Online learning and Blackwell approachability with partial monitoring: optimal convergence rates

Joon Kwon, Vianney Perchet

BLACKWELL APPROACHABILITY WITH PARTIAL MONITORING: OPTIMAL CONVERGENCE RATES

JOON KWON AND VIANNEY PERCHET

Abstract. We study the Blackwell approachability problem with partial monitoring. When the target set is a polytope and is approachable, we construct, for the first time, approaching strategies with convergence rate of order $T^{-1/2}$ in the case of outcome-dependent signals and of order $T^{-1/3}$ in the case of general signals. Those rates are known to be unimprovable without further assumption on the target set or the signalling structure. It therefore establishes the optimal convergence rates for those two cases. Moreover, the proposed strategies are computationally efficient.

Contents

1. Introduction

Online learning has become a standard topic, especially through regret minimisation [CBL06, SS11, BCB12]: the decision maker aims at controlling some cumulative loss against any possible sequence of loss functions that Nature can generate. However, there exists more general frameworks [RST11], such as Blackwell approachability [Bla56], which is the focus of the present work: the decision maker receives vector-valued payoffs (instead of scalar payoffs) and his goal is to make the average payoff converge to a given *target set*. Blackwell approachability contains regret minimization as a special case, as well as many of its variants: internal/swap regret, online combinatorial optimization, etc. (see e.g. [Per14, Kwo16]).

Date: August 28, 2017.

We here study the *partial monitoring* setting, where the decision maker does not observe his (vector-valued) payoffs. Instead, he recieves a random signal, whose law may depend on his decision and on the state of Nature.

1.1. **Previous work.** In the full information setting, both the regret minimization and approachability problems have a worst-case convergence of rate of order $T^{-1/2}$. The rate deals respectively with the average regret in regret minimization, and the distance of the average payoff to the target set in approachability.

In regret minimization with partial monitoring, depending on the signalling structure, the decision maker may or may not be able to minimize the regret. This has given rise to two main directions of research.

The first one, initiated by [PS01] identifies the signaling structures which allow the average regret to be minimized and aims at constructing strategies in those cases: [PS01] constructed a strategy which guarantees a convergence rate of order $T^{-1/4}$ and [CBLS06] proposed an improved strategy with a $T^{-1/3}$ guarantee as well as a general lower bound of the same order. Later, [BPS10, BFP$^+$14] gave a classification of signaling structures according to convergence rates: they established that the optimal convergence rate is either $T^{-1/2}$, $T^{-1/3}$ or 1—this last rate corresponds to the case where the average regret cannot be minimized.

The second line of research was proposed by [Rus99] and focus on the case where average regret cannot be minimized. In that case, he introduced a weaker variant of the regret, which involves the best performance that the Decision Maker could have achieved in hindsight (had he known the sequence of signal laws, but not the sequence of decisions of Nature), for a given signalling structure. His notion of regret, however, coincide with the standard average regret when the latter can be minimized. [Rus99] however did not provide an explicit strategy nor convergence rates. [MS03] constructed approachability-based strategies in the special case where the law of the signal only depends on Natures action (the so-called *outcome-dependent* case). [LMS08] proposed strategies with convergence rates of order $T^{-1/4}\sqrt{\log T}$ in the case of outcome-dependent signals and of order $T^{-1/5}\sqrt{\log T}$ in the case of general signals. The optimal rate of order $T^{-1/3}$ in the case of general signals (for both internal and external regret) was achieved by [Per11b] using calibration-based algorithms.

More recently, the problem of approachability with partial monitoring was considered by [Per11a]. The regret minimization problem from [Rus99] and the internal regret from [LS07, Per11b] turn out to be special cases of this very general framework. However, the convergence rate of the strategy provided in [Per11a] had the drawback of deteriorating quickly with the dimension of the payoff space, as it scales as $T^{-1/(I+3)}$ where I is the number of actions of the decision maker. A dimension-free rate of order $T^{-1/5}$ was given in [MPS14]—see also [MPS13]. However, the optimal rate of convergence was conjectured to be of order $T^{-1/3}$, as for regret minimization.

Other works on the topic include [LS16] which focus on internal regret minimization with partial monitoring, and [MSZ15] in the context of repeated games with incomplete information.

1.2. **Main contributions.** We construct, for the first time, approachability strategies for polytope target sets with convergence rates of order $T^{-1/3}$ in the case of general signals and of order $T^{-1/2}$ in the case of outcome-dependent signals. Those rates are known to be unimprovable without further assumption on the target set or the signalling structure: in the case of general signals, a lower bound of order $T^{-1/3}$ was given in [CBLS06], and the $T^{-1/2}$ rate is already optimal in the full information setting. It therefore establishes the optimal convergence rates for those two cases. Moreover, the proposed strategies are computationally efficient.

1.3. **Outline.** In Section 2, we present the model of two-player games with vector-valued payoffs and with partial monitoring. In Section 3, we recall the dual characterizations of approachability, both in partial monitoring and in full information. In Section 4, we make sure that the known lower bound of order $T^{-1/3}$ applies to our setting. In Section 5, we first construct an auxiliary full information game which we then use to define the strategy for the initial game. The efficiency of the strategy is discussed. In Section 6 we state and prove Theorem 6.1 which is our main result. It establishes an $T^{-1/3}$ rate of convergence for the strategy. In Section 7, we deal with the special case of outcome-dependent signals for which we propose a modified strategy which is proved in Theorem 7.2 to have an $T^{-1/2}$ rate of convergence.

1.4. **Notation.** Exponents will be used to denote the components of a vector: for instance $x = (x^i)_{i\in\mathcal{I}} \in \mathbb{R}^{\mathcal{I}}$. Bold letters will denote maps and calligraphic letters will denote sets. $\langle \cdot \,|\, \cdot \rangle$ will denote the scalar product.

2. The game

2.1. **Ingredients.** We consider a repeated two-player game with vector-valued payoffs and partial monitoring between the *decision maker* and *Nature*. The decision maker (resp. Nature) has a finite set of *pure actions* $\mathcal{I}$ (resp. $\mathcal{J}$). Denote by

$$\Delta(\mathcal{I}) := \left\{ x = (x^i)_{i\in\mathcal{I}} \in \mathbb{R}_+^{\mathcal{I}} \,\middle|\, \sum_{i\in\mathcal{I}} x^i = 1 \right\}$$

the simplex which represents the set of probability distributions over $\mathcal{I}$. $\Delta(\mathcal{I})$ is also called the set of *mixed actions* of the decision maker. $\Delta(\mathcal{J})$ is defined similarly. Let $\mathbf{g} : \mathcal{I} \times \mathcal{J} \to \mathbb{R}^d$ be the vector-valued payoff function which we bilinearly extend to $\mathbf{g} : \Delta(\mathcal{I}) \times \Delta(\mathcal{J}) \to \mathbb{R}^d$:

$$\mathbf{g}(x, y) := \mathbb{E}_{\substack{i\sim x\\ j\sim y}} [\mathbf{g}(i, j)] = \sum_{\substack{i\in\mathcal{I}\\ j\in\mathcal{J}}} x^i y^j \mathbf{g}(i, j)$$
$$\text{where} \quad x = (x^i)_{i\in\mathcal{I}} \in \Delta(I) \quad \text{and} \quad y = (y^j)_{j\in\mathcal{J}} \in \Delta(\mathcal{J}).$$

Denote by $\left\|\mathbf{g}\right\|_2 := \max_{\substack{i\in\mathcal{I}\\ j\in\mathcal{J}}} \left\|\mathbf{g}(i, j)\right\|_2$ its maximum Euclidean norm. Let $\mathcal{S}$ be a finite set of *signals* and $\mathbf{s} : \mathcal{I} \times \mathcal{J} \to \Delta(\mathcal{S})$ the signal distribution function, which we also bilinearly extend to $\Delta(\mathcal{I}) \times \Delta(\mathcal{J})$. All the above elements are assumed to be known to the decision maker. The special case where the law of the signal $\mathbf{s}(i, j)$ does not depend on i is called the *outcome-dependent signals* case, and will be treated in Section 7.

2.2. **The play.** The game is played as follows. At time $t \geqslant 1$,

- the decision maker and Nature simultaneously choose pure actions $i_t \in \mathcal{I}$ and $j_t \in \mathcal{J}$, possibly at random according to mixed actions $x_t \in \Delta(\mathcal{I})$ and $y_t \in \Delta(\mathcal{J})$;
- the decision maker gets (but does not observe) vector payoff $g_t := \mathbf{g}(i_t, j_t) \in \mathbb{R}^d$;
- the decision maker observes signal $s_t \in \mathcal{S}$ which is drawn according to $\mathbf{s}(i_t, j_t) \in \Delta(\mathcal{S})$.

Formally, a *strategy* for the decision maker is a sequence of measurable maps $\sigma = (\sigma_t)_{t\geqslant 1}$ where $\sigma_t : (\Delta(\mathcal{I}) \times \mathcal{I} \times \mathcal{S})^{t-1} \to \Delta(\mathcal{I})$ indicates the mixed action x_t at time t as a function of the information available to the decision maker. In other words:

$$x_t = \sigma_t(x_1, i_1, s_1, \dots, x_{t-1}, i_{t-1}, s_{t-1}).$$

Similarly, a strategy for Nature is a sequence $(\tau_t)_{t\geqslant 1}$ where $\tau_t : (\Delta(\mathcal{I}) \times \mathcal{I} \times \mathcal{S} \times \Delta(\mathcal{J}) \times \mathcal{J})^{t-1} \to \Delta(\mathcal{J})$, so that

$$y_t = \tau_t(x_1, i_1, s_1, y_1, j_1, \dots, x_{t-1}, i_{t-1}, s_{t-1}, y_{t-1}, j_{t-1}).$$

For $T \geqslant 1$, denote $\overline{g}_T := \frac{1}{T}\sum_{t=1}^{T} g_t$ the average vector payoff up to time T.

2.3. **Flags.** The flag function $\mathbf{f} : \Delta(\mathcal{J}) \to \Delta(\mathcal{S})^{\mathcal{I}}$ is defined by

$$\mathbf{f}(y) = (\mathbf{s}(i,y))_{i\in\mathcal{I}}, \quad y \in \Delta(\mathcal{J}).$$

For $t \geqslant 1$, denote $f_t := \mathbf{f}(y_t)$ the flag associated with y_t. Denote $\mathcal{F} = \mathbf{f}(\Delta(\mathcal{J}))$ the set of all possible flags, which is a polytopial subset of $\mathbb{R}^{\mathcal{S}\times\mathcal{I}}$. The notion of flags is fundamental in games with partial monitoring. Although the decision maker does not directly observe it, he can, as will be shown, estimate it. Moreover, it is the maximal information available to him: two mixed actions $y, y \in \Delta(\mathcal{J})$ from Nature that generate the same flag $\mathbf{f}(y) = \mathbf{f}(y)$ are indistinguishable by the decision maker. For $x \in \Delta(\mathcal{I})$ and $f \in \mathcal{F}$, let $\mathbf{m}(x, f) := \mathbf{g}(x, \mathbf{f}^{-1}(f))$ be the set of all payoffs that are compatible with mixed action x and flag f. The set-valued map $\mathbf{m} : \Delta(\mathcal{I}) \times \mathcal{F} \rightrightarrows \mathbb{R}^d$ will be essential in the statement of the characterization of approachable sets (Proposition 3.2) and in the construction of the strategies.

3. Approachability

We recall the definition of approachability and the characterizations of approachable convex sets in both the partial monitoring and full information cases.

Definition 3.1. A closed convex set $\mathcal{C} \subset \mathbb{R}^d$ is *approachable* if there exists a strategy of the decision maker which guarantees

$$\mathbb{E}\left[\mathbf{d}_2\left(\overline{g}_T,\ \mathcal{C}\right)\right] \xrightarrow[T\to+\infty]{} 0,$$

uniformly in the strategy τ of Nature, where $\overline{g}_T = \sum_{t=1}^{T} g_t$, where $\mathbf{d}_2(\,\cdot\,,\ \mathcal{C})$ denotes the Euclidean distance to $\mathcal{C}$, and where the expectation corresponds to the randomization introduced by the strategies and the signals.

Proposition 3.2 (Characterization of approachable convex sets in games with partial monitoring [Per11a])**.** *A closed convex set $\mathcal{C} \subset \mathbb{R}^d$ is approachable if and only if*

$$\forall f \in \mathcal{F},\ \exists x \in \Delta(\mathcal{I}), \quad \mathbf{m}(x, f) \subset \mathcal{C}.$$

3.1. **In games with full information.** The construction of our strategies in Section 5 will involve an auxiliary full information game. We quickly review the characterizations of approachability in full information games with convex compact action sets and bilinear payoff functions. Let $\mathcal{X}$ and $\mathcal{Y}$ be convex compact action sets and $\mathbf{g} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^d$ a bilinear payoff function. The special case of target sets which are closed convex cones will be of particular importance in the subsequent sections. A few facts about closed convex cones are gathered in Appendix B.

Proposition 3.3 (Characterization of approachability in full information games)**.** *A closed convex set $\mathcal{C} \subset \mathbb{R}^d$ is approachable if and only if one of the following properties hold.*

(i) $\forall g \in \mathbb{R}^d,\ \exists x \in \mathcal{X},\ \forall y \in \mathcal{Y}, \quad \langle \mathbf{g}(x,y) - \mathbf{P}_{\mathcal{C}}(g) | g - \mathbf{P}_{\mathcal{C}}(g) \rangle \leqslant 0$*, where* $\mathbf{P}_{\mathcal{C}}$ *denotes the Euclidean projection on* $\mathcal{C}$*;*

(ii) $\forall y \in \mathcal{Y},\ \exists x \in \mathcal{X}, \quad \mathbf{g}(x,y) \in \mathcal{C}$*.*

Moreover, if $\mathcal{C}$ *is a closed convex cone, the above is also equivalent to*

(iii) $\forall z \in \mathcal{C}^\circ,\ \exists x \in \mathcal{X},\ \forall y \in \mathcal{Y}, \quad \langle \mathbf{g}(x,y) | z \rangle \leqslant 0$*,*

where $\mathcal{C}^\circ$ *denotes the polar cone of* $\mathcal{C}$ *(see Appendix B for definitions and properties about closed convex cones).*

Proof. The first two characterizations are classic [Bla56]. Let us assume that $\mathcal{C}$ is a closed convex cone. Let us prove that (ii) and (iii) are equivalent. $\mathcal{C}$ being a closed convex cone, $\mathbf{g}(x,y) \in \mathcal{C}$ is equivalent to $\max_{z \in \mathcal{C}^\circ} \langle \mathbf{g}(x,y) | z \rangle \leqslant 0$. Then, (ii) can be rewritten

$$\max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} \max_{z \in \mathcal{C}^\circ} \langle \mathbf{g}(x,y) | z \rangle \leqslant 0.$$

$\mathcal{X}$ being compact and the quantity $\langle \mathbf{g}(x,y) | z \rangle$ being linear in x, y and z, we can apply Sions minimax theorem twice to get

$$\max_{z \in \mathcal{C}^\circ} \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \langle \mathbf{g}(x,y) | z \rangle \leqslant 0,$$

which is exactly (iii). □

4. On the lower bound

We recall the label-efficient matching pennies problem, for which a lower bound of order $T^{-1/3}$ is known [CBLS06]. For the sake of completeness, we show that this problem is a special case of our general setting from Section 2, and make sure that the corresponding target set is indeed an *approachable polytope*. Together with the upper bound of order $T^{-1/3}$ that we will obtain in Theorem 6.1, this ensures that the convergence rate of order $T^{-1/3}$ is optimal.

The set of pure actions is $\mathcal{I} = \{a, b, c\}$ for the decision maker and $\mathcal{J} = \{a, b\}$ for Nature. We denote $\ell : \mathcal{I} \times \mathcal{J} \to \{0, 1\}$ the loss function for the decision maker which is defined as follows. If the decision maker and Nature play the same action, the loss is 0; otherwise the loss is 1:

$$\ell(a,a) = \ell(b,b) = 0 \quad \text{and} \quad \ell(a,b) = \ell(b,a) = \ell(c,a) = \ell(c,b) = 1.$$

The set of signals is $\mathcal{S} = \{a, b, d\}$. If the decision maker plays action c, the signal corresponds to the action of Nature. Otherwise, the signal is d, which is uninformative:

$$\mathbf{s}(a,a) = \mathbf{s}(a,b) = \mathbf{s}(b,a) = \mathbf{s}(b,b) = d, \quad \mathbf{s}(c,a) = a \quad \text{and} \quad \mathbf{s}(a,b) = b.$$

The goal of the decision maker is to make its average expected *regret* asymptotically nonpositive:

$$\limsup_{T\to+\infty} \mathbb{E}\left[\frac{1}{T}\left(\sum_{t=1}^{T}\ell(i_t,j_t)-\min_{i\in\mathcal{I}}\sum_{t=1}^{T}\ell(i,j_t)\right)\right]\leqslant 0.$$

We classically rewrite this problem as an approachability problem, by introducing suitable vector-valued payoff function and target set. We consider the following vector-valued payoff function:

$$\mathbf{g}(i,j)=(\ell(i,j)-\ell(i',j))_{i'\in\mathcal{I}}\in\mathbb{R}^{\mathcal{I}},\qquad i\in\mathcal{I},\ j\in\mathcal{J},$$

which bilinearly extends to $\Delta(\mathcal{I})\times\Delta(\mathcal{J})$ as:

$$\mathbf{g}(x,y)=(\ell(x,y)-\ell(i',y))_{i'\in\mathcal{I}},\qquad x\in\Delta(\mathcal{I}),\ y\in\Delta(\mathcal{J}),$$

and consider the negative orthant $\mathbb{R}_-^{\mathcal{I}}$ as the target set, which is indeed a polytope.

Proposition 4.1. *The target set $\mathbb{R}_-^{\mathcal{I}}$ is approachable.*

Proof. First note that by definition of $\mathbf{s}$, the flag function $\mathbf{f}$ writes:

$$\mathbf{f}(y)=\begin{pmatrix}\mathbf{s}(a,y)\\ \mathbf{s}(b,y)\\ \mathbf{s}(c,y)\end{pmatrix}=\begin{pmatrix}\delta_d\\ \delta_d\\ y^a\delta_a+y^b\delta_b\end{pmatrix},\qquad y=(y^a,y^b)\in\Delta(\mathcal{J}),$$

which shows in particular that $\mathbf{f}$ is injective.

Let us prove the result by using the characterization of approachability given by Proposition 3.2. Let $f\in\mathcal{F}$. By definition of $\mathcal{F}$, there exists $y=(y^a,y^b)\in\Delta(\mathcal{J})$ such that $\mathbf{f}(y)=f$. We consider $x\in\Delta(\mathcal{I})$ defined as

$$x=(x^a,x^b,x^c):=\begin{cases}(0,1,0) & \text{if } y^a\leqslant y^b\\ (1,0,0) & \text{if } y^b<y^a.\end{cases}$$

Let us prove that $\mathbf{m}(x,f)\subset\mathbb{R}_-^{\mathcal{I}}$. We have

$$\mathbf{m}(x,f)=\{g(x,y')\,|\,\mathbf{f}(y')=f\}=\{\mathbf{g}(x,y)\},$$

where the last equality stands because $\mathbf{f}$ is injective as we saw. Therefore, it remains to prove that $\mathbf{g}(x,y)\in\mathbb{R}_-^{\mathcal{I}}$. Using the definition of $\mathbf{g}$, this is equivalent to

$$\ell(x,y)\leqslant\ell(i,y),\quad \text{for all } i\in\mathcal{I},$$

which easily follows from the definition of ℓ and the construction of x. □

Besides, the expected average regret is bounded from above by the distance of the average vector payoff $\overline{g}_T:=\sum_{t=1}^{T}\mathbf{g}(i_t,j_t)$ to the target set $\mathbb{R}_-^{\mathcal{I}}$:

$$\begin{aligned}\mathbb{E}\left[\max_{i\in\mathcal{I}}\frac{1}{T}\sum_{t=1}^{T}(\ell(i_t,j_t)-\ell(i,j_t))\right]&=\mathbb{E}\left[\max_{i\in\mathcal{I}}\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}(i_t,j_t)^i\right]=\mathbb{E}\left[\max_{i\in\mathcal{I}}\overline{g}_T^i\right]\\ &\leqslant\mathbb{E}\left[\max_{i\in\mathcal{I}}(\overline{g}_T^i)_+\right]\leqslant\mathbb{E}\left[\sqrt{\sum_{i\in\mathcal{I}}\left(\overline{g}_T^i\right)_+^2}\right]\\ &=\mathbb{E}\left[\mathbf{d}_2\left(\overline{g}_T,\ \mathbb{R}_-^d\right)\right].\end{aligned}$$

The above proves that the lower bound of order $T^{-1/3}$ from [CBLS06] does apply the our case of approachable polytope target sets.

