally provide a whole class of achievable targets, relying on a parameter: a response function Ψ.

An unachievable target function. We denote by $\varphi^\star : K \to [0, +\infty)$ the function that associates with a vector of vector payoffs $m \in K$ the index of the smallest ℓ_p–expansion of $\mathcal{C}$ containing a convex combination of its components:

$$\varphi^\star(m) = \min \Bigl\{ \alpha \geqslant 0 : \;\; \exists\, x \in \Delta(\mathcal{A}) \text{ s.t. } x \odot m \in \mathcal{C}_\alpha \Bigr\} = \min_{x\in\Delta(\mathcal{A})} \mathrm{d}_p(x \odot m, \,\mathcal{C})\,, \tag{4}$$

the infimum being achieved by continuity. That is, for all $m \in K$, there exists $x^\star(m)$ such that $x^\star(m) \odot m \in \mathcal{C}_{\varphi^\star(m)}$. The defining equalities of $\varphi^\star$ show that this function is continuous (it is even a Lipschitz function with constant 1 in the ℓ_p–norm).

Definition 3 *A continuous target function $\varphi : K \to [0, +\infty)$ is achievable if the decision-maker has a strategy ensuring that, against all strategies of the opponent player,*

$$\mathrm{d}_p\big(\overline{r}_T,\, \mathcal{C}_{\varphi(\overline{m}_T)}\big) \longrightarrow 0 \qquad \textit{as } T \to \infty\,. \tag{5}$$

More generally, a (possibly non-continuous) target function $\varphi : K \to [0, +\infty)$ is achievable if the following convergence to a set takes place in $\mathbb{R}^{dA} \times \mathbb{R}^d$ as $T \to \infty$:

$$(\overline{m}_T,\, \overline{r}_T) \longrightarrow \mathcal{G}_\varphi \qquad \text{where} \quad \mathcal{G}_\varphi = \Big\{(m, r) \in \mathbb{R}^{dA} \times \mathbb{R}^d \ \text{ s.t. } \ r \in \mathcal{C}_{\varphi(m)}\Big\}\,. \tag{6}$$

The set $\mathcal{G}_\varphi$ is the graph of the set-valued mapping $m \in K \to \mathcal{C}_{\varphi(m)}$. The second part of the definition above coincides with the first one in the case of a continuous φ, as we prove in Mannor et al. (2014, Section B.1). (In general, it is weaker, though.) It is useful in the case of non-continuous target functions to avoid lack of convergence due to errors at early stages. The following lemma is proved by means of two examples in Mannor et al. (2014, Section C).

Lemma 4 *The target function $\varphi^\star$ is not achievable in general.*

An achievable, but not ambitious enough, target function. We resort to a classical relaxation, known as a convex relaxation (see, e.g., Mannor et al., 2009): we only ask for convergence of $(\overline{m}_T,\, \overline{r}_T)$ to the convex hull of $\mathcal{G}_{\varphi^\star}$, not to $\mathcal{G}_{\varphi^\star}$ itself. This convex hull is exactly the graph $\mathcal{G}_{\mathrm{cav}[\varphi^\star]}$, where $\mathrm{cav}[\varphi^\star]$ is the so-called concavification of $\varphi^\star$, defined as the least concave function $K \to [0, +\infty]$ above $\varphi^\star$. Its variational expression reads

$$\mathrm{cav}[\varphi^\star](m) = \sup\left\{\sum_{i \leqslant N} \lambda_i\, \varphi^\star(m_i) : \ \ N \geqslant 1 \ \text{ and } \ \sum_{i \leqslant N} \lambda_i m_i = m\right\}, \tag{7}$$

for all $m \in K$, where the supremum is over all finite convex decompositions of m as elements of K (i.e., the λ_i factors are nonnegative and sum up to 1). By a theorem by Fenchel and Bunt (see Hiriart-Urruty and Lemaréchal, 2001, Theorem 1.3.7) we could actually impose that $1 \leqslant N \leqslant dA + 1$. In general, $\mathrm{cav}[\varphi^\star]$ is not continuous; it however is so when, e.g., K is a polytope.

Definition 5 *A target function $\varphi : K \to [0, +\infty)$ is strictly smaller than another target function φ' if $\varphi \leqslant \varphi'$ and there exists $m \in K$ with $\varphi(m) < \varphi'(m)$. We denote this fact by $\varphi \prec \varphi'$.*

Lemma 6 *The target function $\mathrm{cav}[\varphi^\star]$ is achievable. However, in general, there exist easy-to-construct achievable target functions φ with $\varphi \prec \mathrm{cav}[\varphi^\star]$.*

The first part of the lemma is proved in Mannor et al. (2014, Section B.2); its second part is a special case of Lemma 7 below.

