HAL open science

Approachability, Regret and Calibration; implications and equivalences.

Vianney Perchet

To cite this version:

Vianney Perchet. Approachability, Regret and Calibration; implications and equivalences.. 2013. hal-00773218

HAL Id: hal-00773218

https://hal.science/hal-00773218v1

Preprint submitted on 12 Jan 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Approachability, Regret and Calibration implications and equivalences.

Vianney Perchet

January 12, 2013

Abstract

Blackwell approachability, regret minimization and calibration are three criteria evaluating a strategy (or an algorithm) in different sequential decision problems, or repeated games between a player and Nature. Although they have at first sight nothing in common, links between have been discovered: both consistent and calibrated strategies can be constructed by following, in some auxiliary game, an approachability strategy.

We gathered famous or recent results and provide new ones in order to develop and generalize Blackwell's elegant theory. The final goal is to show how it can be used as a basic powerful tool to exhibit a new class of intuitive algorithms, based on simple geometric properties. In order to be complete, we also prove that approachability can be seen as a byproduct of the very existence of consistent or calibrated strategies.

Introduction

Sequential decision problems can be represented as repeated games between a player and Nature. At each stage the player (also called agent, decision maker or predictor depending on the context) chooses an element of his decision set. At the same time, Nature chooses on her side a state of the world. Those sequences of choices generate a sequence of *outcomes* that induces an overall *payoff* to the player.

The opponent is called *Nature* as we do not precise her payoff, her objectives or her rationality; absolutely no assumptions is made on her behavior, and future states of the world cannot be inferred from the past. Typically the environment is not stochastic or Bayesian but *adversarial*; for instance, Nature can represent one malignant opponent, or a set of independent (or correlated) players. A crucial requirement of these model is that a strategy of the player must be *good* (i.e., it must fulfill some exogenous *criterion*) against every possible sequence of states of the world (or simply against any strategy of Nature).

Depending on the structure of outcomes mappings, overall objectives of the player might vary. Hannan [30] studied the case where an outcome is actually a real payoff. The player's goal is to maximize his average (or cumulative) payoff. As we made no assumption on Nature's behavior, a player can not ensure to himself a given exogenous

amount, unlike in traditional zero sum game where a value can be guaranteed: assume for instance that Nature decides to give a payoff of zero (or one, minus one, etc) to the player at each stage, no matter what he does.

The criterion Hannan introduced is called *regret* and measures the difference between the average payoff the player got and what *he would have got if he had chosen the same action repeatedly*. It is somehow related to convex optimization (if Nature chooses repeatedly the same loss function), or more precisely to *online convex optimization*.

Main results of Hannan [30] are that such a *consistent* strategy, i.e., a strategy without regret exists, and he constructed one. This has been widely refined and improved using different techniques and ideas by notably (providing an exhaustive list seems almost impossible as the subject has been developed by many different communities) Foster & Vohra [23], Hart & Mas-Colell [31], Fudenberg & Levine [28], Lehrer [43], Auer, Cesa-Bianchi & Gentile [3], Cesa-Bianchi & Lugosi [14] (see also references therein), Sorin [71]...

When outcomes are *vectorial* (and not scalar) *payoffs*, the problem is closely related to *multicriteria optimization*, each coordinate representing a different sub-objective. Instead of considering some exogenous convex combination of these objectives or optimizing them in a given order (to encompass this framework into the precedent one), Blackwell [9] introduced another concept. He considered that some *target set* is given and the player's goal is that the average outcome converges to it; on the contrary, Nature tries to push it away. Formally, a given closed set is *approachable*, if the player has a strategy such that the average payoffs remains, after some maybe large stage, arbitrarily closed to this target set, no matter the sequence of moves of Nature.

Blackwell's approachability theory is quite elegant as it relies on simple geometric properties. They allowed him to characterize explicitly approachable convex sets and to provide a simple sufficient approachability condition for non-convex set (such sets are called, in reference to Blackwell, *B-sets*). Spinat [73] proved later that this was in fact almost a necessary condition.

Maybe the first and most important use of this whole theory is due to Kohlberg [39]. He constructed, using this simple tool, an optimal strategy for the uninformed player in zero-sum games with incomplete information, introduced by Aumann and Maschler [5] (see for instance Mertens, Sorin & Zamir [55] and references given for more details on this subject). Approachability gained also a recent interest, both from the game theory and machine learning community, with works of – again non-exhaustively – Vieille [77], Hart & Mas-Colell [32], Spinat [73], Lehrer [42], Benaïm, Hofbauer & Sorin [7], Mannor & Shimkin [49], Lehrer & Solan [44, 46], As Soulaimani, Quincampoix & Sorin [2], Mannor & Tsitsiklis [53], Perchet [60, 61], Rakhlin, Sridharan & Tewari [66], Perchet & Quincampoix [63]...

Another (and the last to be considered here) criterion is *calibration*, written within this framework by Dawid [16] and extended thereafter by, in many others, Foster & Vohra [24], Fudenberg & Levine [27], Lehrer [41], Sandroni, Smorodinsky & Vohra [68], Sorin [71], Perchet [60], Foster, Rakhlin, Sridharan & Tewari [22], and so on.

Here, a stage outcome is not some payoff (either scalar or vectorial) but the actual *state of the world* chosen by Nature. The overall objective of the player is to predict, sequentially, the whole sequence of states so that the average prediction and the empirical distribution of states are asymptotically arbitrarily close. Without any other restrictions, this is in fact fairly easy: one just has to predict at some stage the outcome of the precedent one.

Additional requirements can be, for instance, that predictions can only belong to some finite (yet maybe large) set and that the empirical distribution of states *on the set of stages where a specific prediction is made* is closer to this prediction than to any other possible one. An usual and celebrated example consists in a meteorologist that predicts, each day, the probability of rain the following day. Predictions belongs to 0%, 10%, 20%, etc. and it is asked that that when a meteorologist says that the probability of rain is, say, 30%, it rains in average between 35% and 45% of the times.

Oakes [57] and Dawid [17] proved that no deterministic algorithm can be calibrated (yet this strong assessment could be discussed) while random algorithm can, as proved by Foster & Vohra [24]. The existence of such algorithms can be seen as a negative result, as it claims that a strategic non-informed meteorologist can mimic an expert one (that knows the true underlying process, if it exists); a whole literature studied this aspect and recent results are gathered into the survey of Olszewski [58]. On the other hand, it can also be seen as a positive result, as it states that the long term behavior of Nature can asymptotically be predicted, and this might lead to another class of algorithms and results, as in Foster & Vohra [23] or Perchet [59, 62].

A common feature of regret minimization and calibration is that they can be written as a specific case of approachability of a well chosen target set in some auxiliary vectorial payoff game . The first to notice this property is Blackwell [10] (this idea is already mentioned at the end of the seminal paper of Hannan [30] or in Luce & Raiffa [48]) and then by Foster [21], Hart & Mas-Colell [32], Lehrer & Solan [45], Sorin [71], Perchet [60], Mannor & Stoltz [50], Abernathy, Barltlett & Hazan [1]...

We assumed implicitly that the player observes the sequence of states of the world; this is in fact a crucial hypothesis here, sometimes referred to as *full monitoring*. In particular, we will not consider the case of partial monitoring (or *bandit* problems), or stochastic games (where, for instance, the whole sequence of outcomes could depend of a unique choice at some stage). Those are also interesting subjects, yet far from the current scope.

Objectives and Structure of the paper.
Describing explicit *interactions* and *equivalences* between the notions of approachability, calibration and regret is the central point of this paper, the final argument being that that explicit constructions of consistent and calibrated strategies (even for more precise or refined notions that the ones introduced here) are possible and provided thanks to approachability theory. The remaining is organized as follows:

In Section 1, we introduce the concept of *approachability*, centerpiece of this work.

We first recall (in Subsection 1.1) a sufficient and necessary condition under which an arbitrary set is approachable. The specific case of convex sets, for which a complete characterization is available, is studied in Subsection 1.2. First extensions and generalizations of the framework (e.g., in infinite dimension, with variable stage durations, unbounded payoffs, etc.) are given in Subsection 1.3. Last Subsection 1.4 is concerned with other possible proofs and techniques of approachability. In particular, we show that approachability with respect to the supremum norm can be achieved using some potential minimization, generalizing the *exponential weight algorithm*; we also prove that the usual Euclidian (or Hilbertian) framework is not necessary for approachability.

Proofs are almost always provided, as long as they bring something new to the literature (yet some technical lemmas are delayed to the Appendix).

Regret minimization is introduced in Section 2. Several refinements are introduced and links with game theory (as well as famous algorithms called *exponential weight algorithm* and *follow the perturbed leader*) are given in Subsection 2.3. Since our purpose is to provide reduction to some auxiliary approachability problems, proofs are only sketched in this section and delayed to the last one. An example of regret minimization, with *expert advice* is given for illustration at the end; however, this subject is very well studied in the book of Cesa-Bianchi & Lugosi [14].

Calibration and its generalizations are formalized in Section 3; for the same reasons, proofs are essentially delayed to the last section. We provide there a discussion on wether calibration (yet a weaker but maybe more intuitive notion) can or can not be obtained using deterministic algorithms.

Final Section 4 contains all the reductions to approachability. We prove (or recall) how regret minimizations (either with finite or infinite action spaces) and calibration (either finite or with checking rules) can be obtained using approachability results from the first section.

Maybe the most general results are, on regret minimization, Theorems 4.1 and 4.2 that provide (explicit for the first one) strategy minimizing swap regret if action space are, respectively, finite or infinite. Proposition 4.2, due to Blackwell [10] himself, shows how minimization of the supremum norm of regret is exactly approachability.

Concerning calibration, most striking results might be Proposition 4.5, its consequence Theorem 4.4 and Theorem 4.5. They refine and generalize recent results of Mannor & Stoltz [50] as well as Rakhlin, Sridharan and Tewari [66].

We conclude this Section by explaining how the circle is complete: if regret minimization and calibration can be seen as specific instances of approachability, the converse is also true. Indeed, using some generalized notions of regret and/or calibration, one can construct approachability strategies (in the case of convex sets).

1 Blackwell's approachability

1.1 Approachability of arbitrary sets

Consider a two-person repeated game between a player and Nature. Their actions set are respectively denoted by $\mathcal{A}$ and $\mathcal{B}$ (of respective cardinality A and B) and payoffs are defined trough some vectorial mapping $g : \mathcal{A} \times \mathcal{B} \to \mathbb{R}^d$. The game is repeated in discrete time, and we denote actions chosen at stage $n \in \mathbb{N}$ by $a_n \in \mathcal{A}$ and $b_n \in \mathcal{B}$; they induce a payoff $g_n := g(a_n, b_n) \in \mathbb{R}^d$. Formally, a_n and b_n are functions of the history, *i.e.*, the past observations $h^{n-1} = (a_1, b_1, \ldots, a_{n-1}, b_{n-1}) \in (\mathcal{A} \times \mathcal{B})^{n-1} =: H_{n-1}$.

Explicitly, a strategy σ of the player is a mapping from $H := \bigcup_{n\in\mathbb{N}} H_n$, the set of finite histories, into $\Delta(\mathcal{A})$, the set of probability distributions over $\mathcal{A}$. Similarly, a strategy τ of Nature is a mapping from H into $\Delta(\mathcal{B})$. Kolmogorov's extension theorem implies that a pair (σ, τ) induces a probability distribution $\mathbb{P}_{\sigma,\tau}$ over $\mathcal{H} = (\mathcal{A} \times \mathcal{B})^{\mathbb{N}}$, the set of infinite histories of the game embedded with the product topology.

Before defining the concept of approachability, we introduce some notations. Given a closed set $\mathcal{E} \subset \mathbb{R}^d$, we denote by $d_{\mathcal{E}}(x) = \inf_{z\in\mathcal{E}}\{\|x - z\|\}$ the distance from x to $\mathcal{E}$, by $\mathcal{E}^\delta = \{z \in \mathbb{R}^d \text{ s.t. } d_{\mathcal{E}}(x) < \delta\}$ the δ-open neighborhood of $\mathcal{E}$, and by $\Pi_{\mathcal{E}}(x) = \{z \in \mathcal{E} \text{ s.t. } \|x - z\| = d_{\mathcal{E}}(x)\}$ the projection of x onto $\mathcal{E}$, which is in general non single-valued. We also denote by $\mathrm{co}\Big(\mathcal{E}\Big)$ the convex hull of a set. The mapping g defined on $\mathcal{A} \times \mathcal{B}$ (and more generally any such mapping) is extended to $\Delta(\mathcal{A}) \times \Delta(\mathcal{B})$ by $g(x, y) = \mathbb{E}_{x\otimes y}\Big[g(a, b)\Big]$. The average of a sequence $s = \{s_m\}_{m\in\mathbb{N}}$ up to stage $n \in \mathbb{N}$ is denoted by $\overline{s}_n := \sum_{m=1}^n s_m/n$.

Definition 1.1 *A closed set $\mathcal{E} \subset \mathbb{R}^d$ is approachable by the player if he has a strategy σ ensuring, for every $\varepsilon > 0$, the existence of some integer $N_\varepsilon \in \mathbb{N}$ such that, no matter the strategy τ of Nature,*

$$\sup_{n\geq N_\varepsilon} \mathbb{E}_{\sigma,\tau}\Big(d_{\mathcal{E}}(\overline{g}_n)\Big) \leq \varepsilon \quad \textit{and} \quad \mathbb{P}_{\sigma,\tau}\left(\sup_{n\geq N_\varepsilon} d_{\mathcal{E}}(\overline{g}_n) \geq \varepsilon\right) \leq \varepsilon\,. \tag{1}$$

A set $\mathcal{E}$ is excludable by Nature if she can approach the complement of $\mathcal{E}^\delta$ for some $\delta > 0$.

Informally, a given set $\mathcal{E} \subset \mathbb{R}^d$ is approachable by the player if he has a strategy such that the average payoff converges almost-surely to $\mathcal{E}$, uniformly with respect of the strategies of Nature. The right hand side of Equation (1) clearly implies the first one, which is actually the most commonly used (and rates of convergences, i.e. smallest mappings $\varepsilon \mapsto N_\varepsilon$ satisfying each condition, might differ).

1.1.1 Approachable arbitrary set : Blackwell's sufficient condition

Blackwell [9] provided a simple geometrical condition under which a set $\mathcal{E}$ is approachable. This sufficient condition is in fact almost necessary (as proved in Section 1.1.2, following Spinat [73]).

Definition 1.2 *A closed set $\mathcal{E} \subset \mathbb{R}^d$ is a B-set if for every $z \in \mathbb{R}^d$, there exists a projection $\pi \in \Pi_{\mathcal{E}}(z)$ and $x := x(z) \in \Delta(\mathcal{A})$ such that the hyperplane perpendicular to $z - \pi$ at z separates z from $\Big\{g(x,y)\,,\, y \in \Delta(\mathcal{B})\Big\}$, or formally:*

$$\forall\, z \in \mathbb{R}^d, \exists\, \pi \in \Pi_{\mathcal{E}}(z), \exists\, x \in \Delta(\mathcal{A}), \forall\, y \in \Delta(\mathcal{B}) : \quad \langle g(x,y) - \pi, z - \pi \rangle \leq 0\,. \tag{2}$$

Blackwell [9] proved that being a B-set is sufficient for approachability ; he also exhibited a specific strategy, from now on referred to as Blackwell (approachability) strategy.