5. Construction of the strategy

We study the case where the target set is the negative orthant $\mathbb{R}^d_-$ and we assume it to be approachable. Since a polytope can be represented as an orthant in a higher dimension space, the extension to polytope target sets can be easily carried out as in e.g. [MPS14, Section 5.4.2]. Most of the proofs are postponed to Section A.

5.1. Bi-piecewise affinity.

We aim in this section at constructing a vector-valued map $\mathbf{r} : \Delta(\mathcal{I}) \times \mathcal{F} \to \mathbb{R}^d$ which can be seen as a simplified version of the set-valued map $\mathbf{m} : \Delta(\mathcal{I}) \times \mathcal{F} \rightrightarrows \mathbb{R}^d$. Its properties will be gathered at the end of the section in Proposition 5.4.

Definition 5.1. Let $\mathcal{U}$ be a convex set and $\mathcal{V}$ a vector space. Let $\mathbf{a} : \mathcal{U} \rightrightarrows \mathcal{V}$ be a set-valued function. $\mathbf{a}$ is *affine* if for all $u, u' \in \mathcal{U}$ and $\lambda \in [0, 1]$,

$$\mathbf{a}(\lambda u + (1 - \lambda)u') = \lambda \mathbf{a}(u) + (1 - \lambda)\mathbf{a}(u').$$

The map $\mathbf{f}$ being affine on $\Delta(\mathcal{J})$ by definition, [RZ96, Proposition 2.4] guarantees the existence of a polytopial decomposition of $\mathcal{F}$ such that $\mathbf{f}^{-1}$ is affine on each of those polytopes. The decomposition can then be refined so that each point of $\mathcal{F}$ can be written as a unique convex combination of the vertices of the polytope to which it belongs. This is formalized by the following lemma.

Lemma 5.2. *There exists a finite family $(\mathcal{F}^k)_{k \in \mathcal{K}}$ of polytopes (denote $\mathcal{B}^k$ the set of vertices of $\mathcal{F}^k$ and $\mathcal{B} = \bigcup_{k \in \mathcal{K}} \mathcal{B}^k$) such that*
- *(i) $\mathcal{F} = \bigcup_{k \in \mathcal{K}} \mathcal{F}^k$;*
- *(ii) for each $k \in \mathcal{K}$, $\mathbf{f}^{-1}$ is affine on $\mathcal{F}^k$;*
- *(iii) for all $f \in \mathcal{F}$, there exists a unique $\mu = (\mu^b)_{b \in \mathcal{B}} \in \Delta(\mathcal{B})$ such that*
  - *(a) $f = \sum_{b \in \mathcal{B}} \mu^b \cdot b$;*
  - *(b) for $k \in \mathcal{K}$, $f \in \mathcal{F}^k \Longrightarrow \operatorname{supp} \mu \subset \mathcal{B}^k$.*

From now on, we assume given such a decomposition.

We are going to construct the map $\mathbf{r} = (\mathbf{r}^n)_{1 \leqslant n \leqslant d}$ component by component, and first on $\Delta(\mathcal{I}) \times \mathcal{B}$ before extending it to $\Delta(\mathcal{I}) \times \mathcal{F}$. Denote $(\mathbf{g}^n)_{1 \leqslant n \leqslant d}$ the components of $\mathbf{g}$. For $x \in \Delta(\mathcal{I})$ and $b \in \mathcal{B}$, we set $\mathbf{r}^n(x, b)$ as being the maximum real number of the set $\mathbf{g}^n(x, \mathbf{f}^{-1}(b))$:

$$\mathbf{r}^n(x, b) := \max \mathbf{g}^n(x, \mathbf{f}^{-1}(b)). \tag{1}$$

We then extend $\mathbf{r}$ to $\Delta(\mathcal{I}) \times \mathcal{F}$ as follows. Using property (iii) from Lemma 5.2, a given flag $f \in \mathcal{F}$ can be uniquely written

$$f = \sum_{b \in \mathcal{B}} \mu^b \cdot b,$$

with $\operatorname{supp} \mu$ contained in one of the polytopes $\mathcal{F}^k$. We then use the above coefficients $(\mu^b)_{b \in \mathcal{B}}$ to define

$$\mathbf{r}^n(x, f) := \sum_{b \in \mathcal{B}} \mu^b \cdot \mathbf{r}^n(x, b). \tag{2}$$

This construction will lead to piecewise affinity of $\mathbf{r}(x, f)$ in f – see Proposition 5.4 below. We now turn to the piecewise affinity in x.

Lemma 5.3. *There exists a finite family of polytopes $(\mathcal{X}^\ell)_{\ell \in \mathcal{L}}$ such that*

(i) $\Delta(\mathcal{I}) = \bigcup_{\ell\in\mathcal{L}} \mathcal{X}^\ell$;
(ii) *For each* $\ell \in \mathcal{L}$ *and* $f \in \mathcal{F}$, $\mathbf{r}(\,\cdot\,, f)$ *is affine on* $\mathcal{X}^\ell$.

Let $\mathcal{A}$ be the set of the vertices of the polytopes $\mathcal{X}^\ell$ given the above lemma. The following proposition summarizes some properties of $\mathbf{r}$.

Proposition 5.4. *(i) For all* $x \in \Delta(\mathcal{I})$, $y \in \Delta(\mathcal{J})$ *and* $1 \leqslant n \leqslant d$, *we have* $\mathbf{g}^n(x, y) \leqslant \mathbf{r}^n(x, \mathbf{f}(y))$;
(ii) *For all* $f \in \mathcal{F}$, *there exists* $x \in \Delta(\mathcal{I})$ *such that* $\mathbf{r}(x, f) \in \mathbb{R}^d_-$;
(iii) *For all* $x \in \Delta(\mathcal{I})$, $\mathbf{r}(x, \cdot\,)$ *is affine on each* $\mathcal{F}^k$ $(k \in \mathcal{K})$;
(iv) *For all* $f \in \mathcal{F}$, $\mathbf{r}(\,\cdot\,, f)$ *is affine on each* $\mathcal{X}^\ell$ $(\ell \in \mathcal{L})$.

5.2. **From bi-piecewise affinity to linearity.** In Section 5.1, we constructed a map $\mathbf{r} : \Delta(\mathcal{I}) \times \mathcal{F} \to \mathbb{R}^d$ which is bi-piecewise affine. In this section, we aim at constructing a *linear map* $\mathbf{R} : (\mathbb{R}^{\mathcal{S}\times\mathcal{I}})^{\mathcal{K}\times\mathcal{A}} \to \mathbb{R}^d$ which encodes the map $\mathbf{r}$ in the following sense. From all pairs $(x, f) \in \Delta(\mathcal{I}) \times \mathcal{F}$, there is a simple construction of a vector $\tilde{g} \in (\mathbb{R}^{\mathcal{S}\times\mathcal{I}})^{\mathcal{K}\times\mathcal{A}}$ such that $\mathbf{R}(\tilde{g}) = \mathbf{r}(x, f)$.

Lemma 5.5. *For every* $k \in \mathcal{K}$, *there exists a map* $\mathbf{r}^{[k]} : \Delta(\mathcal{I}) \times \mathbb{R}^{\mathcal{S}\times\mathcal{I}} \to \mathbb{R}^d$ *such that*
(i) *for all* $x \in \Delta(\mathcal{I})$, *the map* $\mathbf{r}^{[k]}(x, \cdot\,) : \mathbb{R}^{\mathcal{S}\times\mathcal{I}} \to \mathbb{R}^d$ *is linear;*
(ii) *for all* $x \in \Delta(\mathcal{I})$ *and* $f \in \mathcal{F}^k$, $\mathbf{r}^{[k]}(x, f) = \mathbf{r}(x, f)$.

Define $L_{\mathbf{r}}$ as the maximal operator norm of the linear maps $\mathbf{r}^{[k]}(a, \cdot\,)$:

$$L_{\mathbf{r}} := \max_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}} \max_{\substack{f\in\mathbb{R}^{\mathcal{S}\times\mathcal{I}}\\ f\neq 0}} \frac{\left\|\mathbf{r}^{[k]}(a, f)\right\|_2}{\|f\|_2}.$$

Lemma 5.6. $L_{\mathbf{r}}$ *is a common Lipschitz constant to* $\mathbf{r}(a, \cdot\,)$ *and* $\mathbf{r}^{[k]}(a, \cdot\,)$ $(k \in \mathcal{K}$ *and* $a \in \mathcal{A})$. *In other words, for all* $k \in \mathcal{K}$ *and* $a \in \mathcal{A}$, *we have*
(i) *for all* $f, f' \in \mathbb{R}^{\mathcal{S}\times\mathcal{I}}$, $\left\|\mathbf{r}^{[k]}(a, f) - \mathbf{r}^{[k]}(a, f')\right\|_2 \leqslant L_{\mathbf{r}} \|f - f'\|_2$;
(ii) *for all* $f, f' \in \mathcal{F}$, $\|\mathbf{r}(a, f) - \mathbf{r}(a, f')\|_2 \leqslant L_{\mathbf{r}} \|f - f'\|_2$.

For each $k \in \mathcal{K}$, define the linear map $\mathbf{R}_k : (\mathbb{R}^{\mathcal{S}\times\mathcal{I}})^{\mathcal{A}} \to \mathbb{R}^d$ as follows

$$\mathbf{R}_k\left((\tilde{g}^{ka})_{a\in\mathcal{A}}\right) := \sum_{a\in\mathcal{A}} \mathbf{r}^{[k]}(a, \tilde{g}^{ka}), \quad \text{for all } (\tilde{g}^{ka})_{a\in\mathcal{A}} \in (\mathbb{R}^{\mathcal{S}\times\mathcal{I}})^{\mathcal{A}}.$$

Then, define the linear map $\mathbf{R} : (\mathbb{R}^{\mathcal{S}\times\mathcal{I}})^{\mathcal{K}\times\mathcal{A}} \to \mathbb{R}^d$ by setting

$$\begin{aligned}\mathbf{R}(\tilde{g}) &:= \sum_{k\in\mathcal{K}} \mathbf{R}_k\left(\left(\tilde{g}^{ka}\right)_{a\in\mathcal{A}}\right)\\ &= \sum_{k\in\mathcal{K}}\sum_{a\in\mathcal{A}} \mathbf{r}^{[k]}(a, \tilde{g}^{ka}), \quad \text{for all } \tilde{g} = (\tilde{g}^{ka})_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}} \in (\mathbb{R}^{\mathcal{S}\times\mathcal{I}})^{\mathcal{K}\times\mathcal{A}}.\end{aligned}$$

The following proposition shows that $\mathbf{R}$ does indeed encode $\mathbf{r}$.

Proposition 5.7. *Let* $x \in \Delta(\mathcal{I})$, $f \in \mathcal{F}$, $\ell \in \mathcal{L}$ *such that* $x \in \mathcal{X}^\ell$, *and* $k_0 \in \mathcal{K}$ *such that* $f \in \mathcal{F}^{k_0}$. *Moreover, let*

$$x = \sum_{a\in\mathcal{A}} \lambda^a \cdot a \quad \textit{where} \quad \begin{cases}(\lambda^a)_{a\in\mathcal{A}} \in \Delta(\mathcal{A})\\ \operatorname{supp}(\lambda^a)_{a\in\mathcal{A}} \subset \mathcal{X}^\ell.\end{cases}$$

be an expression of x as a convex combination of the vertices of $\mathcal{X}^\ell$. Then,

$$\mathbf{R}\left(\left(\mathbb{1}_{\{k_0=k\}}\lambda^a \cdot f\right)_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}}\right) = \mathbf{r}(x,f).$$

Proof. Using the definition of $\mathbf{R}$,

$$\begin{aligned}\mathbf{R}\left(\left(\mathbb{1}_{\{k_0=k\}}\lambda^a \cdot f\right)_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}}\right) &= \sum_{k\in\mathcal{K}}\sum_{a\in\mathcal{A}} \mathbf{r}^{[k]}(a, \mathbb{1}_{\{k_0=k\}}\lambda^a \cdot f) = \sum_{a\in\mathcal{A}} \lambda^a \cdot \mathbf{r}^{[k_0]}(a,f) \\ &= \sum_{a\in\mathcal{A}} \lambda^a \cdot \mathbf{r}(a,f) = \mathbf{r}(x,f),\end{aligned}$$

where the second equality holds because by linearity of $\mathbf{r}^{[k]}(a,\,\cdot\,)$ (property (i) in Lemma 5.5), the fourth because $\mathbf{r}^{[k_0]}(x,\,\cdot\,)$ and $\mathbf{r}(x,\,\cdot\,)$ coincide on $\mathcal{F}^{k_0}$ (property (ii) in Lemma 5.5), and the last by affinity of $\mathbf{r}(\,\cdot\,,f)$ on $\mathcal{X}^\ell$ (property (iv) in Proposition 5.4). □

5.3. **The auxiliary full information game.** We now construct an auxiliary approachability game. The important point will be that the target set is approachable. This fact will be used in the construction and the analysis of the strategy for the initial game.

The payoff space for this auxiliary game is $(\mathbb{R}^{\mathcal{S}\times\mathcal{I}})^{\mathcal{K}\times\mathcal{A}}$. An element $\tilde{g} \in (\mathbb{R}^{\mathcal{S}\times\mathcal{I}})^{\mathcal{K}\times\mathcal{A}}$ will often be written as

$$\tilde{g} = \left(\tilde{g}^{ka}\right)_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}}, \quad \text{where} \quad \tilde{g}^{ka} \in \mathbb{R}^{\mathcal{S}\times\mathcal{I}}.$$

Then, if $\tilde{z} = \left(\tilde{z}^{ka}\right)_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}}$ also belongs to $(\mathbb{R}^{\mathcal{S}\times\mathcal{I}})^{\mathcal{K}\times\mathcal{A}}$, the scalar product $\langle\tilde{g}|\tilde{z}\rangle$ can obviously be written as the sum of the scalar products $\left\langle\tilde{g}^{ka}\middle|\tilde{z}^{ka}\right\rangle$, and a similar expression holds for the square Euclidean norm:

$$\langle\tilde{g}|\tilde{z}\rangle = \sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}} \left\langle\tilde{g}^{ka}\middle|\tilde{z}^{ka}\right\rangle \quad \text{and} \quad \|\tilde{g}\|_2^2 = \sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}} \left\|\tilde{g}^{ka}\right\|_2^2.$$

The auxiliary game is defined as follows. Let $\mathcal{K}\times\mathcal{A}$ be the set of pure actions for the decision maker and $\mathcal{F}$ the convex action set for Nature. The payoff function $\tilde{\mathbf{g}}$ takes values in $(\mathbb{R}^{\mathcal{S}\times\mathcal{I}})^{\mathcal{K}\times\mathcal{A}}$ and is defined as follows. For $(k,a)\in\mathcal{K}\times\mathcal{A}$ and $f\in\mathcal{F}$,

$$\tilde{\mathbf{g}}((k,a),f) := \left(\mathbb{1}_{\{k=k'\}}\mathbb{1}_{\{a=a'\}}\cdot f\right)_{\substack{k'\in\mathcal{K}\\ a'\in\mathcal{A}}} \quad \in (\mathbb{R}^{\mathcal{S}\times\mathcal{I}})^{\mathcal{K}\times\mathcal{A}}.$$

This payoff function is be bilinearly extended to $\Delta(\mathcal{K}\times\mathcal{A})\times\mathbb{R}^{\mathcal{S}\times\mathcal{I}}$. For each $k\in\mathcal{K}$, let $\mathcal{F}_c^k := \mathbb{R}_+\mathcal{F}^k = (\mathcal{F}^k)^{\circ\circ}$ be the smallest closed convex cone containing the convex compact set $\mathcal{F}^k$ (see Appendix B for definitions and properties about closed convex cones), and consider the following subset of $(\mathbb{R}^{\mathcal{S}\times\mathcal{I}})^{\mathcal{A}}$:

$$\tilde{\mathcal{C}}^k := \mathbf{R}_k^{-1}(\mathbb{R}_-^d)\cap(\mathcal{F}_c^k)^{\mathcal{A}} \quad \subset (\mathbb{R}^{\mathcal{S}\times\mathcal{I}})^{\mathcal{A}}.$$

We then define the target set $\tilde{\mathcal{C}}$ as the Cartesian product of the sets $\tilde{\mathcal{C}}^k$:

$$\tilde{\mathcal{C}} := \prod_{k\in\mathcal{K}} \tilde{\mathcal{C}}^k \quad \subset (\mathbb{R}^{\mathcal{S}\times\mathcal{I}})^{\mathcal{A}\times\mathcal{K}}.$$

Lemma 5.8. *(i) The sets $\tilde{\mathcal{C}}^k$ and $\tilde{\mathcal{C}}$ are closed convex cones.*

(ii) $\tilde{\mathcal{C}} \subset \mathbf{R}^{-1}(\mathbb{R}^d_-) \cap \left(\prod_{k\in\mathcal{K}} (\mathcal{F}^k_c)^{\mathcal{A}} \right)$.