A general class of achievable target functions. By (4) we can rewrite (7) as

$$\mathrm{cav}[\varphi^\star](m) = \sup\left\{\sum_{i \leqslant N} \lambda_i\, \mathrm{d}_p\big(x^\star(m_i) \odot m_i,\, \mathcal{C}\big) : \ \ N \geqslant 1 \ \text{ and } \ \sum_{i \leqslant N} \lambda_i m_i = m\right\}.$$

Now, whenever $\mathcal{C}$ is convex, the function $\mathrm{d}_p(\,\cdot\,,\mathcal{C})$ is convex as well over $\mathbb{R}^d$; see, e.g., Boyd and Vandenberghe (2004, Example 3.16). Therefore, denoting by $\varphi^{x^\star}$ the function defined as

$$\varphi^{x^\star}(m) = \sup\left\{ \mathrm{d}_p\!\left(\sum_{i \leqslant N} \lambda_i \, x^\star(m_i) \odot m_i, \ \mathcal{C}\right) : \quad N \geqslant 1 \ \text{ and } \ \sum_{i \leqslant N} \lambda_i m_i = m \right\} \tag{8}$$

for all $m \in K$, we have $\varphi^{x^\star} \leqslant \mathrm{cav}[\varphi^\star]$. The two examples considered in Mannor et al. (2014, Section C) show that this inequality can be strict at some points. We summarize these facts in the lemma below.

Lemma 7 *The inequality* $\varphi^{x^\star} \leqslant \mathrm{cav}[\varphi^\star]$ *always holds; and sometimes* $\varphi^{x^\star} \prec \mathrm{cav}[\varphi^\star]$.

More generally, let us introduce individual response functions Ψ as functions $K \to \Delta(\mathcal{A})$. The target function naturally associated with Ψ in light of (8) is defined, for all $m \in K$, as

$$\varphi^{\Psi}(m) = \sup\left\{ \ \mathrm{d}_p\!\left(\sum_{i \leqslant N} \lambda_i \, \Psi(m_i) \odot m_i, \ \mathcal{C}\right) : \quad N \geqslant 1 \ \text{ and } \ \sum_{i \leqslant N} \lambda_i m_i = m \right\}. \tag{9}$$

Lemma 8 *For all response functions* Ψ*, the target functions* φ^{Ψ} *are achievable. However, in general, there exist easy-to-construct achievable target functions* φ *(possibly of the form* φ^{Ψ}*) with* $\varphi \prec \varphi^{x^\star}$.

The second part of the lemma indicates that there are cleverer choices for the response function Ψ than $x^\star$. This is illustrated by Mannor et al. (2014, Section C, Example 2). We also provide some elements towards a theory of optimality in Mannor et al. (2014, Section D); e.g., there always exist "admissible" functions, i.e., functions that are achievable and such that no strictly smaller function is achievable. But we were unable so far to prove or disprove that a target of the form φ^{Ψ} could be optimal, let alone to provide guidelines on how to choose Ψ in practice.

The first part of the lemma will follow from Theorem 9 below, which provides an explicit and efficient strategy to achieve any φ^{Ψ}. However, we provide in Mannor et al. (2014, Section B.3) a proof based on calibration, which further explains the intuition behind (9). It also advocates why the φ^{Ψ} functions are reasonable targets: resorting to some auxiliary calibrated strategy outputting accurate predictions $\hat{m}_t$ (in the sense of calibration) of the vectors m_t almost amounts to knowing in advance the m_t.

4. A strategy by regret minimization in blocks

In this section we exhibit a strategy to achieve the desired convergence (5) with the target functions φ^{Ψ} advocated in the previous section. The algorithm is efficient, as long as calls to Ψ are (a full discussion of the complexity issues is provided in Section 5). The considered strategy—see Figure 1—relies on some auxiliary regret-minimizing strategy $\mathcal{R}$, with the following property.

Assumption 1 *The strategy* $\mathcal{R}$ *sequentially outputs mixed actions* u_t *such that for all ranges* $B > 0$ *(not necessarily known in advance), for all* $T \geqslant 1$ *(not necessarily known in advance), for all*

```
*Parameters*: a regret-minimizing strategy $\mathcal{R}$ (with initial action $u_1$), a response function $\Psi : K \to \Delta(\mathcal{A})$
*Initialization*: play $x_1 = u_1$ and observe $m_1 \in \mathbb{R}^{dA}$
*For all blocks* $n = 2, 3 \ldots$,
    1. compute the total discrepancy at the beginning of block $n$ (i.e., till the end of block $n-1$),
       $\delta_n = \sum_{t=1}^{n(n-1)/2} x_t \odot m_t - \sum_{k=1}^{n-1} k\, \Psi\bigl(\overline{m}^{(k)}\bigr) \odot \overline{m}^{(k)} \in \mathbb{R}^d \,,$   where   $\overline{m}^{(k)} = \frac{1}{k} \sum_{t=k(k-1)/2+1}^{k(k+1)/2} m_t$
       is the average vector of vector payoffs obtained in block $k \in \{1, \ldots, n-1\}$;
    2. run a fresh instance $\mathcal{R}_n$ of $\mathcal{R}$ for $n$ rounds as follows: set $u_{n,1} = u_1$; then, for $t = 1, \ldots, n$,
        (a) play $x_{n(n-1)/2+t} = u_{n,t}$ and observe $m_{n(n-1)/2+t} \in \mathbb{R}^{dA}$;
        (b) feed $\mathcal{R}_n$ with the vector payoff $m'_{n,t} \in \mathbb{R}^A$ with components given by
            $m'_{n,t,a} = -\bigl\langle \delta_n,\, m_{n(n-1)/2+t,a} \bigr\rangle \in \mathbb{R},$   where  $a \in \mathcal{A}\,,$
            where $\langle \,\cdot\,,\,\cdot\, \rangle$ denotes the inner product in $\mathbb{R}^d$;
        (c) obtain from $\mathcal{R}_n$ a mixed action $u_{n,t+1}$.
```