Theorem 1.1 *If $\mathcal{E}$ is a B-set, then $\mathcal{E}$ is approachable by the player. Moreover, the strategy σ defined by $\sigma(h^n) = x(\overline{g}_n)$ ensures that, for every $\eta > 0$ and against any strategy τ of Nature:*

$$\mathbb{E}_{\sigma,\tau}\Big[d_{\mathcal{E}}(\overline{g}_n)\Big] \leq 2\sqrt{\frac{\kappa_0}{n}} \quad \textit{and} \quad \mathbb{P}_{\sigma,\tau}\left(\sup_{m\geq n} d_{\mathcal{E}}(\overline{g}_m) \geq \eta\right) \leq \frac{8}{\eta^2}\frac{\kappa_0}{n}\,, \tag{3}$$

where $\kappa_0 = \|g\|_\infty := \sup_{x,y} \mathbb{E}_{x,y}\Big[\|g(a,b)\|^2\Big] = \sup_{a,b} \|g(a,b)\|^2$.

Blackwell [9] and Mertens, Sorin & Zamir [55] obtained respectively the bounds in expectation and in probability. The very definition of $\|g\|_\infty$ allows each $g(a,b)$ to be random variables with bounded second moment.

We propose in the following Corollary 1.1 a slight variant that improves the constants (in the deterministic case or when $\mathcal{E}$ is compact); for instance, they are divided by two if $\mathcal{E} = \{0\}$, as in Section 1.3.6.

Corollary 1.1 *A closed set $\mathcal{E}$ is approachable if and only if $\mathcal{E}_g := \mathcal{E} \cap \mathrm{co}\left\{g(a,b)\,;a \in \mathcal{A}, b \in \mathcal{B}\right\}$ is also approachable. Blackwell's strategy applied to $\mathcal{E}_g$ ensures that*

$$\mathbb{E}_{\sigma,\tau}\Big[d_{\mathcal{E}}(\overline{g}_n)\Big] \leq \sqrt{\frac{\kappa}{n}} \quad \textit{and} \quad \mathbb{P}_{\sigma,\tau}\left(\sup_{m\geq n} d_{\mathcal{E}}(\overline{g}_m) \geq \eta\right) \leq \frac{2}{\eta^2}\frac{\kappa}{n}\,,$$

where $\kappa = (\|g\|_\infty + \|\mathcal{E}_g\|)^2$ and $\|\mathcal{E}_g\| := \sup\Big\{\|z\|\,;\, z \in \mathcal{E}_g\Big\}$ is smaller than $\|g\|_\infty$.

Proof: An approachability strategy of $\mathcal{E}$ ensures that any accumulation point of $\overline{g}_n$ must belong to both the closed set $\mathcal{E}$ and to the compact set $\mathrm{co}\left\{g(a,b)\,;a \in \mathcal{A}, b \in \mathcal{B}\right\}$, hence to $\mathcal{E}_g$. Reciprocally, any approachability strategy of $\mathcal{E}_g$ approaches its super-set $\mathcal{E}$.

Let σ be Blackwell's strategy applied to $\mathcal{E}_g$, define $\delta_n := d_{\mathcal{E}}(\overline{g}_n)$ and denote by π_n any element of $\Pi_{\mathcal{E}}(\overline{g}_n)$ given by Equation (2). Definition of $d_{\mathcal{E}}$ implies that

$$\begin{aligned}\delta_{n+1}^2 &\leq \left\|\overline{g}_{n+1} - \pi_n\right\|^2 = \left\|\frac{n}{n+1}(\overline{g}_n - \pi_n) + \frac{1}{n+1}(g_{n+1} - \pi_n)\right\|^2 \\ &= \frac{n^2}{(n+1)^2}\delta_n^2 + \frac{1}{(n+1)^2}\left\|g_{n+1} - \pi_n\right\|^2 + \frac{2n}{(n+1)^2}\left\langle \overline{g}_n - \pi_n, g_{n+1} - \pi_n\right\rangle.\end{aligned}$$

Conditioning on the finite history h^n and using Equation (2) as well as the definitions of $\|g\|_\infty$ and $\|\mathcal{E}_g\|$, the last inequality becomes

$$\mathbb{E}_{\sigma,\tau}\left[\delta_{n+1}^2 \,\middle|\, h^n\right] \leq \frac{n^2}{(n+1)^2}\delta_n^2 + \frac{(\|g\|_\infty + \|\mathcal{E}_g\|)^2}{(n+1)^2}$$

and, with a simple induction, $\mathbb{E}_{\sigma,\tau}[\delta_n^2] \leq \kappa/n$. Thus $\overline{g}_n$ converges in probability towards $\mathcal{E}$. The almost sure convergence is a consequence of the facts that

$$Z_n := \delta_n^2 + \mathbb{E}_{\sigma,\tau}\left[\sum_{k=n}^{\infty} \frac{\|g_{k+1} - \pi_k\|^2}{(k+1)^2}\right] \text{ is a supermartingale and } \mathbb{E}_{\sigma,\tau}[Z_n] \leq \frac{2\kappa}{n}.$$

Indeed, Doobs' inequality (see Neveu [56], prop. IV.5.2) implies then that

$$\mathbb{P}_{\sigma,\tau}(\exists m \geq b \text{ s.t. } Z_m \geq \eta^2) \leq \frac{\mathbb{E}_{\sigma,\tau}[Z_n]}{\eta^2} \leq \frac{2\kappa}{\eta^2 n},$$

which gives the result. ■

Blackwell's strategy depends only on the sequence $\{g_n\}_{n\in\mathbb{N}}$ so these results do not require the finiteness of $\mathcal{B}$ or $\mathcal{A}$, nor that Nature's actions are observed. In fact, we could as well assume the following model that we call the *compact case* (in opposition to the finite case).

Action sets are compact and convex sets, denoted by $\mathcal{X} \subset \mathbb{R}^A$ and $\mathcal{U} \subset \left(\mathbb{R}^d\right)^A$. At stage $n \in \mathbb{N}$, Nature chooses an *outcome* $U_n = (U_n^a)_{a\in\mathcal{A}} \in \left(\mathbb{R}^d\right)^A$ in $\mathcal{U}$ and the player chooses $x_n \in \mathcal{X}$. Those choices incur the vector payoff $g_n = x_n.U_n \in \mathbb{R}^d$, the standard inner product between x_n and U_n. Condition (2) that defines B-set becomes then

$$\forall\, z \in \mathbb{R}^d,\ \exists\, \pi \in \Pi_{\mathcal{E}}(z), \quad \inf_{x\in\mathcal{X}} \sup_{U\in\mathcal{U}} \langle x.U - \pi, z - \pi\rangle \leq 0\,.$$

It is also possible to incorporate randomness in this model. The compact and convex sets $\mathcal{X}$ and $\mathcal{U}$ can be sets of probability distribution (this was the case when $\mathcal{X} = \Delta(\mathcal{A})$) and in that case $x.U$ is the expectation of a random payoff associated with x and U (that must have a second moment).

1.1.2 Equivalent formulations and necessary condition

Blackwell defined geometrically a B-set from outside. As Soulaimani, Quincampoix & Sorin [2] noticed that it can also be defined similarly from inside. Informally, one can interpret these definitions slightly differently: instead of viewing approachability as the convergence of average payoffs to $\mathcal{E}$, it can be understood as preventing average payoffs from escaping $\mathcal{E}$.

First, we need to recall the notion of proximal normals to $\mathcal{E}$.

Definition 1.3 *The set of normal proximal to some closed set $\mathcal{E} \subset \mathbb{R}^d$ at $e \in \mathcal{E}$ is denoted by $NP_{\mathcal{E}}(e) \subset \mathbb{R}^d$ and is defined by:*

$$NP_{\mathcal{E}}(e) := \Big\{p \in \mathbb{R}^d,\ d_{\mathcal{E}}(e+p) = \|p\|\Big\} = \Big\{p \in \mathbb{R}^d,\ B\big(e+p, \|p\|\big) \cap \mathcal{E} = \emptyset\Big\},$$

where $B\big(e+p, \|p\|\big)$ is the open ball of center $e+p$ and radius $\|p\|$.

The equivalent definition of a B-set, which is closely related to the notion of discriminant set in differential games, is given by the following lemma whose proof is immediate and omitted.

Lemma 1.2 *A set $\mathcal{E}$ is a B-set if and only if:*

$$\forall\, e \in \mathcal{E},\ \forall p \in NP_{\mathcal{E}}(e), \qquad \min_{x \in \Delta(\mathcal{A})} \max_{y \in \Delta(\mathcal{B})} \langle p, g(x,y) - e\rangle \leq 0. \tag{4}$$

Interesting results on a slightly different (but equivalent as we shall see) notion of approachability that can be found in the literature can be easily derived from this alternative definition of B-set.

Definition 1.4 *Given $\varepsilon > 0$, a closed set $\mathcal{E} \subset \mathbb{R}^d$ is ε-approachable by the player if he has a strategy σ_ε ensuring that, after some stage $N_\varepsilon \in \mathbb{N}$, no matter the strategy τ of Nature,*

$$\sup_{n \geq N_\varepsilon} \mathbb{E}_{\sigma_\varepsilon,\tau}\Big(d_{\mathcal{E}}(\overline{g}_n)\Big) \leq \varepsilon \quad \textit{and} \quad \mathbb{P}_{\sigma_\varepsilon,\tau}\left(\sup_{n \geq N_\varepsilon} d_{\mathcal{E}}(\overline{g}_n) \geq \varepsilon\right) \leq \varepsilon\,. \tag{5}$$

And a set $\mathcal{E}$ is 0-approachable if it is ε-approachable for every $\varepsilon > 0$.

The difference between approachability and ε-approachability is wether the strategy can depend on ε or not. It is clear that an approachable set is 0-approachable but the converse is not immediate. It is easier to show – following Spinat [73] and thanks to Lemma 1.3 – that a 0-approachable set must contain a B-set and so both notions coincide.

Lemma 1.3 *Let $\{\mathcal{E}_n\}_{n\in\mathbb{N}}$ be a decreasing sequence of compact non-empty 0-approachable sets, then $\mathcal{E}_\infty := \cap_{n\in\mathbb{N}}\mathcal{E}_n$ is also a compact non-empty 0-approachable set.*

Proof: One just has to notice that, for every $\varepsilon > 0$, the $\varepsilon/2$ neighborhood of $\mathcal{E}_\infty$ is included in some $\mathcal{E}_n$ which is $\varepsilon/2$-approachable. And an $\varepsilon/2$-approachability strategy of $\mathcal{E}_n$ will ε-approach $\mathcal{E}$. ■

Lemma 1.3 is not trivially true for approachability[1]. Indeed, one must find an approachability strategy that is independent of ε and a simple concatenation of those σ_ε might not work (except in the specific case of convex sets).

[1] S. Mannor pointed out this interesting property.

Proposition 1.4 *If a closed set $\mathcal{E}$ is 0-approachable, it contains a B-set.*

We only provide a sketch of the proof, complete details can be found in Spinat [73].

Proof: Consider the family of every compact subset of $\mathcal{E}$ that are 0-approachable. It is a non-empty family, ordered by inclusion and, because of Lemma 1.3, every fully ordered subset has a minorant (the intersection of all elements of this subset) which belongs to this family. Thus Zorn's lemma yield that a minimal element $\mathcal{E}_\infty$ exists and we claim that $\mathcal{E}_\infty$ is a B-set.

Indeed, assume the converse: condition (4) does not hold for some $e \in \mathcal{E}_\infty$ and some proximal normal $p \in NP_{\mathcal{E}_\infty}(e)$. So there exists $y_0 \in \Delta(\mathcal{B})$ such that

$$0 < \min_{x\in\Delta(\mathcal{A})} \max_{y\in\Delta(\mathcal{B})} \big\langle p, g(x,y)-e\big\rangle = \max_{y\in\Delta(\mathcal{B})} \min_{x\in\Delta(\mathcal{A})} \big\langle p, g(x,y)-e\big\rangle =: \min_{x\in\Delta(\mathcal{A})} \big\langle p, g(x,y_0)-e\big\rangle.$$

In particular, Definition 1.3 of proximal normals implies that, at least for some small $\lambda \in (0,1)$, $(1-\lambda)e+\lambda g(x,y_0)$ belongs, for every $x \in \Delta(\mathcal{A})$, to $B\left(e+p, \|p\|\right)$. Therefore,

$$\exists\, \delta > 0,\ \forall x \in \Delta(\mathcal{A}), \quad d_{\mathcal{E}_\infty}\Big((1-\lambda)e + \lambda g(x,y_0)\Big) \geq \delta\,. \tag{6}$$

By continuity, Equation (6) holds (up to $\delta/2$ instead of δ) on a small open neighborhood V of e. We shall prove that this implies that $\mathcal{E}_\infty \backslash V$ is still 0-approachable ; it is a contradiction with the minimality of $\mathcal{E}_\infty$ which must therefore be a B-set.

Assume that at some stage $n \in \mathbb{N}$, $\overline{g}_n$ belongs to V and that Nature plays repeatedly accordingly to y_0 after. Then if n is large enough, there exists some large $m \in \mathbb{N}$ such that $\frac{m}{n+m}$ and $\overline{g}_{n+m}$ are, respectively and with arbitrarily high probability, arbitrarily close to λ and to some $(1-\lambda)\overline{g}_n + \lambda g(x,y_0)$, which is at $\delta/2$ from $\mathcal{E}_\infty$.

Consider a $\delta/4$-approachability strategy of $\mathcal{E}_\infty$ denoted by $\sigma_{\delta/4}$. For some large $N \in \mathbb{N}$ independent of τ, the $\mathbb{P}_{\sigma_{\delta/4},\tau}$-probability that $\overline{g}_n$ belongs to V for some $n \geq N$ must therefore be smaller than $\delta/4$. In particular, this implies that $\overline{g}_n$ stays within δ of $\mathcal{E}_\infty \backslash V$ with probability greater than $1-\delta$. Thus, for every $\delta > 0$, there exists a δ-approachability strategy of $\mathcal{E}\infty\backslash V$. ■

A direct consequence of Theorem 1.1 and Proposition 1.4 is the characterization of approachable sets.

Theorem 1.2 *A closed set $\mathcal{E}$ is approachable if and only if it contains a B-set.*

1.2 Specific case of convex sets

In the specific case of convex sets, there exists a *dual and complete characterization* of approachability and excludability due to Blackwell [9]. It is somehow a consequence of the fact that, for any z in some closed and convex set $\mathcal{C} \subset \mathbb{R}^d$ one has:

$$NP_{\mathcal{C}}(z) = \Big\{p \in \mathbb{R}^d \text{ s.t. } \langle p, c-z\rangle \leq 0,\ \forall c \in \mathcal{C}\Big\}\,; \tag{7}$$

in particular this implies that $NP_{\mathcal{C}}(z)$ is a cone, referred to as the normal cone.

1.2.1 Complete characterization of approachable convex set.

Theorem 1.3 *A closed and convex set $\mathcal{C} \subset \mathbb{R}^d$ is approachable by the player if and only if:*

$$\forall\, y \in \Delta(\mathcal{B}),\ \exists\, x \in \Delta(\mathcal{A}), \quad g(x,y) \in \mathcal{C}. \tag{8}$$

And a convex set is either approachable by the player or excludable by Nature.