Proposition 5.9. *The set $\tilde{\mathcal{C}}$ is approachable in the auxiliary game. In other words, for all $\tilde{z} \in \tilde{\mathcal{C}}^\circ$, there exists $\tilde{x} := \tilde{\mathbf{x}}(\tilde{z}) \in \Delta(\mathcal{K}\times\mathcal{A})$ such that*

$$\forall f \in \mathcal{F}, \quad \langle \tilde{\mathbf{g}}(\tilde{x}, f) | \tilde{z} \rangle \leqslant 0.$$

Proof. This full information game has convex compact action sets and a bilinear payoff function. Thanks to Proposition 3.3, the statement of the proposition is then equivalent to Blackwell's condition:

$$\forall f \in \mathcal{F},\ \exists\, \tilde{x} \in \Delta(\mathcal{K}\times\mathcal{A}), \quad \tilde{\mathbf{g}}(\tilde{x}, f) \in \tilde{\mathcal{C}},$$

which we now aim at proving. Let $f \in \mathcal{F}$ and $k_0 \in \mathcal{K}$ such that $f \in \mathcal{F}^{k_0}$. According to property (ii) in Proposition 5.4, there exists $x \in \Delta(\mathcal{I})$ such that such that $\mathbf{r}(x, f) \in \mathbb{R}^d_-$. By Lemma 5.3, there exists $\ell \in \mathcal{L}$ such that $x \in \mathcal{X}^\ell$ and we can write x as a convex combination of the vertices of $\mathcal{X}^\ell$:

$$x = \sum_{a\in\mathcal{A}} \lambda^a \cdot a \quad \text{where} \quad \begin{cases} (\lambda^a)_{a\in\mathcal{A}} \in \Delta(\mathcal{A}) \\ \operatorname{supp}(\lambda^a)_{a\in\mathcal{A}} \subset \mathcal{X}^\ell. \end{cases}$$

Now consider the mixed action

$$\tilde{x} := \left(\mathbb{1}_{\{k=k_0\}} \lambda^a \right)_{\substack{k\in\mathcal{K} \\ a\in\mathcal{A}}} \quad \in \Delta(\mathcal{K}\times\mathcal{A})$$

and let us prove that $\tilde{\mathbf{g}}(\tilde{x}, f) \in \tilde{\mathcal{C}}$. We have by definition of $\tilde{\mathbf{g}}$:

$$\tilde{\mathbf{g}}(\tilde{x}, f) = \left(\mathbb{1}_{\{k=k_0\}} \lambda^a \cdot f \right)_{\substack{k\in\mathcal{K} \\ a\in\mathcal{A}}},$$

and since $\tilde{\mathcal{C}} = \prod_{k\in\mathcal{K}} \tilde{\mathcal{C}}^k$, we only have to check that $(\lambda^a f)_{a\in\mathcal{A}}$ belongs to $\tilde{\mathcal{C}}^{k_0} = \mathbf{R}^{-1}_{k_0}(\mathbb{R}^d_-) \cap (\mathcal{F}^{k_0}_c)^{\mathcal{A}}$. First, because $f \in \mathcal{F}^{k_0}$, $\lambda^a f$ belongs to the closed convex cone $\mathcal{F}^{k_0}_c = \mathbb{R}_+ \mathcal{F}^{k_0}$ and we have indeed $(\lambda^a f)_{a\in\mathcal{A}} \in (\mathcal{F}^{k_0}_c)^{\mathcal{A}}$. Then, let us prove that $\mathbf{R}_{k_0}\left((\lambda^a f)_{a\in\mathcal{A}} \right) \in \mathbb{R}^d_-$. Using Proposition 5.7,

$$\mathbf{R}_{k_0}((\lambda^a f)_{a\in\mathcal{A}}) = \mathbf{R}\left(\left(\mathbb{1}_{\{k=k_0\}} \lambda^a \cdot f \right)_{\substack{k\in\mathcal{K} \\ a\in\mathcal{A}}} \right) = \mathbf{r}(x, f) \in \mathbb{R}^d_-.$$

Therefore, we have proved that $(\lambda^a f)_{a\in\mathcal{A}}$ belongs to $\tilde{\mathcal{C}}^{k_0} = \mathbf{R}^{-1}_{k_0}(\mathbb{R}^d_-) \cap (\mathcal{F}^{k_0}_c)^{\mathcal{A}}$, and thus, that $\tilde{\mathbf{g}}(\tilde{x}, f) \in \tilde{\mathcal{C}}$, which concludes the proof. □

5.4. **The strategy for the initial game.** Let $\tilde{\mathcal{Z}} := \tilde{\mathcal{C}}^\circ \cap \mathcal{B}_2$ where $\mathcal{B}_2$ denotes the closed unit Euclidean ball on $(\mathbb{R}^{\mathcal{S}\times\mathcal{I}})^{\mathcal{K}\times\mathcal{A}}$. The strategy is defined as follows. Let $(\eta_t)_{t\geqslant 1}$ be a positive and nonincreasing sequence and $(\gamma_t)_{t\geqslant 1}$ be a nonincreasing sequence with values in $(0, 1]$. For $t \geqslant 1$,

- compute $\tilde{z}_t := \mathbf{P}_{\tilde{\mathcal{Z}}}\left(\eta_{t-1} \sum_{s=1}^{t-1} \tilde{g}_s \right)$, where $\mathbf{P}_{\tilde{\mathcal{Z}}}$ denotes the Euclidean projection onto $\tilde{\mathcal{Z}}$ (the $\tilde{g}_t$'s are defined below);
- compute $\tilde{x}_t := \tilde{\mathbf{x}}\left(\tilde{z}_t\right) \in \Delta(\mathcal{K}\times\mathcal{A})$, where $\tilde{\mathbf{x}}$ is defined in Proposition 5.9;
- draw $(k_t, a_t) \sim \tilde{x}_t$ and then $i_t \sim (1-\gamma_t) a_t + \gamma_t u$, where $u := (\frac{1}{|\mathcal{I}|}, \ldots, \frac{1}{|\mathcal{I}|})$ is the uniform distribution over $\mathcal{I}$;

- observe signal $s_t \sim \mathbf{s}(i_t, j_t)$ and compute estimator
$$\hat{f}_t = \left(\frac{\mathbb{1}_{\{i_t = i\}}}{(1-\gamma_t)a_t^i + \gamma_t/|\mathcal{I}|} \delta_{s_t} \right)_{i \in \mathcal{I}} \quad \in \mathbb{R}^{\mathcal{S} \times \mathcal{I}},$$
where δ_{s_t} is the Dirac mass associated with $s_t \in \mathcal{S}$ and seen as an element of $\mathbb{R}^{\mathcal{S}}$;
- set $\tilde{g}_t = \tilde{\mathbf{g}}((k_t, a_t), \hat{f}_t)$.

Let $(\mathcal{G}_t)_{t \geqslant 1}$ be the filtration where for each $t \geqslant 1$,
$$\mathcal{G}_t \quad \text{is generated by} \quad (k_1, a_1, i_1, s_1, \dots, k_{t-1}, a_{t-1}, i_{t-1}, s_{t-1}, k_t, a_t).$$
The following lemma gathers the properties of estimator $\hat{f}_t$.

Lemma 5.10. *For all $t \geqslant 1$,*

(i) $\mathbb{E}\left[\hat{f}_t \,\middle|\, \mathcal{G}_t\right] = \mathbb{E}\left[f_t \,|\, \mathcal{G}_t\right]$*;*

(ii) $\mathbb{E}\left[\left\|\hat{f}_t\right\|_2^2 \,\middle|\, \mathcal{G}_t\right] \leqslant \dfrac{|\mathcal{I}|^2}{\gamma_t}$*;*

(iii) $\left\|\hat{f}_t\right\|_2^2 \leqslant \dfrac{|\mathcal{I}|^2}{\gamma_t^2}$.

6. Main result

We now state our main result which establishes that the strategy defined in Section 5.4 guarantees that the average payoff $\overline{g}_T$ (of the initial game) converges in expectation to the negative orthant $\mathbb{R}_-^d$ at rate $T^{-1/3}$. This is an improvement over the convergence at rate $T^{-1/5}$ guaranteed by the strategy from [MPS14].

Theorem 6.1. *Against any strategy of Nature, the strategy defined in Section 5.4 with parameters*
$$\gamma_t = \min\left\{1,\ \gamma_0 t^{-1/3}\right\}, \quad t \geqslant 1, \qquad \textit{where} \qquad \gamma_0 = \left(\frac{7L_{\mathbf{r}}\,|\mathcal{I}|\,|\mathcal{K}|\,|\mathcal{A}|}{3\,\|\mathbf{g}\|_2}\right)^{2/3}$$
$$\eta_t = \eta_0 t^{-2/3}, \quad t \geqslant 1, \qquad \textit{where} \qquad \eta_0 = \sqrt{\frac{\gamma_0}{3\,|\mathcal{I}|^2}}.$$
guarantees for all $T \geqslant \gamma_0^3$,
$$\mathbb{E}\left[\mathbf{d}_2\left(\overline{g}_T,\ \mathbb{R}_-^d\right)\right] \leqslant \frac{16\,\|\mathbf{g}\|_2^{1/3}\,(L_{\mathbf{r}}\,|\mathcal{I}|\,|\mathcal{K}|\,|\mathcal{A}|)^{2/3}}{T^{1/3}} + \frac{4\,\|\mathbf{g}\|_2}{\sqrt{T}} + \frac{6\,\|\mathbf{g}\|_2^{2/3}\,(L_{\mathbf{r}}\,|\mathcal{I}|\,|\mathcal{K}|\,|\mathcal{A}|)^{1/3}}{T^{2/3}},$$
where $\mathbf{d}_2(\,\cdot\,,\ \mathbb{R}_-^d)$ denotes the Euclidean distance to the negative orthant $\mathbb{R}_-^d$.

Remark 6.2. Since $L_{\mathbf{r}}$ scales linearly with $\|\mathbf{g}\|_2$, so does the dominant term of the above bound, as expected.

Let us introduce some notation. Let $\overline{\tilde{g}}_T$ be the average for $t = 1, \dots, T$ of auxiliary payoffs $\tilde{g}_t$. In the analysis we will partition the set of stages $\{1, \dots, T\}$ with respect to the realized values of $k_t \in \mathcal{K}$ and $a_t \in \mathcal{A}$. For $k \in \mathcal{K}$ and $a \in \mathcal{A}$, let

$N_T(k,a)$ be the set of stages $t \in \{1,\dots,T\}$ where $k_t = k$ and $a_t = a$, and $\lambda_T(k,a)$ the corresponding proportion of stages:

$$N_T(k,a) := \{1 \leqslant t \leqslant T \,|\, k_t = k,\ a_t = a\}$$
$$\lambda_T(k,a) := \frac{|N_T(k,a)|}{T}.$$

Then, for any sequence $(u_t)_{1\leqslant t\leqslant T}$, we denote $\overline{u}_T(k,a)$ its average over $t \in N_T(k,a)$:

$$\overline{u}_T(k,a) := \begin{cases} \dfrac{1}{|N_T(k,a)|} \displaystyle\sum_{t\in N_T(k,a)} u_t & \text{if } N_T(k,a) \neq \varnothing \\ 0 & \text{otherwise.} \end{cases}$$

The proof is divided into the subsections below which are mostly independent. Here is a overview of the main steps:

$$\begin{aligned} & \overline{\tilde{g}}_T \\ \text{is close to} \quad & \frac{1}{T}\sum_{t=1}^T g(a_t,y_t) \quad \text{(Lemma 6.11)}, \\ \text{which is equal to} \quad & \sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}} \lambda_T(k,a)\cdot \mathbf{g}(a,\overline{y}_T(k,a)) \quad \text{(Lemma 6.10)}, \\ \text{which is closer to } \mathbb{R}^d_- \text{ than} \quad & \sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}} \lambda_T(k,a)\cdot \mathbf{r}(a,\overline{f}_T(k,a)) \quad \text{(Lemma 6.9)}, \\ \text{which is close to} \quad & \sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}} \lambda_T(k,a)\cdot \mathbf{r}^{[k]}(a,\overline{\hat{f}}_T(k,a)) \quad \text{(Lemma 6.8)}, \\ \text{which is equal to} \quad & \mathbf{R}(\overline{\tilde{g}}_T) \quad \text{(Lemma 6.5)}, \\ \text{which is close to} \quad & \mathbb{R}^d_- \quad \text{(Lemmas 6.4 and 6.3)}. \end{aligned}$$

6.1. Average auxiliary payoff $\overline{\tilde{g}}_T$ is close to auxiliary target set $\tilde{\mathcal{C}}$.

Lemma 6.3.

$$\mathbb{E}\left[\mathbf{d}_2\left(\overline{\tilde{g}}_T,\ \tilde{\mathcal{C}}\right)\right] \leqslant \frac{1}{2\eta_T T} + \frac{|\mathcal{I}|^2}{2T\gamma_T}\sum_{t=1}^T \eta_{t-1},$$

Proof. For $t \geqslant 1$, we can write

$$\begin{aligned} \tilde{z}_t = \mathbf{P}_{\tilde{\mathcal{Z}}}\left(\eta_{t-1}\sum_{s=1}^{t-1}\tilde{g}_s\right) &= \operatorname*{argmin}_{\tilde{z}\in\tilde{\mathcal{Z}}} \left\| \tilde{z} - \eta_{t-1}\sum_{s=1}^{t-1}\tilde{g}_s \right\|_2^2 \\ &= \operatorname*{argmax}_{\tilde{z}\in\tilde{\mathcal{Z}}} \left\{ \left\langle \eta_{t-1}\sum_{s=1}^{t-1}\tilde{g}_s \middle| \tilde{z} \right\rangle - \frac{1}{2}\|\tilde{z}\|_2^2 \right\}. \end{aligned}$$

Then, Theorem C.1 together with the fact that $\left\|\tilde{\mathcal{Z}}\right\|_2 = \left\|\tilde{\mathcal{C}}^\circ \cap \mathcal{B}_2\right\|_2 \leqslant 1$ gives

$$\max_{\tilde{z}\in\tilde{\mathcal{Z}}} \sum_{t=1}^T \langle \tilde{g}_t|\tilde{z}\rangle - \sum_{t=1}^T \langle \tilde{g}_t|\tilde{z}_t\rangle \leqslant \frac{1}{2\eta_T} + \frac{1}{2}\sum_{t=1}^T \eta_{t-1}\|\tilde{g}_t\|_2^2.$$

By taking the expectation and dividing by T, we get

$$\mathbb{E}\left[\max_{\tilde{z}\in\tilde{\mathcal{Z}}}\left\langle \overline{\tilde{g}}_T \middle| \tilde{z}\right\rangle\right] \leqslant \frac{1}{2\eta_T T} + \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\langle \tilde{g}_t|\tilde{z}_t\rangle\right] + \frac{1}{2T}\sum_{t=1}^{T}\eta_{t-1}\mathbb{E}\left[\|\tilde{g}_t\|_2^2\right].$$

We first analyze the first sum of the right-hand side. Let us prove that each scalar product $\langle \tilde{g}_t|\tilde{z}_t\rangle$ is nonpositive in expectation. For all $1 \leqslant t \leqslant T$, we replace $\tilde{g}_t$ by its definition:

$$\mathbb{E}\left[\langle \tilde{g}_t|\tilde{z}_t\rangle\right] = \mathbb{E}\left[\left\langle \tilde{\mathbf{g}}((k_t,a_t),\hat{f}_t)\middle|\tilde{z}_t\right\rangle\right].$$

We then consider the conditional expectation with respect to $\mathcal{G}_t$. The application $\tilde{\mathbf{g}}((k_t,a_t),\,\cdot\,)$ being linear, and the variables k_t, a_t and $\tilde{z}_t$ being measurable with respect to $\mathcal{G}_t$, we can make $\mathbb{E}\left[\hat{f}_t \,\middle|\, \mathcal{G}_t\right]$ appear as follows:

$$\begin{aligned}
\mathbb{E}\left[\langle \tilde{g}_t|\tilde{z}_t\rangle\right] &= \mathbb{E}\left[\mathbb{E}\left[\left\langle \tilde{\mathbf{g}}((k_t,a_t),\hat{f}_t)\middle|\tilde{z}_t\right\rangle \,\middle|\, \mathcal{G}_t\right]\right] = \mathbb{E}\left[\left\langle \tilde{\mathbf{g}}\left((k_t,a_t),\mathbb{E}\left[\hat{f}_t\,\middle|\,\mathcal{G}_t\right]\right)\middle|\tilde{z}_t\right\rangle\right] \\
&= \mathbb{E}\left[\langle \tilde{\mathbf{g}}((k_t,a_t),\mathbb{E}\left[f_t\,|\,\mathcal{G}_t\right])|\tilde{z}_t\rangle\right] = \mathbb{E}\left[\langle \tilde{\mathbf{g}}((k_t,a_t),f_t)|\tilde{z}_t\rangle\right],
\end{aligned}$$

where we used Lemma 5.10 to replace the conditional expectation of $\hat{f}_t$ by the conditional expectation of f_t. Now consider the sigma-algebra $\mathcal{H}_t$ generated by

$$(k_1,a_1,i_1,s_1,\ldots,k_{t-1},a_{t-1},i_{t-1},s_{t-1}).$$

By definition of the strategy, the law of random variable (k_t,a_t) knowing $\mathcal{H}_t$ is $\tilde{x}_t$. We now summarize the above computation by introducing the conditional expectation with respect to $\mathcal{H}_t$ and f_t:

$$\begin{aligned}
\mathbb{E}\left[\langle \tilde{g}_t|\tilde{z}_t\rangle\right] &= \mathbb{E}\left[\langle \tilde{\mathbf{g}}((k_t,a_t),f_t)|\tilde{z}_t\rangle\right] = \mathbb{E}\left[\mathbb{E}\left[\langle \tilde{\mathbf{g}}((k_t,a_t),f_t)|\tilde{z}_t\rangle\,|\,\mathcal{H}_t,f_t\right]\right] \\
&= \mathbb{E}\left[\langle \tilde{\mathbf{g}}\left(\mathbb{E}\left[(k_t,a_t)\,|\,\mathcal{H}_t,f_t\right],f_t\right)|\tilde{z}_t\rangle\right] = \mathbb{E}\left[\langle \tilde{\mathbf{g}}\left(\mathbb{E}\left[(k_t,a_t)\,|\,\mathcal{H}_t\right],f_t\right)|\tilde{z}_t\rangle\right] \\
&= \mathbb{E}\left[\langle \tilde{\mathbf{g}}(\tilde{x}_t,f_t)|\tilde{z}_t\rangle\right].
\end{aligned}$$

By definition of the strategy, $\tilde{x}_t = \tilde{\mathbf{x}}(\tilde{z}_t)$. In other words (see Proposition 5.9), for all $f \in \mathcal{F}$, the scalar product $\langle \tilde{\mathbf{g}}(\tilde{x}_t,f)|\tilde{z}_t\rangle$ is nonpositive. This is in particular true for $f = f_t$. Therefore, $\mathbb{E}\left[\langle \tilde{g}_t|\tilde{z}_t\rangle\right] \leqslant 0$.

We now turn to the second sum that involves the squared norms $\|\tilde{g}_t\|_2^2$. For $1 \leqslant t \leqslant T$, using the definition of $\tilde{\mathbf{g}}$,

$$\begin{aligned}
\|\tilde{g}_t\|_2^2 = \left\|\tilde{\mathbf{g}}((k_t,a_t),\hat{f}_t)\right\|_2^2 &= \left\|\left(\mathbb{1}_{\{k=k_t\}}\mathbb{1}_{\{a=a_t\}}\hat{f}_t\right)_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}}\right\|_2^2 \\
&= \sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}}\left\|\mathbb{1}_{\{k=k_t\}}\mathbb{1}_{\{a=a_t\}}\hat{f}_t\right\|_2^2 = \left\|\hat{f}_t\right\|_2^2.
\end{aligned}$$

Using (ii) from Lemma 6.3, we have

$$\mathbb{E}\left[\|\tilde{g}_t\|_2^2\right] = \mathbb{E}\left[\left\|\hat{f}_t\right\|_2^2\right] = \mathbb{E}\left[\mathbb{E}\left[\left\|\hat{f}_t\right\|_2^2\,\middle|\,\mathcal{G}_t\right]\right] \leqslant \frac{|\mathcal{I}|^2}{\gamma_t} \leqslant \frac{|\mathcal{I}|^2}{\gamma_T},$$

where for the last inequality, we used the assumption that sequence $(\gamma_t)_{t\geqslant 1}$ is nonincreasing.

