Online Learning and Blackwell Approachability with Partial Monitoring: Optimal Convergence Rates

Joon Kwon
Centre de Mathématiques Appliquées
École Polytechnique
Université Paris–Saclay

Vianney Perchet
Centre de Mathématiques et de Leurs Applications
École Normale Supérieure Paris–Saclay
& Criteo Research, Paris

Abstract

Blackwell approachability is an online learning setup generalizing the classical problem of regret minimization by allowing for instance multi-criteria optimization, global (online) optimization of a convex loss, or online linear optimization under some cumulative constraint. We consider *partial monitoring* where the decision maker does not necessarily observe the outcomes of his decision (unlike the traditional regret/bandit literature). Instead, he receives a random signal correlated to the decision–outcome pair, or only to the outcome.

We construct, for the first time, approachability algorithms with convergence rate of order $O(T^{-1/2})$ when the signal is independent of the decision and of order $O(T^{-1/3})$ in the case of general signals. Those rates are optimal in the sense that they cannot be improved without further assumption on the structure of the objectives and/or the signals.

1 Introduction

Online learning has become a standard topic, especially through regret minimisation [Cesa-Bianchi and Lugosi, 2006, Shalev-Shwartz, 2011, Bubeck and Cesa-Bianchi, 2012]: 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 [Rakhlin et al., 2011], such as Blackwell approachability [Blackwell, 1956], which is the focus of the present work: the decision maker recieves vector-valued payoffs (instead of scalar payoffs/losses) 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. [Kwon, 2016]). Further applications are mentionned in Section 2.2.

The full information setting, which corresponds to the case where the decision maker does observe his vector-valued payoffs, is well understood and has a worst-case convergence rate of order $O(T^{-1/2})$ (see e.g. [Perchet, 2014]) We here study the *partial monitoring* setting, where the decision maker does not necessaritly observe his (vector-valued) payoffs. Instead, he recieves a random signal, whose law may depend on his decision and on the state of Nature.

The partial monitoring setting was first studied in the special case of regret minimization. Unlike the full information setting, the decision maker may not be able to minimize the regret, depending on the signaling structure. This has given rise to two main directions of research.

The first one, initiated by Piccolboni and Schindelhauer [2001] identifies the signaling structures which allow the average regret to be minimized and aims at constructing algorithms in those cases: Piccolboni and Schindelhauer [2001] constructed an algorithm guaranteeing a convergence rate of order $O(T^{-1/4})$ and Cesa-Bianchi et al. [2006] proposed an improved algorithm with a $O(T^{-1/3})$ guarantee as well as a general lower bound of the same order. Later, Bartók et al. [2010, 2014] gave a classification of signaling structures according to convergence rates: they established that the optimal convergence rate is either $O(T^{-1/2})$, $O(T^{-1/3})$ or $O(1)$—this last rate corresponds to the case where the average regret cannot be minimized.

The second line of research was proposed by Rustichini [1999] focus on the case where average regret cannot

Proceedings of the 20^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2017, Fort Lauderdale, Florida, USA. JMLR: W&CP volume 54. Copyright 2017 by the author(s).

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. Rustichini [1999] however did not provide an explicit algorithm nor convergence rates. Mannor and Shimkin [2003] constructed approachability-based algorithms in the special case where the law of the signal only depends on Nature's decision (the so-called *outcome-dependent* case). Lugosi et al. [2008] proposed algorithms with convergence rates of order $O(T^{-1/4}\sqrt{\log T})$ in the case of outcome-dependent signals and of order $O(T^{-1/5}\sqrt{\log T})$ in the case of general signals. The optimal rate of order $O(T^{-1/3})$ in the case of general signals was achieved by Perchet [2011b] using calibration-based algorithms.

More recently, the problem of approachability with partial monitoring has been considered by Perchet [2011a]. The regret minimization problem from Rustichini [1999] and the internal regret from Lehrer and Solan [2007], Perchet [2011b] turn out to be special cases of this very general framework. However, the convergence rate of the algorithm provided in Perchet [2011a] had the drawback of deteriorating quickly with the dimension of the payoff space, as it scales as $O(T^{-1/(I+3)})$ where I is the number of actions of the decision maker. A dimension-free rate of order $O(T^{-1/5})$ was given in Mannor et al. [2014b]—see also Mannor et al. [2013]. However, the optimal rate of convergence was conjectured to be of order $O(T^{-1/3})$, as for regret minimization.

Main contributions and Outline

We construct, for the first time, approachability algorithms for polytope target sets with convergence rates of order $O(T^{-1/3})$ in the case of general signals and of order $O(T^{-1/2})$ 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 $O(T^{-1/3})$ was given in Cesa-Bianchi et al. [2006], and the $O(T^{-1/2})$ rate is already optimal in the full information setting (since they both hold in the case where standard average regret can be minimized, these lower bounds hold for both lines of research on partial monitoring). It therefore establishes the optimal convergence rates for those two cases. Moreover, the proposed algorithms are computationally efficient.

In Section 2, we present the model of repeated decision process with vector payoffs and with partial monitoring; we recall some well-known results on Blackwell approachability (with full and partial monitoring) that will be useful. In Section 3, we first introduce an auxiliary full information game which we then use to construct the algorithm for the initial game. The efficiency of the algorithm is discussed. In Section 4.2 we state and prove Theorem 4.1 which is our main result. It establishes an $O(T^{-1/3})$ rate of convergence for the algorithm. In Section 4.3, we deal with the special case of outcome-dependent signals for which we propose a modified algorithm which is proved in Theorem 4.3 to guarantee an $O(T^{-1/2})$ rate of convergence.

2 Framework

We consider a repeated decision process between the *decision maker* and *Nature*. The finite set of decisions of the decision maker (resp. Nature) is denoted by $\mathcal{I}$ (resp. $\mathcal{J}$). It is usually necessary in adversarial settings to consider random algorithms; we denote by $\Delta(\mathcal{I})$ the simplex of probability distributions over $\mathcal{I}$, i.e.,

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

and $\Delta(\mathcal{J})$ is defined similarly.

At each stage $t \geqslant 1$, the decision maker and Nature simultaneously choose decisions $i_t \in \mathcal{I}$ and $j_t \in \mathcal{J}$, possibly at random according to the probability distributions $x_t \in \Delta(\mathcal{I})$ and $y_t \in \Delta(\mathcal{J})$. These choices generate a vectorial payoff $g_t := \mathbf{g}(i_t, j_t) \in \mathbb{R}^d$ to the decision maker. The mapping $\mathbf{g}$ is extended bi-linearly to $\Delta(\mathcal{I}) \times \Delta(\mathcal{J})$ by:

$$\mathbf{g}(x,y) := \mathbb{E}_{\substack{i\sim x \\ j\sim y}} \left[\mathbf{g}(i,j)\right] = \sum_{\substack{i\in\mathcal{I} \\ j\in\mathcal{J}}} x^i y^j \mathbf{g}(i,j)$$

where $x = (x^i)_{i\in\mathcal{I}} \in \Delta(I)$ and $y = (y^j)_{j\in\mathcal{J}} \in \Delta(\mathcal{J})$, and we define $\|\mathbf{g}\|_2 := \max_{\substack{i\in\mathcal{I} \\ j\in\mathcal{J}}} \|\mathbf{g}(i,j)\|_2$.