Figure 1: The proposed strategy, which plays in blocks of increasing lengths 1, 2, 3, . . .

sequences of vectors $m'_t \in \mathbb{R}^A$ of one-dimensional payoffs lying in the bounded interval $[-B, B]$, possibly chosen by some adversary, where $t = 1, \ldots, T$,

$$\max_{u \in \Delta(\mathcal{A})} \sum_{t=1}^{T} u \odot m'_t \leqslant 4B\sqrt{T \ln A} + \sum_{t=1}^{T} u_t \odot m'_t \,.$$

Note in particular that the auxiliary strategy $\mathcal{R}$ adapts automatically to the range B of the payoffs and to the number of rounds T, and has a sublinear worst-case guarantee. (The adaptation to B will be needed because K is unknown.) Such auxiliary strategies indeed exist, for instance, the polynomially weighted average forecaster of Cesa-Bianchi and Lugosi (2003). Other ones with a larger constant factor in front of the $B\sqrt{T \ln A}$ term also exist, for instance, exponentially weighted average strategies with learning rates carefully tuned over time, as in Cesa-Bianchi et al. (2007); de Rooij et al. (2014).

For the sake of elegance (but maybe at the cost of not providing all the intuitions that led us to this result), we only provide in Figure 1 the time-adaptive version of our strategy, which does not need to know the time horizon T in advance. The used blocks are of increasing lengths 1, 2, 3, ... Simpler versions with fixed block length L require a tuning of L of the order of $\sqrt{T}$ to optimize the theoretical bound.

4.1. Performance bound for the strategy

We denote by $\| \cdot \|$ the Euclidian norm and let $K_{\max} = \max \Bigl\{ \max_{m \in K} \|m\|, \ \max_{m,m' \in K} \|m - m'\| \Bigr\}$ be a bound on the range of the norms of the (differences of) elements in K. Note that the strategy itself does not rely on the knowledge of this bound $K_{\max}$ as promised in Remark 1; only its performance bound does. Also, the convexity of $\mathcal{C}$ is not required. The proof is in Section A.

Theorem 9 *For all response functions Ψ, for all $T \geqslant 1$, for all sequences $m_1, \ldots, m_T \in \mathbb{R}^{dA}$ of vectors of vector payoffs, possibly chosen by an adversary,*

$$\mathrm{d}_p\big(\overline{r}_T, \mathcal{C}_{\varphi^\Psi(\overline{m}_T)}\big) = O\big(T^{-1/4}\big)\,.$$

More precisely, with the notation of Figure 1, denoting in addition by N the largest integer such that $N(N+1)/2 \leqslant T$, by

$$\overline{m}^{\text{part.}} = \frac{1}{T - N(N-1)/2} \sum_{t=N(N-1)/2+1}^{T} m_t$$

the partial average of the vectors of vector payoffs m_t obtained during the last block, and by $c_T \in \mathcal{C}_{\varphi^\Psi(\overline{m}_T)}$ the following convex combination,

$$c_T = \frac{1}{T}\left(\sum_{k=1}^{N-1} k\,\Psi\big(\overline{m}^{(k)}\big) \odot \overline{m}^{(k)} + \left(T - \frac{N(N-1)}{2}\right)\Psi\big(\overline{m}^{\text{part.}}\big) \odot \overline{m}^{\text{part.}}\right),$$

we have $$\left\| \frac{1}{T}\sum_{t=1}^{T} x_t \odot m_t - c_T \right\|_2 \leqslant \big(8K_{\max}\sqrt{\ln A}\big)\,T^{-1/4} + \sqrt{2}\,K_{\max}\,T^{-1/2}\,. \tag{10}$$

4.2. Discussion

In this section we gather comments, remarks, and pointers to the literature. We discuss in particular the links and improvements over the concurrent (and independent) works by Bernstein and Shimkin (2014) and Azar et al. (2014).

Do we have to play in blocks? Our strategy proceeds in blocks, unlike the ones exhibited for the case of known games, as the original strategy by Blackwell (1956) or the more recent one by Bernstein and Shimkin (2014). This is because of the form of the aim φ^Ψ we want to achieve: it is quite demanding. Even the calibration-based strategy considered in the proof of Lemma 8 performs some grouping, according to the finitely many possible values of the predicted vectors of vector payoffs. Actually, it is easy to prove that the following quantity, which involves no grouping in rounds, cannot be minimized in general:

$$\left\| \frac{1}{T}\sum_{t=1}^{T} x_t \odot m_t - \frac{1}{T}\sum_{t=1}^{T} \Psi(m_t) \odot m_t \right\|_1 . \tag{11}$$