Putting everything together, we obtain in expectation the following bound on the distance from $\bar{\tilde{g}}_T$ to $\tilde{\mathcal{C}}$:

$$\mathbb{E}\left[\mathbf{d}_2\left(\bar{\tilde{g}}_T,\ \tilde{\mathcal{C}}\right)\right] = \mathbb{E}\left[\max_{\tilde{z}\in\tilde{\mathcal{Z}}}\left\langle \bar{\tilde{g}}_T \middle| \tilde{z}\right\rangle\right] \leqslant \frac{1}{2\eta_T T} + \frac{|\mathcal{I}|^2}{2\gamma_T T}\sum_{t=1}^{T}\eta_{t-1},$$

where the above equality comes from the expression of the Euclidean distance to $\tilde{\mathcal{C}}$ given by Proposition B.6. □

6.2. From $\bar{\tilde{g}}_T$ in the auxiliary space to $\mathbf{R}(\bar{\tilde{g}}_T)$ in the initial space.

Lemma 6.4.

$$\mathbf{d}_2\left(\mathbf{R}(\bar{\tilde{g}}_T),\ \mathbb{R}^d_-\right) \leqslant \left(L_{\mathbf{r}}\sqrt{|\mathcal{K}|\,|\mathcal{A}|}\right)\cdot \mathbf{d}_2\left(\bar{\tilde{g}}_T,\ \tilde{\mathcal{C}}\right).$$

Proof. It follows from property (ii) in Lemma 5.8 that $\tilde{\mathcal{C}} \subset \mathbf{R}^{-1}(\mathbb{R}^d_-)$. Therefore, we can write

$$\begin{aligned}
\mathbf{d}_2(\mathbf{R}(\bar{\tilde{g}}_T),\ \mathbb{R}^d_-) &= \min_{g'\in\mathbb{R}^d_-}\left\|\mathbf{R}(\bar{\tilde{g}}_T) - g'\right\|_2 \leqslant \min_{\tilde{g}\in\mathbf{R}^{-1}(\mathbb{R}^d_-)}\left\|\mathbf{R}(\bar{\tilde{g}}_T) - \mathbf{R}(\tilde{g})\right\|_2 \\
&\leqslant \min_{\tilde{g}\in\tilde{\mathcal{C}}}\left\|\mathbf{R}(\bar{\tilde{g}}_T) - \mathbf{R}(\tilde{g})\right\|_2 \leqslant \|\mathbf{R}\|\cdot\min_{\tilde{g}\in\tilde{\mathcal{C}}}\left\|\bar{\tilde{g}}_T - \tilde{g}\right\|_2 \\
&= \|\mathbf{R}\|\cdot\mathbf{d}_2\left(\bar{\tilde{g}}_T,\ \tilde{\mathcal{C}}\right),
\end{aligned}$$

where $\|\mathbf{R}\|$ is the operator norm of $\mathbf{R}$. To conclude the proof, let us prove that the latter is bounded from above by $L_{\mathbf{r}}\sqrt{|\mathcal{K}|\,|\mathcal{A}|}$. Let $\tilde{g} \in (\mathbb{R}^{\mathcal{S}\times\mathcal{I}})^{\mathcal{K}\times\mathcal{A}}$. By definition of $\mathbf{R}$, and using the Lipschitz constant $L_{\mathbf{r}}$ from Lemma 5.6 which is common to the linear applications $\mathbf{r}^{[k]}(a,\,\cdot\,)$, we have

$$\begin{aligned}
\|\mathbf{R}(\tilde{g})\|_2 &= \left\|\sum_{\substack{k\in\mathcal{K}\\a\in\mathcal{A}}}\mathbf{r}^{[k]}(a,\tilde{g}^{ka})\right\|_2 \leqslant \sum_{\substack{k\in\mathcal{K}\\a\in\mathcal{A}}}\left\|\mathbf{r}^{[k]}(a,\tilde{g}^{ka})\right\|_2 \leqslant \sum_{\substack{k\in\mathcal{K}\\a\in\mathcal{A}}} L_{\mathbf{r}}\left\|\tilde{g}^{ka}\right\|_2 \\
&\leqslant L_{\mathbf{r}}\sqrt{|\mathcal{K}|\,|\mathcal{A}|\sum_{\substack{k\in\mathcal{K}\\a\in\mathcal{A}}}\|\tilde{g}^{ka}\|_2^2} = L_{\mathbf{r}}\sqrt{|\mathcal{K}|\,|\mathcal{A}|}\cdot\|\tilde{g}\|_2\,,
\end{aligned}$$

where we used the Cauchy-Schwarz inequality for the third inequality, and the proof is complete. □

6.3. Decomposition of $\mathbf{R}(\bar{\tilde{g}}_T)$.

We have the following expression of the image by $\mathbf{R}$ of the average auxiliary payoff $\bar{\tilde{g}}_T$.

Lemma 6.5.

$$\mathbf{R}(\bar{\tilde{g}}_T) = \mathbf{R}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{g}_t\right) = \sum_{\substack{k\in\mathcal{K}\\a\in\mathcal{A}}}\lambda_T(k,a)\cdot\mathbf{r}^{[k]}(a,\bar{\hat{f}}_T(k,a)).$$

Proof. Using the definitions of $\mathbf{R}$, $\tilde{g}_t$, $\tilde{\mathbf{g}}$, and the linearity of $\mathbf{R}$ and $\mathbf{r}^{[k]}(a,\,\cdot\,)$, we can write

$$\begin{aligned}\mathbf{R}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{g}_t\right) &= \frac{1}{T}\sum_{t=1}^{T}\mathbf{R}(\tilde{g}_t) = \frac{1}{T}\sum_{t=1}^{T}\sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}}\mathbf{r}^{[k]}(a,\tilde{g}_t^{ka}) \\ &= \frac{1}{T}\sum_{t=1}^{T}\sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}}\mathbf{r}^{[k]}\left(a,\mathbb{1}_{\{k=k_t\}}\mathbb{1}_{\{a=a_t\}}\hat{f}_t\right) \\ &= \sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}}\lambda_T(k,a)\cdot\mathbf{r}^{[k]}(a,\overline{\hat{f}}_T(k,a)).\end{aligned}$$

□

6.4. **Average estimator $\overline{\hat{f}}_T(k,a)$ is close to average flag $\overline{f}_T(k,a)$.**

Lemma 6.6.

$$\mathbb{E}\left[\sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}}\lambda_T(k,a)\left\|\overline{\hat{f}}_T(k,a)-\overline{f}_T(k,a)\right\|_2\right] \leqslant |\mathcal{I}|\,|\mathcal{K}|\,|\mathcal{A}|\left(\frac{8}{\sqrt{T\gamma_T}}+\frac{8}{3T\gamma_T}\right).$$

Proof. Let $k\in\mathcal{K}$ and $a\in\mathcal{A}$. Consider the random process $(X_t(k,a))_{t\geqslant 1}$ defined by

$$X_t(k,a) := \mathbb{1}_{\{k_t=k,\ a_t=a\}}\left(\hat{f}_t-f_t\right),$$

and to which we are aiming at applying Corollary D.4. $(X_t(k,a))_{t\geqslant 1}$ is a martingale difference sequence with respect to filtration $(\mathcal{G}_t)_{t\geqslant 1}$. Indeed, since $\mathbb{1}_{\{k_t=k,\ a_t=a\}}$ is measurable with respect to $\mathcal{G}_t$,

$$\mathbb{E}\left[\mathbb{1}_{\{k_t=k,\ a_t=a\}}\left(\hat{f}_t-f_t\right)\,\middle|\,\mathcal{G}_t\right] = \mathbb{1}_{\{k_t=k,\ a_t=a\}}\mathbb{E}\left[\hat{f}_t-f_t\,\middle|\,\mathcal{G}_t\right] = 0.$$

where the last equality follows from (i) in Lemma 5.10. Moreover, using (iii) from Lemma 5.10, we bound each $X_t(k,a)$ as follows.

$$\begin{aligned}\|X_t(k,a)\|_2 &\leqslant \left\|\hat{f}_t-f_t\right\|_2 \leqslant \left\|\hat{f}_t\right\|_2+\|f_t\|_2 \leqslant \frac{|\mathcal{I}|}{\gamma_t}+\left\|(\mathbf{s}(i,y_t))_{i\in\mathcal{I}}\right\|_2 \\ &= \frac{|\mathcal{I}|}{\gamma_t}+\sqrt{\sum_{i\in\mathcal{I}}\|\mathbf{s}(i,y_t)\|_2^2} \leqslant \frac{|\mathcal{I}|}{\gamma_t}+\sqrt{|\mathcal{I}|} \leqslant \frac{2\,|\mathcal{I}|}{\gamma_t} \leqslant \frac{2\,|\mathcal{I}|}{\gamma_T},\end{aligned}$$

where we used the fact that $0<\gamma_t\leqslant 1$ for the penultimate inequality, and the assumption that sequence $(\gamma_t)_{t\geqslant 1}$ is nonincreasing for the last inequality. As far as the conditional variances are concerned, we have

$$\begin{aligned}\mathbb{E}\left[\|X_t(k,a)\|_2^2\,\middle|\,\mathcal{G}_t\right] &= \mathbb{E}\left[\mathbb{1}_{\{k_t=k,\ a_t=a\}}\left\|\hat{f}_t-f_t\right\|_2^2\,\middle|\,\mathcal{G}_t\right] \leqslant \mathbb{E}\left[\left\|\hat{f}_t-f_t\right\|_2^2\,\middle|\,\mathcal{G}_t\right] \\ &\leqslant \mathbb{E}\left[\left\|\hat{f}_t\right\|_2^2\,\middle|\,\mathcal{G}_t\right]+\mathbb{E}\left[\|f_t\|_2^2\,\middle|\,\mathcal{G}_t\right] \leqslant \frac{|\mathcal{I}|^2}{\gamma_t}+|\mathcal{I}| \\ &\leqslant \frac{2\,|\mathcal{I}|^2}{\gamma_t} \leqslant \frac{2\,|\mathcal{I}|^2}{\gamma_T}.\end{aligned}$$

where the first term of the second line has been bounded using property (ii) from Lemma 5.10, whereas the second term is bounded by $|\mathcal{I}|$ since

$$\|f_t\|_2^2 = \left\|(\mathbf{s}(i,y_t))_{i\in\mathcal{I}}\right\|_2^2 = \sum_{i\in\mathcal{I}} \|\mathbf{s}(i,y_t)\|_2^2 \leqslant |\mathcal{I}|\,.$$

Therefore we have

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\|X_t(k,a)\|_2^2 \,\middle|\, \mathcal{G}_t\right] \leqslant \frac{2\,|\mathcal{I}|^2}{\gamma_T}.$$

We can now apply Corollary D.4 with $M = 2\,|\mathcal{I}|\,/\gamma_T$ and $V = 2\,|\mathcal{I}|^2\,/\gamma_T$ to get:

$$\mathbb{E}\left[\left\|\frac{1}{T}\sum_{t=1}^{T} X_t(k,a)\right\|_2\right] \leqslant \frac{8\,|\mathcal{I}|}{\sqrt{T\gamma_T}} + \frac{8\,|\mathcal{I}|}{3T\gamma_T}.$$

Besides, it follows from the definition of $X_t(k,a)$ that

$$\frac{1}{T}\sum_{t=1}^{T} X_t(k,a) = \lambda_T(k,a)\left(\overline{\hat{f}}_T(k,a) - \overline{f}_T(k,a)\right).$$

Finally, by summing over k and a, we obtain:

$$\mathbb{E}\left[\sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}} \lambda_T(k,a)\left\|\left(\overline{\hat{f}}_T(k,a) - \overline{f}_T(k,a)\right)\right\|_2\right] \leqslant |\mathcal{I}|\,|\mathcal{K}|\,|\mathcal{A}|\left(\frac{8}{\sqrt{T\gamma_T}} + \frac{8}{3T\gamma_T}\right).$$

□

6.5. **Average estimator $\overline{\hat{f}}_T(k,a)$ is close to $\mathcal{F}_c^k$.**

Lemma 6.7.

$$\mathbb{E}\left[\sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}} \mathbf{d}_2\left(\overline{\tilde{g}}_T^{ka},\ \mathcal{F}_c^k\right)\right] \leqslant \sqrt{|\mathcal{K}|\,|\mathcal{A}|}\left(\frac{1}{2\eta_T T} + \frac{|\mathcal{I}|^2}{2\gamma_T T}\sum_{t=1}^{T}\eta_{t-1}\right)$$

Proof. Consider the set $\tilde{\mathcal{Z}}_0$ defined by

$$\tilde{\mathcal{Z}}_0 := \prod_{k\in\mathcal{K}}\left((\mathcal{F}_c^k)^\circ \cap \mathcal{B}_2\right)^{\mathcal{A}},$$

and let us assume for the moment that the following inclusion holds:

$$\tilde{\mathcal{Z}}_0 \subset \sqrt{|\mathcal{K}|\,|\mathcal{A}|}\cdot\tilde{\mathcal{Z}}. \tag{3}$$

For each $k\in\mathcal{K}$ and $a\in\mathcal{A}$, $\mathcal{F}_c^k$ being a closed convex cone, Proposition B.6 gives the following expression of the distance of $\overline{\tilde{g}}_T^{ka}$ to $\mathcal{F}_c^k$:

$$\mathbf{d}_2\left(\overline{\tilde{g}}_T^{ka},\ \mathcal{F}_c^k\right) = \max_{\tilde{z}^{ka}\in(\mathcal{F}_c^k)^\circ\cap\mathcal{B}_2}\left\langle \overline{\tilde{g}}_T^{ka} \,\middle|\, \tilde{z}^{ka}\right\rangle.$$

By summing over k and a, we have:

$$\begin{aligned}\sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}} \mathbf{d}_2\left(\overline{\tilde{g}}_T^{ka},\ \mathcal{F}_c^k\right) &= \sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}} \max_{\tilde{z}^{ka}\in(\mathcal{F}_c^k)^\circ\cap\mathcal{B}_2}\left\langle \overline{\tilde{g}}_T^{ka} \,\middle|\, \tilde{z}^{ka}\right\rangle = \max_{\tilde{z}\in\tilde{\mathcal{Z}}_0}\sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}}\left\langle \overline{\tilde{g}}_T \middle| \tilde{z}\right\rangle\\ &\leqslant \sqrt{|\mathcal{K}|\,|\mathcal{A}|}\cdot\max_{\tilde{z}\in\tilde{\mathcal{Z}}}\left\langle \overline{\tilde{g}}_T \middle| \tilde{z}\right\rangle = \sqrt{|\mathcal{K}|\,|\mathcal{A}|}\cdot\mathbf{d}_2\left(\overline{\tilde{g}}_T,\ \tilde{\mathcal{C}}\right),\end{aligned}$$

where for the inequality we used inclusion (3), and for the last equality Proposition B.6 together with the fact that $\tilde{\mathcal{Z}} = \tilde{\mathcal{C}}^\circ \cap \mathcal{B}_2$ by definition. Taking the expectation and substituting distance $\mathbf{d}_2(\overline{\tilde{g}}_T,\ \tilde{\mathcal{C}})$ by the bound from Lemma 6.3 yields the result.

Let us now prove inclusion (3). Let $\tilde{z} = (\tilde{z}^{ka})_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}} \in \tilde{\mathcal{Z}}_0$. First, let us prove that $\tilde{z} \in \tilde{\mathcal{C}}^\circ$. Let $\tilde{g} \in \tilde{\mathcal{C}}$. We can write

$$\langle \tilde{g} | \tilde{z} \rangle = \sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}} \left\langle \tilde{z}^{ka} \middle| \tilde{g}^{ka} \right\rangle.$$

But for each $k \in \mathcal{K}$ and $a \in \mathcal{A}$, by definition of $\tilde{\mathcal{Z}}_0$, we have $\tilde{z}^{ka} \in (\mathcal{F}_c^k)^\circ$, and since $\tilde{\mathcal{C}} \subset \prod_{k\in\mathcal{K}} (\mathcal{F}_c^k)^{\mathcal{A}}$ by definition, we also have $\tilde{g}^{ka} \in \mathcal{F}_c^k$. Therefore, $\left\langle \tilde{g}^{ka} \middle| \tilde{z}^{ka} \right\rangle \leqslant 0$ and consequently, $\langle \tilde{g} | \tilde{z} \rangle \leqslant 0$. This proves $\tilde{\mathcal{Z}}_0 \subset \tilde{\mathcal{C}}^\circ$.

Let $\tilde{z} \in \tilde{\mathcal{Z}}_0$. By definition of $\tilde{\mathcal{Z}}_0$, we have $\left\| \tilde{z}^{ka} \right\|_2 \leqslant 1$ for all $k \in \mathcal{K}$ and $a \in \mathcal{A}$. Thus

$$\left\| \tilde{z} \right\|_2 = \sqrt{\sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}} \left\| \tilde{z}^{ka} \right\|_2^2} \leqslant \sqrt{|\mathcal{K}|\,|\mathcal{A}|},$$

and therefore $\tilde{\mathcal{Z}}_0 \subset \sqrt{|\mathcal{K}|\,|\mathcal{A}|} \cdot \mathcal{B}_2$. Finally, we have

$$\tilde{\mathcal{Z}}_0 \subset \tilde{\mathcal{C}}^\circ \cap \sqrt{|\mathcal{K}|\,|\mathcal{A}|} \cdot \mathcal{B}_2 = \sqrt{|\mathcal{K}|\,|\mathcal{A}|} \cdot \tilde{\mathcal{Z}}.$$

□

6.6. $\mathbf{r}^{[k]}(a, \overline{\hat{f}}_T(k,a))$ **is close to** $\mathbf{r}(a, \overline{f}_T(k,a))$**.**

Lemma 6.8.

$$\mathbb{E}\left[\sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}} \lambda_T(k,a) \left\| \mathbf{r}(a, \overline{f}_T(k,a)) - \mathbf{r}^{[k]}(a, \overline{\hat{f}}_T(k,a)) \right\|_2 \right] \leqslant L_{\mathbf{r}}\, |\mathcal{I}|\, |\mathcal{K}|\, |\mathcal{A}| \left(\frac{8}{\sqrt{T\gamma_T}} + \frac{8}{3T\gamma_T} \right) + L_{\mathbf{r}} \sqrt{|\mathcal{K}|\,|\mathcal{A}|} \left(\frac{1}{\eta_T T} + \frac{|\mathcal{I}|^2}{\gamma_T T} \sum_{t=1}^{T} \eta_{t-1} \right).$$

Proof. Let $(k,a) \in \mathcal{K} \times \mathcal{A}$ and denote $f := \overline{f}_T(k,a)$ and $\hat{f} := \overline{\hat{f}}_T(k,a)$ in this proof only to alleviate notation. Denote $\mathbf{P}^{[k]}$ the Euclidean projection onto $\mathcal{F}_c^k$. Then of course $\mathbf{P}^{[k]}(\hat{f})$ belongs to $\mathcal{F}_c^k$, and since $\mathbf{r}(a,\,\cdot\,)$ and $\mathbf{r}^{[k]}(a,\,\cdot\,)$ coincide on $\mathcal{F}_c^k$ by Lemma 5.5, we can write

$$\begin{aligned}\mathbf{r}(a,f) - \mathbf{r}^{[k]}(a,\hat{f}) = \mathbf{r}(a,f) - \mathbf{r}(a,\hat{f}) + \mathbf{r}(a,\hat{f}) &- \mathbf{r}(a, \mathbf{P}^{[k]}(\hat{f})) \\ &+ \mathbf{r}^{[k]}(a, \mathbf{P}^{[k]}(\hat{f})) - \mathbf{r}^{[k]}(a,\hat{f}).\end{aligned}$$

Thus, by taking the norm and using the triangle inequality and the Lipschitz constant $L_{\mathbf{r}}$ which is common to $\mathbf{r}(a,\,\cdot\,)$ and $\mathbf{r}^{[k]}(a,\,\cdot\,)$, we get

$$\left\| \mathbf{r}(a,f) - \mathbf{r}^{[k]}(a,\hat{f}) \right\|_2 \leqslant L_{\mathbf{r}} \left(\left\| f - \hat{f} \right\|_2 + 2 \cdot \mathbf{d}_2\left(\hat{f},\ \mathcal{F}_c^k \right) \right).$$