Proof: Let $\mathcal{C} \subset \mathbb{R}^d$ be a convex set and $p \in \mathbb{R}^d$ be a normal proximal of $\mathcal{C}$ at some $z \in \mathcal{C}$. Because of Property (7), Condition (8) can be immediately rewritten into

$$\max_{y \in \Delta(\mathcal{B})} \min_{x \in \Delta(\mathcal{A})} \langle p, g(x,y) - z \rangle \leq 0.$$

The mapping $(x,y) \mapsto \langle p, g(x,y) - z \rangle$ is linear in both of its argument, so von Neumann minmax theorem implies that operator min and max can be switched, i.e.,

$$\min_{x \in \Delta(\mathcal{A})} \max_{y \in \Delta(\mathcal{B})} \langle p, g(x,y) - z \rangle = \max_{y \in \Delta(\mathcal{B})} \min_{x \in \Delta(\mathcal{A})} \langle p, g(x,y) - z \rangle \leq 0, \tag{9}$$

thus $\mathcal{C}$ is a B-set and is approachable by the player.

On the contrary, if Condition (8) is not satisfied, there exists some $y_0 \in \Delta(\mathcal{B})$ such that $g(x, y_0) \notin \mathcal{C}$ for every $x \in \Delta(\mathcal{A})$. By continuity, there exists $\delta > 0$ such that $d_{\mathcal{C}}(g(x, y_0)) \geq \delta$. If Nature plays repeatedly accordingly to y_0, then the law of large numbers implies that $\overline{g}_n$ converges uniformly to the set of $\{g(x, y_0),\ x \in \Delta(\mathcal{A})\}$ which is included in the complement of $\mathcal{C}^\delta$. So $\mathcal{C}$ is excludable by Nature and, of course, is not approachable by the player. ■

Proof of Theorem 1.3 relies on the Hilbertian structure of $\mathbb{R}^d$. However, using different arguments, it can be generalized to any normed space, see Theorem 1.7.

Remark 1.1 *In the specific case of a convex set, Blackwell strategy at stage $n+1 \in \mathbb{N}$ can be decomposed as follows:*

i) Given $\overline{g}_n \in \mathbb{R}^d$, compute its projection $\Pi_{\mathcal{C}}(\overline{g}_n)$ on the closed and convex set $\mathcal{C}$;

ii) Solve the projected zero-sum game defined by Equation (9), i.e., find $x_{n+1} \in \Delta(\mathcal{A})$ that minimizes this problem and choose a_{n+1} accordingly to it.

These steps ensure that $x_{n+1} = x(\overline{g}_n)$ as introduced in Definition 1.2. So Blackwell strategy reduces to a projection onto a convex set and the resolution of some linear program (solving a zero-sum game can be reduced to the latter, see Sorin [70], appendix A).

On the other hand, checking wether a convex set is approachable or not, i.e., if it satisfies Condition (8) (or equivalently the more complicate Condition (2)) is NP-hard, even with $\mathcal{C} = \{0\}$. Mannor & Tsilikis [53] has indeed reduced this to the 3-SAT problem.

In the compact case where action set are $\mathcal{X} \subset \mathbb{R}^A$ and $\mathcal{U} \subset \left([0,1]^d\right)^A$, a closed convex $\mathcal{C} \subset \mathbb{R}^d$ is approachable if and only if

$$\forall\, U \in \mathcal{U},\ \exists\, x \in \Delta(\mathcal{A}),\ x.U \in \mathcal{C}.$$

1.2.2 Sharper high probability bounds

In this section, we use the convexity of $\mathcal{C}$ to exhibit high probability bounds improving Corollary 1.1.

Corollary 1.5 *If $\mathcal{C} \subset \mathbb{R}^d$ is a closed and convex approachable set, Blackwell strategy ensures that for every $\eta > 0$ and against any strategy τ of Nature :*

$$\mathbb{P}_{\sigma,\tau}\left(\sup_{m\geq n} d_{\mathcal{C}}(\overline{g}_m) - \frac{2\|g\|_\infty}{\sqrt{m}} \geq \eta\right) \leq 4\exp\Big(-\frac{\eta^2 n}{32\|g\|_\infty^2}\Big). \tag{10}$$

Proof: Distance to a convex set is Lipschitz and convex, so

$$\begin{aligned}\left\{\sup_{m\geq n} d_{\mathcal{C}}(\overline{g}_m) - \frac{2\|g\|_\infty}{\sqrt{m}} \geq \eta\right\} &\subset \left\{\sup_{m\geq n} d_{\mathcal{C}}\Big(\mathbb{E}\big[\overline{g}_m\big]\Big) + \Big\|\overline{g}_m - \mathbb{E}\big[\overline{g}_m\big]\Big\| - \frac{2\|g\|_\infty}{\sqrt{m}} \geq \eta\right\}\\ &\subset \left\{\sup_{m\geq n} \mathbb{E}\Big[d_{\mathcal{C}}\big(\overline{g}_m\big)\Big] + \Big\|\overline{g}_m - \mathbb{E}\big[\overline{g}_m\big]\Big\| - \frac{2\|g\|_\infty}{\sqrt{m}} \geq \eta\right\}\\ &\subset \left\{\sup_{m\geq n} \Big\|\overline{g}_m - \mathbb{E}\big[\overline{g}_m\big]\Big\| \geq \eta\right\},\end{aligned}$$

where the third inclusion is a consequence of the rate of convergence of Blackwell strategy. We conclude using Lemma 5.3. ■

This result must be put in perspective with Corollary 1.1 that states that, for any arbitrary approachable set $\mathcal{E}$ and every $\eta > 0$, $\mathbb{P}_{\sigma,\tau}\left(\sup_{m\geq n} d_{\mathcal{E}}(\overline{g}_m) \geq \eta\right) \leq (\eta^2 n/8\|g\|_\infty^2)^{-1}$.

1.2.3 Biased approachability

We assume in this section that the closed and convex set $\mathcal{C} \subset \mathbb{R}^d$ is not approachable by the player. In that case, the natural extension of Blackwell strategy would be defined by $\sigma(h^n) = x_{n+1} \in \Delta(\mathcal{A})$, where x_n is optimal in the projected zero-sum game with payoffs

$$\langle g(x,y) - \Pi_{\mathcal{C}}(\overline{g}_n), \overline{g}_n - \Pi_{\mathcal{C}}(\overline{g}_n)\rangle.$$

Corollary 1.6 *Even if a closed and convex set $\mathcal{C} \subset \mathbb{R}^d$ is not approachable by the player, Blackwell's strategy σ ensures that*

$$\mathbb{E}_{\sigma,\tau}\Big[d_{\mathcal{C}}\left(\overline{g}_n\right) - \delta\Big] \leq \sqrt{\frac{\kappa}{n}} + \frac{\delta}{\sqrt{n}}, \text{ where } \delta = \sup_{y\in\Delta(\mathcal{B})} \inf_{x\in\Delta(\mathcal{A})} d_{\mathcal{C}}(g(x,y).$$

Proof: We only need to prove that σ is in fact exactly Blackwell's approachability strategy of the closure of $\mathcal{C}^\delta$ (the δ-neighborhood of $\mathcal{C}$) which is by definition and Condition (8) approachable. This is simply due to the fact that:

$$\forall z \notin \mathcal{C}^\delta, \quad \Pi_{\mathcal{C}^\delta}(z) = \Pi_{\mathcal{C}}(z) + \delta\frac{z - \Pi_{\mathcal{C}}(z)}{\|z - \Pi_{\mathcal{C}}(z)\|}.$$

Indeed, $\mathcal{C}^\delta = \mathcal{C}+\delta B(0,1)$, so $\Pi_{\mathcal{C}^\delta}(z)$ minimizes $\|z-(c+\delta e)\|^2 = \|z-c\|^2 - 2\delta\langle z-c, e\rangle + \delta^2$ over $(c,e) \in \mathcal{C}\times B(0,1)$. And necessarily, one must have $e = (z-c)/\|z-c\|$ and $c = \Pi_{\mathcal{C}}(z)$.

The results follows from the fact that $d_{\mathcal{C}^\delta}(z) \leq d_{\mathcal{C}}(z) + \delta$ and $\|\mathcal{C}^\delta\| \leq \|\mathcal{C}\| + \delta$. ■

The key ingredient of Corollary 1.6 is not the rates of convergence (which are a direct consequence of the fact that $\mathcal{C}^\delta$ is approachable), but the fact that it does not require the computation of δ and $\mathcal{C}^\delta$ (we recall that determining if a convex set is NP-hard, thus determining the smallest approachable extension is even more complex). Notice that if $\mathcal{C}$ is approachable, rates of Condition 8 and Corollary 1.6 and of Theorem 1.1 match.

This result has to be put in perspective with the following proposition that also deals with biased approachability, yet on different level.

Proposition 1.7 *Assume that player and Nature strategies generates a sequence of payoffs such that, at every stage n,*

$$\langle \overline{g}_n - \pi_{\mathcal{E}}(\overline{g}_n), \mathbb{E}_{\sigma,\tau}\left[g_{n+1}\middle|h^n\right] - \pi_{\mathcal{E}}(\overline{g}_n)\rangle \leq \varepsilon_n,$$

for some sequence ε_n. Then

$$\mathbb{E}_{\sigma,\tau}\left[d_{\mathcal{E}}(\overline{g}_n)^2\right] \leq \frac{\kappa}{n} + \frac{2\sum_{m=1}^n m\varepsilon_m}{(n+1)^2} \text{ and}$$

$$\mathbb{P}_{\sigma,\tau}\left(\sup_{m\geq n} d_{\mathcal{E}}(\overline{g}_m) \geq \eta\right) \leq \frac{2\kappa + 2\sum_{m=n}^\infty \frac{\varepsilon_m}{m}}{\eta^2 n}.$$

In particular, if ε_n converges to 0, then $\overline{g}_n$ converges in expectation to $\mathcal{E}$; the convergence is almost sure as soon as $\sum_{n\in\mathbb{N}} \frac{\varepsilon_n}{n} < \infty$.

Proof: The proof is identical to the one of Corollary 1.1. ■

Actually, the result is stated for arbitrary sets and holds for non-deterministic sequences of ε_n. On the other hand, for convex sets, concentration inequalities introduced in the previous section show that

$$\mathbb{P}_{\sigma,\tau}\Big\{\sup_{m\geq n} d^2_{\mathcal{C}}(\overline{g}_m) - \frac{2\sum_{s=1}^m s\varepsilon_s}{(m+1)^2} \geq \eta\Big\} \leq 3\exp\Big(-\frac{M_{\eta,n}}{2}\Big)$$

thus $\overline{g}_n$ converges almost surely to $\mathcal{C}$ as soon as ε_n goes (in expectation) to 0.

1.3 Generalizations and extensions

1.3.1 Deterministic approachability and procedures in law

As mentioned in Section 1.1.1, Blackwell's approachability strategy does not use the fact that actions chosen by Nature are observed, as it is only required to observe the sequence

of payoffs. In fact, it is not even required that the random variable $g_n = g(a_n, b_n)$ is perfectly observed.

Indeed, denote by γ_n the observation made after stage n, and assume it is equal to either $g(x_n, b_n)$ or $g(x_n, y_n)$, where x_n and y_n are mixed action of stage n (i.e., laws of a_n or b_n). Blackwell's strategy applied to the sequence of γ_n ensures that the sequence of deterministic averages $\overline{\gamma}_n$ converges to $\mathcal{E}$, uniformly with respect to Nature's strategy.

To conclude that this describes an approachability strategy, it remains to notice that $d_{\mathcal{E}}(\overline{g}_n) \leq d_{\mathcal{E}}(\overline{\gamma}_n) + \left\|\overline{g}_n - \overline{\gamma}_n\right\|$ and that the norm of $\overline{g}_n - \overline{\gamma}_n$ converges almost surely to zero, because it is an average of bounded martingale differences (using classical concentration arguments to get rates of convergence independent of strategies).

1.3.2 Approachability in infinite dimension spaces

We assume in this section that g no longer takes value in some Euclidian space. Formally, there exists a probability space $(\Omega, \mu, \mathcal{F})$ such that, for every $a \in \mathcal{A}$ and $b \in \mathcal{B}$, $g(a, b) \in L_2(\Omega, \mu, \mathcal{F})$ – g is extended to $\Delta(\mathcal{A}) \times \Delta(\mathcal{B})$ as before. The finite case can be easily embedded into this framework by defining, $\Omega = \{1, \ldots, d\}$ and $\mu = \frac{1}{d}\sum_{k=1}^{d} \delta_k$.

In this context, notions of approachability slightly differ, as the uniform convergence with respect to Nature's strategy is not required:

Definition 1.5 *A closed set $\mathcal{E} \subset \mathbb{R}^d$ is approachable by the player if he has a strategy σ ensuring that, no matter the strategy τ of Nature, $\overline{g}_n$ converges μ-almost surely to $\mathcal{E}$, for $\mathbb{P}_{\sigma,\tau}$-almost every histories.*

A set $\mathcal{E}$ is excludable by Nature if she can approach the complement of $\mathcal{E}^\delta$ for some $\delta > 0$.

Lehrer [42] has proved that the natural inner product of $L_2(\Omega, \mu, \mathcal{F})$ allows to extend the definition of B-sets and Blackwell's characterization of approachable convex sets still holds (Equation (8), in the previous section).

Theorem 1.4 *A closed convex set $\mathcal{C}$ is approachable if and only if*

$$\forall\, y \in \Delta(\mathcal{B}),\ \exists\, x \in \Delta(\mathcal{A}), \quad g(x, y) \in \mathcal{C}. \tag{11}$$

The proof relies on the following geometric principle, adapted from Lehrer [42].

Lemma 1.8 *Let $\mathcal{C}$ be a closed convex subset of $L_2(\Omega, \mu, \mathcal{F})$. If, for every $n \in \mathbb{N}$, g_n is bounded μ-as by $M \in L_2(\Omega, \mu, \mathcal{F})$ and $\langle \overline{g}_n - \Pi_{\mathcal{C}}(\overline{g}_n), g_{n+1} - \Pi_{\mathcal{C}}(\overline{g}_n)\rangle \leq 0$, then $\overline{g}_n$ converges μ-as to $\mathcal{C}$.*

Proof: Let us denote $f_n = \overline{g}_n - \Pi_{\mathcal{C}}(\overline{g}_n)$. The finite dimensional arguments of the proof of Corollary 1.1 imply that $\left\|f_n\right\| \leq 2\|M\|/\sqrt{n}$ thus $\overline{g}_n$ converges in probability to $\mathcal{C}$.

The almost sure convergence is a consequence of the fact that

$$\left\|f_{n+1} - f_n\right\| \leq \left\|\left(\overline{g}_{n+1} - \overline{g}_n\right) - \left(\Pi_{\mathcal{C}}(\overline{g}_{n+1}) - \Pi_{\mathcal{C}}(\overline{g}_n)\right)\right\| \leq 2\left\|\overline{g}_{n+1} - \overline{g}_n\right\| \leq \frac{4\|M\|}{n+1}$$

so f_n has small increments and we conclude using the technical Lemma 5.4. ■

Convexity of $\mathcal{C}$ is only used to get a Lipschitzian projection.

Proof of Theorem 1.4: Every arguments behind the proof of Theorem 1.3 hold in $L_2(\Omega, \mu, \mathcal{F})$. Therefore, a closed convex set satisfying Blackwell condition remains a B-set with respect to the natural inner product of $L_2(\Omega, \mu, \mathcal{F})$.

Assume that $\mathcal{C}$ is a B-set and consider Blackwell's strategy, denoted as usual by σ (and τ is Nature's strategy). Let $\mu \otimes \mathbb{P}_{\sigma,\tau}$ be the product measure on $\Omega \times \mathcal{H}$ on which we define the random variable $\widetilde{g}_n$ by $\widetilde{g}_n[\omega, h] = \widetilde{g}_n(a_n, b_n)[\omega]$ where (a_n, b_n) is the pair of actions played at stage n accordingly to h. Since $\mathcal{A}$ and $\mathcal{B}$ are finite, g_n and $\widetilde{g}_n$ are uniformly bounded and the sequence $\widetilde{g}_n$ satisfies the geometric principle.