It remains to describe the overall objectives, introducing the concept of Blackwell approachability.

2.1 Blackwell Approachability

Given a fixed, closed and convex target set $\mathcal{C} \subset \mathbb{R}^d$, the overarching aim of the decision maker is to guarantee that the average vector payoff $\bar{g}_T := \frac{1}{T}\sum_{t=1}^T g_t$ converges to $\mathcal{C}$. This set is said to be *approachable* by the decision maker if he has a algorithm such that

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

uniformly with respect to the choices of Nature, 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 algorithm of the decision maker.

Before introducing the partial monitoring setup, we recall some useful results with full monitoring, i.e., when the decision maker gets to observe j_t (or at least g_t) at the end of stage t.

Characterization of approachability with full monitoring A closed convex set $\mathcal{C} \subset \mathbb{R}^d$ is approachable by the decision maker [Blackwell, 1956] if and only if one of the following properties hold.

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

(ii) $\forall y \in \Delta(\mathcal{J}),\ \exists\, x \in \Delta(\mathcal{I}), \quad \mathbf{g}(x,y) \in \mathcal{C}$.

If $\mathcal{C}$ is a closed convex cone, the above is equivalent to

(iii) $\forall z \in \mathcal{C}^\circ,\ \exists\, x \in \Delta(\mathcal{I})\ \forall y \in \Delta(\mathcal{J}), \langle \mathbf{g}(x,y) | z \rangle \leqslant 0$,

where $\mathcal{C}^\circ$ is the polar cone of $\mathcal{C}$—see Appendix C.

We emphasize that $\Delta(\mathcal{I})$ and $\Delta(\mathcal{J})$ being simplices is irrelevant. The same result holds with any bilinear reward function $\mathbf{g} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^d$ defined over two convex compact sets $\mathcal{X}$ and $\mathcal{Y}$ of any Euclidean space. The special case where $\mathcal{C}$ is a closed convex cone will be of particular importance in the subsequent sections, we hence gather a few well-known facts on the topic in Appendix C.

2.2 Possible applications of Blackwell approachability

Regret minimization can be easily recast as an approachability problem—see Blackwell [1954], Abernethy et al. [2011], Perchet [2014]. We here mention some other possible applications of Blackwell approachability.

Regret minimization with adversarial constraints & global cost In the setting of regret minimization with long term constraints (see Jenatton et al. [2016]), the sequence of decisions i_t must not only minimize the average loss but also satisfy, asymptotically, some external constraints as in linear programming. Typically, the benchmark of an algorithm is the following

$$\min_{x \in \Delta(\mathcal{I})} \Big\{ \frac{1}{T} \sum_{t=1}^{T} \ell(x, j_t);\ \bar{A}_T x \leqslant \bar{b}_T \Big\}$$

where $\bar{A}_T = \frac{1}{T}\sum_{t=1}^T A_t \in \mathbb{R}^{k \times \mathcal{I}}$ and $\bar{b}_T = \frac{1}{T}\sum_{t=1}^T b_t \in \mathbb{R}^k$, and where the inequality is to be understood component-wise. The sequences of matrices A_t and vectors b_t that define the constraint set are chosen sequentially by Nature.

The approachable equivalent target is naturally defined as

$$\Big\{ (y, z, A, b) \in \mathbb{R}^{\mathcal{J}+1+k \times \mathcal{I}+k};\ z \leqslant \min_{x \in \Delta(\mathcal{I}), Ax \leqslant b} \ell(x, y) \Big\}.$$

This set is not necessarily convex and a decision maker might not be able to compete with the best decision in hindsight that satisfies the average constraints; there are ways to circumvent that issue (by considering a convex super-set that contains it or other techniques Bernstein et al. [2013], Mannor et al. [2014a]).

This problem is actually strongly related to the global cost minimization Even-Dar et al. [2009]. In that setting, the global regret is defined as

$$\mathcal{L}\Big(\frac{1}{T}\sum_{t=1}^{T} g(x_t, j_t)\Big) - \min_{x \in \Delta(\mathcal{I})} \mathcal{L}\Big(\frac{1}{T}\sum_{t=1}^{T} g(x, j_t)\Big),$$

where $g : \Delta(\mathcal{I}) \times \Delta(\mathcal{J}) \to \mathbb{R}^d$ is some vectorial reward and $\mathcal{L} : \mathbb{R}^d \to \mathbb{R}$ is some non-linear loss mapping.

Varying stage duration Another application of Blackwell approachability is when stages have different *duration* or *weights* Mannor and Shimkin [2008]. At stage $t \geqslant 1$, the decisions $i_t \in \mathcal{I}$ and $j_t \in \mathcal{J}$ generate a reward vector $g(i_t, j_t) \in \mathbb{R}^d$, but some of them might be more important than others (or last longer); this is represented by a scalar $\omega(i_t, j_t) \in \mathbb{R}_+$. Given a target set $\mathcal{C} \subset \mathbb{R}^d$, the goal of the decision maker is that the weighted average reward vector converges to $\mathcal{C}$:

$$\frac{\sum_{t=1}^T \omega(i_t, j_t) g(i_t, j_t)}{\sum_{t=1}^T \omega(i_t, j_t)} \xrightarrow[T \to +\infty]{} \mathcal{C}$$

This can be rewritten as a traditional approachability problem where the reward vector is $(g(i_t, j_t)\omega(i_t, j_t), \omega(i_t, j_t))$ and the target set is

$$\mathcal{C}^\omega = \Big\{ (z, w) \in \mathbb{R}^{d+1};\ \frac{1}{w} z \in \mathcal{C} \Big\},$$

which is a convex cone as soon as $\mathcal{C}$ is convex.

Other applications Approachability can also be seen as a powerful generic tool to solve other online learning problems such as constructing calibrated predictions [Foster, 1999, Foster et al., 2011, Perchet, 2014] as well as constructing optimal algorithms in repeated zero-sum games with imperfect information [Aumann and Maschler, 1995] or as a building block in constructing Nash equilibria in repeated multi-players games [Tomala, 2013].

2.3 Partial Monitoring

With partial monitoring (as well as in the classical multi-armed bandit scenario), the decision maker does not necessarily observe j_t nor g_t at the end of stage t but he instead receives some *signal* $s_t \in \mathcal{S}$, where $\mathcal{S}$ is a finite set. More precisely, there exists a mapping $\mathbf{s} : \mathcal{I} \times \mathcal{J} \to \Delta(\mathcal{S})$, which is known to the decision maker, that indicates the (conditional) law of signal s_t as a function of decisions i_t and j_t, i.e., s_t is drawn according to probability distribution $\mathbf{s}(i_t, j_t) \in \Delta(\mathcal{S})$. $\mathbf{s}$ is also bilinearly extended to $\Delta(\mathcal{I}) \times \Delta(\mathcal{J})$.

The special case where the law of the signal $\mathbf{s}(i, j)$ does not depend on i is called *outcome-dependent*, and will be treated in its dedicated Section 4.3.