Indeed, for the simplest case of regret minimization, the m_t consist of scalar components $\ell_{a,t} \geqslant 0$, where $a \in \mathcal{A}$, each representing the nonnegative loss associated with action a at round t. The cumulative loss is to be minimized, that is, the set $\mathcal{C} = (-\infty, 0]$ is to be approached, and its expansions are given by $\mathcal{C}_\alpha = (-\infty, \alpha]$, for $\alpha \geqslant 0$. The target function φ^Ψ thus represents what the cumulative loss of the strategy is compared to. Considering $\Psi\big((\ell_a)_{a\in\mathcal{A}}\big) \in \arg\min_{a\in\mathcal{A}} \ell_a$, we see that (11) boils down to controlling

$$\left| \sum_{t=1}^{T}\sum_{a\in\mathcal{A}} x_{a,t}\ell_{a,t} - \sum_{t=1}^{T} \min_{a'_t\in\mathcal{A}} \ell_{a'_t,t} \right| ,$$

which is impossible (see, e.g., Cesa-Bianchi and Lugosi, 2006). In this example of regret minimization, the bound (10) corresponds to the control (from above and from below) of some shifting regret for $\sqrt{T}$ blocks; the literature thus shows that the obtained $T^{-1/4}$ rate to do so is optimal (again, see, e.g., Cesa-Bianchi and Lugosi, 2006, Chapter 5 and the references therein).

In a nutshell, what we proved in this paragraph is that if we are to ensure the convergence (1) by controlling a quantity of the form (10), then we have to proceed in blocks and convergence cannot hold at a faster rate than $T^{-1/4}$. However, the associated strategy is computationally efficient.

Trading efficiency for a better rate. Theorem 9 shows that some set is approachable here, namely, $\mathcal{G}_{\varphi^\Psi}$: it is thus a B–set in the terminology of Spinat (2002), see also Hou (1971). Therefore, there exists some (abstract and possibly computationally extremely inefficient) strategy which approaches it at a $1/\sqrt{T}$–rate. Indeed, the proof of existence of such a strategy does not rely on any constructive argument.

Links with the strategy of Bernstein and Shimkin (2014). We explain here how our strategy and proof technique compare to the ones described in the mentioned reference. The setting is the one of a known game with a known target set $\mathcal{C}$, which is known to be approachable. The latter assumption translates in our more general case into the existence of a response function $\Psi_\mathcal{C}$ such that $\Psi_\mathcal{C}(m) \odot m \in \mathcal{C}$ for all $m \in K$. In that case, one wishes to use the null function $\varphi = 0$ as a target function. A straightforward generalization of the arguments of Bernstein and Shimkin (2014) then corresponds to noting that to get the desired convergence $\mathrm{d}_p\big(\overline{r}_T, \mathcal{C}\big) \to 0$, it suffices to show that there exist vectors $\widetilde{m}_t$ such that

$$\left\| \frac{1}{T}\sum_{t=1}^{T} x_t \odot m_t - \frac{1}{T}\sum_{t=1}^{T} \Psi_\mathcal{C}\big(\widetilde{m}_t\big) \odot \widetilde{m}_t \right\|_1 \longrightarrow 0\,; \tag{12}$$

of course, this is a weaker statement than trying to force convergence of the quantity (11) towards 0. Section A.2 recalls how to prove the convergence (12), which takes place at the optimal $1/\sqrt{T}$–rate.

On the related framework of Azar et al. (2014). The setting considered therein is exactly the one described in Section 2: the main difference with our work lies in the aim pursued and in the nature of the results obtained. The quality of a strategy is evaluated therein based on some quasi-concave and Lipschitz function $f : \mathbb{R}^d \to \mathbb{R}$. With the notation of Theorem 9, the extension to an unknown horizon T of their aim would be to guarantee that

$$\liminf_{T\to\infty}\ f\!\left(\frac{1}{T}\sum_{t=1}^{T} x_t \odot m_t\right) - \min_{k\in\{1,\ldots,N-1\}}\ \max_{x\in\Delta(\mathcal{A})} f\big(x \odot \overline{m}^{(k)}\big) \geqslant 0\,, \tag{13}$$

where N is of order $\sqrt{T}$ (e.g., as the N considered in Theorem 9). A direct consequence of our Theorem 9 and of the Lipschitz assumption on f is that

$$\liminf_{T\to\infty}\ f\!\left(\frac{1}{T}\sum_{t=1}^{T} x_t \odot m_t\right) - f\!\left(O(1/\sqrt{T}) + \frac{1}{T}\sum_{k=1}^{N-1} k\,\Psi\big(\overline{m}^{(k)}\big) \odot \overline{m}^{(k)}\right) \geqslant 0\,. \tag{14}$$

The quasi-concavity of f implies that the image by f of a convex combination is larger than the minimum of the images by f of the convex combinations. That is,

$$\liminf_{T\to\infty}\ f\!\left(\frac{1}{T}\sum_{t=1}^{T} x_t \odot m_t\right) - \min_{k=1,\ldots,N-1} f\big(\Psi\big(\overline{m}^{(k)}\big) \odot \overline{m}^{(k)}\big) \geqslant 0\,.$$

Defining Ψ as $\Psi(m) \in \displaystyle\operatorname*{arg\,max}_{x \in \Delta(\mathcal{A})} f(x \odot m)$, we get (13).

However, we need to underline that the aim (13) is extremely weak: assume, for instance, that during some block Nature chooses $\overline{m}^{(k)}$ such that $x \odot \overline{m}^{(k)} = \min f$ for all $x \in \Delta(\mathcal{A})$. Then (13) is satisfied irrespectively of the algorithm. On the contrary, the more demanding aim (14) that we consider is not necessarily satisfied and an appropriate algorithm—as our one—must be used.