As a consequence, $\widetilde{g}_n$ converges $\mu \otimes \mathbb{P}_{\sigma,\tau}$-as to $\mathcal{C}$ which is therefore approachable. ■

1.3.3 Approachability with infinite action space – non-linear approachability

It is also possible to generalize the previous results when actions spaces are not necessarily finite but two subsets of a given topological space, denoted by $\mathcal{X}$ and $\mathcal{Y}$. Payoff mapping g is now a function from $\mathcal{X} \times \mathcal{Y}$ into $L_2(\Omega, \mu, \mathcal{F})$. In particular, it is not required in this section that g is linear in each of its variable.

Theorem 1.5 *Assume the following regularity assumptions on g:*

a) there exists $M \in L_2(\Omega, \mu, \mathcal{F})$ such that $g(x, y) \leq M$, μ-as, for every $x, y \in \mathcal{X} \times \mathcal{Y}$;

b) for every $y \in \mathcal{Y}$, $\mathcal{G}(y)$, the closure of $\{g(x, y), x \in \mathcal{X}\}$, is a compact and convex set.

c) for every $u \in L_2(\Omega, \mu, \mathcal{F})$ such that $\sup_{c \in \mathcal{C}} \langle c, u \rangle < +\infty$, the zero-sum game with payoffs defined by $\langle u, g(x, y) \rangle$ has a value.

Then it holds that

i) Blackwell's characterization of convex approachable set holds :

$$\mathcal{C} \text{ is approachable (in pure strategy) if and only if } \forall\, y \in \mathcal{Y},\ \mathcal{G}(y) \cap \mathcal{C} \neq \emptyset;$$

ii) $\mathcal{C}$ is approachable if and only if for every $z \in L_2(\Omega, \mu, \mathcal{F})$:

$$\sup_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} \langle z - \Pi_{\mathcal{C}}(z), g(x, y) - \Pi_{\mathcal{C}}(z) \rangle \leq 0.$$

iii) If there exists y_0 such that $\mathcal{G}(y_0) \cap \mathcal{C} = \emptyset$, then $\mathcal{C}$ is excludable by Nature;

Proof: The deterministic approachability strategy associated with Blackwell's characterization is defined as follows. Denote as before by $\overline{g}_n \in L_2(\Omega, \mu, \mathcal{F})$ the average payoff up to stage n. Since $\Pi_{\mathcal{C}}$ is the projection onto a convex set, one has

$$\sup_{c \in \mathcal{C}} \langle c, \overline{g}_n - \Pi_{\mathcal{C}}(\overline{g}_n), \rangle \leq \langle \Pi_{\mathcal{C}}(\overline{g}_n), \overline{g}_n - \Pi_{\mathcal{C}}(\overline{g}_n) \rangle < +\infty.$$

Assumption *c)* ensures that the game with payoff $\langle g(x,y) - \Pi_{\mathcal{C}}(\overline{g}_n), \overline{g}_n - \Pi_{\mathcal{C}}(\overline{g}_n) \rangle$ has a value which is, using Blackwell characterization, less or equal than 0. The approachability strategy consists in playing $x_n \in \mathcal{X}$, any 2^{-n}-optimal strategy of the latter game, i.e.,

$$\sup_{y \in \mathcal{Y}} \langle g(x_n, y) - \Pi_{\mathcal{C}}(\overline{g}_n), \overline{g}_n - \Pi_{\mathcal{C}}(\overline{g}_n) \rangle \leq \frac{1}{2^n}.$$

The fact that this describes an approachability strategy follows from arguments used in the proof of Corollary 1.1 and technical Lemma 5.4.

Assume that Blackwell's condition does not hold, i.e., there exists y_0 such that $\mathcal{G}(y_0) \cap \mathcal{C} = \emptyset$; Nature, by playing repeatedly y_0, can ensure that $\overline{g}_n$ belongs to $\mathcal{G}(y_0)$. The intersection between the closed convex set $\mathcal{C}$ and the compact convex set $\mathcal{G}(y_0)$ is empty, so they can be strictly separated. Since Nature can approach $\mathcal{G}(y_0)$, $\mathcal{C}$ is excludable, thus not approachable. ■

Assumption *b)* is required to get point *iii)*. Second conditions of *i)* and *ii)* are sufficient for approachability (but not necessary).

When actions sets $\mathcal{A}$ and $\mathcal{B}$ are finite, the projected game with payoff $\langle g(a,b), u \rangle$ typically does not have a value for some $u \in L_2(\Omega, \mu, \mathcal{F})$; so we considered instead mixed actions and strategies. This can be generalized when actions space are two measurable sets $(\mathcal{A}, A)$ and $(\mathcal{B}, B)$, using the same tools as for procedures in law, see Section 1.3.1.

Denote by $\mathcal{X}$ and $\mathcal{Y}$ the sets of probability distributions onto $(\mathcal{A}, A)$ and $(\mathcal{B}, B)$, embedded with the weak-$\star$ topology ; the mapping g is extended to $\mathcal{X} \times \mathcal{Y}$ multi-linearly as usual. Then, under mild assumptions (for example if $\mathcal{A}$ and $\mathcal{B}$ are compact and g is continuous, see e.g. Sorin [70]), the projected game with payoff $\langle g(x,y), u \rangle$ has a value (at least for every u such that $\sup_{c \in \mathcal{C}} \langle c, u \rangle < +\infty$). So $\mathcal{C}$ is approachable with respect to action sets $\mathcal{X}$ and $\mathcal{Y}$. In particular, there exists an approachability strategy such that the averages of observed payoffs $\gamma_n = g(x_n, b_n)$, where $x_n \in \mathcal{X}$ is the action dictated to be played at stage n, converge to $\mathcal{C}$ – and the rate of convergence is $O\left(1/\sqrt{n}\right)$.

Similarly to Section 1.3.1, this is an approachability strategy of $\mathcal{C}$ since $\overline{g}_n - \overline{\gamma}_n$ is again an average of bounded Martingale differences, and concentration inequalities of sums of bounded martingales differences in any Hilbert spaces, see e.g. Chen & White [15], imply that, in expectation and with great probability, $\left\| \overline{g}_n - \overline{\gamma}_n \right\| \leq O\left(1/\sqrt{n}\right)$. Almost sure convergence is again a consequence of Lemma 5.4.

1.3.4 Approachability with activation

This section is concerned with the case where only a fragment of all coordinates of the payoff vector (belonging to $L_2(\Omega, \mu, \mathcal{F})$) are active at each stage. Formally, there exists a

mapping $\mathcal{X} : H \to L_2(\Omega, \mu, \mathcal{F})$ such that, after any finite history $h^n = (a_1, b_1, \ldots, a_n, b_n)$, $\mathcal{X}[h^n] \in L_2(\Omega, \mu, \mathcal{F})$ has value in $\{0, 1\}$ and only the coordinates $\omega \in \Omega$ with $\mathcal{X}[h^n](\omega) = 1$ are active. In particular, wether a coordinate is active at a stage might depend on choices of actions of this specific stage. We also assume that $\mathcal{X}[h^n]$ increases μ-almost surely to infinity, no matter the pair of strategies.

In this framework, we denote *tilted* averages of payoffs by

$$\overline{g}_{\mathcal{X},n} := \frac{\sum_{m=1}^n \mathcal{X}[h^m] g(a_m, b_m)}{\sum_{m=1}^n \mathcal{X}[h^m]} \quad \text{(with the convention that } \frac{0}{0} = 0\text{)}.$$

A set $\mathcal{E} \subset L_2(\Omega, \mu, \mathcal{F})$ is approachable if the player has a strategy σ such that, for any strategy τ of Nature, the sequence $\overline{g}_{\mathcal{X},n} - \Pi_{\mathcal{E}}(\overline{g}_{\mathcal{X},n})$ converges to zero μ-almost surely, for $\mathbb{P}_{\sigma,\tau}$-almost all infinite histories.

We will only focus on *product sets*, that can be described by

$$\mathcal{C} = \Big\{ f \in L_2(\Omega, \mu, \mathcal{F}) \,,\ f_0 \leq f \text{ on } \Omega_0 \text{ and } f \leq f_1 \text{ on } \Omega_1 \Big\}$$

where Ω_0 and Ω_1 are two measurable subsets of Ω and $f_0, f_1 \in L_2(\Omega, \mu, \mathcal{F})$. The following theorem shows that, in this specific framework, a notion of *tilted* B-set is sufficient for approachability

Theorem 1.6 *Let $\mathcal{C} \subset L_2(\Omega, \mu, \mathcal{F})$ be a product set. Then any strategy σ such that, for any strategy τ of Nature, and for $\mathbb{P}_{\sigma,\tau}$-almost every infinite history,*

$$\left\langle \frac{\mathcal{X}[h^{n+1}]}{\sum_{m=1}^n \mathcal{X}[h^m]} \Big(\overline{g}_{\mathcal{X},n} - \Pi_{\mathcal{C}}(\overline{g}_{\mathcal{X},n}) \Big), g(x_{n+1}, y_{n+1}) - \Pi_{\mathcal{C}}(\overline{g}_{\mathcal{X},n}) \right\rangle \leq 0,$$

where $x_{n+1} = \sigma(h^n)$ and $y_{n+1} = \tau(h^n)$, is an approachability strategy of $\mathcal{C}$.

The proof is similar to the one of Theorem 1.4, except that Lemma 5.5 is used instead of Lemma 1.8, so it is omitted.

The next proposition shows that approachability with activation of a product set $\mathcal{C} = \prod_{k=1}^d \mathcal{C}^k \subset \mathbb{R}^d$ in Euclidian spaces can actually be reduced to usual approachability. The only condition is that activation at stage n depends only of current actions (i.e., $\mathcal{X}[h^n] = \mathcal{X}(a_n, b_n)$ where $\mathcal{X}(a, b)$ might be a random variable); we also assume, without loss of generality, that the origin belongs to $\mathcal{C}$ and even that $\mathcal{C} = \prod_{k=1}^d [0, b^k] \subset \mathbb{R}^d$.

Proposition 1.9 *A product set $\mathcal{C} \subset \mathbb{R}^d$ is approachable with activation depending only on current actions if and only if the following convex set*

$$\widetilde{\mathcal{C}} := \Big\{ (z, \omega) \in \mathbb{R}^d \times \mathbb{R}^d_+ ; \Big(\frac{z^k}{\omega^k} \Big)_{k \in \{1, \ldots, d\}} \in \mathcal{C} \Big\} \text{ with the convention that } \frac{0}{0} = 0$$

is approachable in the game with payoffs defined by

$$\widetilde{g}_{\mathcal{X}}(a, b) = \Big(g_{\mathcal{X}}(a, b), \mathcal{X}(a, b) \Big) \in \mathbb{R}^d \times \mathbb{R}^d_+, \quad \text{where } g_{\mathcal{X}}(a, b) = \big(\mathcal{X}^k(a, b) g^k(a, b) \big)_{k \in \{1, \ldots, d\}}.$$

Moreover, there exists a strategy such that, in expectation,

$$d_{\mathcal{C}}\left(\overline{g}_{\mathcal{X},n}\right) \leq \frac{4\|g\|_\infty^2}{\sqrt{n}} \frac{1}{\underline{\mathcal{X}}_n}, \textit{ with } \underline{\mathcal{X}}_n = \inf\left\{\overline{\mathcal{X}}_n^k \quad s.t. \ \overline{\mathcal{X}}_n^k := \frac{\sum_{m=1}^n \mathcal{X}^k(a_m,b_m)}{n} > 0\right\}.$$

Proof: Consider any fixed $(z,\omega) \in \mathbb{R}^d \times \mathbb{R}^d$; we can always assume that every coordinates of ω are non equal to 0. Indeed, since $\mathcal{C}$ is a product set, $d_{\mathcal{C}}\left((z^k/\omega^k)_k\right) = d_{\mathcal{C}}\left((\widehat{z}^k/\widehat{\omega}^k)_k\right)$ where $(\widehat{z}^k,\widehat{\omega}^k) = (z^k,\omega^k)$ if $\omega^k \neq 0$ and $(\widehat{z}^k,\omega^k) = (c^k,1)$ with c^k arbitrarily chosen in $\mathcal{C}^k$ if $\omega^k = 0$.

Define $(z_e,\omega_e) \in \Pi_{\widetilde{\mathcal{C}}}\big(z,\omega\big)$ and $\underline{\omega}$ the smallest coordinate of ω. Since $\mathcal{C} = \prod_{k=1}^d [0,b^k]$, then $\widetilde{\mathcal{C}} = \prod_{k=1}^d \left\{(z^k,\omega^k); 0 \leq z^k \leq b^k\omega^k\right\}$, thus necessarily

$$\omega^k \leq \omega_e^k \leq \omega^k \frac{\|g\|_\infty + 1}{b^k+1} \leq (\|g\|_\infty + 1)\omega^k.$$

As a consequence,

$$d_{\mathcal{C}}\left(\frac{z}{\omega}\right) \leq \left\|\frac{z}{\omega} - \frac{z_e}{\omega_e}\right\| \leq \frac{1}{\underline{\omega}}\|z - z_e\| + \|z\| \sup_k \left|\frac{1}{\omega^k} - \frac{1}{\omega_e^k}\right| \leq 2\|g\|_\infty \frac{d_{\widetilde{\mathcal{C}}}(z,\omega)}{\underline{\omega}}.$$

Reciprocally,

$$d_{\widetilde{\mathcal{C}}}(z,\omega) \leq \left\|(z,\omega) - \left(\left(\omega^k\, \Pi_{\mathcal{C}}^k\left(\frac{z}{\omega}\right)\right)_k, \omega\right)\right\| \leq \left\|\frac{z}{\omega} - \Pi_{\mathcal{C}}\left(\frac{z}{\omega}\right)\right\| \leq d_{\mathcal{C}}\left(\frac{z}{\omega}\right).$$

Finally, if $\mathcal{C}$ is a product set containing 0, then $\widetilde{\mathcal{C}}$ is a convex cone. The result is a consequence of Blackwell's characterization of approachable sets. ■

Assuming that the origin belongs to the product set $\mathcal{C}$ is of course non-restrictive, one can always choose to transform the origin into any point. Moreover, in some cases, product set property can be relaxed. For instance, if there exists two coordinates ℓ and ℓ' that are always active together, i.e., if $\mathcal{X}(a,b)^\ell = \mathcal{X}(a,b)^{\ell'}$ for every pair (a,b), then the results holds if $\mathcal{C} := \prod_{k \notin \{\ell,\ell'\}} \mathcal{C}^k \times \mathcal{C}^{\ell,\ell'}$ where the convex set $\mathcal{C}^{\ell,\ell'} \subset \mathbb{R}^2$ does not need to be a product set.

1.3.5 Variable stage duration

Cesaro averages of payoffs are considered in the usual definition of approachability. In this section, we make the implicit assumption that all stages does not have the same weights (when computing averages) or, equivalently, that they do not have the same length duration: payoffs obtained on long stages must have more importance than on short stages. We distinguish two classes of variable and random stage duration: wether they depend or not on the actions chosen.