A crucial concept with partial monitoring is *flags*. The flag function $\mathbf{f} : \Delta(\mathcal{J}) \to \Delta(\mathcal{S})^{\mathcal{I}}$ is defined by

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

and $f_t := \mathbf{f}(y_t)$ denotes the flag associated with y_t. Although the decision maker does not directly observe it, he can, as will be shown, estimate it. As a matter of fact, it is the maximal information available: two random choices $y, y' \in \Delta(\mathcal{J})$ that generate the same flag are absolutely indistinguishable by the decision maker.

We denote by $\mathcal{F} = \mathbf{f}(\Delta(\mathcal{J}))$ the set of all possible flags, which is a polytopial subset of $\mathbb{R}^{\mathcal{S} \times \mathcal{I}}$. Moreover, for any $x \in \Delta(\mathcal{I})$ 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 random decision 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 1) and in the construction of the algorithms.

Characterization of approachability with partial monitoring Perchet [2011a] A closed convex set $\mathcal{C} \subset \mathbb{R}^d$ is approachable by the decision maker if and only if

$$\forall f \in \mathcal{F}, \ \exists x \in \Delta(\mathcal{I}), \quad \mathbf{m}(x, f) \subset \mathcal{C}. \tag{1}$$

Notice that with full monitoring, and with bandit monitoring in the case of regret minimization, the flag function is fully informative in the sense that $\mathcal{F} = \Delta(\mathcal{J})$ and $f(y) = y$. As a consequence, in those cases, $(x, f(y)) = \{\mathbf{g}(x, y)\}$ and thus the above characterization and the original one of Blackwell coincide.

Examples of partial monitoring There are many examples where the feedback of the decision maker (or available to any algorithm) is neither full nor bandit. Consider a repeated task, where the decision maker has access to several basic algorithms (or experts) but where it is costly to observe the outcome of any of them, as in the apple tasting problem, or label efficient prediction [Cesa-Bianchi et al., 2005]. Practical examples would be automatic subtitling of a video or classification of a music, etc. The decision maker cannot know whether an expert is correct or not unless he asks a human to manually do the classification, which obviously is costly and cannot be done to often.

Other examples of partial monitoring involve routing in congested networks. A decision maker aims at sending messages through the less congested path, however the congestion is not observed, only the number of lost packets (yet the probability of losing packet increase with congestion). The same phenomenon actually occurs in several instances of learning scenarii with censored or perturbed data.

3 From Partial to Full Monitoring

We focus on the case where the target set is the negative orthant $\mathcal{C} := \mathbb{R}^d_-$ 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. [Mannor et al., 2014b, Section 5.4.2].

Detailed proofs of our claims below can be found in Appendix A.

The main difficulty of partial monitoring is that, even if the flags f_t were observed by the decision maker (which is not the case), the only way to make sure that $\bar{g}_T$ converges to $\mathcal{C}$ is to ensure that the average set $\frac{1}{T}\sum_{t=1}^{T} \mathbf{m}(i_t, f_t)$ is asymptotically contained in $\mathcal{C}$. This would require to handle and control averages of set-valued mappings, which may be tedious[1]. Instead, we introduce a single-valued mapping $\mathbf{R}$ that *represents* $\mathbf{m}$, in the sense that

$$\frac{1}{T}\sum_{t=1}^{T} \mathbf{m}_t \subset \mathcal{C} \quad \Longleftarrow \quad \frac{1}{T}\sum_{t=1}^{T} \mathbf{R}_t \subset \mathcal{C},$$

or at least such that the left term is implied by the right one. Moreover, if we manage to enforce that $\mathbf{R}$ is linear, then we could apply the same techniques as in the classical full monitoring case.

[1] The naïve idea of representing a set of possible payoffs by the compatible payoffs which is the farther away from the target set $\mathcal{C}$ could lead to linear regret, see e.g. Mannor et al. [2014b]

3.1 Bi-piecewise affinity

First, notice that the flag mapping $\mathbf{f}$ is *affine*[2] on $\mathcal{F}$.

This yields (see [Rambau and Ziegler, 1996, Proposition 2.4]) the existence of a simplicial decomposition of $\mathcal{F}$ such that $\mathbf{f}^{-1}$ is affine on each of those simplices.

More precisely, there exists a finite family $(\mathcal{F}^k)_{k\in\mathcal{K}}$ of simplices such that $\mathcal{F} = \bigcup_{k\in\mathcal{K}} \mathcal{F}^k$ and $\mathbf{f}^{-1}$ is affine on each $\mathcal{F}^k$. Moreover, if we denote by $\mathcal{B}^k$ the set of vertices of $\mathcal{F}^k$ and $\mathcal{B} = \bigcup_{k\in\mathcal{K}} \mathcal{B}^k$ then for all $f \in \mathcal{F}^k$, there exists a unique $\mu = (\mu^b)_{b\in\mathcal{B}} \in \Delta(\mathcal{B})$ such that

$$f = \sum_{b\in\mathcal{B}} \mu^b \cdot b \quad \text{and, moreover, } \operatorname{supp}\mu \subset \mathcal{B}^k, \tag{2}$$

where $\operatorname{supp}\mu$ is the support of μ.

From now on, we assume being given such a decomposition.

Example in small dimension Before, describing the algorithms and results in the general case, we first give some intuitions on the above statements in smaller dimension. Assume that Nature has 3 actions, so that $\Delta(\mathcal{J})$ is a triangle, for instance with vertices $(0;0)$, $(0.5;1)$ and $(1;0)$, see e.g., Figure 1, page 219, Rambau and Ziegler [1996]. More over, we are going to assume that mixed actions on a vertical segment are undistinguishable to the decision maker.

The undistinguishable actions sets are therefore not linear over the whole horizontal segment [0,1] but are linear on [0,0.5] and [0.5,1]. The name of *piecewise affine mapping* comes from this property. Rambau and Ziegler [1996] would call "chamber" the two segments [0,0.5] and [0.5,1], of respectives vertices set 0;0.5 and 0.5;1, and the linearity of inverse mappings on a chamber is a consequence of the very last equation page 221 (stated as a Minkowski sum). Proposition 2.4 of Rambau and Ziegler generalizes this toy example to any polytopes and any linear mapping (i.e., projections).

Notice that, because of this lack of linearity, observing 0 half of the stages and 1 on the other half of the stages is intrinsically different than observing $1/2$ at all stages. However, since the chambers have only 3 vertices $\{0, 0.5, 1\}$, it is possible to lift the segment $[0, 1]$ into a 2 dimension simplex to recover linearity. This is precisely the objectives of the next section in higher dimension.

Back to the general case We now explain how we can reduce the problem of approachability with partial monitoring, without linearity, to another auxiliary approachability problem, in a lifted space of higher dimension, but with linearity.