We now multiply by $\lambda_T(k,a)$. The last term in the above right-hand side is transformed as

$$2\lambda_T(k,a) \cdot \mathbf{d}_2\left(\hat{f},\ \mathcal{F}_c^k \right) = 2 \cdot \mathbf{d}_2\left(\lambda_T(k,a)\hat{f},\ \mathcal{F}_c^k \right) = 2 \cdot \mathbf{d}_2\left(\overline{\tilde{g}}_T^{ka},\ \mathcal{F}_c^k \right),$$

where used the fact that $\mathcal{F}_c^k$ is a convex cone to push the factor $\lambda_T(k,a)$ into the distance. Therefore,

$$\lambda_T(k,a)\left\|\mathbf{r}(a,f)-\mathbf{r}^{[k]}(a,\hat{f})\right\|_2 \leqslant L_{\mathbf{r}}\cdot\lambda_T(k,a)\left\|f-\hat{f}\right\|_2+2L_{\mathbf{r}}\cdot\mathbf{d}_2\left(\overline{\tilde{g}}_T^{ka},\ \mathcal{F}_c^k\right).$$

Finally, we get the result by taking the expectation, summing over k and a, and plugging Lemmas 6.6 and 6.7. □

6.7. **g is closer to $\mathbb{R}_-^d$ than r.**

Lemma 6.9.

$$\mathbf{d}_2\left(\sum_{\substack{k\in\mathcal{K}\\a\in\mathcal{A}}}\lambda_T(k,a)\cdot\mathbf{g}(a,\overline{y}_T(k,a)),\ \mathbb{R}_-^d\right)\leqslant\mathbf{d}_2\left(\sum_{\substack{k\in\mathcal{K}\\a\in\mathcal{A}}}\lambda_T(k,a)\cdot\mathbf{r}(a,\overline{f}_T(k,a)),\ \mathbb{R}_-^d\right).$$

Proof. Let $k\in\mathcal{K}$ and $a\in\mathcal{A}$ such that $\lambda_T(k,a)>0$, in other words such that $|N_T(k,a)|\geqslant 1$. First note that $\mathbf{f}(\overline{y}_T(k,a))=\overline{f}_T(k,a)$. Indeed, using the affinity of $\mathbf{f}$,

$$\begin{aligned}\mathbf{f}(\overline{y}_T(k,a))&=\mathbf{f}\left(\frac{1}{|N_T(k,a)|}\sum_{t\in N_T(k,a)}y_t\right)=\frac{1}{|N_T(k,a)|}\sum_{t\in N_T(k,a)}\mathbf{f}(y_t)\\&=\frac{1}{|N_T(k,a)|}\sum_{t\in N_T(k,a)}f_t=\overline{f}_T(k,a).\end{aligned}$$

For each component $n\in\{1,\dots,d\}$, we have $\mathbf{g}^n(a,\overline{y}_T(k,a))\leqslant\mathbf{r}^n(a,\overline{f}_T(k,a))$ by property (i) in Proposition 5.4. Finally, using the explicit expression of the Euclidean distance to $\mathbb{R}_-^d$, we have

$$\begin{aligned}\mathbf{d}_2\left(\sum_{\substack{k\in\mathcal{K}\\a\in\mathcal{A}}}\lambda_T(k,a)\cdot\mathbf{g}(a,\overline{y}_T(k,a)),\ \mathbb{R}_-^d\right)&=\sqrt{\sum_{n=1}^d\left(\sum_{\substack{k\in\mathcal{K}\\a\in\mathcal{A}}}\lambda_T(k,a)\cdot\mathbf{g}^n(a,\overline{y}_T(k,a))\right)_+^2}\\&\leqslant\sqrt{\sum_{n=1}^d\left(\sum_{\substack{k\in\mathcal{K}\\a\in\mathcal{A}}}\lambda_T(k,a)\cdot\mathbf{r}^n(a,\overline{f}_T(k,a))\right)_+^2}\\&=\mathbf{d}_2\left(\sum_{\substack{k\in\mathcal{K}\\a\in\mathcal{A}}}\lambda_T(k,a)\cdot\mathbf{r}(a,\overline{f}_T(k,a)),\ \mathbb{R}_-^d\right).\end{aligned}$$

□

6.8. **Decomposition of $\mathbf{g}(a_t,y_t)$ with respect to the realized auxiliary action (k_t,a_t).**

Lemma 6.10.

$$\frac{1}{T}\sum_{t=1}^T\mathbf{g}(a_t,y_t)=\sum_{\substack{k\in\mathcal{K}\\a\in\mathcal{A}}}\lambda_T(k,a)\cdot\mathbf{g}(a,\overline{y}_T(k,a))$$

Proof. Using the definitions of $N_T(k,a)$ and $\lambda_T(k,a)$, and the linearity of $\mathbf{g}(a,\,\cdot\,)$, we have

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}(a_t,y_t) &= \frac{1}{T}\sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}}\sum_{t\in N_T(k,a)}\mathbf{g}(a,y_t)\\ &= \sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}}\frac{|N_T(k,a)|}{T}\cdot\frac{1}{|N_T(k,a)|}\sum_{t\in N_T(k,a)}\mathbf{g}(a,y_t)\\ &= \sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}}\lambda_T(k,a)\cdot\mathbf{g}(a,\overline{y}_T(k,a)).\end{aligned}$$

□

6.9. **From $\mathbf{g}(i_t,j_t)$ to $\mathbf{g}(a_t,y_t)$.**

Lemma 6.11.

$$\mathbb{E}\left[\left\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}(i_t,j_t)-\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}(a_t,y_t)\right\|_2\right]\leqslant\frac{2\sqrt{\pi}\,\|\mathbf{g}\|_2}{\sqrt{T}}+\frac{2\,\|\mathbf{g}\|_2}{T}\sum_{t=1}^{T}\gamma_t.$$

Proof. Consider the process $(X_t)_{t\geqslant 1}$ defined by

$$X_t=\mathbf{g}(i_t,j_t)-(1-\gamma_t)\mathbf{g}(a_t,y_t)-\gamma_t\mathbf{g}(u,y_t),$$

and the filtration $(\mathcal{G}'_t)_{t\geqslant 1}$ where $\mathcal{G}'_t$ is generated by

$$(k_1,a_1,y_1,i_1,s_1,\dots,k_{t-1},a_{t-1},y_{t-1},i_{t-1},s_{t-1},k_t,a_t,y_t).$$

$(X_t)_{t\geqslant 1}$ is a martingale difference sequence with respect to filtration $(\mathcal{G}'_t)_{t\geqslant 1}$. Indeed, knowing $\mathcal{G}'_t$, the law of i_t is $(1-\gamma_t)a_t+\gamma_t u$ by definition of the strategy, and thus the law of (i_t,j_t) is $((1-\gamma_t)a_t+\gamma_t u)\otimes y_t$. We can then write, by bilinearity of $\mathbf{g}$:

$$\mathbb{E}\left[\mathbf{g}(i_t,j_t)\,|\,\mathcal{G}'_t\right]=(1-\gamma_t)\mathbf{g}(a_t,y_t)+\gamma_t\mathbf{g}(u,y_t).$$

Moreover, $\|X_t\|_2$ is always bounded by $2\,\|\mathbf{g}\|_2$:

$$\begin{aligned}\|X_t\|_2 &= \|(1-\gamma_t)\left(\mathbf{g}(i_t,j_t)-\mathbf{g}(a_t,y_t)\right)+\gamma_t\left(\mathbf{g}(i_t,j_t)-\mathbf{g}(u,y_t)\right)\|_2\\ &\leqslant (1-\gamma_t)\,\|\mathbf{g}(i_t,j_t)-\mathbf{g}(a_t,y_t)\|_2+\gamma_t\,\|\mathbf{g}(i_t,j_t)-\mathbf{g}(u,y_t)\|_2\\ &\leqslant 2\,\|\mathbf{g}\|_2\,.\end{aligned}$$

We can thus apply Corollary D.2 with $M=2\,\|\mathbf{g}\|_2$ to get

$$\mathbb{E}\left[\left\|\frac{1}{T}\sum_{t=1}^{T}X_t\right\|_2\right]\leqslant\frac{2\sqrt{\pi}\,\|\mathbf{g}\|_2}{\sqrt{T}}.$$

Therefore,

$$\begin{aligned}\left\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}(i_t,j_t)-\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}(a_t,y_t)\right\|_2 &= \left\|\frac{1}{T}\sum_{t=1}^{T}\left(X_t+\gamma_t(\mathbf{g}(u,y_t)-\mathbf{g}(a_t,y_t))\right)\right\|_2\\ &\leqslant \left\|\frac{1}{T}\sum_{t=1}^{T}X_t\right\|_2+\left\|\frac{1}{T}\sum_{t=1}^{T}\gamma_t\left(\mathbf{g}(u,y_t)-\mathbf{g}(a_t,y_t)\right)\right\|_2\\ &\leqslant \left\|\frac{1}{T}\sum_{t=1}^{T}X_t\right\|_2+\frac{2\,\|\mathbf{g}\|_2}{T}\sum_{t=1}^{T}\gamma_t,\end{aligned}$$

And taking the expectation:

$$\mathbb{E}\left[\left\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}(i_t,j_t)-\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}(a_t,y_t)\right\|_2\right]\leqslant\frac{2\sqrt{\pi}\left\|\mathbf{g}\right\|_2}{\sqrt{T}}+\frac{2\left\|\mathbf{g}\right\|_2}{T}\sum_{t=1}^{T}\gamma_t.$$

□

6.10. **Final bound.** We now combine the above lemmas in the order specified at the beginning of the section to get:

(4)

$$\mathbb{E}\left[\mathbf{d}_2\left(\bar{g}_T,\ \mathbb{R}^d_-\right)\right]\leqslant\frac{2\sqrt{\pi}\left\|\mathbf{g}\right\|_2}{\sqrt{T}}+\frac{2\left\|\mathbf{g}\right\|_2}{T}\sum_{t=1}^{T}\gamma_t+L_{\mathbf{r}}\left|\mathcal{I}\right|\left|\mathcal{K}\right|\left|\mathcal{A}\right|\left(\frac{8}{\sqrt{T\gamma_T}}+\frac{8}{3T\gamma_T}\right)$$
$$+\frac{3L_{\mathbf{r}}}{2}\sqrt{\left|\mathcal{K}\right|\left|\mathcal{A}\right|}\left(\frac{1}{\eta_T T}+\frac{\left|\mathcal{I}\right|^2}{\gamma_T T}\sum_{t=1}^{T}\eta_{t-1}\right).$$

We take care of the above second term as follows. Using the fact that $\gamma_t\leqslant\gamma_0 t^{-1/3}$, we have:

$$\frac{1}{T}\sum_{t=1}^{T}\gamma_t\leqslant\frac{1}{T}\sum_{t=1}^{T}\gamma_0 t^{-1/3}\leqslant\frac{\gamma_0}{T}\int_0^T t^{-1/3}\,\mathrm{d}t=\frac{3\gamma_0}{2}T^{-1/3}.$$

Besides, we assume from now on that $T\geqslant\gamma_0^3$, so that $\gamma_T=\gamma_0 T^{-1/3}$. Then, the third term from (4) is equal to

$$8L_{\mathbf{r}}\left|\mathcal{I}\right|\left|\mathcal{K}\right|\left|\mathcal{A}\right|\gamma_0^{-1/2}T^{-1/3}+\frac{8}{3}L_{\mathbf{r}}\left|\mathcal{I}\right|\left|\mathcal{K}\right|\left|\mathcal{A}\right|\gamma_0^{-1}T^{-2/3}.$$

Also, the expression $\eta_t=\eta_0 t^{-2/3}$ allows to write

$$\sum_{t=1}^{T}\eta_{t-1}=\eta_0\left(2+\sum_{t=2}^{T-1}t^{-2/3}\right)\leqslant\eta_0\left(\int_0^1 t^{-2/3}\,\mathrm{d}t+\int_1^{T-1}t^{-2/3}\,\mathrm{d}t\right)=3\eta_0 T^{1/3},$$

which gives, after simplification, the following upper bound on the last term from (4):

$$\frac{3L_{\mathbf{r}}\sqrt{\left|\mathcal{K}\right|\left|\mathcal{A}\right|}}{2}\left(\eta_0^{-1}+3\left|\mathcal{I}\right|^2\gamma_0^{-1}\eta_0\right)T^{-1/3}.$$

Therefore, injecting the expression $\eta_0=\sqrt{\gamma_0/3\left|\mathcal{I}\right|^2}$ and rearranging the terms according to their respective dependencies in T, the bound becomes:

$$\mathbb{E}\left[\mathbf{d}_2\left(\bar{g}_T,\ \mathbb{R}^d_-\right)\right]\leqslant\left(3\left\|\mathbf{g}\right\|_2\gamma_0+8L_{\mathbf{r}}\left|\mathcal{I}\right|\left|\mathcal{K}\right|\left|\mathcal{A}\right|\gamma_0^{-1/2}+3\sqrt{3}L_{\mathbf{r}}\sqrt{\left|\mathcal{K}\right|\left|\mathcal{A}\right|}\left|\mathcal{I}\right|\gamma_0^{-1/2}\right)T^{-1/3}$$
$$+2\sqrt{\pi}\left\|\mathbf{g}\right\|_2T^{-1/2}+\frac{8}{3}L_{\mathbf{r}}\left|\mathcal{I}\right|\left|\mathcal{K}\right|\left|\mathcal{A}\right|\gamma_0^{-1}T^{-2/3}.$$

The above first term is further bounded from above by:

$$\left(3\left\|\mathbf{g}\right\|_2\gamma_0+14L_{\mathbf{r}}\left|\mathcal{I}\right|\left|\mathcal{K}\right|\left|\mathcal{A}\right|\gamma_0^{-1/2}\right)T^{-1/3}.$$

Eventually, injecting the expression $\gamma_0 = (7L_{\mathbf{r}} |\mathcal{I}| |\mathcal{K}| |\mathcal{A}| / 3 \|\mathbf{g}\|_0)^{2/3}$ gives our final bound:

$$\mathbb{E}\left[\mathbf{d}_2\left(\overline{g}_T,\ \mathbb{R}^d_-\right)\right] \leqslant 16 \|\mathbf{g}\|_2^{1/3} \left(L_{\mathbf{r}} |\mathcal{I}| |\mathcal{K}| |\mathcal{A}|\right)^{2/3} T^{-1/3} + 4 \|\mathbf{g}\|_2 T^{-1/2} + 6 \|\mathbf{g}\|_2^{2/3} \left(L_{\mathbf{r}} |\mathcal{I}| |\mathcal{K}| |\mathcal{A}|\right)^{1/3} T^{-2/3}.$$

7. Outcome-dependent signals

This section studies the special case where the law $\mathbf{s}(i,j)$ of the signal does not depend on the pure action i of the decision maker. In other words, we assume that

$$\mathbf{s}(\,\cdot\,, j) \quad \text{is constant,} \quad \text{for all } j \in \mathcal{J}.$$

We aim at constructing a strategy which achieves a $T^{-1/2}$ convergence rate. Again, we assume that the target set is the negative orthant $\mathbb{R}^d_-$ and that it is approachable. We will heavily rely on elements from the previous sections. To take advantage of the above assumption, the strategy from Section 5 will be modified in two ways. First, the estimator will be simpler since exploration is unnecessary, and second, the mixed action of the decision maker will not be perturbed with the uniform distribution. Unless stated otherwise, all previous notation and assumptions stand.

The modified strategy is defined as follows. Let $(\eta_t)_{t \geqslant 1}$ be a positive and nonincreasing sequence. For $1 \leqslant t \leqslant T$;

- compute $\tilde{z}_t = \mathbf{P}_{\tilde{\mathcal{Z}}}\left(\eta_{t-1} \sum_{s=1}^{t-1} \tilde{g}_s\right)$ and $\tilde{x}_t := \tilde{\mathbf{x}}(\tilde{z}_t) \in \Delta(\mathcal{K} \times \mathcal{A})$.
- draw $(k_t, a_t) \sim \tilde{x}_t$ and then $i_t \sim a_t$;
- observe signal $s_t \in \mathcal{S}$ and compute estimator
$$\hat{f}_t = \left(\delta_{s_t}\right)_{i \in \mathcal{I}} \quad \in \mathbb{R}^{\mathcal{S} \times \mathcal{I}};$$
- set $\tilde{g}_t = \tilde{\mathbf{g}}((k_t, a_t), \hat{f}_t)$.

The definition of the strategy implies that the law of i_t knowing $\mathcal{G}_t$ is a_t. Let us state the properties of the new estimator.

Lemma 7.1. *For $t \geqslant 1$,*

(i) $\mathbb{E}\left[\hat{f}_t \,\middle|\, \mathcal{G}_t\right] = \mathbb{E}\left[f_t \,|\, \mathcal{G}_t\right]$;

(ii) $\left\|\hat{f}_t\right\|_2^2 = |\mathcal{I}|$.

Theorem 7.2. *Against any strategy of Nature, the above strategy with parameters*

$$\eta_t = \eta_0 t^{-1/2}, \quad t \geqslant 1, \qquad \textit{where} \qquad \eta_0 = \sqrt{\frac{1}{2\,|\mathcal{I}|}}$$

guarantees for all $T \geqslant 1$:

$$\mathbb{E}\left[\mathbf{d}_2\left(\overline{g}_T,\ \mathbb{R}^d_-\right)\right] \leqslant \frac{4\|\mathbf{g}\|_2 + 8L_{\mathbf{r}}\sqrt{|\mathcal{I}|}\,|\mathcal{K}|\,|\mathcal{A}|}{\sqrt{T}}.$$

One can check that statements from Lemmas 6.4, 6.5, 6.9 and 6.10 still hold. We state and prove below new versions of the remaining lemmas, which were affected by the modifications of the estimator and the law of i_t. The analysis can be summarized as follows.

$$\begin{aligned}
\overline{g}_T \quad &\text{is close to} \quad \frac{1}{T}\sum_{t=1}^{T} g(a_t, y_t) \quad \text{(Lemma 7.7)}\\
&\text{is equal to} \quad \sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}} \lambda_T(k,a)\cdot \mathbf{g}(a, \overline{y}_T(k,a)) \quad \text{(Lemma 6.10)}\\
\text{is closer to } \mathbb{R}^d_- &\text{ than} \quad \sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}} \lambda_T(k,a)\cdot \mathbf{r}(a, \overline{f}_T(k,a)) \quad \text{(Lemma 6.9)}\\
&\text{is close to} \quad \sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}} \lambda_T(k,a)\cdot \mathbf{r}^{[k]}(a, \overline{\hat{f}}_T(k,a)) \quad \text{(Lemma 7.6)}\\
&\text{is equal to} \quad \mathbf{R}(\overline{\tilde{g}}_T) \quad \text{(Lemma 6.5)}\\
&\text{is close to} \quad \mathbb{R}^d_- \quad \text{(Lemmas 6.4 and 7.3).}
\end{aligned}$$

7.1. Average auxiliary payoff $\overline{\tilde{g}}_T$ is close to auxiliary target set $\tilde{\mathcal{C}}$.

Lemma 7.3.

$$\mathbb{E}\left[\mathbf{d}_2\left(\overline{\tilde{g}}_T,\ \tilde{\mathcal{C}}\right)\right] \leqslant \frac{1}{2\eta_T T} + \frac{|\mathcal{I}|}{2T}\sum_{t=1}^{T}\eta_{t-1}.$$

Proof. We follow the proof of Lemma 6.4. The regret bound given by Theorem C.1 still holds:

$$\max_{\tilde{z}\in\tilde{\mathcal{Z}}}\sum_{t=1}^{T}\langle \tilde{g}_t|\tilde{z}\rangle - \sum_{t=1}^{T}\langle \tilde{g}_t|\tilde{z}_t\rangle \leqslant \frac{1}{2\eta_T} + \frac{1}{2T}\sum_{t=1}^{T}\eta_{t-1}\left\|\tilde{g}_t\right\|_2^2.$$