In addition, the strategy designed in Azar et al. (2014) still requires some knowledge—the set K of vectors of vector payoffs needs to be known (which is a severe restriction)—and uses projections onto convex sets. The rate they obtain for their weaker aim is $O(T^{-1/4})$, as we get for our improved aim.

An interpretation of the rates. Based on all remarks above, we conclude this section with an intuitive interpretation of the $T^{-1/4}$ rate obtained in Theorem 9, versus the $1/\sqrt{T}$ rate achieved by Blackwell's original strategy or variations of it as the one described above in the case where $\mathcal{C}$ is approachable. The interpretation is in terms of the number of significant computational units N_{comp} (projections, solutions of convex or linear programs, etc.) to be performed. The strategies with the faster rate $1/\sqrt{T}$ perform at least one or two of these units at each round, while our strategy does it only of the order of $\sqrt{T}$ times during T rounds—see the calls to Ψ. In all the cases, the rate is $\sqrt{N_{\text{comp}}/T}$.

5. Applications (worked out)

In this section we work out the applications mentioned in Section 2.2. Some others could be considered, such as global costs (see Even-Dar et al., 2009; Bernstein and Shimkin, 2014) but we omit them for the sake of conciseness.

5.1. Link with classical approachability, opportunistic approachability

We recall that in the setting of known finite games described in Section 2.1, vectors of vector payoffs m actually correspond to the $r(\,\cdot\,, y)$, where y is some mixed action of the opponent. This defines the set K. The response function Ψ will thus be a function of $r(\,\cdot\,, y) \in K$. A natural (but not necessarily optimal, as illustrated by Mannor et al., 2014, Section C, Exemple 2) choice is, for all $y \in \Delta(\mathcal{B})$,

$$x^\star(y) = \Psi\bigl(r(\,\cdot\,, y)\bigr) \in \operatorname*{arg\,min}_{x \in \Delta(\mathcal{A})} \mathrm{d}_2\bigl(r(x,y),\, \mathcal{C}\bigr)\,.$$

A key feature of our algorithm, even based on this non-necessarily optimal response function, is that it is never required to compute the quantity α_{unif} defined in (2), which, depending on whether it is null or positive, indicates whether a convex set $\mathcal{C}$ is approachable or not and in the latter case, suggests to consider the least approachable convex set $\mathcal{C}_{\alpha_{\text{unif}}}$. The latter problem of determining the approachability of a set is actually an extremely difficult problem as even the determination of the approachability of the singleton set $\mathcal{C} = \{0\}$ in known games is NP–hard to perform; see Mannor and Tsitsiklis (2009).

On the other hand, our strategy only needs to compute $\sqrt{T}$ calls to Ψ in T steps. Moreover, each of these queries simply consists of solving the convex program

$$\min \Bigl\| \textstyle\sum_{a \in \mathcal{A}} x_a r(a,y) - c \Bigr\|^2 \qquad \text{s.t.} \quad x \in \Delta(\mathcal{A}), \;\; c \in \mathcal{C}\,,$$

which can be done efficiently. (It even reduces to a quadratic problem when $\mathcal{C}$ is a polytope.) Doing so, our algorithm ensures in particular that the average payoffs $\overline{r}_T$ are asymptotically inside of or on the border of the set $\mathcal{C}_{\alpha_{\text{unif}}}$.

To see that there is no contradiction between these statements, note that our algorithm does not, neither in advance nor in retrospect, issue any statement on the value of α_{unif}. It happens to perform approachability to $\mathcal{C}_{\alpha_{\text{unif}}}$ for the specific sequence of actions chosen by the opponent but does not determine a minimal approachable set which would suited for all sequences of actions. In particular, it does not provide a certificate of whether a given convex set $\mathcal{C}$ is approachable or not.

This is of course a nice feature of our method but it comes at a cost: the main drawback is the lower rate of convergence of $T^{-1/4}$ instead of $T^{-1/2}$. But we recall that the latter superior rates requires in general, to the best of our knowledge, the knowledge of α_{unif}.

Opportunistic approachability? In general, in known games, one has that the target function considered above, $\varphi^{x^\star}$, satisfies $\varphi^{x^\star} \prec \alpha_{\text{unif}}$. That is, easy-to-control sequences of vectors $r(\,\cdot\,, y_t)$ can get much closer to $\mathcal{C}$ than the uniform distance α_{unif}: we get some pathwise refinement of classical approachability. This should be put in correspondance with the recent, but different, notion of opportunistic approachability (see Bernstein et al., 2013). However, quantifying exactly what we gain here with the pathwise refinement would require much additional work (maybe a complete paper as the one mentioned above) and this is why we do not explore further this issue.