We first construct a map $\mathbf{r} = (\mathbf{r}^n)_{1\leqslant n\leqslant d}$ componentwise, and first on $\Delta(\mathcal{I}) \times \mathcal{B}$ before extending it to $\Delta(\mathcal{I}) \times \mathcal{F}$. Denote by $(\mathbf{g}^n)_{1\leqslant n\leqslant d}$ the components of payoff function $\mathbf{g}$. For $x \in \Delta(\mathcal{I})$ and $b \in \mathcal{B}$, we set $\mathbf{r}^n(x, b)$ as the largest element in the set $\mathbf{g}^n(x, \mathbf{f}^{-1}(b))$:

$$\mathbf{r}^n(x, b) := \max \mathbf{g}^n(x, \mathbf{f}^{-1}(b)). \tag{3}$$

This construction ensures that, for any $y \in \mathbf{f}^{-1}(b)$, $\mathbf{g}(x, y) \in \mathbb{R}^d_-$ as soon as $\mathbf{r}(x, b)$ is also in $\mathbb{R}^d_-$.

We then extend $\mathbf{r}$ to $\Delta(\mathcal{I}) \times \mathcal{F}$ as follows. Recall that a given flag $f \in \mathcal{F}$ can be uniquely written as

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

and that the support of μ is contained in one of the polytopes $\mathcal{F}^k$. We then use the above coefficients $(\mu^b)_{b\in\mathcal{B}}$ to define

$$\mathbf{r}^n(x, f) := \sum_{b\in\mathcal{B}} \mu^b \cdot \mathbf{r}^n(x, b). \tag{4}$$

This construction gives the existence of a finite family of polytopes $(\mathcal{X}^\ell)_{\ell\in\mathcal{L}}$ covering $\Delta(\mathcal{I})$ satisfying the following properties.

Proposition 3.1 *(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 $\mathcal{F}^k$;

(iv) For all $f \in \mathcal{F}$, $\mathbf{r}(\cdot, f)$ is affine on each $\mathcal{X}^\ell$.

We denote by $\mathcal{A}$ the set of all vertices of polytopes $\mathcal{X}^\ell$.

3.2 From bi-piecewise affinity to linearity

In this section, we construct the *linear map* $\mathbf{R} : (\mathbb{R}^{\mathcal{S}\times\mathcal{I}})^{\mathcal{K}\times\mathcal{A}} \to \mathbb{R}^d$.

First, we claim (see Proposition A.2) that the construction of $\mathbf{R}$ would be easy if we only consider the restriction of the affine mappings $\mathbf{r}(x, \cdot)$ to $\mathcal{F}^k$. Indeed, since $\mathcal{F}^k$ is included in $\Delta(\mathcal{S})^{\mathcal{I}}$, it would be sufficient to extend those mapping linearly to $\mathbb{R}^{\mathcal{S}\times\mathcal{I}}$. The intuition might be clearer in the trivial case where $|\mathcal{I}| = d = 1$. In that case, consider an affine mapping $\phi : \Delta(\mathcal{S}) \to \mathbb{R}$. Then the mapping $\phi^* : \mathbb{R}^{\mathcal{S}}_+ \to \mathbb{R}$ defined by $\phi^*(x) = \phi(\frac{x}{\|x\|_1})\|x\|_1$ is linear on $\mathbb{R}^{\mathcal{S}}_+$ and can easily extended into a linear mapping on the whole space $\mathbb{R}^{\mathcal{S}}$.

In the general case, 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

[2] We recall that a set-valued function $\mathbf{a} : \mathcal{U} \rightrightarrows \mathcal{V}$ is affine if for all $u, u' \in \mathcal{U}$ and $\lambda \in [0, 1]$, $\mathbf{a}(\lambda u + (1 - \lambda)u') = \lambda\mathbf{a}(u) + (1 - \lambda)\mathbf{a}(u')$.

(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 $f \in \mathcal{F}^k$, $\mathbf{r}^{[k]}(x, f) = \mathbf{r}(x, f)$.

Given $\mathbf{r}^{[k]}$, we now define the linear map $\mathbf{R}_k : (\mathbb{R}^{\mathcal{S}\times\mathcal{I}})^{\mathcal{A}} \to \mathbb{R}^d$ as follows

$$\mathbf{R}_k\left((\tilde{g}^{ka})_{a\in\mathcal{A}}\right) := \sum_{a\in\mathcal{A}} \mathbf{r}^{[k]}(a, \tilde{g}^{ka}),$$

for all $(\tilde{g}^{ka})_{a\in\mathcal{A}} \in (\mathbb{R}^{\mathcal{S}\times\mathcal{I}})^{\mathcal{A}}$. Finally, we define the linear map $\mathbf{R} : (\mathbb{R}^{\mathcal{S}\times\mathcal{I}})^{\mathcal{K}\times\mathcal{A}} \to \mathbb{R}^d$ by setting

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

for all

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

The following proposition shows that $\mathbf{R}$ can indeed be used as a replacement for $\mathbf{r}$.

Proposition 3.2 *Let $x \in \mathcal{X}^\ell \subset \Delta(\mathcal{I})$, $f \in \mathcal{F}^{k_0} \subset \mathcal{F}$, for some $\ell \in \mathcal{L}$ and $k_0 \in \mathcal{K}$. 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,

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

This formulation allows us to represent the original decision process with partial monitoring as another one with full monitoring (with respect to $\mathbf{R}$).

3.3 An auxiliary approachability problem with full monitoring

We now construct an auxiliary approachability problem. The payoff space is $(\mathbb{R}^{\mathcal{S}\times\mathcal{I}})^{\mathcal{K}\times\mathcal{A}}$ and an element $\tilde{g} \in (\mathbb{R}^{\mathcal{S}\times\mathcal{I}})^{\mathcal{K}\times\mathcal{A}}$ will often be written as

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

Thus, 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 $\langle\tilde{g}|\tilde{z}\rangle$ and the Euclidean norm can be decomposed into:

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

The auxiliary problem is the following. Let $\mathcal{K}\times\mathcal{A}$ be the set of decisions for the decision maker and $\mathcal{F}$ be the convex decision 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, for $(k, a) \in \mathcal{K}\times\mathcal{A}$ and $f \in \mathcal{F}$, by

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

This payoff function is 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 C 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 :

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

Proposition 3.3 *(i) The sets $\tilde{\mathcal{C}}^k$ and $\tilde{\mathcal{C}}$ are closed convex cones.*

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

(iii) $\tilde{\mathcal{C}}$ is approachable in the auxiliary approachability problem. In other words, for all $\tilde{z} \in \tilde{\mathcal{C}}^\circ$ (the polar cone of $\tilde{\mathcal{C}}$), there exists $\tilde{x} := \tilde{\mathbf{x}}(\tilde{z}) \in \Delta(\mathcal{K}\times\mathcal{A})$ such that

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

4 Back to the original approachability problem

4.1 The algorithm for the initial problem

The learning algorithm we will construct in the original problem with partial monitoring is based on the approachability algorithm of the auxiliary problem with full monitoring. It depends on two parameters $\eta > 0$ and $0 < \gamma \leqslant 1$. Let $\tilde{\mathcal{Z}} := \tilde{\mathcal{C}}^\circ \cap \mathcal{B}_2$ where $\mathcal{B}_2$ denotes the closed unit Euclidean ball on $(\mathbb{R}^{\mathcal{S}\times\mathcal{I}})^{\mathcal{K}\times\mathcal{A}}$.