In Lemma 6.4, the second sum was nonpositive in expectation thanks to the fact that $\mathbb{E}\left[\hat{f}_t \,\middle|\, \mathcal{G}_t\right] = \mathbb{E}\left[f_t \,|\, \mathcal{G}_t\right]$. The same reasoning can be applied in the present case since the property of the estimator is guaranteed by Lemma 7.1. Therefore, we have

$$\mathbb{E}\left[\mathbf{d}_2\left(\overline{\tilde{g}}_T,\ \tilde{C}\right)\right] \leqslant \frac{1}{2\eta_T T} + \frac{1}{2T}\sum_{t=1}^{T}\eta_{t-1}\mathbb{E}\left[\left\|\tilde{g}_t\right\|_2^2\right].$$

Then, for $1 \leqslant t \leqslant T$, we have

$$\left\|\tilde{g}_t\right\|_2^2 = \left\|\tilde{\mathbf{g}}((k_t,a_t),\hat{f}_t)\right\|_2^2 = \left\|\left(\mathbb{1}_{\{k_t=k,\ a_t=a\}}\hat{f}_t\right)_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}}\right\|_2^2 = \left\|\hat{f}_t\right\|_2^2 = |\mathcal{I}|\,,$$

where we used property (ii) from Lemma 7.1 for the last equality. The result follows. □

7.2. Average estimator $\overline{\hat{f}}_T(k,a)$ is close to average flag $\overline{f}_T(k,a)$.

Lemma 7.4.

$$\mathbb{E}\left[\sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}}\lambda_T(k,a)\left\|\overline{\hat{f}}_T(k,a) - \overline{f}_T(k,a)\right\|_2\right] \leqslant 2\,|\mathcal{K}|\,|\mathcal{A}|\sqrt{\frac{\pi\,|\mathcal{I}|}{T}}.$$

Proof. Let $k \in \mathcal{K}$ and $a \in \mathcal{A}$. As in Lemma 6.6, we consider

$$X_t(k,a) := \mathbb{1}_{\{k_t=k,\ a_t=a\}} \left(\hat{f}_t - f_t\right),$$

which is a sequence of martingale differences with respect to filtration $(\mathcal{G}_t)_{t\geqslant 1}$ thanks to property (i) from Lemma 7.1. But this time, we use Corollary D.2 instead of Corollary D.4. Each X_t is bounded as follows

$$\|X_t(k,a)\|_2 \leqslant \left\|\hat{f}_t\right\|_2 + \|f_t\|_2 = \sqrt{|\mathcal{I}|} + \sqrt{\sum_{i\in\mathcal{I}} \|\mathbf{s}(i,y_t)\|_2^2} \leqslant 2\sqrt{|\mathcal{I}|},$$

where we used property (ii) from Lemma 7.1. Corollary D.2 then gives

$$\mathbb{E}\left[\lambda_T(k,a) \left\|\overline{\hat{f}}_T(k,a) - \overline{f}_T(k,a)\right\|_2\right] = \mathbb{E}\left[\left\|\frac{1}{T}\sum_{t=1}^T X_t(k,a)\right\|_2\right] \leqslant 2\sqrt{\frac{\pi\,|\mathcal{I}|}{T}}.$$

The result follows by summing over $k \in \mathcal{K}$ and $a \in \mathcal{A}$. □

7.3. Average estimator $\overline{\hat{f}}_T(k,a)$ is close to $\mathcal{F}_c^k$.

Lemma 7.5.

$$\mathbb{E}\left[\sum_{\substack{k\in\mathcal{K}\\a\in\mathcal{A}}} \mathbf{d}_2\left(\overline{\tilde{g}}_T^{ka},\ \mathcal{F}_c^k\right)\right] \leqslant \sqrt{|\mathcal{K}|\,|\mathcal{A}|}\left(\frac{1}{2\eta_T T} + \frac{|\mathcal{I}|}{2T}\sum_{t=1}^T \eta_{t-1}\right).$$

Proof. The following inequality from the proof of Lemma 6.7 still holds

$$\sum_{\substack{k\in\mathcal{K}\\a\in\mathcal{A}}} \mathbf{d}_2\left(\overline{\tilde{g}}_T^{ka},\ \mathcal{F}_c^k\right) \leqslant \sqrt{|\mathcal{K}|\,|\mathcal{A}|}\cdot \mathbf{d}_2\left(\overline{\tilde{g}}_T,\ \tilde{\mathcal{C}}\right).$$

Then, taking the expectation and injecting the new bound on $\mathbb{E}\left[\mathbf{d}_2\left(\overline{\tilde{g}}_T,\ \tilde{\mathcal{C}}\right)\right]$ given by Lemma 7.3 yields the result. □

7.4. $\mathbf{r}^{[k]}(a, \overline{\hat{f}}_T(k,a))$ is close to $\mathbf{r}(a, \overline{f}_T(k,a))$.

Lemma 7.6. *For all $k \in \mathcal{K}$ and $a \in \mathcal{A}$,*

$$\mathbb{E}\left[\sum_{\substack{k\in\mathcal{K}\\a\in\mathcal{A}}} \lambda_T(k,a)\left\|\mathbf{r}(a,\overline{f}_T(k,a)) - \mathbf{r}^{[k]}(a,\overline{\hat{f}}_T(k,a))\right\|_2\right] \leqslant 2L_{\mathbf{r}}\,|\mathcal{K}|\,|\mathcal{A}|\sqrt{\frac{\pi\,|\mathcal{I}|}{T}}$$
$$+ L_{\mathbf{r}}\sqrt{|\mathcal{K}|\,|\mathcal{A}|}\left(\frac{1}{\eta_T T} + \frac{|\mathcal{I}|}{T}\sum_{t=1}^T \eta_{t-1}\right).$$

Proof. Let $k \in \mathcal{K}$ and $a \in \mathcal{A}$. Using notation $f = \overline{f}_T(k,a)$ and $\hat{f} = \overline{\hat{f}}_T(k,a)$, the following inequality from the proof of Lemma 6.8 still holds

$$\lambda_T(k,a)\left\|\mathbf{r}(a,f) - \mathbf{r}^{[k]}(a,\hat{f})\right\|_2 \leqslant L_{\mathbf{r}}\cdot\lambda_T(k,a)\left\|f-\hat{f}\right\|_2 + 2L_{\mathbf{r}}\cdot\mathbf{d}_2\left(\overline{\tilde{g}}_T^{ka},\ \mathcal{F}_c^k\right).$$

The result follows from taking the expectation, summing over $k \in \mathcal{K}$ and $\mathcal{A}$, and injecting the bounds from Lemmas 7.4 and 7.5. □

7.5. **From $\mathbf{g}(i_t, j_t)$ to $\mathbf{g}(a_t, y_t)$.**

Lemma 7.7.

$$\mathbb{E}\left[\left\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}(i_t, j_t) - \frac{1}{T}\sum_{t=1}^{T}\mathbf{g}(a_t, y_t)\right\|_2\right] \leqslant \frac{2\sqrt{\pi}\,\|\mathbf{g}\|_2}{\sqrt{T}}.$$

Proof. The process $(\mathbf{g}(i_t, j_t) - \mathbf{g}(a_t, y_t))_{t\geqslant 1}$ is a martingale difference sequence with respect to filtration $(\mathcal{G}'_t)_{t\geqslant 1}$ introduced in the proof of Lemma 6.11. It is moreover bounded by $2\,\|\mathbf{g}\|_2$. Therefore, Corollary D.2 gives:

$$\mathbb{E}\left[\left\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}(i_t, j_t) - \frac{1}{T}\sum_{t=1}^{T}\mathbf{g}(a_t, y_t)\right\|_2\right] \leqslant \frac{2\sqrt{\pi}\,\|\mathbf{g}\|_2}{\sqrt{T}}.$$

□

7.6. **Final bound.** Similarly to the proof of Theorem 6.1, the combination of the above lemmas gives:

$$\mathbb{E}\left[\mathbf{d}_2\left(\bar{g}_T,\ \mathbb{R}^d_-\right)\right] \leqslant \frac{2\sqrt{\pi}\,\|\mathbf{g}\|_2}{\sqrt{T}} + \frac{2\sqrt{\pi}L_{\mathbf{r}}\,|\mathcal{K}|\,|\mathcal{A}|\,\sqrt{|\mathcal{I}|}}{\sqrt{T}} + \frac{3L_{\mathbf{r}}\sqrt{|\mathcal{K}|\,|\mathcal{A}|}}{2}\left(\frac{1}{\eta_T T} + \frac{|\mathcal{I}|}{T}\sum_{t=1}^{T}\eta_{t-1}\right). \tag{5}$$

The expression $\eta_t = \eta_0 t^{-1/2}$ allows to write

$$\sum_{t=1}^{T}\eta_{t-1} = \eta_0\left(2 + \sum_{t=2}^{T-1}t^{-1/2}\right) \leqslant \eta_0\left(\int_0^1 t^{-1/2}\,\mathrm{d}t + \int_1^{T-1}t^{-1/2}\,\mathrm{d}t\right) \leqslant 2\eta_0 T^{1/2}.$$

Then, the last term from (5) is bounded from above by

$$\frac{3L_{\mathbf{r}}\sqrt{|\mathcal{K}|\,|\mathcal{A}|}}{2}\left(\frac{1}{\eta_0} + 2\,|\mathcal{I}|\,\eta_0\right)T^{-1/2}.$$

Injecting the expression $\eta_0 = \sqrt{1/2\,|\mathcal{I}|}$ and simplifying finally yields the result:

$$\mathbb{E}\left[\mathbf{d}_2\left(\bar{g}_T,\ \mathbb{R}^d_-\right)\right] \leqslant \left(4\,\|\mathbf{g}\|_2 + 8L_{\mathbf{r}}\sqrt{|\mathcal{I}|}\,|\mathcal{K}|\,|\mathcal{A}|\right)T^{-1/2}.$$

8. Discussion

8.1. **Computational efficiency.** We discuss the computational efficiency of the strategies studied in Sections 6 and 7. The following arguments hold for both.

The first step of the strategy is the computation of $\tilde{z}_t$ which consists of an Euclidean projection onto $\tilde{\mathcal{Z}} := \tilde{\mathcal{C}}^\circ \cap \mathcal{B}_2$, which is efficient. Indeed, $\tilde{\mathcal{C}}^\circ$ being a closed convex cone, the Euclidean projection onto $\tilde{\mathcal{Z}}$ can be immediately deduced from the Euclidean projection onto $\tilde{\mathcal{C}}^\circ$. The latter projection can be efficiently computed since $\tilde{\mathcal{C}}^\circ$ is a polytope (as it can be easily checked). The second step is the computation of $\tilde{x}_t := \tilde{\mathbf{x}}(\tilde{z}_t)$ which, according to the definition of $\tilde{\mathbf{x}}$ in Proposition 5.9, can be computed by solving the following minimax problem:

$$\min_{\tilde{x}\in\Delta(\mathcal{K}\times\mathcal{A})}\max_{f\in\mathcal{F}}\left\langle\tilde{\mathbf{g}}(\tilde{x}, f)\middle|\tilde{z}_t\right\rangle.$$

The sets $\Delta(\mathcal{K} \times \mathcal{A})$ and $\mathcal{F}$ being polytopes, this can be solved efficiently using e.g. linear programming. Then, the computations of estimator $\hat{f}_t$ and auxiliary payoff $\tilde{g}_t$ are easy.

Therefore, the whole strategy can be efficiently computed. Moreover, the per-step complexity is constant.

8.2. **High probability guarantee and almost-sure convergence.** Theorem 6.1 only provides a convergence guarantee in expectation. We quickly describe how the analysis can be adapted to obtain, for the same strategy, a high probability guarantee as well as almost-sure convergence.

We do not modify Lemmas 6.4, 6.5, 6.9 and 6.10 as they do not involve expectations.

The proof of Lemma 6.3 is modified as follows in order to obtain a high probability guarantee on $\mathbf{d}_2(\overline{\tilde{g}}_T,\ \tilde{\mathcal{C}})$. We can easily see that $(\langle \tilde{g}_t | \tilde{z}_t \rangle)_{t \geqslant 1}$ is a bounded sequence of super-martingale differences with respect to filtration $(\mathcal{H}_t)_{t \geqslant 1}$ and that $(\|\tilde{g}_t\|_2^2 - (|\mathcal{I}| / \gamma_t)^2)_{t \geqslant 1}$ is a bounded sequence of super-martingale differences with respect to $(\mathcal{G}_t)_{t \geqslant 1}$. Applying the Hoeffding–Azuma inequality then gives the high probability version of the lemma.

The modification of Lemmas 6.6 and 6.11 is straightforward. We simply apply the high probability version of the involved concentration inequalities instead of the bounds in expectation: Proposition D.3 instead of Corollary D.4 and Proposition D.1 instead of Corollary D.2, respectively.

The high probability versions of Lemmas 6.7 and 6.8 immediately follow from those of Lemma 6.3, and Lemmas 6.6 and 6.7, respectively.

Then, the almost-sure convergence follow from a standard Borel-Cantelli argument.

8.3. **Using other regret minimizing strategies.** As explained in the proof of Lemma 6.3, the strategy defined in Section 5.4 is based on a regret minimizing strategy, specifically, the Mirror Descent strategy associated with the Euclidean regularizer on $\tilde{\mathcal{Z}}$ and time-varying parameters $(\eta_t)_{t \geqslant 1}$. As detailed in the proof, this strategy guarantees the following regret bound:

$$\max_{\tilde{z} \in \tilde{\mathcal{Z}}} \sum_{t=1}^{T} \langle \tilde{g}_t | \tilde{z} \rangle - \sum_{t=1}^{T} \langle \tilde{g}_t | \tilde{z}_t \rangle \leqslant \frac{1}{2\eta_T} + \frac{1}{2} \sum_{t=1}^{T} \eta_{t-1} \|\tilde{g}_t\|_2^2 .$$

We can easily see that any regret minimizing strategy which guarantees a regret bound of the form

$$\max_{\tilde{z} \in \tilde{\mathcal{Z}}} \sum_{t=1}^{T} \langle \tilde{g}_t | \tilde{z} \rangle - \sum_{t=1}^{T} \langle \tilde{g}_t | \tilde{z}_t \rangle \leqslant \frac{A}{\eta_T} + B \sum_{t=1}^{T} \eta_{t-1} \|\tilde{g}_t\|_2^2$$

could be used to construct an alternative approachability strategy for the initial game, with the same rate of convergence. In particular, any Mirror Descent strategy (see e.g. [SS11, Bub11]) associated with some strongly convex regularizer on $\tilde{\mathcal{Z}}$ would be appropriate.

An interesting question is whether the choice of another regularizer would help improve the dependency in $|\mathcal{I}|$, $|\mathcal{K}|$ and $|\mathcal{A}|$ of the bound from Theorem 6.1. Note however that a general regularizer would not *a priori* retain the computational efficiency of the Euclidean regularizer (see Section 8.1).

8.4. **Comparison with [MPS14].** The strategy proposed in [MPS14] is computationally efficient and has a dimension-independent convergence rate of $T^{-1/5}$. We here highlight a few ideas which were already present in [MPS14], and those we have introduced in the present work to achieve an optimal convergence rate of $T^{-1/3}$.

[MPS14] already used the single-valued map $\mathbf{r}$ which is a simpler version of the set-valued map $\mathbf{m}$, which retains the key property characterizing the approachability of the target set (see Proposition 3.2 and property (ii) in Proposition 5.4). Besides, the decomposition of $\mathcal{F}$ and $\Delta(\mathcal{I})$ into polytopes was considered to obtain the piecewise-affinity of $\mathbf{r}$. This fundamental property was then used in the averaging of the flag estimators. The proposed strategy is constructed by dividing time into blocks of the same length: the decision maker plays a constant mixed action on each time block, which is used to average the flag estimators; and the decision maker changes his mixed action from one block to the other in order to achieve the convergence to the target set.

The strategy constructed in Section 5.4 manages to average the estimators and to approach the target *at the same time*, resulting in an improved (and optimal) convergence rate of $T^{-1/3}$. We enumerate some of the main ideas used to achieve this. First, we introduce the linear map $\mathbf{R}$ which allows to easily relate the auxiliary game and the initial game. In particular, it gives a simple comparison between a) the distance of the average payoff to the target set in the initial game and b) the distance of the average auxiliary payoff to the auxiliary target set (Lemma 6.4). Moreover, it combines well with the use of convex cones. Those are used, in particular, to consider the distance $\mathbf{d}_2(\overline{\tilde{g}}_T^{ka},\ \mathcal{F}_c^k)$ instead of $\mathbf{d}_2(\overline{\hat{f}}_T(k,a),\ \mathcal{F}^k)$: this avoids the difficulty of having a different estimator normalization for each couple (k,a), by simply considering working with sums. Finally, the auxiliary target set $\tilde{\mathcal{C}}$ is defined by

$$\tilde{\mathcal{C}} = \prod_{k\in\mathcal{K}} \tilde{\mathcal{C}}^k \quad \text{where} \quad \tilde{\mathcal{C}}^k = \mathbf{R}_k^{-1}(\mathbb{R}_-^d) \cap (\mathcal{F}_c^k)^{\mathcal{A}}.$$

The set $\mathbf{R}_k^{-1}(\mathbb{R}_-^d)$ corresponds to approaching the negative orthant in the initial game, whereas the set $(\mathcal{F}_c^k)^{\mathcal{A}}$ corresponds to making the sure the average estimator $\overline{\hat{f}}_T(k,a)$ is close to $\mathcal{F}^k$. Considering the intersection therefore allows to manage both *at the same time.*

Appendix A. Proofs of technical lemmas

A.1. **Proof of Lemma 5.3.** Let $1 \leqslant n \leqslant d$ and $b \in \mathcal{B}$. Let us first prove that $\mathbf{r}^n(\,\cdot\,,b)$ is piecewise affine. The map $\mathbf{f}$ being affine and defined on $\Delta(\mathcal{J})$, the set $\mathbf{f}^{-1}(b)$ is a polytope. Denote $y_{b,1},\dots,y_{b,q}$ its vertices. Let $x \in \Delta(\mathcal{I})$. By linearity of $\mathbf{g}(x,\,\cdot\,)$, $\mathbf{r}^n(x,b)$ can then be written

$$\mathbf{r}^n(x,b) = \max \mathbf{g}^n(x,\mathbf{f}^{-1}(b)) = \max_{1\leqslant p\leqslant q} \mathbf{g}^n(x,y_{b,p}).$$

$\mathbf{r}^n(\,\cdot\,,b)$ now appears as the maximum of a finite family $(\mathbf{g}^n(\,\cdot\,,y_{b,p}))_{1\leqslant p\leqslant q}$ of linear functions. It is therefore piecewise affine and so is $\mathbf{r}(\,\cdot\,,b)$. Therefore, for each $b \in \mathcal{B}$ there exists a decomposition of $\Delta(\mathcal{I})$ into polytopes on each of which $\mathbf{r}(\,\cdot\,,b)$ is affine. $\mathcal{B}$ being finite, we can consider the decomposition $(\mathcal{X}^\ell)_{\ell\in\mathcal{L}}$ which refines all of them. $\mathbf{r}(\,\cdot\,,b)$ is therefore affine on each polytope $\mathcal{X}^\ell$ for all $b \in \mathcal{B}$. Let us now prove that $\mathbf{r}(\,\cdot\,,f)$ is affine on each polytope $\mathcal{X}^\ell$ for all $f \in \mathcal{F}$.