5.2. Regret minimization under sample path constraints

We recall that the difficulty of this setting is that there exists a hard constraint, given by the costs having to (asymptotically) lie in Γ. The aim is to get the average of the payoffs as close as possible to $\mathcal{P}$ given this hard constraint. We will choose below a response function Ψ such that for all $m \in K$, one has $G\big(\Psi(m) \odot m\big) \in \Gamma$ and we will adjust (9) to consider only payoffs:

$$\phi^{\Psi}(m) = \sup\left\{ \mathrm{d}_p\!\left(P \sum_{i=1}^{N} \lambda_i \, \Psi(m_i) \odot m_i, \ \mathcal{P} \right) : \ \sum_{i=1}^{N} \lambda_i m_i = m \right\}.$$

As long as Γ is convex, the strategy of Figure 1 and its analysis can then be adapted to get (3):

$$\mathrm{d}_p\big(P\overline{r}_T, \, \mathcal{P}_{\phi^{\Psi}(\overline{m}_T)}\big) \longrightarrow 0 \qquad \text{and} \qquad \mathrm{d}_p\big(G\overline{r}_T, \, \Gamma\big) \longrightarrow 0\,.$$

A reasonable choice of Ψ. We assume that the cost contraint is feasible, i.e., that for all $m \in K$, there exists $x \in \Delta(\mathcal{A})$ such that $G(x \odot m) \in \Gamma$. We then define, for all $m \in K$,

$$x^\star(m) = \Psi(m) \in \arg\min\Big\{ \mathrm{d}_p\big(P(x \odot m), \, \mathcal{P}\big) \,:\ x \in \Delta(\mathcal{A}) \ \text{ s.t. } \ G(x \odot m) \in \Gamma \Big\}\,,$$

where the minimum is indeed achieved by continuity as soon as both $\mathcal{P}$ and Γ are closed sets. At least when P is a linear form (i.e., takes scalar values), Γ is convex, and $\mathcal{P}$ is an interval, the defining equation of $x^\star$ is a linear optimization problem under a convex constraint and can be solved efficiently (see, e.g., Mannor et al., 2009; Bernstein and Shimkin, 2014).

Link with earlier work. The setting of the mentioned references is the one of a known game, with some linear scalar payoff function and vector-valued cost functions $u : \Delta(\mathcal{A}) \times \Delta(\mathcal{B}) \to [0, M]$ and

$c : \Delta(\mathcal{A}) \times \Delta(\mathcal{B}) \to \mathbb{R}^g$. (With no loss of generality we can assume that the payoff function takes values in a bounded nonnegative interval.) The vector m of our general formulation corresponds to

$$m(y) = \left[\begin{array}{c} u(\,\cdot\,, y) \\ c(\,\cdot\,, y) \end{array} \right].$$

The payoff set $\mathcal{P}$ to be be approached given the constraints is $[M, +\infty)$, that is, payoffs are to be maximized given the constraints: $\mathcal{P}_\alpha = [M - \alpha, +\infty)$. Abusing the notation by not distinguishing between $m(y)$ and y, we denote the maximal payoff under the constraint by

$$\phi^\star(y) = \max\bigl\{u(x, y) : \ x \in \Delta(\mathcal{A}) \ \text{ s.t. } \ c(x, y) \in \Gamma\bigr\}\,.$$

This target function corresponds to (4) in the same way as ϕ^Ψ corresponds to (9). Mannor et al. (2009) exactly proceed as we did in Section 3: they first show that $\phi^\star$ is unachievable in general and then show that the relaxed goal $\operatorname{cav}[\phi^\star]$ can be achieved. They propose a computationally complex strategy to do so (based on calibration) but Bernstein and Shimkin (2014) already noted that simpler and more tractable strategies could achieve $\operatorname{cav}[\phi^\star]$ as well.

The target function $\phi^{x^\star}$, which we proved above to be achievable, improves on $\operatorname{cav}[\phi^\star]$, even though, as in the remark concluding Section 5.1, it is difficult to quantify in general how much we gain. One should look at specific examples to quantify the improvement from $\operatorname{cav}[\phi^\star]$ to φ^ψ (as we do in Mannor et al., 2014, Section C). The added value in our approach mostly lies in the versatility: we do not need to assume that some known game is taking place.

Acknowledgments

Vianney Perchet acknowledges funding from the ANR, under grants ANR-10-BLAN-0112 and ANR-13-JS01-0004-01. Shie Mannor was partially supported by the ISF under contract 890015. Gilles Stoltz would like to thank Investissements d'Avenir (ANR-11-IDEX-0003 / Labex Ecodec / ANR-11-LABX-0047) for financial support.

References

Y. Azar, U. Feige, M. Feldman, and M. Tennenholtz. Sequential decision making with vector outcomes. In *Proceedings of ITCS*, 2014.

W. Beibei and K.J.R. Liu. Advances in cognitive radio networks: a survey. *IEEE Journal of Selected Topics in Signal Processing*, 5(1):5–23, 2011.

A. Bernstein and N. Shimkin. Response-based approachability and its application to generalized no-regret algorithms. arXiv:1312.7658 [cs.LG], 2014.

A. Bernstein, S. Mannor, and N. Shimkin. Opportunistic strategies for generalized no-regret problems. In *Proceedings of COLT*, pages 158–171, 2013.

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

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2004.

N. Cesa-Bianchi and G. Lugosi. Potential-based algorithms in on-line prediction and game theory. *Machine Learning*, 3(51):239–261, 2003.

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

N. Cesa-Bianchi, Y. Mansour, and G. Stoltz. Improved second-order bounds for prediction with expert advice. *Machine Learning*, 66(2/3):321–352, 2007.