At stage t, follow the three following steps:

(i) compute $\tilde{z}_t := \mathbf{P}_{\tilde{\mathcal{Z}}}\left(\eta\sum_{s=1}^{t-1}\tilde{g}_s\right)$, where $\mathbf{P}_{\tilde{\mathcal{Z}}}$ denotes the Euclidean projection onto $\tilde{\mathcal{Z}}$, and then $\tilde{x}_t := \tilde{\mathbf{x}}(\tilde{z}_t) \in \Delta(\mathcal{K}\times\mathcal{A})$, where $\tilde{\mathbf{x}}$ is defined in Proposition 3.3;

(ii) draw $(k_t, a_t) \sim \tilde{x}_t$ and then $i_t \sim (1-\gamma)a_t + \gamma u$, where $u := (\frac{1}{|\mathcal{I}|}, \ldots, \frac{1}{|\mathcal{I}|})$ is the uniform distribution over $\mathcal{I}$; receive signal $s_t \sim \mathbf{s}(i_t, j_t)$.

(iii) Let $\hat{f}_t = \left(\frac{\mathbb{1}_{\{i_t=i\}}}{\mathbb{P}[i_t=i\,|\,\mathcal{G}_t]}\delta_{s_t}\right)_{i\in\mathcal{I}} \in \mathbb{R}^{\mathcal{S}\times\mathcal{I}}$, where δ_{s_t} is the Dirac mass associated with $s_t \in \mathcal{S}$, $(\mathcal{G}_t)_{t\geqslant 1}$ is the filtration generated by $(k_1, a_1, i_1, s_1, \dots, k_{t-1}, a_{t-1}, i_{t-1}, s_{t-1}, k_t, a_t)$, and set $\tilde{g}_t := \tilde{\mathbf{g}}((k_t, a_t), \hat{f}_t)$.

The definition of the algorithm implies that

$$\mathbb{P}[i_t = i\,|\,\mathcal{G}_t] = (1-\gamma)a_t^i + \frac{\gamma}{|\mathcal{I}|}, \quad i \in \mathcal{I}.$$

and thus, it is easy to see that (see Lemma A.4) $\hat{f}_t$ is unbiased, i.e., $\mathbb{E}\left[\hat{f}_t \,\middle|\, \mathcal{G}_t\right] = \mathbb{E}[f_t\,|\,\mathcal{G}_t]$, with relatively small variance $\mathbb{E}\left[\left\|\hat{f}_t\right\|_2^2 \,\middle|\, \mathcal{G}_t\right] \leqslant |\mathcal{I}|^2/\gamma$.

4.2 Main result

We now state our main result establishing the rate of convergence of $O(T^{-1/3})$ of the average payoff $\bar{g}_T$ to the negative orthant $\mathbb{R}^d_-$. The constants depend on $L_{\mathbf{r}}$, the maximal Lipschitz constant of the maps $\mathbf{r}(x,\cdot)$.

Theorem 4.1 *Let $T \geqslant 1$ be an integer. Against any choices of Nature, the algorithm defined in Section 4.1 run with $\eta = \sqrt{\frac{\gamma}{T|\mathcal{I}|^2}}$ and*

$$\gamma = \min\left\{\left(\frac{11\ L_{\mathbf{r}}\,|\mathcal{I}|\,|\mathcal{K}|\,|\mathcal{A}|}{4\,\|\mathbf{g}\|_2}\right)^{2/3} T^{-1/3},\ 1\right\}$$

guarantees that $\mathbb{E}\left[\mathbf{d}_2\left(\bar{g}_T,\ \mathbb{R}^d_-\right)\right]$ is upper-bounded by

$$\frac{12\,\|\mathbf{g}\|_2^{1/3}\,(L_{\mathbf{r}}\,|\mathcal{I}|\,|\mathcal{K}|\,|\mathcal{A}|)^{2/3}}{T^{1/3}} + \frac{2\sqrt{\pi}\,\|\mathbf{g}\|_2}{T^{1/2}} + \frac{6\,\|\mathbf{g}\|_2^{2/3}\,(L_{\mathbf{r}}\,|\mathcal{I}|\,|\mathcal{K}|\,|\mathcal{A}|)^{1/3}}{T^{2/3}},$$

where $\mathbf{d}_2(\,\cdot\,,\ \mathbb{R}^d_-)$ denotes the distance to $\mathbb{R}^d_-$.

Remark 4.2 *Since $L_{\mathbf{r}}$ scales linearly with $\|\mathbf{g}\|_2$, so does the dominant term of the above bound, as expected.*

The proof is divided into several independent steps.

First, we need to introduce some notation. Let $\bar{\tilde{g}}_T$ 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 $\bar{u}_T(k,a)$ its average over $t \in N_T(k,a)$:

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

Here is an overview of the main steps and arguments.

1. **$\bar{g}_T$ is close to $\frac{1}{T}\sum_{t=1}^T g(a_t, y_t)$:**
   a_t and y_t are approximately the conditional law of i_t and j_t; concentration inequalities and the bilinearity of $\mathbf{g}$ yield that $\mathbf{g}(i_t, j_t)$ and $\mathbf{g}(a_t, y_t)$ are close in expectation, see Lemma B.9.
2. **which is equal to $\sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}} \lambda_T(k,a)\cdot\mathbf{g}(a, \bar{y}_T(k,a))$:**
   This is a consequence of the definitions of $\lambda_T(k,a)$ and $\bar{y}_T(k,a)$, see Lemma B.8.
3. **closer to $\mathbb{R}^d_-$ than $\displaystyle\sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}} \lambda_T(k,a)\mathbf{r}(a, \bar{f}_T(k,a))$:**
   it follows from Proposition 3.1, see Lemma B.7.
4. **which is close to $\sum_{\substack{k\in\mathcal{K}\\ a\in\mathcal{A}}} \lambda_T(k,a)\cdot\mathbf{r}^{[k]}(a, \bar{\hat{f}}_T(k,a))$**
   concentration inequalities give that $\bar{\hat{f}}_T(k,a)$ is close to $\bar{f}_T(k,a)$ in expectation. Then, the auxiliary average payoff being close the auxiliary target set implies that $\bar{\hat{f}}_T(k,a)$ is close to $\mathcal{F}^k$ on which $\mathbf{r}^{[k]}(a,\cdot)$ and $\mathbf{r}(a,\cdot)$ coincide. This way, we prove that $\mathbf{r}^{[k]}(a, \bar{\hat{f}})_T(k,a)$ is close to $\mathbf{r}(a, \bar{f}_T(k,a))$. See Lemma B.6.
5. **which is equal to $\mathbf{R}(\bar{\tilde{g}}_T)$:**
   this follows from the definition of $\mathbf{R}$, see Lemma B.3.
6. **which is close to $\mathbb{R}^d_-$:**
   this follows from the fact that if the average auxiliary payoff $\bar{\tilde{g}}_T$ is close to the auxiliary target set $\tilde{\mathcal{C}}$ then the average payoff $\bar{g}_T$ is close to the target $\mathbb{R}^d_-$. See Lemmas B.2 and B.1.

4.3 Outcome-dependent signals

This section studies the case where $\mathbf{s}(i,j)$ does not depend on the decision i of the decision maker, i.e., $\mathbf{s}(i,j) = \mathbf{s}(i',j)$ for all $i, i' \in \mathcal{I}$.