Let $f \in \mathcal{F}$, $\ell \in \mathcal{L}$, $x_1, x_2 \in \mathcal{X}^\ell$ and $\lambda \in [0,1]$. Using property (iii) from Lemma 5.2, we consider the unique decomposition $f = \sum_{b \in \mathcal{B}} \mu^b \cdot b$ and $k \in \mathcal{K}$ such that $\operatorname{supp} \mu \subset \mathcal{F}^k$. Using the definition of $\mathbf{r}$ and the affinity of $\mathbf{r}(\,\cdot\,, b)$ on $\mathcal{X}^\ell$, we have

$$\begin{aligned}
\mathbf{r}(\lambda x_1 + (1-\lambda)x_2, f) &= \sum_{b \in \mathcal{B}} \mu^b \cdot \mathbf{r}(\lambda x_1 + (1-\lambda)x_2, b) \\
&= \sum_{b \in \mathcal{B}} \mu^b \left(\lambda \mathbf{r}(x_1, b) + (1-\lambda)\mathbf{r}(x_2, b)\right) \\
&= \lambda \sum_{b \in \mathcal{B}} \mu^b \cdot \mathbf{r}(x_1, b) + (1-\lambda) \sum_{b \in \mathcal{B}} \mu^b \cdot \mathbf{r}(x_2, b) \\
&= \lambda \mathbf{r}(x_1, f) + (1-\lambda)\mathbf{r}(x_2, f),
\end{aligned}$$

where the last equality stands because of the uniqueness of the decomposition of f lets us recognize the definitions of $\mathbf{r}(x_1, b)$ and $\mathbf{r}(x_2, b)$ from Equation (2).

A.2. **Proof of Proposition 5.4.** (i) Let $x \in \Delta(\mathcal{I})$ and $y \in \Delta(\mathcal{J})$. Denote $f = \mathbf{f}(y)$. We use property (iii) from Lemma 5.2 to get the unique decomposition $f = \sum_{b \in \mathcal{B}} \mu^b \cdot b$ and $k \in \mathcal{K}$ such that $\operatorname{supp} \mu \subset \mathcal{F}^k$. $\mathbf{f}^{-1}$ being affine on $\mathcal{F}^k$ by property (ii) in Lemma 5.2, we have

$$\mathbf{f}^{-1}\left(\sum_{b \in \operatorname{supp} \mu} \mu^b \cdot b\right) = \sum_{b \in \mathcal{B}} \mu^b \cdot \mathbf{f}^{-1}(b).$$

Therefore we can write

$$\begin{aligned}
\mathbf{g}(x, y) \in \mathbf{g}(x, \mathbf{f}^{-1}(f)) = \mathbf{g}\left(x, \mathbf{f}^{-1}\left(\sum_{b \in \operatorname{supp} \mu} \mu^b \cdot b\right)\right) = \mathbf{g}\left(x, \sum_{b \in \mathcal{B}} \mu^b \cdot \mathbf{f}^{-1}(b)\right) \\
= \sum_{b \in \operatorname{supp} \mu} \mu^b \cdot \mathbf{g}(x, \mathbf{f}^{-1}(b)).
\end{aligned}$$

Then for each $1 \leqslant n \leqslant d$,

$$\begin{aligned}
\mathbf{g}^n(x, y) \leqslant \max \sum_{b \in \operatorname{supp} \mu} \mu^b \cdot \mathbf{g}^n(x, \mathbf{f}^{-1}(b)) = \sum_{b \in \mathcal{B}} \mu^b \cdot \max \mathbf{g}^n(x, \mathbf{f}^{-1}(b)) \\
= \sum_{b \in \mathcal{B}} \mu^b \cdot \mathbf{r}^n(x, b) = \mathbf{r}^n(x, f),
\end{aligned}$$

where for the second equality, we recognized the definition of $\mathbf{r}^n(x, b)$ from Equation (1) on page 7, and the the last equality, the definition of $\mathbf{r}^n(x, f)$ from Equation (2).

(ii) Let $f \in \mathcal{F}$. Thanks to the characterization of approachability from Proposition 3.2, there exists $x \in \Delta(\mathcal{I})$ such that $\mathbf{m}(x, f) \in \mathbb{R}^d_-$. Let $f = \sum_{b \in \mathcal{B}} \mu^b \cdot b$ be the unique decomposition of f given by Lemma 5.2. With the same arguments as above, we have for each $1 \leqslant n \leqslant d$,

$$\begin{aligned}\mathbf{r}^n(x,f) &= \sum_{b\in\mathcal{B}} \mu^b \cdot \mathbf{r}^n(x,b) = \sum_{b\in\mathcal{B}} \mu^b \cdot \max \mathbf{g}^n(x,\mathbf{f}^{-1}(b))\\ &= \max \sum_{b\in\mathcal{B}} \mu^b \cdot \mathbf{g}^n(x,\mathbf{f}^{-1}(b)) = \max \mathbf{g}^n\left(x, \mathbf{f}^{-1}\left(\sum_{b\in\mathcal{B}} \mu^b\cdot b\right)\right)\\ &= \max \mathbf{g}^n(x,\mathbf{f}^{-1}(f)) = \max \mathbf{m}^n(x,f) \leqslant 0.\end{aligned}$$

Therefore, $\mathbf{r}(x,f) \in \mathbb{R}^d_-$.

(iii) Let $x \in \Delta(\mathcal{I})$, $k \in \mathcal{K}$, $f_1, f_2 \in \mathcal{F}^k$ and $\lambda \in [0,1]$. We use property (iii) from Lemma 5.2 to write $f_1 = \sum_{b\in\mathcal{B}} \mu_1^b \cdot b$ and $f_2 = \sum_{b\in\mathcal{B}} \mu_2^b \cdot b$ with $\operatorname{supp}\mu_1 \subset \mathcal{F}^k$ and $\operatorname{supp}\mu_2 \subset \mathcal{F}^k$. The unique decomposition of $\lambda f_1 + (1-\lambda) f_2$ given by Lemma 5.2 is then

$$\lambda f_1 + (1-\lambda) f_2 = \sum_{b\in\mathcal{B}} (\lambda \mu_1^b + (1-\lambda)\mu_2^b)\cdot b.$$

Therefore, using the definition of $\mathbf{r}$ and the affinity of $\mathbf{r}(x,\cdot)$ on $\mathcal{F}^k$,

$$\begin{aligned}\mathbf{r}(x,\lambda f_1 + (1-\lambda) f_2) &= \mathbf{r}\left(x, \sum_{b\in\mathcal{B}} (\lambda \mu_1^b + (1-\lambda)\mu_2^b)\cdot b\right)\\ &= \sum_{b\in\mathcal{B}} (\lambda \mu_1^b + (1-\lambda)\mu_2^b)\cdot \mathbf{r}(x,b)\\ &= \lambda \sum_{b\in\mathcal{B}} \mu_1^b \cdot \mathbf{r}(x,b) + (1-\lambda)\sum_{b\in\mathcal{B}} \mu_2^b\cdot \mathbf{r}(x,b)\\ &= \lambda \mathbf{r}(x,f_1) + (1-\lambda)\cdot \mathbf{r}(x,f_2).\end{aligned}$$

(iv) is already proved in Lemma 5.3.

A.3. **Proof of Lemma 5.5.** Let $k \in \mathcal{K}$ and $x \in \Delta(\mathcal{I})$. Let us consider $\operatorname{span}(\mathcal{F}^k) \subset \mathbb{R}^{\mathcal{S}\times\mathcal{I}}$, the linear span of $\mathcal{F}^k$. There exists a basis $(f_1,\ldots,f_q)$ of $\operatorname{span}(\mathcal{F}^k)$ such that f_p belongs to $\mathcal{F}^k$ for each $1 \leqslant p \leqslant q$. We now define $\mathbf{r}^{[k]}(x,\cdot)$ on $\operatorname{span}(\mathcal{F}^k)$ by setting

$$\mathbf{r}^{[k]}(x,f_p) := \mathbf{r}(x,f_p), \quad \text{for each element } f_p \text{ of the basis,}$$

and extending linearly. $\mathbf{r}^{[k]}(x,\cdot)$ can then be further extended to the whole space $\mathbb{R}^{\mathcal{S}\times\mathcal{I}}$ by setting its value to zero on some complementary subspace of $\operatorname{span}(\mathcal{F}^k)$.

Let us now prove that $\mathbf{r}^{[k]}(x,\cdot)$ coincides with $\mathbf{r}(x,\cdot)$ on $\mathcal{F}^k$. Let $f \in \mathcal{F}^k$. In particular, f belongs to $\operatorname{span}(\mathcal{F}^k)$ and can be uniquely written

$$f = \sum_{p=1}^{q} \lambda_p f_p, \quad \text{where} \quad \lambda_1,\ldots,\lambda_q \in \mathbb{R}.$$

The application $\mathbf{r}^{[k]}(x,\cdot)$ being linear by definition, we have

$$\mathbf{r}^{[k]}(x,f) = \sum_{p=1}^{q} \lambda_p \mathbf{r}(x,f_p).$$

We now aim at proving that the above sum is equal to $\mathbf{r}(x,f)$. This cannot be done by directly applying the affinity of $\mathbf{r}(x,\cdot)$ (property (iii) in Lemma 5.4) because some of the above coefficients λ_p may be negative. To overcome this, we first separate the terms according to the signs of the coefficients λ_p. We denote Λ^+ (resp. Λ^-) the sum of all positive (resp. negative) coefficients λ_p and write

$$\begin{aligned}\mathbf{r}^{[k]}(x,f) &= \sum_{\lambda_p>0} \lambda_p \mathbf{r}(x,f_p) + \sum_{\lambda_p<0} \lambda_p \mathbf{r}(x,f_p) \\ &= \Lambda^+ \sum_{\lambda_p>0} \left(\frac{\lambda_p}{\Lambda^+}\right) \mathbf{r}(x,f_p) + \Lambda^- \sum_{\lambda_p<0} \left(\frac{\lambda_p}{\Lambda^-}\right) \mathbf{r}(x,f_p).\end{aligned}$$

Since each of the above sum is now a convex combination, we can apply the affinity of $\mathbf{r}(x,\cdot)$:

$$\mathbf{r}^{[k]}(x,f) = \Lambda^+ \cdot \mathbf{r}\left(x, \sum_{\lambda_p>0}\left(\frac{\lambda_p}{\Lambda^+}\right) f_p\right) + \Lambda^- \mathbf{r}\left(x, \sum_{\lambda_p<0}\left(\frac{\lambda_p}{\Lambda^-}\right) f_p\right).$$

Let us prove that

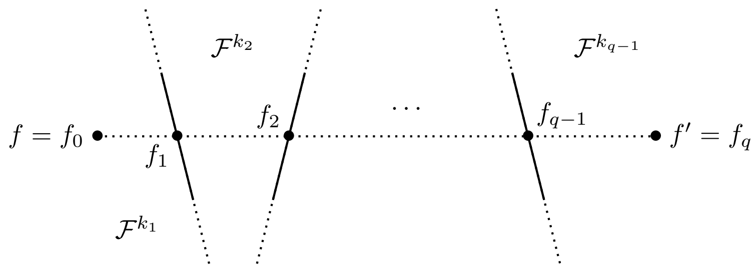
$$\mathbf{r}(x,f) - \Lambda^- \mathbf{r}\left(x, \sum_{\lambda_p<0}\left(\frac{\lambda_p}{\Lambda^-}\right) f_p\right) = \Lambda^+ \cdot \mathbf{r}\left(x, \sum_{\lambda_p>0}\left(\frac{\lambda_p}{\Lambda^+}\right) f_p\right). \tag{6}$$

This will prove that $\mathbf{r}^{[k]}(x,f) = \mathbf{r}(x,f)$.

$$\begin{aligned}\mathbf{r}(x,f) - \Lambda^- \mathbf{r}\left(x, \sum_{\lambda_p<0}\left(\frac{\lambda_p}{\Lambda^-}\right) f_p\right) &= (1-\Lambda^-)\left(\frac{1}{1-\Lambda^-}\mathbf{r}(x,f)\right. \\ &\qquad \left. + \frac{-\Lambda^-}{1-\Lambda^-}\mathbf{r}\left(x, \sum_{\lambda_p<0}\left(\frac{\lambda_p}{\Lambda^-}\right) f_p\right)\right) \\ &= (1-\Lambda^-) \\ &\qquad \times \mathbf{r}\left(x, \frac{1}{1-\Lambda^-} f + \sum_{\lambda_p<0}\left(-\frac{\lambda_p}{1-\Lambda^-}\right) f_p\right) \\ &= (1-\Lambda^-)\cdot \mathbf{r}\left(x, \frac{1}{1-\Lambda^-}\left(f - \sum_{\lambda_p<0} \lambda_p f_p\right)\right) \\ &= (1-\Lambda^-)\cdot \mathbf{r}\left(x, \sum_{\lambda_p>0}\left(\frac{\lambda_p}{1-\Lambda^-}\right) f_p\right).\end{aligned}$$

For relation (6) to be true, it is now enough to prove that $\Lambda^+ + \Lambda^- = 1$. Since $\mathcal{F}^k \subset \mathcal{F} \subset \Delta(\mathcal{S})^{\mathcal{I}}$, for any $f_0 = (f_0^{is})_{\substack{s\in\mathcal{S}\\ i\in\mathcal{I}}} \in \mathcal{F}^k$, we have

$$\sum_{\substack{s\in\mathcal{S}\\ i\in\mathcal{I}}} f_0^{is} = \sum_{i\in\mathcal{I}}\sum_{s\in\mathcal{S}} f_0^{is} = \sum_{i\in\mathcal{I}} 1 = |\mathcal{I}|\,.$$



By applying the above to f and the f_p, we get

$$\begin{aligned}|\mathcal{I}| = \sum_{\substack{s\in\mathcal{S}\\ i\in\mathcal{I}}} f^{is} &= \sum_{\substack{s\in\mathcal{S}\\ i\in\mathcal{I}}} \left(\sum_{\lambda_p>0} \lambda_p f_p^{is} + \sum_{\lambda_p<0} \lambda_p f_p^{is} \right) \\ &= \sum_{\lambda_p>0} \lambda_p \sum_{\substack{s\in\mathcal{S}\\ i\in\mathcal{I}}} f_p^{is} + \sum_{\lambda_p<0} \lambda_p \sum_{\substack{s\in\mathcal{S}\\ i\in\mathcal{I}}} f_p^{is} \\ &= \Lambda^+ |\mathcal{I}| + \Lambda^- |\mathcal{I}| ,\end{aligned}$$

and we indeed get $\Lambda^+ + \Lambda^- = 1$ by dividing by $|\mathcal{I}|$, which concludes the proof.

A.4. **Proof of Lemma 5.6.** Property (i) follows from the definition of $L_{\mathbf{r}}$ and the linearity of the map $\mathbf{r}^{[k]}(a, \cdot)$.

(ii) Let $k \in \mathcal{K}$, $a \in \mathcal{A}$ and $f, f' \in \mathcal{F}$. $(\mathcal{F}^k)_{k\in\mathcal{K}}$ being a finite decomposition of $\mathcal{F}$ into convex polytopes, there exists a finite sequence $(k_1, k_2, \ldots, k_q)$ in $\mathcal{K}$ such that the k_p's are all different and a sequence $(f_0 = f, f_1, f_2, \ldots, f_q = f')$ in the affine segment $[f, f']$ such that $[f_{p-1}, f_p] \subset \mathcal{F}^{k_p}$ for each $1 \leqslant p \leqslant q$. Therefore, using the fact that $\mathbf{r}^{[k']}(a, \cdot)$ and $\mathbf{r}(a, \cdot)$ coincide on $\mathcal{F}^{k'}$ for all $k' \in \mathcal{K}$, we can write

$$\begin{aligned}\|\mathbf{r}(a, f) - \mathbf{r}(a, f')\|_2 &= \left\| \sum_{p=1}^{q} (\mathbf{r}(a, f_{p-1}) - \mathbf{r}(a, f_p)) \right\|_2 \\ &= \left\| \sum_{p=1}^{q} \mathbf{r}^{[k_p]}(a, f_{p-1}) - \mathbf{r}^{[k_p]}(a, f_p) \right\|_2 \\ &\leqslant \sum_{p=1}^{q} \left\| \mathbf{r}^{[k_p]}(a, f_{p-1}) - \mathbf{r}^{[k_p]}(a, f_p) \right\|_2 \\ &\leqslant L_{\mathbf{r}} \sum_{p=1}^{q} \|f_{p-1} - f_p\|_2 \\ &= L_{\mathbf{r}} \|f - f'\|_2 ,\end{aligned}$$

where the last equality holds because the points $f_0, \ldots, f_q$ are aligned and ordered.

A.5. **Proof of Lemma 5.8.** (i) Let $k \in \mathcal{K}$. $\mathbf{R}_k^{-1}(\mathbb{R}_-^d)$ is a closed convex cone as the inverse image via a linear application of the closed convex cone $\mathbb{R}_-^d$ (Proposition B.5). $\mathcal{F}_c^k$ is a closed convex cone by definition, and $(\mathcal{F}_c^k)^{\mathcal{A}}$ is thus a closed convex cone as a Cartesian product of closed convex cones. Therefore,

$\tilde{\mathcal{C}}^k = \mathbf{R}_k^{-1}(\mathbb{R}^d_-) \cap (\mathcal{F}_c^k)^{\mathcal{A}}$ is also a closed convex cone as the intersection of two closed convex cones. Then, $\tilde{\mathcal{C}}$ is also a closed convex cone as a Cartesian product of closed convex cones.

(ii) Let $\tilde{g} = (\tilde{g}^{ka})_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}} \in \tilde{\mathcal{C}}$. By definition of $\tilde{\mathcal{C}}$, for each $k \in \mathcal{K}$, $(\tilde{g}^{ka})_{a\in\mathcal{A}}$ belongs to $\tilde{\mathcal{C}}^k$ and thus to $(\mathcal{F}_c^k)^{\mathcal{A}}$. Therefore, $\tilde{g} \in \prod_{k\in\mathcal{K}}(\mathcal{F}_c^k)^{\mathcal{A}}$. Moreover,

$$\mathbf{R}(\tilde{g}) = \sum_{k\in\mathcal{K}} \mathbf{R}_k\left((\tilde{g}^{ka})_{a\in\mathcal{A}}\right)$$

belongs to $\mathbb{R}^d_-$. Indeed, each term of the above sum belongs to $\mathbb{R}^d_-$ because for all $k \in \mathcal{K}$, $(\tilde{g}^{ka})_{a\in\mathcal{A}} \in \tilde{\mathcal{C}}^k \subset \mathbf{R}_k^{-1}(\mathbb{R}^d_-)$.

A.6. **Proof of Lemma 5.10.** (i) Let $i \in \mathcal{I}$. Note that by definition of the strategy, we have

$$\mathbb{P}\left[i_t = i \,|\, \mathcal{G}_t\right] = (1-\gamma_t)a_t^i + \frac{\gamma_t}{|\mathcal{I}|},$$

and therefore, estimator $\hat{f}_t$ rewrites as

$$\hat{f}_t = \left(\frac{\mathbb{1}_{\{i_t=i\}}}{\mathbb{P}\left[i_t = i \,|\, \mathcal{G}_t\right]}\right)_{i\in\mathcal{I}}.$$

Then, using the conditional expectation with respect to event $\{i_t = i\}$, we have

$$\begin{aligned}
\mathbb{E}\left[\hat{f}_t^i \,\middle|\, \mathcal{G}_t\right] &= \mathbb{E}\left[\frac{\mathbb{1}_{\{i_t=i\}}}{\mathbb{P}\left[i_t = i \,|\, \mathcal{G}_t\right]}\delta_{s_t} \,\middle|\, \mathcal{G}_t\right] \\
&= \mathbb{P}\left[i_t = i \,|\, \mathcal{G}_t\right] \times \mathbb{E}\left[\frac{\delta_{s_t}}{\mathbb{P}\left[i_t = i \,|\, \mathcal{G}_t\right]} \,\middle|\, \mathcal{G}_t, \{i_t = i\}\right] \\
&= \mathbb{E}\left[\delta_{s_t} \,|\, \mathcal{G}_t, \{i_t = i\}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\delta_{s_t} \,|\, y_t, \mathcal{G}_t, \{i_t = i\}\right] \,|\, \mathcal{G}_t, \{i_t = i\}\right] \\
&= \mathbb{E}\left[\mathbf{s}(i, y_t) \,|\, \mathcal{G}_t, \{i_t = i\}\right] \\
&= \mathbb{E}\left[\mathbf{s}(i, y_t) \,|\, \mathcal{G}_t\right] \\
&= \mathbb{E}\left[f_t^i \,|\, \mathcal{G}_t\right],
\end{aligned}$$

hence the result.