Assume for the moment that ω_n, the maybe random length (or weight) of the n-th stage, is independent of actions chosen by player and Nature. In this context, σ is an

approachability strategy of a closed set $\mathcal{E}$ if $\overline{g}_{\omega,n} := \sum_{m=1}^n \omega_m \overline{g}_m / \sum_{m=1}^n \omega_m$ converges to $\mathcal{E}$, $\mathbb{P}_{\sigma,\tau}$-almost surely, uniformly with respect to the strategy τ of Nature. It will be convenient to define $\Omega_n = \sum_{m=1}^n \omega_m$.

Proposition 1.10 *Let $\mathcal{E} \subset \mathbb{R}^d$ be a closed B-set. Then Blackwell's strategy applied to the sequence of weighted averages $\overline{g}_{\omega,n}$ ensures that for every $n \in \mathbb{N}$ and $\eta > 0$*

$$\mathbb{E}_{\sigma,\tau}\Big[d_{\mathcal{E}}\left(\overline{g}_{\omega,n}\right)\Big] \leq \sqrt{\frac{\sum_{m=1}^n \omega_m^2}{\Omega_n^2}\kappa} \quad \textit{and}$$

$$\mathbb{P}_{\sigma,\tau}\Big\{\exists m \geq n\,,\ d_{\mathcal{E}}\left(\overline{g}_{\omega,m}\right) \geq \eta\Big\} \leq \left[\frac{\sum_{m=1}^n \omega_m^2}{\Omega_n^2} + \sum_{k=n+1}^{\infty}\left(\frac{\omega_k}{\Omega_k}\right)^2\right]\frac{\kappa}{\eta^2}\ .$$

The proof is absolutely identical with Cesaro averages (when $\omega_n = 1$ for every $n \in \mathbb{N}$) thus omited. In particular, for any polynomial weights, i.e. if $\omega_n = n^\alpha$ with $\alpha > -1$, a B-set is approachable at the rate of convergence of $O\left(1/\sqrt{n}\right)$, which is independent of α – only the constant depends on α, see e.g. Mannor, Perchet & Stoltz [52].

In fact, as we shall see in the following Section 1.4.1, a B-set is approachable as soon as the usual Robbins-Monroe assumptions are satisfied almost surely:

$$\sum_{n\in\mathbb{N}} \frac{\omega_n}{\Omega_n} = +\infty \quad \text{and} \quad \sum_{n\in\mathbb{N}}\left(\frac{\omega_n}{\Omega_n}\right)^2 < \infty\ .$$

We now turn to the case where a stage length might depend on the actions of the player and Nature. For simplicity, we assume that there exists a mapping $\omega : \mathcal{A}\times\mathcal{B} \to [\underline{\omega},\overline{\omega}] \subset (0,1]$ such that $\omega_n := \omega(a_n,b_n)$. Approachability in this framework can be reduced to regular approachability, similarly to what has be done with activation.

Proposition 1.11 *A closed set $\mathcal{E} \subset \mathbb{R}^d$ is approachable with respect to weighted averages if and only if the following cone $\widetilde{\mathcal{E}}$ is approachable with Cesaro averages*

$$\widetilde{\mathcal{E}} = \Big\{(z,\omega) \in \mathbb{R}^d \times [\underline{\omega},\overline{\omega}]\,;\ \frac{z}{\omega} \in \mathcal{E}\Big\}\,.$$

Moreover, if $\mathcal{E}$ is convex then $\widetilde{\mathcal{E}}$ is also convex, thus $\mathcal{E}$ is approachable with respect to weighted averages if and only if

$$\forall\, y \in \Delta(\mathcal{B}),\ \exists\, x \in \Delta(\mathcal{A}),\ \frac{\mathbb{E}_{x,y}\Big[\omega(a,b)g(a,b)\Big]}{\mathbb{E}_{x,y}\Big[\omega(a,b)\Big]} \in \mathcal{E}\ .$$

Proof: Let $(z,\omega) \in \mathbb{R}^d \times [\underline{\omega},\overline{\omega}]$ and $(z_e,\omega_e) \in \Pi_{\widetilde{\mathcal{E}}}\big(z,\omega\big)$, then

$$d_{\mathcal{E}}\left(\frac{z}{\omega}\right) \leq \left\|\frac{z}{\omega} - \frac{z_e}{\omega_e}\right\| \leq \frac{1}{\omega}\left\|z - z_e\right\| + \|z\|\left|\frac{1}{\omega} - \frac{1}{\omega_e}\right| \leq \left(\frac{1}{\underline{\omega}} + \frac{\|g\|_\infty}{\underline{\omega}^2}\right) d_{\widetilde{\mathcal{E}}}(z,\omega).$$

As before, one has reciprocally,

$$d_{\widetilde{\mathcal{E}}}(z,\omega) \leq \left\|(z,\omega) - \left(\omega\,\Pi_{\mathcal{E}}\left(\frac{z}{\omega}\right),\omega\right)\right\| \leq \overline{\omega}\left\|\frac{z}{\omega} - \Pi_{\mathcal{E}}\left(\frac{z}{\omega}\right)\right\| \leq \overline{\omega}\, d_{\mathcal{E}}\left(\frac{z}{\omega}\right) .$$

If $\mathcal{E}$ is convex, then $\widetilde{\mathcal{E}}$ is a convex cone and the characterization of approachable convex set (in this framework due to Mannor and Shimkin [49]) is a simple consequence of Condition (8). ■

1.3.6 Unbounded payoffs and strong law of large numbers

At the end of Section 1.1.1, we noticed that we can assume that Nature choose outcomes U in some given compact set $\mathcal{U} \subset \left(\mathbb{R}^d\right)^A$ such that the player's payoff is, in expectation, $x.U$. The fact that $\mathcal{U}$ is a fixed compact set can be weakened (similarly to Stoltz [74]), and we can assume that U_n belongs to $\mathcal{U}_n \subset \left(\mathbb{R}^d\right)^A$ as long as

$$\sum_{n\in\mathbb{N}} \frac{\left\|U_n\right\|^2}{(n+1)^2} < \infty \quad \text{or even} \quad \sum_{n\in\mathbb{N}} \frac{\mathbb{E}_{\sigma,\tau}\left[\|a_n.U_n\|^2\right]}{(n+1)^2} < \infty ,$$

with a convergence uniform with respect to Nature's strategy.

Indeed, under this assumption, the proof of Theorem 1.1 does not change, i.e., if condition (2) is satisfied at every stage, then the same arguments yield that

$$\mathbb{E}_{\sigma,\tau}\left[\delta_n^2\right] \leq 2\frac{\sum_{m=1}^n \mathbb{E}_{\sigma,\tau}\left[\|a_m.U_m\|^2\right] + \|\mathcal{E}_g\|^2}{n^2}$$

and Z_n is a supermartingale such that

$$\mathbb{E}_{\sigma,\tau}\left[Z_n\right] \leq 4\frac{\|\mathcal{E}_g\|^2}{n} + 2\left(\frac{\sum_{m=1}^n \mathbb{E}_{\sigma,\tau}\left[\|a_m.U_m\|^2\right]}{n^2} + \sum_{m=n+1}^{\infty} \frac{\mathbb{E}_{\sigma,\tau}\left[\|a_m.U_m\|^2\right]}{m^2}\right) .$$

By assumption, every terms goes to zero uniformly with respect to Nature's strategy, hence $\mathcal{E}$ is approachable.

This sheds new lights on approachability theory: it can be seen as a generalization of Kolmogorov strong law of large numbers (see Feller [20], chapter X.7 and Mertens, Sorin & Zamir [55], exercice 4, page 104). Indeed, Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of independent random variable in $\mathbb{R}^d$ and define $v_n := \mathbb{E}\left[\left\|X_n - \mathbb{E}[X_n]\right\|^2\right]$. As soon as $\sum_{n\in\mathbb{N}} v_n/n^2$ is bounded, $\overline{X}_n - \mathbb{E}\left[\overline{X}_n\right]$ converges almost-surely to 0; moreover

$$\mathbb{P}\left\{\exists\, m \geq n; \left\|\overline{X}_m - \mathbb{E}[\overline{X}_m]\right\| \geq \eta\right\} \leq \frac{1}{\eta^2}\left(\frac{\sum_{m=1}^n v_m}{n^2} + \sum_{m=n+1}^{\infty} \frac{v_m}{m^2}\right)$$

or even with an exponential decay (since $\{0\}$ is convex, see Section 1.2.2).

Finally, the approachability bound (in expectation) matches the optimal bound in the law of large number and thus is in some sense optimal. Indeed, if X_n is an i.i.d. sequence such that $X_n = \pm 1$ with probability 1/2, then by denoting $\mathcal{E} = \{0\}$, one has

$$\mathbb{E}\left[\left|\overline{X}_n\right|^2\right] = \mathbb{E}\left[d_{\mathcal{E}}(\overline{X}_n)\right] = \frac{1}{n} = \frac{\kappa}{n}.$$

1.3.7 Bounded memory

Blackwell's approachability strategy does not require to know at each stage the whole sequence of past payoffs, but only the current average. Nonetheless, to update this average either stage number of the complete history must be kept in memory which takes of course an increasing required size of memory. This is why the question of wether it is possible to approach a closed set $\mathcal{E}$ using simpler strategies, for example implementable by a finite automata or with a finite memory, arises.

A strategy σ has a bounded memory of size $M \in \mathbb{N}$ if, for every finite history $h^n \in H_n$, $\sigma(h^n)$ depends only on $\left(a_{n-M+1}, b_{n-M+1}, \ldots, a_n, b_n\right)$, i.e. the last M profiles of actions played. Lehrer & Solan [44, 46] proved that an approachable convex set $\mathcal{C}$ remains approachable by a player if it is restricted to use strategies with a bounded memory of size $M \in \mathbb{N}$; indeed, the average payoff converges to some $O(1/\sqrt{M})$-neighborhood of $\mathcal{C}$.

The basic idea is relatively natural; play Blackwell's strategy on a block of size M, then erase the memory and start over. It is only necessary to encode the beginning (and the end) of a block, but this can be done using $\sqrt{M}$ stages, for examples by playing always the same action and by ensuring that no such sequence appears in the same bloc. The average payoff on each block will be $1/\sqrt{M}$ close to $\mathcal{C}$ which is convex, hence the overall average payoff is also $1/\sqrt{M}$ of $\mathcal{C}$.

On the other hand, Zapechelnyuk [82] considered the strategy with bounded memory directly adapted from Blackwell's, that is defined by $\sigma(h^n) = x\left(\overline{g}_n^M\right)$, where $x(\cdot)$ is given by the definition of a B-set and $\overline{g}_n^M$ is the average payoff on the last M stages. For instance, we are interested by this strategy in the game where payoffs of player (that chooses a row) are given by the following matrix:

	L	R
T	(0,-1)	(0,1)
B	(1,0)	(-1,0)

and $\mathcal{C} = \mathbb{R}^2_-$. For M big enough, there exists a strategy of Nature such that the sequence $\left(\overline{g}_n^M\right)_{n\in\mathbb{N}}$ enters a cycle (of length either $2M$ or $2M+2$). Roughly speaking, this latter is defined by four successive blocks of lengths $M/2$ (or $M/2+1$) where within a block, the same pair of actions (except on at most one stage) is played. And one can show that the order of these actions is (T, L), (T, R), (B, R) et (B, L).

At the end of the blocks (B, R) and (T, L), $\overline{g}_n^M$ is close, respectively to $(-1/2, 1/2)$ or $(1/2, -1/2)$. So it is at a distance of around 1/2 from C, and the sequence $(\overline{g}_n^M)_{n\in\mathbb{N}}$ of averages of payoffs on the M last stages does not converge to $\mathcal{C}$.

However, nothing indicates wether the sequence $\overline{g}_n$ does or does not converge to $\mathcal{C}$ (which is the case in this example).

1.4 Alternative techniques and proofs of approachability

1.4.1 Approachability in continuous time

Benaïm, Hofbauer & Sorin [7] noticed that Blackwell's approachability strategy of a B-set $\mathcal{E}$ satisfies the following recurrence relation: condionnaly to h^n,

$$\frac{1}{1/n+1}\left(\mathbb{E}_{\sigma,\tau}\left[\overline{g}_{n+1}\middle| h^n\right]-\overline{g}_n\right)\in T(\overline{g}_n)-\overline{g}_n,$$

where $T(z) = \left\{\omega\in\mathbb{R}^d; \|\omega\|\leq\|g\|_\infty \text{ and } \exists\, p\in\Pi_{\mathcal{E}}(z), \langle z-p,\omega-p\rangle\leq 0\right\}$. Therefore, the sequence of averages payoff $\{\overline{g}_n\}_{n\in\mathbb{N}}$ is a *Discrete Stochasitic Approximation* (a DSA for short) of $\mathbf{g}$, solution of the associated ordinary differential inclusion

$$\dot{\mathbf{g}}\in T(\mathbf{g})-\mathbf{g},\quad \mathbf{g}(0)=\mathbf{g}_0\in\mathbb{R}^d.$$

The derivative of the mapping $\delta(t) = d^2_{\mathcal{C}}(\mathbf{g}(t))$ satisfies $\delta'(t)\leq -2\delta(t)/t$ thus it is a Lyapounov function and $\delta(t)\leq\delta(0)t^{-2}$. As a consequence, $\mathbf{g}$ converges to $\mathcal{E}$ and, as an DSA, the sequence $\{\overline{g}_n\}_{n\in\mathbb{N}}$ converges a.s. to $\mathcal{E}$. However, rates of convergence of DSA are usually not explicit and might not be uniform.

To circumvent this issue, one might consider procedures in law, as defined in Section 1.3.1, that are deterministic and thus can be represented as an Euler Scheme of the associated ordinary differential inclusion. They might provide explicit rates as the difference between the average payoff and its expectation converges to zero, and is controlled by concentration inequalities (see Sorin [72] or Kwon [40]).

As Soulaimani, Quincampoix & Sorin [2] have considered an auxiliary differential game $\mathcal{D}$ where control spaces of the player and Nature are respectively $\mathcal{X}=\Delta(\mathcal{A})$ and $\mathcal{Y}=\Delta(\mathcal{B})$) and the game dynamic is given by:

$$\frac{d}{dt}\overline{\mathbf{g}}(t)=\frac{-\overline{\mathbf{g}}(t)+g(x(t),y(t))}{t},\quad \overline{\mathbf{g}}(0)=0.$$

The intuition is that $\overline{\mathbf{g}}(t)=\frac{1}{t}\int_0^t g(x(s),y(s))ds$ is the average payoff at time t. The change of variables $t=e^s$ and $\overline{\mathbf{g}}(s)=\mathbf{g}(e^s)$ modifies the dynamic into

$$\frac{d}{dt}\mathbf{g}(s)=-\mathbf{g}(s)+g(x(s),y(s)):=f(\mathbf{g}(s),x(s),y(s)),\quad \mathbf{g}(0)=\overline{\mathbf{g}}(1).$$

This transformation proves the characterization of a B-set given in Equation (4). Indeed, a set $\mathcal{E}$ is approachable if the player can force the dynamic to stay within $\mathcal{E}$. Therefore, a closed set E is a B-set if and only if it is a discriminating domain for the player with respect to the dynamic f, i.e. if

$$\forall p\in\mathcal{C},\forall q\in \mathrm{NC}_{\mathcal{C}}(p), \sup_{y\in\mathcal{Y}}\inf_{x\in\mathcal{X}}\langle f(p,x,y),q\rangle\leq 0.$$

1.4.2 Information-based strategies

Blackwell's strategy is a payoff-based strategy as the running relevant state variable is the sequence of average payoffs. We develop in this section a conceptually completely different kind of strategy based on the sequence of observed profile of action played, as in Perchet & Quincampoix [63] or Mannor, Perchet & Stoltz [51].