We aim at constructing an approachability algorithm of the negative orthant $\mathbb{R}^d_-$ with a $O(T^{-1/2})$ convergence rate. The algorithm from Section 3 will be modified in two ways. First, the estimate $\hat{f}_t$ will be simpler as exploration is unnecessary, and second, the random decision of the decision maker will not be perturbed. All previous notation and assumptions stand.

Let $\eta > 0$ be a parameter. For $1 \leqslant t \leqslant T$;

(i) let $\tilde{z}_t = \mathbf{P}_{\tilde{\mathcal{Z}}}\left(\eta \sum_{s=1}^{t-1} \tilde{g}_s\right)$, $\tilde{x}_t := \tilde{\mathbf{x}}(\tilde{z}_t) \in \Delta(\mathcal{K} \times \mathcal{A})$.

(ii) draw $(k_t, a_t) \sim \tilde{x}_t$ and $i_t \sim a_t$; receive $s_t \in \mathcal{S}$

(iii) Let $\hat{f}_t = (\delta_{s_t})_{i \in \mathcal{I}} \quad \in \mathbb{R}^{\mathcal{S} \times \mathcal{I}}$, $\tilde{g}_t = \tilde{\mathbf{g}}((k_t, a_t), \hat{f}_t)$.

The definition of the algorithm implies that the law of i_t knowing $\mathcal{G}_t$ is a_t. The new estimate is also unbiased $\mathbb{E}\left[\hat{f}_t \,\middle|\, \mathcal{G}_t\right] = \mathbb{E}\left[f_t \,|\, \mathcal{G}_t\right]$ with fixed variance $\left\|\hat{f}_t\right\|_2^2 = |\mathcal{I}|$.

Theorem 4.3 *Let $T \geqslant 1$. Against any choices of Nature, the above algorithm with parameter $\eta = (T\,|\mathcal{I}|)^{-1/2}$ guarantees*

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

The proof is omitted as it follows the same steps as the one of Theorem 4.1. One can check that Lemmas B.2, B.3, B.7 and B.8 still hold. The modification of the other lemmas is quite straightforward using the simpler estimate $\hat{f}_t$, without exploration parameter γ.

5 Discussions

5.1 Computational efficiency

The first step of our algorithm is the computation of $\tilde{z}_t$ which follows from projecting onto $\tilde{\mathcal{Z}} := \tilde{\mathcal{C}}^\circ \cap \mathcal{B}_2$, which can be done efficiently as $\tilde{\mathcal{C}}^\circ$ is a polyhedral cone.

The second step is to compute $\tilde{x}_t := \tilde{\mathbf{x}}(\tilde{z}_t)$ which by solving the following minimax problem:

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

As $\Delta(\mathcal{K} \times \mathcal{A})$ and $\mathcal{F}$ are polytopes, this boils down to a linear program. So, the per-step complexity is constant and sums up to a projection and a linear programming. The construction of the simplicial subdivisions is tedious, but only needed once, before learning.

5.2 Almost-sure convergence

Theorem 4.1 only provides a convergence guarantee in expectation. We quickly describe how to adapt the analysis to obtain high probability guarantees.

The proof of Lemma B.1 can be modified in order to obtain a high probability guarantee on $\mathbf{d}_2(\bar{\tilde{g}}_T,\ \tilde{\mathcal{C}})$. 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 $(\|\tilde{g}_t\|_2^2 - (|\mathcal{I}|/\gamma)^2)_{t \geqslant 1}$ 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.

Modifications of Lemmas B.4 and B.9 are straightforward. We apply high probability versions of the concentration inequalities, Propositions E.1 and E.3.

The high probability versions of Lemmas B.5 and B.6 immediately follow from those of Lemma B.1, and Lemmas B.4 and B.5, respectively. The almost-sure convergence follow from a Borel-Cantelli argument.

5.3 Comparison with Mannor et al. [2014b]

The algorithm proposed in Mannor et al. [2014b] has a dimension-independent convergence rate of $O(T^{-1/5})$. We highlight a few ideas already present in Mannor et al. [2014b], and those we have introduced here to obtain the optimal convergence rate of $O(T^{-1/3})$.

Mannor et al. [2014b] also used the single-valued map $\mathbf{r}$ and the decomposition of $\mathcal{F}$ and $\Delta(\mathcal{I})$ into polytopes, introduced by Perchet [2011b]. This allowed them to get the piecewise-affinity of $\mathbf{r}$. This fundamental property was then used in the averaging of the flag estimates on time blocks of fixed lengths, on which the decision maker uses the same random decision.

The algorithm constructed in Section 4.1 manages to average the estimators and to approach the target *at the same time*, without requiring such blocks, resulting in an improved optimal rate of $O(T^{-1/3})$.

Acknowledgements

The authors are grateful to I. Kortchemski, R. Laraki, S. Sorin and G. 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

Jacob Abernethy, Peter L. Bartlett, and Elad Hazan. Blackwell approachability and low-regret learning are equivalent. In *JMLR: Workshop and Conference Proceedings (COLT)*, volume 19, pages 27–46, 2011.

R. J. Aumann and M. B. Maschler. *Repeated Games with Incomplete Information.* MIT Press, Cam-

bridge, MA, 1995. With the collaboration of Richard E. Stearns.

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.

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.

A. Bernstein, S. Mannor, and N. Shimkin. Opportunistic strategies for generalized no-regret problems. *J. Mach. Learn. Res.: Workshop Conf. Proc.*, 30:158–171, 2013.

David Blackwell. Controlled random walks. In *Proceedings of the International Congress of Mathematicians*, volume 3, pages 336–338, 1954.

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

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

S. Bubeck and N. Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5:1–122, 2012.

N. Cesa-Bianchi, G. Lugosi, and G. Stoltz. Minimizing regret with label-efficient prediction. *IEEE Transactions on Information Theory, Institute of Electrical and Electronics Engineers*, 51:2152–2162, 2005.

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

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

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

D. P. Foster, A. Rakhlin, K. Sridharan, and A. Tewari. Complexity-based approach to calibration with checking rules. *J. Mach. Learn. Res.: Workshop Conf. Proc.*, 19:293–314, 2011.

Dean P. Foster. A proof of calibration via Blackwell's approachability theorem. *Games and Economic Behavior*, 29(1):73–78, 1999.

R. Jenatton, J. Huang, and C. Archambeau. Adaptive algorithms for online convex optimization with long-term constraints. *Journal of Machine Learning Research: Workshop and Conference Proceedings (ICML)*, 48:1–10, 2016.

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

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

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

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

S. Mannor and N. Shimkin. Regret minimization in repeated matrix games with variable stage duration. *Games Econom. Behav.*, 63:227–258, 2008.

S. Mannor, V. Perchet, and G. Stoltz. Approachability in unknown games: Online learning meets multi-objective optimization. *Journal of Machine Learning Research: Workshop and Conference Proceedings (COLT)*, 35:339–355, 2014a.

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

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.

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, 2014b.

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

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

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

A. Piccolboni and C. Schindelhauer. Discrete prediction games with arbitrary feedback and loss. In *Proceedings of the 14th Annual Conference on Compu-*

tational Learning Theory (COLT), pages 208–223. Springer, 2001.