S. de Rooij, T. van Erven, P.D. Grünwald, and W. Koolen. Follow the leader if you can, hedge if you must. *Journal of Machine Learning Research*, 2014. In press.

E. Even-Dar, R. Kleinberg, S. Mannor, and Y. Mansour. Online learning for global cost functions. In *Proceedings of COLT*, 2009.

J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of Convex Analysis*. Springer-Verlag, 2001.

T.-F. Hou. Approachability in a two-person game. *The Annals of Mathematical Statistics*, 42: 735–744, 1971.

C.-L. Hwang and A.S. Md Masud. *Multiple Objective Decision Making, Methods and Applications: a state-of-the-art survey*. Springer-Verlag, 1979.

S. Mannor and J. N. Tsitsiklis. Approachability in repeated games: Computational aspects and a Stackelberg variant. *Games and Economic Behavior*, 66(1):315–325, 2009.

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

S. Mannor, V. Perchet, and G. Stoltz. Approachability in unknown games: Online learning meets multi-objective optimization. arXiv:1402.2043 [stat.ML], 2014.

K. Miettinen. *Nonlinear Multiobjective Optimization*. Springer, 1999.

H. Simon. Cognitive radio: brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications*, 23(2):201–220, 2005.

X. Spinat. A necessary and sufficient condition for approachability. *Mathematics of Operations Research*, 27:31–44, 2002.

Appendix A.

We prove Theorem 9, as well as the convergence (12), which was a key point in the comparison of our work with the one by Bernstein and Shimkin (2014).

A.1. Proof of Theorem 9

The first part of the theorem follows from its second part, together with the definition of φ^{Ψ} as a supremum and the equivalence between ℓ_p– and ℓ_2–norms.

It thus suffices to prove the second part of the theorem, which we do by induction. We use a self-confident approach: we consider a function $\beta : \{1, 2, \ldots\} \to [0, +\infty)$ to be defined by the analysis and assume that we have proved that our strategy is such that for some $n \geqslant 1$ and for all sequences of vectors of vector payoffs $m_t \in K$, possibly chosen by some adversary,

$$\|\delta_{n+1}\|_2 = \left\| \sum_{t=1}^{n(n+1)/2} x_t \odot m_t - \sum_{k=1}^{n} k\,\Psi\bigl(\overline{m}^{(k)}\bigr) \odot \overline{m}^{(k)} \right\|_2 \leqslant \beta(n)\,.$$

We then study what we can guarantee for $n+2$. We have

$$\begin{aligned}
\|\delta_{n+2}\|_2^2 &= \left\| \delta_n + \left(\sum_{t=n(n+1)/2+1}^{(n+1)(n+2)/2} x_t \odot m_t - (n+1)\,\Psi\bigl(\overline{m}^{(n+1)}\bigr) \odot \overline{m}^{(n+1)} \right) \right\|_2^2 \\
&= \|\delta_n\|_2^2 + \left\| \sum_{t=n(n+1)/2+1}^{(n+1)(n+2)/2} x_t \odot m_t - (n+1)\,\Psi\bigl(\overline{m}^{(n+1)}\bigr) \odot \overline{m}^{(n+1)} \right\|_2^2 \\
&\qquad +2 \left\langle \delta_n, \sum_{t=n(n+1)/2+1}^{(n+1)(n+2)/2} x_t \odot m_t - (n+1)\,\Psi\bigl(\overline{m}^{(n+1)}\bigr) \odot \overline{m}^{(n+1)} \right\rangle . \qquad (15)
\end{aligned}$$

We upper bound the two squared norms by $\beta(n)^2$ and $(n+1)^2 K_{\max}^2$, respectively. The inner product can be rewritten, with the notation of Figure 1, as

$$\begin{aligned}
&\left\langle \delta_n, \sum_{t=n(n+1)/2+1}^{(n+1)(n+2)/2} x_t \odot m_t - (n+1)\,\Psi\bigl(\overline{m}^{(n+1)}\bigr) \odot \overline{m}^{(n+1)} \right\rangle \\
&\qquad\qquad = -\sum_{t=1}^{n+1} u_{n+1,t} \odot m'_{n+1,t} + \sum_{t=1}^{n+1} u^{(n+1)} \odot m'_{n+1,t} \qquad (16)
\end{aligned}$$

where we used the short-hand notation $u^{(n+1)} = \Psi\bigl(\overline{m}^{(n+1)}\bigr)$. Now, the Cauchy–Schwarz inequality indicates that for all a and t,

$$\bigl|m'_{n+1,t,a}\bigr| \leqslant \|\delta_n\|_2 \, \|m_{n(n+1)/2+t,a}\|_2 \leqslant K_{\max}\,\beta(n)\,,$$

where we used the induction hypothesis. Assumption 1 therefore indicates that the quantity (16) can be bounded by $4K_{\max}\,\beta(n)\sqrt{(n+1)\ln A}$.