The basic idea follows from the following simple fact. Define $\theta_n = \delta_{a_n,b_n} \in \Delta(\mathcal{A}\times\mathcal{B})$ as the Dirac mass on $(a_n, b_n) \in \mathcal{A}\times\mathcal{B}$ and let $\overline{\theta}_n = \sum_{m=1}^n \theta_m/n$ be their average. By definition, $\overline{g}_n = \mathbb{E}_{\overline{\theta}_n}[g(a,b)]$ belongs to $\mathcal{E}$ if and only if $\overline{\theta}_n$ belongs to the following set

$$\widetilde{\mathcal{E}} := \Big\{\theta \in \Delta(\mathcal{A}\times\mathcal{B}) \text{ s.t. } \mathbb{E}_\theta[g(a,b)] \in \mathcal{E}\Big\} \subset \Delta(\mathcal{A}\times\mathcal{B}).$$

If $\mathcal{E}$ is closed and convex, then $\widetilde{\mathcal{E}}$ (seen as a subset of $\mathbb{R}^{A\times B}$) is also closed and convex; it remains to compare distance between $\mathcal{E}$ and $\widetilde{\mathcal{E}}$.

Lemma 1.12 *There exists $\gamma > 0$ such that, for any probability measure $\theta \in \Delta(\mathcal{A}\times\mathcal{B})$ and any set $\mathcal{E}$*

$$\gamma d_{\widetilde{\mathcal{E}}}\Big(\theta\Big) \le d_{\mathcal{E}}\Big(\mathbb{E}_\theta[g(a,b)]\Big) \le \|g\|_\infty \sqrt{AB}\, d_{\widetilde{\mathcal{E}}}\Big(\theta\Big).$$

Proof: For any $\theta \in \Delta(\mathcal{A}\times\mathcal{B})$, define $g(\theta) = \mathbb{E}_\theta[g(a,b)]$. Let $\underline{\theta} \in \Pi_{\widetilde{\mathcal{E}}}(\theta)$, so $g(\underline{\theta}) \in \mathcal{E}$ and

$$\begin{aligned} d_{\mathcal{E}}\Big(g(\theta)\Big) &\le \Big\|g(\theta) - g(\underline{\theta})\Big\| = \left\|\sum_{a,b}\Big(\theta(a,b) - \underline{\theta}(a,b)\Big) g(a,b)\right\| \\ &\le \|g\|_\infty \big\|\theta - \underline{\theta}\big\|_1 \le \|g\|_\infty \sqrt{AB}\big\|\theta - \underline{\theta}\big\|_2 \end{aligned}$$

This gives the second inequality.

For the first inequality, notice that $g : \Delta(\mathcal{A}\times\mathcal{B}) \subset \mathbb{R}^{A\times B} \to \mathrm{co}\{g(a,b)\}$ is a linear mapping, so its inverse g^{-1} is piecewise linear thus Lipschitz, see e.g., Billera & Sturmfels [8], bottom of page 530, or Walkup & Wets [81]. As a consequence, there exists $\lambda > 0$ such that for every $z, z' \in \mathrm{co}\{g(a,b)\}$ and any points θ such that $g(\theta) = z$, there exists θ' such that $g(\theta') = z'$ and $\|\theta - \theta'\| \le \lambda\|z - z'\|$. In particular, for every $\theta \in \Delta(\mathcal{A}\times\mathcal{B})$, if $\underline{z} \in \Pi_{\mathcal{E}}(g(\theta))$ there exists $\underline{\theta}$ such that $g(\underline{\theta}) = \underline{z}$, thus $\underline{\theta} \in \widetilde{\mathcal{E}}$ and

$$d_{\widetilde{\mathcal{E}}}(\theta) \le \Big\|\theta - \underline{\theta}\Big\| \le \lambda\Big\|g(\theta) - g(\underline{\theta})\Big\| = \lambda\Big\|g(\theta) - \Pi_{\mathcal{E}}(g(\theta))\Big\| = \lambda d_{\mathcal{E}}(g(\theta)) \ ,$$

and one just has to take $\gamma = 1/\lambda$. ■

The consequence of this lemma is that an approachability strategy for $\widetilde{\mathcal{E}}$ is an approachability strategy for $\mathcal{E}$ (and reciprocally); apart from the requirement to compute $\widetilde{\mathcal{E}}$, only constants in rates of convergence deteriorate.

The main advantage of this new kind of algorithms is that they do not rely on the observed sequences of payoffs. For example, consider the cases where payoffs are not

vectors in some Euclidian space but in some arbitrarily normed space (or even payoffs can be subsets of this space). If the image space is not Hilbertian, then Blackwell's proofs do no longer hold; on the other hand, the transformation sequences of payoff into sequences of profile of action remains true. Therefore, we get this very general version of characterization of approachable convex sets.

Theorem 1.7 *Let* $\Big(\mathbf{H}, \mathcal{N}(\cdot)\Big)$ *be any normed space (not necessarily Hilbertian) and* $g : \Delta(\mathcal{A}) \times \Delta(\mathcal{B}) \to \mathbf{H}$ *(or* $g : \mathcal{A} \times \mathcal{B} \to \mathbf{H}$ *is* $\mathcal{A}$ *and* $\mathcal{B}$ *are some compact convex sets) any continuous bi-linear mapping. Then Blackwell's characterization of approachable convex sets holds:*

$$\textit{A convex set } \mathcal{C} \subset \mathbf{H} \textit{ is approachable if and only if } \forall y \in \Delta(\mathcal{B}), \exists x \in \Delta(\mathcal{A}), g(x,y) \in \mathcal{C}.$$

The result is already proved if $\mathcal{A}$ and $\mathcal{B}$ are finite. If they are some compact convex sets and g is continuous, then one can discretize them to get ε-approachability strategy. Since $\mathcal{C}$ is convex, they can be concatenate into an approachability strategy (using the doubling trick).

From the point of view of computational geometry, this result is rather intuitive. Indeed, no matter the image space, $\mathrm{co}\left\{g(a,b); a \in \mathcal{A}, b \in \mathcal{B}\right\}$ is a polytope with at most AB vertices which belongs to a relative space of finite dimension at most $AB - 1$. Up to a renormalization, this gives Theorem 1.7. However, in case where $\mathbf{H} = L_2(\Omega, \mu, \mathcal{F})$, this does not directly imply previous results as the approachability is only in probability and not μ-almost surely.

1.4.3 Potential-based and uniform-norm approachability

Approachability was first defined with respect to the ℓ^2 distance. Roughly speaking, this induce a repeated game (see also the next subsection) between the player and Nature where the first player minimizes the distance to the set $\mathcal{E}$ and Nature maximizes it. This can be generalized to a more general class of mappings $\Phi : \mathbb{R}^d \to \mathbb{R}$, called potentials, that are twice continuously differentiable (although this condition can be fairly weakened).

An illustration of the interest of potential based approachability is given in the following Corollary 1.16. It yields fastest rates of convergence when distances to sets are defined with respect to the uniform norm $\|\cdot\|_\infty$ instead of the Euclidian norm $\|.\|_2$.

Let us denote by δ the minimum level of Φ that player can guarantee in expectation if he plays second, i.e.

$$\delta = \inf\Big\{\lambda \in \mathbb{R} \text{ s.t. } \forall y \in \Delta(\mathcal{B}), \exists x \in \Delta(\mathcal{A}), \Phi\big(g(x,y)\big) \leq \lambda\Big\} \text{ and } \mathcal{E}_\delta := \Phi^{-1}\big((-\infty, \delta]\big).$$

Theorem 1.8 *Assume that, for every* z *outside* $\mathcal{E}_\delta$*, the gradient* $\nabla\Phi(z)$ *points sufficiently towards* z*, i.e., there exists* $\beta > 0$ *such*

$$\forall z \notin \mathcal{E}_\delta,\ \exists x := x(z) \in \Delta(\mathcal{A}) \textit{ s.t. } \langle \nabla\Phi(z), g(x,y) - z\rangle \leq -\beta(\Phi(z) - \delta),\ \forall y \in \Delta(\mathcal{B}). \quad (12)$$

Then, no matter the strategy of Nature, choosing $x_{n+1} = x(\overline{g}_n)$ yields, in expectation,

$$\Phi(\overline{g}_n) - \delta \leq \frac{\kappa_\Phi}{(\beta-1)}\frac{1}{n} \text{ if } \beta > 1 \quad \text{and} \quad \Phi(\overline{g}_n) - \delta \leq \kappa_\Phi \frac{\log(n)+1}{n^\beta} \text{ if } 0 < \beta \leq 1,$$

where κ_Φ is a constant depending uniquely on Φ.

If $\beta = 0$ but the inequality is strict in (12), then uniform convergence still holds yet at a non-explicit rate.

Proof: First, notice that we can focus on the case where $\delta = 0$. The proof follows from Hart & Mas-Colell [32] and Sorin [71] (see also Cesa-Bianchi & Lugosi [13, 14]) and is based on a Taylor expansion of Φ. Indeed, since $\overline{g}_{n+1} = \overline{g}_n + (g_{n+1} - \overline{g}_n)/(n+1)$ and Φ is $\mathcal{C}^2$, there exists some $\xi_n \in [\overline{g}_{n+1}, \overline{g}_n]$ such that

$$\Phi(\overline{g}_{n+1}) = \Phi(\overline{g}_n) + \frac{1}{n+1}\langle \nabla\Phi(\overline{g}_n), g_{n+1} - \overline{g}_n\rangle + \frac{1}{2(n+1)^2}(g_{n+1} - \overline{g}_n)' D^2\Phi(\xi_n)(g_{n+1} - \overline{g}_n)$$

where $\nabla\Phi$ and $D^2\Phi$ are respectively the gradient and the Hessian of Φ; since the latter is $\mathcal{C}^2$ and every g_n belongs to the same compact set, there exists κ_Φ such that $(g_{n+1} - \overline{g}_n)' D^2\Phi(\xi_n)(g_{n+1} - \overline{g}_n) \leq 2\kappa_\Phi$, for every $n \in \mathbb{N}$. As a consequence, one has

$$\mathbb{E}\Big[\Phi(\overline{g}_{n+1})\Big] \leq \left(1 - \frac{\beta}{n+1}\right)\mathbb{E}\Big[\Phi(\overline{g}_n)\Big] + \frac{\kappa_\Phi}{(n+1)^2}$$

and the result follows from simple induction when $\beta \geq 1$. When $0 < \beta < 1$, the bound is a consequence of the fact that

$$\left(1 - \frac{\beta}{n+1}\right)\frac{\log(n)+1}{n^\beta} + \frac{1}{(n+1)^2} \leq \frac{\log(n+1)+1}{(n+1)^\beta}.$$

The proof is a bit more intricate for $\beta = 0$ (along with a strict inequality in (12)), but we can use the fact that $\overline{g}_n$ is a D.S.A. of the following differential inclusion

$$\dot{\mathbf{g}} \in A_\Phi(\mathbf{g}) - \mathbf{g} \text{ with } A_\Phi(z) = \Big\{\omega \in \mathbb{R}^d; \|\omega\| \leq \|g\|_\infty \text{ and } \langle \nabla\Phi(z), \omega - z\rangle > 0\Big\}.$$

The mapping $t \mapsto \Phi(\mathbf{g}(t))$ is a Lyapounov function since if $\mathbf{g}(t) \notin \mathcal{C}_\delta$

$$\frac{d}{dt}\Phi(\mathbf{g}(t)) = \langle \nabla\Phi(\mathbf{g}(t)), \dot{\mathbf{g}}(t)\rangle \in \langle \nabla\Phi(\mathbf{g}(t)), A_\Phi(\mathbf{g}(t)) - \mathbf{g}(t)\rangle < 0,$$

therefore $\mathbf{g}$ converges to $\mathcal{C}_\delta$ and so does $\overline{g}_n$. ■

If $\mathcal{C}$ and Φ are convex, then Equation (12) always holds with $\beta = 1$, and we recover Theorem 7.6 of Cesa-Bianchi & Lugosi [14] (due to Hart & Mas-Colell [32]):

Corollary 1.13 *If $\mathcal{C} = \Phi^{-1}\big((-\infty, 0]\big)$ for some convex, twice continuously differentiable mapping Φ whose Hessian is bounded in norm by κ_Φ on $\mathrm{co}\big\{g(a,b)\big\}$, there exists a strategy such that, in expectation and no matter the strategy of Nature, $\Phi(\overline{g}_n) \leq \frac{2\kappa_\Phi(\log(n)+1)}{n}$.*

The assumption that Φ is twice continuously differentiable can be easily weakened, in particular as soon as the constant κ_Φ exists. The next proposition is concerned with the sequence of sums of payoffs $G_n = \sum_{m=1}^n g_m$ instead of averages. It will be used, in some cases, to improve rates of convergence.

Proposition 1.14 *Assume that*

$$\forall z \notin \mathcal{E}_\delta,\ \exists x := x(z) \in \Delta(\mathcal{A}) \text{ s.t. } \langle \nabla\Phi(z), g(x,y)\rangle \leq 0,\ \forall y \in \Delta(\mathcal{B})$$

and there exists $\kappa_\Phi > 0$ such that $g(x,y)' D^2\Phi(z) g(x,y) \leq \kappa_\Phi$ for every $x \in \Delta(\mathcal{A})$, $y \in \Delta(\mathcal{B})$ and $z \notin \mathcal{E}_\delta$. Then, no matter Nature's strategy, choosing $x_{n+1} = x(G_n)$ yields $\mathbb{E}\big[\Phi(G_n)\big] \leq n\kappa_\Phi + \Phi(0)$.

Proof: This is a consequence of the fact that, for some $\xi_n \in [G_n, G_{n+1}]$,

$$\Phi(G_{n+1}) = \Phi(G_n) + \langle \nabla\Phi(G_n), g_{n+1}\rangle + g'_{n+1} D^2\Phi(\xi_n) g_{n+1}$$

followed by an immediate induction. ■

This result can be immediately extended if Φ is not $\mathcal{C}^2$ but such that

$$\Phi(G_{n+1}) - (\Phi(G_n) + \langle \nabla\Phi(G_n), g_{n+1}\rangle) \leq \kappa_\Phi.$$

As mentioned before, the following corollary shows a faster convergence if $\mathcal{C}$ is an approachable cone. Proposition 1.14 is even used more deeply to get optimal rates of convergence (both in the number of stages and the dimension) below to obtain approachability with respect to the uniform norm.

Corollary 1.15 *If $\mathcal{C}$ is an approachable closed and convex cone, then Blackwell's strategy ensures that, no matter Nature's strategy and for every $N \in \mathbb{N}$*

$$\mathbb{E}_{\sigma,\tau}\left[d_{\mathcal{C}}(\overline{g}_n)\right] \leq \frac{\|g\|_\infty}{\sqrt{n}}.$$

Proof: First, if $\mathcal{C}$ is a cone then necessarily $\langle z - \Pi_{\mathcal{C}}(z), \Pi_{\mathcal{C}}(z)\rangle \leq 0$, therefore the first condition of Proposition 1.14 is the characterization of the fact that $\mathcal{C}$ is a B-set. Second, if $\Phi(\cdot) = d^2_{\mathcal{C}}(\cdot)$ then Φ satisfies the second condition of Proposition 1.14 – or at least its straightforward extension – with $\kappa_\Phi = \|g\|^2_\infty$.