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

A. Rakhlin, K. Sridharan, and A. Tewari. Online learning: Beyond regret. *J. Mach. Learn. Res.: Workshop Conf. Proc.*, 19:559–594, 2011.

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

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

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

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.

T. Tomala. Belief-free communication equilibria. *Mathematics of Operations Research*, 38:617–637, 2013.

A Proofs of technical results

A.1 Proof of Proposition 3.1

Lemma A.1 *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 $\mathcal{X}^\ell$.

Proof. 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 $\lambda \in [0, 1]$. 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

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

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 (4). □

Proof. of Proposition 3.1 (i) Let $x \in \Delta(\mathcal{I})$ and $y \in \Delta(\mathcal{J})$. Denote $f = \mathbf{f}(y)$. We consider the unique decomposition $f = \sum_{b\in\mathcal{B}} \mu^b \cdot b$ and $k \in \mathcal{K}$ such that $\operatorname{supp}\mu \subset \mathcal{F}^k$. $\mathbf{f}^{-1}$ being affine on $\mathcal{F}^k$, we have

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

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

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

where for the second equality, we recognized the definition of $\mathbf{r}^n(x, b)$ from Equation (3) on page 5, and the the last equality, the definition of $\mathbf{r}^n(x, f)$ from Equation (4).

(ii) Let $f \in \mathcal{F}$. Thanks to the characterization of approachability from Proposition 1, 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 3.1. 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 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 3.1 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 $\mathcal{F}^k$,

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

(iv) is already proved in Lemma A.1. □

A.2 Existence of $\mathbf{r}^{[k]}$

Proposition A.2 *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 $f \in \mathcal{F}^k$, $\mathbf{r}^{[k]}(x, f) = \mathbf{r}(x, f)$.

Proof. Let $k \in \mathcal{K}$ and $x \in \Delta(\mathcal{I})$. Let us consider $\mathrm{span}(\mathcal{F}^k) \subset \mathbb{R}^{\mathcal{S}\times\mathcal{I}}$, the linear span of $\mathcal{F}^k$. There exists a basis $(f_1, \dots, f_q)$ of $\mathrm{span}(\mathcal{F}^k)$ 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 $\mathrm{span}(\mathcal{F}^k)$ by setting

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

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 $\mathrm{span}(\mathcal{F}^k)$.

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 $\mathrm{span}(\mathcal{F}^k)$ and can be uniquely written

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

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 3.1) 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

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

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

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

Let us prove that

$$\begin{aligned}
&\mathbf{r}(x, f) - \Lambda^- \mathbf{r}\left(x, \sum_{\lambda_p<0} \left(\frac{\lambda_p}{\Lambda^-}\right) f_p\right) \\
&\qquad = \Lambda^+ \cdot \mathbf{r}\left(x, \sum_{\lambda_p>0} \left(\frac{\lambda_p}{\Lambda^+}\right) f_p\right). \quad (5)
\end{aligned}$$

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

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

For relation (5) 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.3 A lemma on $L_{\mathbf{r}}$

Recall that $L_{\mathbf{r}}$ was defined as the maximal Lipschitz constant of mappings $\mathbf{r}(x,\cdot)$. Then it is also 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 A.3 *$L_{\mathbf{r}}$ 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*

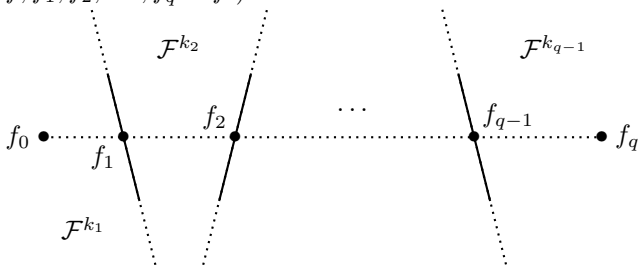
(i) for all $f, f' \in \mathbb{R}^{\mathcal{S}\times\mathcal{I}}$, $\left\|\mathbf{r}^{[k]}(a,f) - \mathbf{r}^{[k]}(a,f')\right\|_2 \leqslant L_{\mathbf{r}} \|f - f'\|_2$;

(ii) for all $f, f' \in \mathcal{F}$, $\|\mathbf{r}(a,f) - \mathbf{r}(a,f')\|_2 \leqslant L_{\mathbf{r}} \|f - f'\|_2$.

Proof. 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 $1 \leqslant p \leqslant q$, see Figure 1. Therefore, using the fact that $\mathbf{r}^{[k']}(a,\,\cdot\,)$ and

Figure 1: An illustrative figure of the sequence $(f_0 = f, f_1, f_2, \ldots, f_q = f')$



$\mathbf{r}(a,\,\cdot\,)$ coincide on $\mathcal{F}^{k'}$ for all $k' \in \mathcal{K}$, we can write

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

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

A.4 Proof of Proposition 3.2

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

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

where the second equality holds because by linearity of $\mathbf{r}^{[k]}(a,\,\cdot\,)$ (Proposition A.2), the fourth because $\mathbf{r}^{[k_0]}(x,\,\cdot\,)$ and $\mathbf{r}(x,\,\cdot\,)$ coincide on $\mathcal{F}^{k_0}$ (property (ii) in Proposition A.2, and the last by affinity of $\mathbf{r}(\,\cdot\,,f)$ on $\mathcal{X}^\ell$ (property (iv) in Proposition 3.1). □

A.5 Proof of Proposition 3.3

(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 C.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 $k \in \mathcal{K}$, $(\tilde{g}^{ka})_{a\in\mathcal{A}} \in \tilde{\mathcal{C}}^k \subset \mathbf{R}_k^{-1}(\mathbb{R}_-^d)$.

(iii) This full information game has convex compact decision sets and a bilinear payoff function. Thanks to the characterization of approachable closed convex cones in full information, the statement of the proposition is then equivalent to Blackwell 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 3.1, there exists $x \in \Delta(\mathcal{I})$ such that such that $\mathbf{r}(x, f) \in \mathbb{R}_-^d$. By Proposition 3.1, 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 random decision

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

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 $\tilde{\mathcal{C}}^{k_0} = \mathbf{R}_{k_0}^{-1}(\mathbb{R}_-^d) \cap (\mathcal{F}_c^{k_0})^{\mathcal{A}}$. First, because $f \in \mathcal{F}^{k_0}$, $\lambda^a f$ belongs to the closed convex cone $\mathcal{F}_c^{k_0} = \mathbb{R}_+\mathcal{F}^{k_0}$ and we have indeed $(\lambda^a f)_{a\in\mathcal{A}} \in (\mathcal{F}_c^{k_0})^{\mathcal{A}}$. Then, let us prove that $\mathbf{R}_{k_0}\left((\lambda^a f)_{a\in\mathcal{A}}\right) \in \mathbb{R}_-^d$. Using Proposition 3.2,

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

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

A.6 Properties of the estimate $\hat{f}_t$.

Lemma A.4 *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) $\mathbb{E}\left[\left\|\hat{f}_t\right\|_2^2 \,\middle|\, \mathcal{G}_t\right] \leqslant \dfrac{|\mathcal{I}|^2}{\gamma}$;

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

Proof. (i) Let $i \in \mathcal{I}$. Using the conditional expectation with respect to event $\{i_t = i\}$, we have

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

hence the result.