(ii) We write

$$\begin{aligned}
\mathbb{E}\left[\left\|\hat{f}_t\right\|_2^2 \,\middle|\, \mathcal{G}_t\right] &= \mathbb{E}\left[\sum_{i\in\mathcal{I}}\left\|\frac{\mathbb{1}_{\{i_t=i\}}}{\mathbb{P}\left[i_t = i \,|\, \mathcal{G}_t\right]}\delta_{s_t}\right\|_2^2 \,\middle|\, \mathcal{G}_t\right] \\
&= \mathbb{P}\left[i_t = i \,|\, \mathcal{G}_t\right] \times \mathbb{E}\left[\sum_{i\in\mathcal{I}}\left\|\frac{\delta_{s_t}}{\mathbb{P}\left[i_t = i \,|\, \mathcal{G}_t\right]}\right\|_2^2 \,\middle|\, \mathcal{G}_t, \{i_t = i\}\right] \\
&= \sum_{i\in\mathcal{I}} \frac{1}{\mathbb{P}\left[i_t = i \,|\, \mathcal{G}_t\right]}\mathbb{E}\left[\left\|\delta_{s_t}\right\|_2^2 \,\middle|\, \mathcal{G}_t, \{i_t = i\}\right] \\
&= \sum_{i\in\mathcal{I}} \frac{1}{\mathbb{P}\left[i_t = i \,|\, \mathcal{G}_t\right]} \\
&\leqslant \frac{|\mathcal{I}|^2}{\gamma_t},
\end{aligned}$$

where the last inequality stands because $\mathbb{P}[i_t = i \,|\, \mathcal{G}_t] \geqslant \gamma_t/|\mathcal{I}|$ by definition of the strategy.

(iii) We have

$$\left\| \hat{f}_t \right\|_2^2 = \sum_{i \in \mathcal{I}} \left\| \frac{\mathbb{1}_{\{i_t = i\}}}{\mathbb{P}[i_t = i \,|\, \mathcal{G}_t]} \delta_{s_t} \right\|_2^2 = \sum_{i \in \mathcal{I}} \mathbb{1}_{\{i_t = i\}} \frac{\|\delta_{s_t}\|_2^2}{\mathbb{P}[i_t = i \,|\, \mathcal{G}_t]^2}$$
$$\leqslant \frac{|\mathcal{I}|^2}{\gamma_t^2} \sum_{i \in \mathcal{I}} \mathbb{1}_{\{i_t = i\}} = \frac{|\mathcal{I}|^2}{\gamma_t^2}.$$

A.7. **Proof of Lemma 7.1.** (i) For $i \in \mathcal{I}$, we write

$$\mathbb{E}\left[\hat{f}_t^i \,\middle|\, \mathcal{G}_t \right] = \mathbb{E}\left[\mathbb{E}\left[\delta_{s_t} \,|\, \mathcal{G}_t, y_t \right] \,|\, \mathcal{G}_t \right] = \mathbb{E}\left[\mathbf{s}(i, y_t) \,|\, \mathcal{G}_t \right] = \mathbb{E}\left[f_t^i \,\middle|\, \mathcal{G}_t \right].$$

(ii) The Euclidean norm of a Dirac being equal to 1,

$$\left\| \hat{f}_t \right\|_2^2 = \|(\delta_{s_t})_{i \in \mathcal{I}}\|_2^2 = \sum_{i \in \mathcal{I}} \|\delta_{s_t}\|_2^2 = |\mathcal{I}|.$$

Appendix B. Closed convex cones

Throughout the section, $\mathcal{W}$ will be a finite-dimensional vector space and $\mathcal{W}^*$ its dual.

Definition B.1. A nonempty subset $\mathcal{C}$ of $\mathcal{W}$ is a closed convex cone if it is closed and if for all $w, w' \in \mathcal{C}$ and $\lambda \in \mathbb{R}_+$, we have $w + w' \in \mathcal{C}$ and $\lambda w \in \mathcal{C}$.

The following proposition gathers a few immediate properties.

Proposition B.2. *(i) A closed convex cone is convex.*
(ii) An intersection of closed convex cones is a closed convex cone.
(iii) A Cartesian product of closed convex cones is a closed convex cone.
(iv) A half-space of the form $\{w \in \mathcal{W} \,|\, \langle z | w \rangle \leqslant 0\}$ (for some $z \in \mathcal{W}^*$) is a closed convex cone.

Definition B.3. Let $\mathcal{A}$ be a subset of $\mathcal{W}$. The polar cone of $\mathcal{A}$ is a subset of the dual space $\mathcal{W}^*$ defined by

$$\mathcal{A}^\circ = \{z \in \mathcal{W}^* \,|\, \forall w \in \mathcal{A},\ \langle w | z \rangle \leqslant 0\}.$$

The following proposition is an immediate consequence of the Bipolar theorem — see e.g. Theorem 3.3.14 in [BL10].

Proposition B.4. *Let $\mathcal{A}$ be a subset of $\mathcal{W}$.*

(i) $\mathcal{A}^{\circ\circ}$ is the smallest closed convex cone containing $\mathcal{A}$.
(ii) If $\mathcal{A}$ is closed and convex, then $\mathcal{A}^{\circ\circ} = \mathbb{R}_+\mathcal{A}$.
(iii) If $\mathcal{A}$ is a closed convex cone, then $\mathcal{A}^{\circ\circ} = \mathcal{A}$.

Proposition B.5. *Let $\varphi : \mathcal{W} \to \tilde{\mathcal{W}}$ be a linear application between two finite-dimensional vector spaces $\mathcal{W}$ and $\tilde{\mathcal{W}}$, φ^* its transpose, $\mathcal{C}$ and $\tilde{\mathcal{C}}$ closed convex cones in $\mathcal{W}$ and $\tilde{\mathcal{W}}$ respectively.*

(i) $\varphi(\mathcal{C})$ is a closed convex cone.
(ii) Then $\varphi^{-1}(\tilde{\mathcal{C}}) = \varphi^*(\tilde{\mathcal{C}}^\circ)^\circ$. In particular, $\varphi^{-1}(\tilde{\mathcal{C}})$ is a closed convex cone.

Proof. Property (i) is obvious. We prove property (ii) as follows. For $w \in \mathcal{W}$,

$$\begin{aligned} w \in \varphi^{-1}(\tilde{\mathcal{C}}) \quad &\iff \quad \varphi(w) \in \tilde{\mathcal{C}} \quad \iff \quad \varphi(w) \in \tilde{\mathcal{C}}^{\circ\circ} \\ &\iff \quad \forall \tilde{z} \in \tilde{\mathcal{C}}^{\circ}, \quad \langle \tilde{z} | \varphi(w) \rangle \leqslant 0 \\ &\iff \quad \forall z \in \tilde{\mathcal{C}}^{\circ}, \quad \langle \varphi^*(\tilde{z}) | w \rangle \leqslant 0 \\ &\iff \quad w \in \varphi^*(\tilde{\mathcal{C}}^{\circ})^{\circ}. \end{aligned}$$

Therefore, $\varphi^{-1}(\tilde{\mathcal{C}})$ is a closed convex cone because it is a polar cone. □

Proposition B.6 (see e.g. [AHR12]). *Let $\mathcal{C}$ be a closed convex cone in $\mathbb{R}^n$. For all point $w \in \mathbb{R}^n$, its Euclidean distance to $\mathcal{C}$ can be written*

$$\mathbf{d}_2(w,\ \mathcal{C}) = \max_{z \in \mathcal{C}^\circ \cap \mathcal{B}_2} \langle w | z \rangle .$$

where $\mathcal{B}_2$ denotes the closed unit Euclidean ball.

Appendix C. Regret minimization

The construction of the strategies from Sections 5 and 7 are based on regret minimizing strategies. The following statement is very similar to well-known regret bounds (see e.g. [RT09, Proposition 11], [SS11, Lemma 2.20] or [BCB12, Theorem 5.4]) but allows for time-varying parameters $(\eta_t)_{t \geqslant 1}$.

Theorem C.1 (Theorem 5.1 in [KM14], Theorem I.3.1 in [Kwo16]). *Let $n \geqslant 1$, $\mathbb{R}^n$ endowed with its canonical Euclidean structure, $\mathcal{Z}$ a nonempty convex compact subset of $\mathbb{R}^d$, $(u_t)_{t \geqslant 1}$ a sequence in $\mathbb{R}^n$, $(\eta_t)_{t \geqslant 1}$ a positive and nonincreasing sequence, and*

$$z_t = \operatorname*{argmax}_{z \in \mathcal{Z}} \left\{ \left\langle \eta_{t-1} \sum_{s=1}^{t-1} u_s \middle| z \right\rangle - \frac{1}{2} \|z\|_2^2 \right\}, \quad t \geqslant 1.$$

Then, for all $T \geqslant 1$,

$$\max_{z \in \mathcal{Z}} \sum_{t=1}^{T} \langle u_t | z \rangle - \sum_{t=1}^{T} \langle u_t | z_t \rangle \leqslant \frac{\|\mathcal{Z}\|_2^2}{2\eta_T} + \frac{1}{2} \sum_{t=1}^{T} \eta_{t-1} \|u_t\|_2^2 .$$

Appendix D. Concentration inequalities

Proposition D.1 (Corollary 3.5 in [KS91]). *Let $(U_t)_{t \geqslant 1}$ be a sequence of martingale differences in $\mathbb{R}^d$, bounded almost-surely by $M > 0$:*

$$\forall t \geqslant 1, \quad \|U_t\|_2 \leqslant M, \quad a.s.$$

Then, for every $\varepsilon > 0$ and $T \geqslant 1$,

$$\mathbb{P}\left[\left\| \frac{1}{T} \sum_{t=1}^{T} U_t \right\|_2 \geqslant \varepsilon \right] \leqslant 2 \exp\left(-\frac{T\varepsilon^2}{4M^2} \right).$$

Corollary D.2. *Under the assumptions of Proposition D.1, we have:*

$$\mathbb{E}\left[\left\| \frac{1}{T} \sum_{t=1}^{T} U_t \right\|_2 \right] \leqslant M \sqrt{\frac{\pi}{T}}.$$

Proof. The result follows from Proposition D.1 by integrating the tail of the distribution:

$$\begin{aligned}\mathbb{E}\left[\left\|\overline{U}_T\right\|_2\right] &= \int_0^{+\infty} \mathbb{P}\left[\left\|\overline{U}_T\right\|_2 \geqslant \varepsilon\right] \,\mathrm{d}\varepsilon \leqslant \int_0^{+\infty} 2e^{-T\varepsilon^2/4M^2}\,\mathrm{d}\varepsilon \\ &= 2\int_0^{+\infty} e^{-\varepsilon^2(T/4M^2)}\,\mathrm{d}\varepsilon = M\sqrt{\frac{\pi}{T}}.\end{aligned}$$

□

The following Bernstein-like inequality is proved in [Pin94] — see also Corollary A.2 in [TY14].

Proposition D.3. *Let $(X_t)_{t\geqslant 1}$ be a martingale difference sequence in a Hilbert space with respect to a filtration $(\mathcal{G}_t)_{t\geqslant 0}$. Suppose that $\|X_t\| \leqslant M$ almost-surely, and*

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[\|X_t\|^2 \,\middle|\, \mathcal{G}_{t-1}\right] \leqslant V.$$

Then,

$$\mathbb{P}\left[\max_{1\leqslant t\leqslant T}\left\|\sum_{t'=1}^{t} X_{t'}\right\| \geqslant \varepsilon\right] \leqslant 2\exp\left(-\frac{\varepsilon^2}{2TV + 2M\varepsilon/3}\right).$$

Corollary D.4. *Under the assumptions of Proposition D.3,*

$$\mathbb{E}\left[\left\|\frac{1}{T}\sum_{t=1}^{T} X_t\right\|\right] \leqslant 4\sqrt{2}\sqrt{\frac{V}{T}} + \frac{4M}{3T}.$$

Proof. Let $A \geqslant 0$ to be chosen later.

$$\begin{aligned}\mathbb{E}\left[\left\|\overline{X}_T\right\|\right] &= \int_0^{+\infty} \mathbb{P}\left[\left\|\overline{X}_T\right\| \geqslant \varepsilon\right]\,\mathrm{d}\varepsilon \\ &\leqslant 2\int_0^{+\infty} \exp\left(-\frac{\varepsilon^2 T^2}{2VT + 2M\varepsilon T/3}\right)\mathrm{d}\varepsilon \\ &= 2\int_0^{+\infty} \exp\left(-\frac{\varepsilon^2 T}{2V + 2M\varepsilon/3}\right)\mathrm{d}\varepsilon \\ &\leqslant 2\left(A + \int_A^{+\infty} \exp\left(-\frac{\varepsilon^2 T}{2\varepsilon(V/A + M/3)}\right)\mathrm{d}\varepsilon\right) \\ &= 2\left(A + \int_A^{+\infty} \exp\left(-\frac{\varepsilon T}{2(V/A + M/3)}\right)\mathrm{d}\varepsilon\right) \\ &= 2\left(A + \left[-\frac{2}{T}\left(\frac{V}{A} + \frac{M}{3}\right)\exp\left(-\frac{\varepsilon T}{2(V/A + M/3)}\right)\right]_A^{+\infty}\right) \\ &\leqslant 2A + \frac{4}{T}\left(\frac{V}{A} + \frac{M}{3}\right).\end{aligned}$$

Choosing $A = \sqrt{2V/T}$ gives:

$$\mathbb{E}\left[\left\|\overline{X}_T\right\|\right] \leqslant 4\sqrt{2}\sqrt{\frac{V}{T}} + \frac{4M}{3T}.$$

□

Acknowledgements. The authors are grateful to Igor Kortchemski, Rida Laraki, Sylvain Sorin and particularly Gilles Stoltz for careful proofreading and numerous remarks which have helped improve this work. Both authors are supported by a public grant as part of the *Investissement d'avenir* project, `ANR-11-LABX-0056-LMH`, LabEx LMH. The second author is also partially funded by the ANR grant `ANR-13-JS01-0004-01` and he benefited from the support of the *FMJH Program Gaspard Monge in optimization and operations research* (supported in part by EDF) and from the support of the CNRS through the PEPS projects.

References

[AHR12] Jacob Abernethy, Elad Hazan, and Alexander Rakhlin. Interior-point methods for full-information and bandit online learning. *IEEE Transactions on Information Theory*, 58(7):4164–4175, 2012.

[BCB12] Sébastien Bubeck and Nicolo Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Machine Learning*, 5(1):1–122, 2012.

[BFP+14] Gábor Bartók, Dean P Foster, Dávid Pál, Alexander Rakhlin, and Csaba Szepesvári. Partial monitoring – classification, regret bounds, and algorithms. *Mathematics of Operations Research*, 39(4):967–997, 2014.

[BL10] Jonathan M. Borwein and Adrian S. Lewis. *Convex analysis and nonlinear optimization: theory and examples*. Springer, 2010.

[Bla56] David Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6(1):1–8, 1956.

[BPS10] Gábor Bartók, Dávid Pál, and Csaba Szepesvári. Toward a classification of finite partial-monitoring games. In *Proceedings of the 21st International Conference on Algorithmic Learning Theory (ALT)*, pages 224–238. Springer, 2010.

[Bub11] Sébastien Bubeck. *Introduction to Online Optimization: Lecture Notes*. Princeton University, 2011.

[CBL06] Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge University Press, 2006.

[CBLS06] Nicolo Cesa-Bianchi, Gábor Lugosi, and Gilles Stoltz. Regret minimization under partial monitoring. *Mathematics of Operations Research*, 31(3):562–580, 2006.

[KM14] Joon Kwon and Panayotis Mertikopoulos. A continuous-time approach to online optimization. *arXiv:1401.6956*, 2014.

[KS91] Olav Kallenberg and Rafal Sztencel. Some dimension-free features of vector-valued martingales. *Probability Theory and Related Fields*, 88(2):215–247, 1991.

[Kwo16] Joon Kwon. *Mirror descent strategies for regret minimization and approachability*. PhD thesis, Université Pierre-et-Marie-Curie, 2016.

[LMS08] Gábor Lugosi, Shie Mannor, and Gilles Stoltz. Strategies for prediction under imperfect monitoring. *Mathematics of Operations Research*, 33(3):513–528, 2008.

[LS07] Ehud Lehrer and Eilon Solan. Learning to play partially-specified equilibrium. *Levine's Working Paper Archive*, 2007.

[LS16] Ehud Lehrer and Eilon Solan. A general internal regret-free strategy. *Dynamic Games and Applications*, 6(1):112–138, 2016.

[MPS13] Shie Mannor, Vianney Perchet, and Gilles Stoltz. A primal condition for approachability with partial monitoring. *Journal of Dynamics and Games*, 1(3):447–469, 2013.

[MPS14] Shie Mannor, Vianney Perchet, and Gilles Stoltz. Set-valued approachability and online learning with partial monitoring. *The Journal of Machine Learning Research*, 15(1):3247–3295, 2014.

[MS03] Shie Mannor and Nahum Shimkin. On-line learning with imperfect monitoring. In *Learning Theory and Kernel Machines*, pages 552–566. Springer, 2003.

[MSZ15] Jean-François Mertens, Sylvain Sorin, and Shmuel Zamir. *Repeated games*, volume 55. Cambridge University Press, 2015.

[Per11a] Vianney Perchet. Approachability of convex sets in games with partial monitoring. *Journal of Optimization Theory and Applications*, 149(3):665–677, 2011.

[Per11b] Vianney Perchet. Internal regret with partial monitoring: Calibration-based optimal algorithms. *The Journal of Machine Learning Research*, 12:1893–1921, 2011.

[Per14] Vianney Perchet. Approachability, regret and calibration: Implications and equivalences. *Journal of Dynamics and Games*, 1(2):181–254, 2014.

[Pin94] Iosif Pinelis. Optimum bounds for the distributions of martingales in Banach spaces. *The Annals of Probability*, 22(4):1679–1706, 1994.

[PS01] Antonio Piccolboni and Christian Schindelhauer. Discrete prediction games with arbitrary feedback and loss. In *Proceedings of the 14th Annual Conference on Computational Learning Theory (COLT)*, pages 208–223. Springer, 2001.

[RST11] Alexander Rakhlin, Karthik Sridharan, and Ambuj Tewari. Online learning: Beyond regret. In *JMLR: Workshop and Conference Proceedings (COLT)*, volume 19, pages 559–594, 2011.

[RT09] Alexander Rakhlin and Ambuj Tewari. Lecture notes on online learning. 2009.

[Rus99] Aldo Rustichini. Minimizing regret: The general case. *Games and Economic Behavior*, 29(1):224–243, 1999.

[RZ96] Jörg Rambau and Günter M. Ziegler. Projections of polytopes and the generalized Baues conjecture. *Discrete & Computational Geometry*, 16(3):215–237, 1996.

[SS11] Shai Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4(2):107–194, 2011.

[TY14] Pierre Tarres and Yuan Yao. Online learning as stochastic approximation of regularization paths: optimality and almost-sure convergence. *IEEE Transactions on Information Theory*, 60(9):5716–5735, 2014.