Since $\mathcal{C}$ is a cone, $d^2_{\mathcal{C}}(\overline{g}_n) = \Phi(G_{n+1})/n^2$ and the result follows. ■

For simplicity, we will assume that $\|g(a,b)\|_\infty \leq 1$ and we only consider target sets such that, for some $b_k, c_k \leq 1$,

$$\mathcal{C} := \left\{ z \in \mathbb{R}^d \text{ s.t. } b_k \leq z_k \leq c_k, \forall k \in \{1,\ldots,d\} \right\}.$$

The ℓ^∞-distance to this set is denoted by $d^\infty_{\mathcal{C}}$, i.e. $d^\infty_{\mathcal{C}}(z) = \inf_{c\in\mathcal{C}} \|z - c\|_\infty$.

Corollary 1.16 *There exists a strategy σ of the player such that, against any strategy τ of Nature and every $n \in N$ and $\delta > 0$, with probability at least $1-\delta$,*

$$\mathbb{E}_{\sigma,\tau}\left[d_{\mathcal{C}}^{\infty}(\overline{g}_n)\right] \leq 14\sqrt{\frac{\log(2d)}{n}} \text{ and } \mathbb{P}_{\sigma,\tau}\left\{d_{\mathcal{C}}^{\infty}(\overline{g}_n) \geq \delta\right\} \leq 16\sqrt{\frac{\log\left(\frac{2d}{\delta}\right)}{n}}.$$

Proof: We first prove a similar result in the specific case where $\mathcal{C} = \mathbb{R}^d_-$ is the negative orthant and if an horizon N is known in advance. Then we will use a *doubling trick* to conclude for the orthant; we will finally show how to reduce approachability of any product set $\mathcal{C}$.

Let Φ be the following potential, depending on a parameter $\eta > 0$ to be fixed later:

$$\Phi(z) := \frac{1}{\eta}\log\left[\sum_{k=1}^{d} e^{\eta z_k}\right] \text{ so that } d_{\mathcal{C}}^{\infty}(z) \leq \Phi(z) \text{ and } \Phi(0) \leq \log(d) \text{ ; moreover}$$

$$\nabla\Phi(z)^i = \frac{e^{\eta z_i}}{\sum_{k=1}^{d} e^{\eta z_k}} \text{ and } D^2\Phi(z) = \eta\,\mathrm{diag}\left(\frac{e^{\eta z_i}}{\sum_{k=1}^{d} e^{\eta z_k}}\right) - \eta\nabla\Phi(z)\nabla\Phi(z)',$$

where $\mathrm{diag}(\lambda_i)$ is the matrix whose diagonal is $\lambda_1, \ldots, \lambda_d$ and zero everywhere. As a consequence, since $\mathcal{C}$ is approachable, the first condition of Proposition 1.14 is satisfied and $\omega' D^2\Phi(z)\omega \leq \eta\omega'\,\mathrm{diag}(\lambda_i)\omega - \eta\|\nabla\Phi(z)\omega\|_2^2 \leq \eta\|\omega\|_\infty^2$ implies the second one.

So, convexity of $d_{\mathcal{C}}^{\infty}(\cdot)$, Proposition 1.14 and the choice $\eta = \sqrt{\log(d)/N}$ imply that

$$\mathbb{E}_{\sigma,\tau}\left[d_{\mathcal{C}}^{\infty}(\overline{g}_N)\right] \leq \mathbb{E}_{\sigma,\tau}\left[\frac{d_{\mathcal{C}}^{\infty}(\overline{G}_N)}{N}\right] \leq \frac{1}{N}\mathbb{E}[\Phi(G_N)] \leq \eta + \frac{\log(d)}{\eta N} = 2\sqrt{\frac{\log(d)}{N}}.$$

We now make appeal to the *doubling trick*, that is, we consider the strategy consisting in playing by blocks of lengths 2^k, following the potential associated with $\eta_k := \sqrt{\log(d)/2^k}$ on the k-th block and reseting everything at the beginning of a new block. A simple induction, based on the convexity of $d_{\mathcal{C}}^{\infty}(\cdot)$, shows that, at the end of any block,

$$\mathbb{E}_{\sigma,\tau}\left[d_{\mathcal{C}}^{\infty}(\overline{g}_{2^k-1})\right] \leq 2(1+\sqrt{2})\sqrt{\frac{\log(d)}{2^k-1}}.$$

Hence it remains to control distances within blocks. Yet, using the previous bound obtained for ends of blocks, one has for $n = 2^k - 1 + m$ with $m \leq 2^k$,

$$\begin{aligned}
\mathbb{E}_{\sigma,\tau}\left[d_{\mathcal{C}}^{\infty}(\overline{g}_n)\right] &\leq \frac{2^k-1}{n}\mathbb{E}_{\sigma,\tau}\left[d_{\mathcal{C}}^{\infty}(\overline{g}_{2^k-1})\right] + \frac{m}{n}\left(\eta_k + \frac{\log(d)}{\eta_k m}\right) \\
&\leq \frac{\sqrt{\log(d)}}{n}\left[2(1+\sqrt{2})\sqrt{2^k-1} + \frac{m}{\sqrt{2^k}} + \sqrt{2^k}\right] \\
&\leq (4+2\sqrt{2})\sqrt{\frac{\log(d)}{n}}.
\end{aligned}$$

Concentration arguments give the bound in high probability. Indeed, the union bound implies

$$\mathbb{P}\Big\{\exists k \leq d \text{ s.t. } \left|\overline{g}_n^k - \mathbb{E}[\overline{g}_n^k]\right| \geq \varepsilon\Big\} \leq 2d \exp\left(-\frac{n\varepsilon}{2}\right),$$

thus the probability that $\|\overline{g}_n - \mathbb{E}[\overline{g}_n]\|_\infty$ is smaller than $\sqrt{\frac{2}{n}\log\left(\frac{2d}{\delta}\right)}$ is bigger than $1-\delta$. The result for the orthant is a direct consequence of the triangle inequality.

We no longer assume that $\mathcal{C}$ is an orthant, but is defined by

$$\mathcal{C} := \Big\{z \in \mathbb{R}^d \text{ s.t. } b_k \leq z_k \leq c_k, \forall k \in \{1, \ldots, d\}\Big\}.$$

Let $h(x,y) = \Big(g_k(x,y) - c_k, b_k - g_k(x,y)\Big)_{k\in\{1,\ldots,d}} \in \mathbb{R}^{2d}$, then approachability of $\overline{g}_n$ to $\mathcal{C}$ is equivalent to the approachability of $\overline{h}_n$ to the negative orthant, since $d_{\mathcal{C}}^\infty(\overline{g}_n) = d_{\mathbb{R}_-^{2d}}^\infty(\overline{g}_n)$ and $g(x,y) \in \mathcal{C}$ if and only if $h(x,y) \in \mathbb{R}_-^{2d}$. The result follows from the bound exhibited for the orthant.

■

1.4.4 From weak approachability to approachability

Recall that a closed set $\mathcal{E}$ is approachable if the player has a strategy such that after some (maybe large) stage N, the payoffs remains in a small neighborhood of $\mathcal{E}$. Similarly, it is excludable if Nature can enforce the dual: after some stage N, the payoffs remains outside some neighborhood of $\mathcal{E}$. Blackwell proved that there exists a dichotomy for convex sets: they are either approachable or excludable. This is not true for any set, as illustrated in the following example, due to Blackwell.

Consider the set and payoff matrix defined by, with $\mathcal{A} = \{T, B\}$ and $\mathcal{B} = \{L, R\}$,

$$\mathcal{E} = \Big\{(1/2, y), y \in [0, 1/4]\Big\} \cup \Big\{(1, y), y \in [1/4, 1]\Big\} \quad \text{and} \quad \begin{array}{ccc} & L & R \\ T & (1,0) & (1,1) \\ B & (0,0) & (0,0) \end{array}.$$

Assume that the strategy of the player dictates to play T during N stages (with N a large even number) then to play either always T or always B during the following N stages, depending on wether Nature has played more than half of the time R during the first N stages.

In the former case, the player got after N stages, an average payoff of $(1, y)$ with $y \geq 1/2$ thus by keeping to play T for N stages, he ensures that its average payoff after $2N$ stages is $(1, y')$ with $y' \geq y/2 \geq 1/4$. In the latter case, the payoff after N stages is $(1, y)$ with $y \leq 1/2$, thus the payoff after $2N$ stages is $(1/2, y')$ with $y' = y/2 \leq 1/4$.

As a consequence, this strategy guarantees that, after $2N$ stages, the payoff is exactly in $\mathcal{E}$. So if this procedure is applied during $2N_1$ stages, then started over for $2N_2 = 2e^{N_1}$

stages, then started again over for $2N_3 = 2e^{N_2}$ stages and so on, the payoff is infinitely often arbitrarily closed to $\mathcal{E}$ which is therefore not excludable.

Unfortunately, $\mathcal{E}$ is not approachable; indeed, this would imply that at least one of the two connected (and convex) component of $\mathcal{E}$ is approachable. But neither of them satisfies Blackwell's characterization.

In this example, the player cannot enforce the payoff to remain close to $\mathcal{E}$, but if he knows in advance that there are only N stages in the game, then he can ensure that, at the terminal stage, the payoff is in $\mathcal{E}$ (or at least, for odd integer, $1/N$-close to $\mathcal{E}$). A natural weaker concept of approachability emerges: a set $\mathcal{E} \subset \mathbb{R}^d$ is weakly-approachable if, given some fixed large length of the game, the player has a strategy such that the terminal average payoff is close to $\mathcal{E}$.

Definition 1.6 *A closed set $\mathcal{E} \subset \mathbb{R}^d$ is weakly approachable if for every $\varepsilon > 0$, there exists $N_\varepsilon \in \mathbb{N}$ such that, in any game of length $n \geq N$, the player has a strategy σ_n such that, no matter the strategy τ of Nature, $\mathbb{E}_{\sigma_n,\tau}\Big[d_{\mathcal{E}}(\overline{g}_n)\Big] \leq \varepsilon$.*

Similarly, $\mathcal{E}$ is weakly excludable if Nature can weakly approach the complement of E^δ for some $\delta > 0$.

We emphasize the fact that in weak-approachability, strategies can depend on the length of the game n, which is not allowed for regular approachability. The question rose by Blackwell [9] and solved by Vieille [77] is wether there exists a dichotomy between weakly-approachable and weakly-excludable sets.

Theorem 1.9 *Any closed set is either weakly-approachable or weakly-approachable.*

Proof: We only sketch here the proof of Vieille[77].

Consider the differential zero-sum game where the player and Nature choose action $x(t) \in \Delta(\mathcal{A})$ and $y(t) \in \Delta(\mathcal{B})$ in continuous time (actually, even the formal definition of strategies might require precise notations and concepts). In this game, a state variable which represents the accumulated payoff, evolves following the dynamic $\dot{\mathbf{G}}(t) = g\big(x(t), y(t)\big)$ and $\mathbf{G}(0) = 0$ during the time $t = 0$ and $t = 1$.

In this game, the overall objective of the player is to minimize the terminal payoff $d_{\mathcal{E}}\left(\mathbf{G}(1)\right)$, while Nature maximizes it. The important fact is that one can prove, using techniques and results from differential games, that this game has a value v. If $v = 0$, then the player has a strategy such that the cumulated payoff at time $t = 1$ is exactly in $\mathcal{E}$ whereas if $v > 0$, Nature has a strategy such that this cumulated payoff is bounded away from $\mathcal{E}$.

It remains to understand that a game in discrete time with N stages is a discretization (or an approximation) of this differential game and as N goes to infinity, this approximation is more and more precise. Therefore, if the player can enforce that $\mathbf{G}(1)$ belongs to $\mathcal{E}$, then he can ensure that $\overline{g}_N$ is arbitrarily close to $\mathcal{E}$ when N is large enough. The converse holds for Nature, hence the result. ■

Actually, the focus of this section if not only this important (and elegant) result but also the following properties, inspired from Cesa-Bianchi & Lugosi [14] or Rakhlin, Sridharan & Tewari [66]. Given an approachable convex set $\mathcal{C}$, let σ_N be an optimal strategy in the N-stage zero-sum game with terminal payoff $\mathbb{E}_{\sigma,\tau}\Big[d_{\mathcal{C}}(\overline{g}_N)\Big]$ and denote by v_N the value of this game (its existence is not difficult).

We know that $\mathbb{E}_{\sigma,\tau}\Big[d_{\mathcal{C}}(\overline{g}_n)\Big]$ can be upper bounded, using some adequate approachability strategy, by $O\left(\|g\|_\infty/\sqrt{n}\right)$; but it is also obviously lower-bounded by v_n. So the computation of v_n could indicate wether the rate is tight or not. On the other hand, exact computation of v_n might be challenging, yet if satisfies

$$v_n = \min_{\sigma_n}\max_{\tau}\mathbb{E}_{\sigma,\tau}\Big[d_{\mathcal{C}}(\overline{g}_n)\Big] = \max_{\tau}\min_{\sigma_n}\mathbb{E}_{\sigma,\tau}\Big[d_{\mathcal{C}}(\overline{g}_n)\Big] \leq \max_{\tau}\min_{\sigma(\tau)}\mathbb{E}_{\sigma,\tau}\Big[d_{\mathcal{C}}(\overline{g}_n)\Big]$$

where $\sigma(\tau)$ is the strategy that chooses, given τ and after the finite history h^n, $x_{n+1} = x(\tau(h^n))$. In particular, $\mathbb{E}_{\tau,\sigma(\tau)}[\widehat{g}_n] := \sum_{m=1}^n \mathbb{E}_{\tau,\sigma(\tau)}[g_m|h^{m-1}]/n$ belongs to $\mathcal{C}$ and therefore (removing the dependency in σ)

$$v_n \leq \max_{\tau}\mathbb{E}\Big[\,\|\overline{g}_n - \mathbb{E}_\tau[\overline{g}_n]\|\,\Big] \leq \sup_{\widetilde{g}}\mathbb{E}\left[\left\|\frac{\sum_{m=1}^n \widetilde{g}_m - \mathbb{E}[\widetilde{g}_m|h^{m-1}]}{n}\right\|\right] \leq O\left(\frac{\|g\|_\infty}{\sqrt{n}}\right)$$

where the supremum is taken over all sequences $\widetilde{g}_m - \mathbb{E}[\widetilde{g}_m|h^{m-1}]$ of martingale differences with $\widetilde{g}_m \in \{g(a,b), a\in\mathcal{A}, b\in\mathcal{B}\}$. Last inequality is a consequence of Hoeffding-Azuma's inequality in Euclidian spaces.

A question that naturally arises is wether we can concatenate – using the doubling trick – optimal strategies in games of length 2^k to construct an approachability strategy of $\mathcal{C}$ (i.e., independent of any horizon n). The answer is both no and yes: no with the current definition of v_n. Indeed the only guarantee is that terminal payoff is v_{2^k}-close to $\mathcal{C}$ but, for instance, payoff at middle stages could be arbitrarily away.

On the other hand, since $\mathcal{C}$ is a convex set, we can modify the definition of v_n as follows so that the answer is yes. Define

$$v'_n = \max_{\tau}\min_{\sigma_n}\mathbb{E}_{\sigma,\tau}\Big[\sup_{m\leq n}\frac{m}{n}d_{\mathcal{C}}(\overline{g}_m)\Big] \leq \max_{\tau}\min_{\sigma(\tau)}\mathbb{E}_{\sigma,\tau}\Big[\sup_{m\leq n}\frac{m}{n}d_{\mathcal{C}}(\overline{g}_m)\Big],$$

so that, using the same arguments and Doobs (or Hoeffding) maximal inequality

$$v'_n \leq \max_{\tau}\mathbb{E}\Big[\sup_{m\leq n}\frac{m}{n}\,\|\overline{g}_n - \mathbb{E}_\tau[\overline{g}_n]\|\,\Big] \leq \sup_{\widetilde{g}}\mathbb{E}\left[\frac{1}{n}\left\|\sum_{s=1}^m \widetilde{g}_s - \mathbb{E}[\widetilde{g}_s|h^{s-1}]\right\|\right] \leq O\left(\frac{\|g\|_\infty}{\sqrt{n}}\right).$$

Finally, the doubling trick works with this definition of v'_n, see the proof of Corollary 1.16.