(ii) We write

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

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

(iii) We have

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

□

B Proof of Theorem 4.1

B.1 Average auxiliary payoff $\bar{\tilde{g}}_T$ is close to auxiliary target set $\tilde{\mathcal{C}}$

Lemma B.1

$$\mathbb{E}\left[\mathbf{d}_2\left(\bar{\tilde{g}}_T,\ \tilde{\mathcal{C}}\right)\right] \leqslant \frac{1}{2\eta T} + \frac{\eta\,|\mathcal{I}|^2}{2\gamma}.$$

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

$$\tilde{z}_t = \mathbf{P}_{\tilde{\mathcal{Z}}}\left(\eta \sum_{s=1}^{t-1} \tilde{g}_s\right) = \arg\min_{\tilde{z}\in\tilde{\mathcal{Z}}} \left\| \tilde{z} - \eta \sum_{s=1}^{t-1} \tilde{g}_s \right\|_2^2$$
$$= \arg\max_{\tilde{z}\in\tilde{\mathcal{Z}}} \left\{ \left\langle \eta \sum_{s=1}^{t-1} \tilde{g}_s \middle| \tilde{z} \right\rangle - \frac{1}{2}\|\tilde{z}\|_2^2 \right\}.$$

Then, Theorem D.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

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

By taking the expectation and dividing by T, we get

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

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:

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

where we used Lemma A.4 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 algorithm, the law of random variable (k_t, a_t) knowing $\mathcal{H}_t$ is $\tilde{x}_t$. We now resume the above computation by introducing the conditional expectation with respect to $\mathcal{H}_t$ and f_t:

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

By definition of the algorithm, $\tilde{x}_t = \tilde{\mathbf{x}}(\tilde{z}_t)$. In other words (see Proposition 3.3), 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 $\tilde{\mathbf{g}}$,

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

Using (ii) from Lemma B.1, we have

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

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

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

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

B.2 From $\bar{\tilde{g}}_T$ in the auxiliary space to $\mathbf{R}(\bar{\tilde{g}}_T)$ in the initial space

Lemma B.2

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

Proof. It follows from property (ii) in Proposition 3.3 that $\tilde{\mathcal{C}} \subset \mathbf{R}^{-1}(\mathbb{R}^d_-)$. Therefore, we can write

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

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 A.3 which is common to the linear applications $\mathbf{r}^{[k]}(a,\,\cdot\,)$, we have

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

which concludes the proof. □

B.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 B.3

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

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

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

□

B.4 Average estimator $\bar{\hat{f}}_T(k,a)$ is close to average flag $\bar{f}_T(k,a)$

Lemma B.4

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

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 E.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 $\mathcal{G}_t$,

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

where the last equality follows from (i) in Lemma A.4. Moreover, using (iii) from Lemma A.4, we bound each $X_t(k,a)$ as follows.

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

where we used the fact that $\gamma \geqslant 1$ for the last inequality. As far as the conditional variances are concerned, we have

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

where the first term of the second line has been bounded using property (ii) from Lemma A.4, 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

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

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

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

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

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

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

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

□

B.5 Average estimator $\bar{\hat{f}}_T(k,a)$ is close to $\mathcal{F}_c^k$

Lemma B.5

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

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:

$$\tilde{\mathcal{Z}}_0 \subset \sqrt{|\mathcal{K}|\,|\mathcal{A}|}\cdot\tilde{\mathcal{Z}}. \tag{6}$$

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

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

By summing over k and a, we have:

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

where for the inequality we used inclusion (6), and for the last equality Proposition C.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 $\mathbf{d}_2(\bar{\tilde{g}}_T,\ \tilde{\mathcal{C}})$ by the bound from Lemma B.1 yields the result.

Let us now prove inclusion (6). 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

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

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

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

□

B.6 $\mathbf{r}^{[k]}(a, \bar{\hat{f}}_T(k,a))$ is close to $\mathbf{r}(a, \bar{f}_T(k,a))$

Lemma B.6

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

Proof. Let $(k,a) \in \mathcal{K}\times\mathcal{A}$ and denote $f := \bar{f}_T(k,a)$ and $\hat{f} := \bar{\hat{f}}_T(k,a)$ 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 Proposition A.2, we can write

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

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 $\mathbf{r}^{[k]}(a,\cdot)$ to 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

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

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,

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

Finally, we get the result by taking the expectation, summing over k and a, and plugging Lemmas B.4 and B.5. □

B.7 $\mathbf{g}$ is closer to $\mathbb{R}^d_-$ than $\mathbf{r}$

Lemma B.7

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

Proof. Let $k\in\mathcal{K}$ and $a\in\mathcal{A}$. First note that $\mathbf{f}(\bar{y}_T(k,a)) = \bar{f}_T(k,a)$. Indeed, using the affinity of $\mathbf{f}$,

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

For each component $n\in\{1,\dots,d\}$, we have $\mathbf{g}^n(a,\bar{y}_T(k,a)) \leqslant \mathbf{r}^n(a,\bar{f}_T(k,a))$ by property (i) in Proposition 3.1. Finally, using the explicit expression of the Euclidean distance to $\mathbb{R}^d_-$, we have

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

□

B.8 Decomposition of $\mathbf{g}(a_t,y_t)$ with respect to the realized auxiliary decision (k_t,a_t)

Lemma B.8

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

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

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

□

B.9 From $\mathbf{g}(i_t,j_t)$ to $\mathbf{g}(a_t,y_t)$

Lemma B.9

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

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

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

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

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

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

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

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

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

We can thus apply Corollary E.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,

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

And taking the expectation:

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

□

B.10 Final bound

We now combine the above lemmas in the order specified at the beginning of the section to get:

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

Injecting the values of η and γ yields the result.

C Closed Convex Cones

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

Definition C.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 C.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 C.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 Borwein and Lewis [2010].

Proposition C.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 C.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 $w \in \mathcal{W}$,

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

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

Proposition C.6 *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*

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

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

D A regret minimization bound

The following statement is classic in the regret minimization literature—see e.g. Shalev-Shwartz [2011, Theorem 2.4].

Theorem D.1 *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 $\mathbb{R}^n$, $\eta > 0$, and*

$$z_t = \arg\max_{z\in\mathcal{Z}} \left\{ \left\langle \eta \sum_{s=1}^{t-1} u_s \middle| z \right\rangle - \frac{1}{2}\left\|z\right\|_2^2 \right\}, \quad t \geqslant 1.$$

Then, for all $T \geqslant 1$,

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

E Concentration inequalities

The following result is a generalization to vector-valued martingale differences of Hoeffding–Azuma's inequality and is due to Kallenberg and Sztencel [1991].

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

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

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 E.2 *Under the assumptions of Proposition E.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 E.1 by integrating the tail of the distribution:

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

□

The following Bernstein-like inequality is proved in Pinelis [1994]—see also [Tarres and Yao, 2014, Corollary A.2].

Proposition E.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 E.4 *Under the assumptions of Proposition E.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.

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

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

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

□