Putting everything together, we have proved that the induction holds provided that β is defined, for instance, for all $n \geqslant 1$, as

$$\beta(n+1)^2 = \beta(n)^2 + 8K_{\max}\beta(n)\sqrt{(n+1)\ln A} + K_{\max}^2(n+1)^2 \,.$$

In addition, we have that $\beta(1)^2 = K_{\max}^2$ is a suitable value, by definition of $K_{\max}$. By the lemma below, taking $\gamma_1 = 4K_{\max}\sqrt{\ln A}$ and $\gamma_2 = K_{\max}^2$, we thus get first

$$\beta(n)^2 \leqslant 8K_{\max}^2(\ln A)\, n^3 \qquad \text{or} \qquad \beta(n) \leqslant 2K_{\max}\sqrt{2n^3 \ln A}$$

for all $n \geqslant 1$, hence the final bound

$$\|\delta_{n+1}\|_2 \leqslant 2K_{\max}\sqrt{2n^3 \ln A}$$

still for all $n \geqslant 1$.

It only remains to relate the quantity at hand in (10) to the δ_{n+1}. Actually, T times the quantity whose norm is taken in (10) equals δ_N plus at most N differences of elements in K. Therefore,

$$\left\| \frac{1}{T}\sum_{t=1}^{T} x_t \odot m_t - c_T \right\|_2 \leqslant \frac{1}{T}\big(\|\delta_N\|_2 + NK_{\max}\big) \,.$$

In addition, $N(N+1)/2 \leqslant T$ implies $N \leqslant \sqrt{2T}$, which concludes the proof of the theorem.

Lemma 10 *Consider two positive numbers γ_1, γ_2 and form the positive sequence (u_n) defined by $u_1 = \gamma_2$ and*

$$u_{n+1} = u_n + 2\gamma_1\sqrt{(n+1)\,u_n} + \gamma_2(n+1)^2$$

for all $n \geqslant 1$. Then, for all $n \geqslant 1$,

$$u_n \leqslant \max\big\{2\gamma_1^2,\, \gamma_2\big\}\, n^3 \,.$$

Proof We proceed by induction and note that the relation is satisfied by construction for $n = 1$. Assuming now it holds for some $n \geqslant 1$, we show that it is also true for $n+1$. Denoting $C = \max\big\{2\gamma_1^2,\, \gamma_2\big\}$, we get

$$u_{n+1} = u_n + 2\gamma_1\sqrt{(n+1)\,u_n} + \gamma_2(n+1)^2 \leqslant C\,n^3 + 2\gamma_1\sqrt{C}\sqrt{(n+1)\,n^3} + \gamma_2(n+1)^2 \,.$$

It suffices to show that the latter upper bound is smaller than $C\,(n+1)^3$, which follows from

$$2\gamma_1\sqrt{C}\sqrt{(n+1)\,n^3} + \gamma_2(n+1)^2 \leqslant \big(2\gamma_1\sqrt{2C} + \gamma_2\big)\, n^2 + 2\gamma_2\, n + \gamma_2 \leqslant 3C\,n^2 + 3C\,n + C\,;$$

indeed, the first inequality comes from bounding $n+1$ by 2 and expanding the $(n+1)^2$ term, while the second inequality holds because $C \geqslant \gamma_2$ and $2C \geqslant 2\gamma_1\sqrt{2C}$ by definition of C. ■

A.2. Proof of the convergence (12)

The construction of the strategy at hand and the proof of its performance bound also follow some self-confident approach: denote, for $t \geqslant 1$,

$$\delta_{t+1} = \sum_{s=t}^{t} x_s \odot m_s - \sum_{s=1}^{t} \Psi_{\mathcal{C}}\big(\widetilde{m}_s\big) \odot \widetilde{m}_s \,.$$

No blocks are needed and we proceed as in (15) by developing the square Euclidian norm; we show that the inner product can be forced to be non-positive, which after an immediate recurrence shows that $\|\delta_{T+1}\|_2$ is less than something of the order of $1/\sqrt{T}$, which is the optimal rate for approachability. Indeed, the claimed inequality

$$\langle \delta_{t+1},\, x_{t+1} \odot m_{t+1} \rangle \leqslant \big\langle \delta_{t+1},\, \Psi_{\mathcal{C}}\big(\widetilde{m}_{t+1}\big) \odot \widetilde{m}_{t+1} \big\rangle \tag{17}$$

follows from the following choices, defining the strategy:

$$x_{t+1} \in \mathop{\arg\min}_{x \in \Delta(\mathcal{A})} \max_{m \in K} \langle \delta_{t+1},\, x \odot m \rangle \qquad \text{and} \qquad \widetilde{m}_{t+1} \in \mathop{\arg\max}_{m \in K} \min_{x \in \Delta(\mathcal{A})} \langle \delta_{t+1},\, x \odot m \rangle \,.$$

Then, by von Neumann's minmax theorem, for all $m' \in K$ and $x'\Delta(\mathcal{A})$,

$$\langle \delta_{t+1},\, x_{t+1} \odot m' \rangle \leqslant \min_{x \in \Delta(\mathcal{A})} \max_{m \in K} \langle \delta_{t+1},\, x \odot m \rangle = \max_{m \in K} \min_{x \in \Delta(\mathcal{A})} \langle \delta_{t+1},\, x \odot m \rangle \leqslant \big\langle \delta_{t+1},\, x' \odot \widetilde{m}_{t+1} \big\rangle \,.$$

Choosing $m' = m_{t+1}$ and $x' = \Psi_{\mathcal{C}}\big(\widetilde{m}_{t+1}\big)$ entails (17).