This technique seems void at first sight, but might be useful in some specific examples, as in Proposition 4.2 in Section 4 (see also Remark 4.2). In this case, because of the geometry of $\mathcal{C}$, one has $d_{\mathcal{C}}(z) \leq 2\|z-c\|_\infty$, for every $c\in\mathcal{C}$. Then the same tools yield that v_n is smaller than $O\left(\sqrt{\log(d)/n}\right)$ which is negligible compared to $\|g\|_\infty/\sqrt{n} \simeq \sqrt{d/n}$ as the dimension d increases.

2 Regret minimization

Hannan [30] introduced the concept of *external regret* in repeated two-player games (between a player and Nature, with scalar payoff) in order to define an exogenous criterium to evaluate a strategy in a non-Bayesian framework. In words, the player has no external regret (or his strategy is *externally consistent*) if, asymptotically, he could not have gained strictly more if he had known, before the beginning of the game, the empirical distribution of moves of Nature. This notion has notably been refined by Foster & Vohra [23] (see also Fudenberg & Levine [28]) into *internal regret*: a player has no internal regret (or his strategy is *internally consistent*) if he has no external regret on the set of stages where he played a specific action, as soon as this set is big enough.

2.1 Finite action spaces

Consider a two-person repeated game where action spaces of the player and Nature are $\mathcal{A}$ and $\mathcal{B}$ (of cardinalities A and B) and $\rho : \mathcal{A} \times \mathcal{B} \to [0,1]$ is a real payoff mapping. Extension of ρ to $\Delta(\mathcal{A}) \times \Delta(\mathcal{B})$ and strategies are defined as in the previous section.

2.1.1 External regret

Choices of actions $a_n \in \mathcal{A}$ and $b_n \in \mathcal{B}$ generate a regret $r_n \in \mathbb{R}^A$ defined by

$$r_n = r(a_n, b_n) := \Big(\rho(1, b_n) - \rho(a_n, b_n), \ldots, \rho(A, b_n) - \rho(a_n, b_n)\Big) \in \mathbb{R}^A.$$

Intuitively, the regret r_n represents the differences between what the player could have got and what he actually got. And a player has no external regret if, asymptotically, every component of the average regret is non positive. In words, this means that the player could not think *" if I had known* [the empirical distribution of Nature's actions], *I would have always played action a^* "*, hence the terminology of *regret*. Indeed, by linearity of ρ,

$$\overline{r}_n = \Big(\rho(1, \overline{b}_n) - \overline{\rho}_n, \ldots, \rho(A, \overline{b}_n) - \overline{\rho}_n\Big) \in \mathbb{R}^A.$$

Given a vector $U \in \mathbb{R}^d$, the notation U^+ will stand for the positive part of U, i.e., $U^+ = \big(\max\{U^i, 0\}\big)_{1 \le i \le d}$. Similarly, U^- is the negative part of U.

Definition 2.1 *A strategy σ of the player has no external regret if, for all strategy τ of Nature, $\mathbb{P}_{\sigma,\tau}$-almost surely,*

$$\limsup_{n\to\infty} \max_{a^* \in \mathcal{A}} \rho(a^*, \overline{b}_n) - \overline{\rho}_n \le 0, \ \textit{or equivalently,} \ \limsup_{n\to\infty} \|\overline{r}_n^+\|_\infty \le 0. \tag{13}$$

The existence of externally consistent strategies goes back to Hannan [30]. However, the following theorem, with rates of convergence independent of Nature's strategy, is due to Cesa-Bianchi & Lugosi [14].

Theorem 2.1 *There exists an externally consistent strategy σ, such that, no matter the strategy τ of Nature and for every $n \in \mathbb{N}$,*

$$\mathbb{E}_{\sigma,\tau}\Big[\max_{a^*\in\mathcal{A}} \rho(a^*, \overline{b}_n) - \overline{\rho}_n\Big] = \mathbb{E}_{\sigma,\tau}\Big[\|\overline{r}_n^+\|_\infty\Big] \leq 2\sqrt{\frac{\log(A)}{n}}.$$

We will not yet provide proofs of this result; instead, we will show a weaker result, following Zinkevich [83]. The basic idea is to notice that the overall objective is to maximize the convex function $\rho(\cdot, \overline{b}_n)$ and therefore to apply any convex-maximization techniques, for example a gradient descent.

Proof: First, we claim that for every $n \in \mathbb{N}$, there exists a strategy σ_n (that depends on n), such that

$$\mathbb{E}_{\sigma,\tau}\Big[\max_{a^*\in\mathcal{A}} \rho(a^*, \overline{b}_n) - \overline{\rho}_n\Big] \leq \sqrt{\frac{A}{n}}.$$

Let η be a parameter to be fixed later and define, for every $m \leq n$ the strategy σ following an usual gradient descend:

$$x'_{m+1} = x_m + \eta\rho(\cdot, b_m) \text{ and } x_{m+1} = \Pi_{\Delta(\mathcal{A})}\left(x'_{m+1}\right), \quad \text{with } x_1 = x'_1 = \left(\frac{1}{A}, \ldots, \frac{1}{A}\right),$$

the projection step ensures that x_{m+1} stays in $\Delta(\mathcal{A})$. Simple calculations show that, for every $a \in \mathcal{A}$,

$$\begin{aligned}
\mathbb{E}_{\sigma,\tau}\left[\sum_{m=1}^n \rho(a, b_m) - \rho_m\right] &= \mathbb{E}_{\sigma,\tau}\left[\sum_{m=1}^n (a - x_m)'\rho(\cdot, b_m)\right] = \sum_{m=1}^n (a - x_m)'\frac{(x'_{m+1} - x_m)}{\eta} \\
&= \frac{1}{2\eta}\sum_{m=1}^n \|a - x_m\|^2 + \|x_m - x'_{m+1}\|^2 - \|a - x'_{m+1}\|^2 \\
&\leq \frac{1}{2\eta}\sum_{m=1}^n \|a - x'_m\|^2 + \|x_m - x'_{m+1}\|^2 - \|a - x'_{m+1}\|^2 \\
&= \frac{1}{2\eta}\|a - x_1\|^2 + \frac{1}{2\eta}\sum_{m=1}^n \eta^2\|\rho(\cdot, b_m)\|^2 \leq \frac{1}{2\eta} + \frac{\eta n A}{2}.
\end{aligned}$$

Balancing the two terms by choosing $\eta = 1/\sqrt{nA}$ proves the claim. We stress out the fact that this strategy ensures that, at stage t, the regret is bounded as

$$\mathbb{E}_{\sigma,\tau}\left[\sum_{m=1}^t \rho(a, b_m) - \rho_m\right] \leq \frac{\sqrt{nA}}{2} + \frac{t\sqrt{A}}{2\sqrt{n}},$$

which might be considerably bigger than $\sqrt{tA}$ for small t, but this uniform guarantee allows the use of a doubling trick, as in Corollary 1.16, to conclude. ■

To get the $\log(A)$ term instead of A in the upper bound, one just has to follow the algorithm known as *exponential weight algorithm*, defined by:

$$x_{n+1}[a] = \frac{\exp\left(\eta_n \rho(a, \overline{b}_n)\right)}{\sum_{a' \in \mathcal{A}} \exp\left(\eta_n \rho(a', \overline{b}_n)\right)} \text{ where } \eta_n = \sqrt{8n \log(A)} \ ,$$

see, e.g., Littlestone & Warmuth [47], Vovk [79] or Auer, Cesa-Bianchi & Gentile [3].

The following corollary shows that the previous result can be extended to the compact case. Actually, the proof is exactly the same, since it did not use the fact that $\mathcal{B}$ is finite, thus is omitted.

Corollary 2.1 *Assume that Nature chooses at every stage an outcome vector U_n in a compact set $\mathcal{U} \subset [0;1]^A$ such that the players payoff at this stage is $U_n^{a_n}$. Then there exists a strategy σ, such that, no matter the strategy τ of Nature and for every $n \in \mathbb{N}$,*

$$\mathbb{E}_{\sigma,\tau}\left[\max_{a^* \in \mathcal{A}} U_n^{a^*} - U_n^{a_n}\right] \leq 2\sqrt{\frac{\log(A)}{n}}.$$

Theorem 2.1 and Corollary 2.1 can actually be proved using more complex optimization procedures, as mirror descent instead of gradient descent (see e.g., Rakhlin [64] or Bubeck [12] for a survey on the use of these techniques in machine learning) and without using doubling trick. We will, on the contrary, prove them using approachability theory.

Remark 2.1 *In Section 1.4.4, we claimed that we could not use a doubling trick. It was possible here because any strategy σ_n, although only optimal at the final stage n, ensures relatively good performance at all stages. For instance, at the specific stage $t = n/2$, the regret is bounded in $\frac{3\sqrt{2}}{4}\sqrt{A/t} \simeq 1.06\sqrt{A/t}$. This was not the case in the previous section, where the distance to the set could be of the order of a constant.*

More specifically, strategies σ_n are somehow equivalent to weak approachability (only the final stage matters). If we could always concatenate strategies using a doubling trick to output a strategy that behaves well at all stages, then we could construct approachability strategy from weak approachability strategies. This would mean that any set is either approachable or excludable, which is not true in general (in fact, as proved in Section 4.1, regret corresponds more to the approachability of convex sets, on which weak and regular approachability coincide).

An usual criticism to the notion of regret in games (and this could lead to long and probably unfruitful debates) is that a player compares his payoff with the payoff he would have got if he had always played the pure action a^*. However, if he had played something else, then Nature would (or at least could) have chosen a totally different sequence b_n so the comparison is meaningless. An easy and unsatisfactory answer is to say that a player's action does not change the behavior of Nature (as in the learning with experts advices literature, see Cesa-Bianchi & Lugosi [14]). A less unsatisfactory answer consists in stating that since there is absolutely no prior on Nature, it is impossible to infer whatsoever on her strategy if the world had been different. So we should compare the payoff with respect to best information available, which is the current sequence.

Let us develop a third point of view, based on game theoretic perspectives. The basic idea is that regret is *not* a criterion to compare different strategies: it does not say that a strategy without regret is a better strategy than always playing a^*. In our repeated game, the player maximizes his cumulated payoff without any structural assumption on Nature. Therefore, he can just sequentially formulate *predictions* upon her behavior (we purposely remain vague on this subject) and play a *best response* to it. Regret is a simple measure on how much a sequence of predictions is *correct* or not. A large regret would mean that the player was most of the time wrong.

2.1.2 Internal and Φ-regret

The notion of external regret has been refined by Foster & Vohra [23] into the so-called *internal regret*. In words, a player has no internal regret (or his strategy is internally consistent) if he has no external regret on the set of stages on which he chose a specific given action.

Formally, choices of action $a_n \in \mathcal{A}$ and $b_n \in \mathcal{B}$ generate, besides an external regret r_n, an internal regret R_n which is an $A \times A$-matrix whose raw are null except the a_n-th one which is r'_n; stated otherwise

$$R_n^{a,a'} = R(a_n, b_n)^{a,a'} := \begin{cases} \rho(a', b_n) - \rho(a_n, b_n) & \text{if } a = a_n \\ 0 & \text{otherwise} \end{cases}.$$

Let us introduce here some notations. Given two sequences $g_n \in \mathbb{R}^d$ and $a_n \in \mathcal{A}$, recall that $\overline{g}_n$ denote the average up to stage n. We define, for every $a \in \mathcal{A}$, the following subset of stages and conditional averages

$$\mathbb{N}_n[a] := \Big\{ m \in \{1, \dots, n\} \text{ s.t. } a_m = a \Big\} \text{ and } \overline{g}_n[a] = \frac{\sum_{m \in \mathbb{N}_n[a]} g_m}{|\mathbb{N}_n[a]|}.$$

Definition 2.2 *A strategy σ is internally consistent if, no matter the strategy τ of Nature, $\mathbb{P}_{\sigma,\tau}$-almost surely,*

$$\limsup_{n\to\infty} \|\overline{R}_n^+\|_\infty \le 0 \text{ or equivalently } \limsup_{n\to\infty} \max_{a\in\mathcal{A}} \frac{|\mathbb{N}_n[a]|}{n} \left(\max_{a^*\in\mathcal{A}} \rho\big(a^*, \overline{b}_n[a]\big) - \overline{\rho}_n[a] \right) \le 0.$$

It is compulsory to multiply the regret accumulated on $\mathbb{N}_n[a]$ by the frequency of action a, namely $|\mathbb{N}_n[a]|/n$. Otherwise internally consistent strategies would not exist. However, another possible formulation (see Lehre & Solan [45]) is to require that

$$\limsup_{n\to\infty} \max_{a^*\in\mathcal{A}} \rho\big(a^*, \overline{b}_n[a]\big) - \overline{\rho}_n[a] \le 0, \quad \text{for every action } a \in \mathcal{A} \text{ s.t. } \lim_{n\to\infty} |\mathbb{N}_n[a]| = \infty,$$

but, unfortunately, this definition does not allow to measure internal regret at a given finite stage n.

Existence part of the following theorem is first due to Foster & Vohra [23]; rates of convergence (constant are not optimal, see e.g. Stoltz & Lugosi [75]) can be inferred from rates of external regret, as showed in Section 2.1.3 where proof is postponed.

Theorem 2.2 *There exist internally consistent strategies such that, for every $n \in \mathbb{N}$*

$$\mathbb{E}_{\sigma,\tau}\left[\max_{a\in\mathcal{A}} \frac{|\mathbb{N}_n[a]|}{n}\left(\max_{a^*\in\mathcal{A}} \rho\big(a^*, \overline{b}_n[a]\big) - \overline{\rho}_n[a]\right)\right] = \mathbb{E}_{\sigma,\tau}\Big[\left\|\overline{R}_n^+\right\|_\infty\Big] \leq 3\sqrt{\frac{\log(A)}{n}}.$$

Regret has been refined further by Blum & Mansour [11] into swap-regret (or Φ-regret). Define, for every mapping $\phi : \mathcal{A} \to \mathcal{A}$, family $\Phi \subset \{\phi : \mathcal{A} \to \mathcal{A}\}$ and $n \in \mathbb{N}$,

$$\overline{R^\phi}_n = \frac{1}{n}\sum_{m=1}^{n} \rho(\phi(a_m), b_m) - \rho(a_m, b_m) \text{ and } \overline{R^\Phi}_n = \left(\overline{R^\phi}_n\right)_{\phi\in\Phi} \in \mathbb{R}^{|\Phi|}.$$

Definition 2.3 *A strategy σ has no Φ-regret if, no matter the strategy τ of Nature, $\mathbb{P}_{\sigma,\tau}$-almost surely,*

$$\limsup_{n\to\infty} \left\|\overline{R^\Phi}_n^+\right\|_\infty \leq 0 \text{ or equivalently } \limsup_{n\to\infty} \max_{\phi\in\Phi} \frac{\sum_{m=1}^n \rho(\phi(a_m), b_m)}{n} - \overline{\rho}_n \leq 0.$$

Existence of such strategies is due to Blum & Mansour [11]; proofs are again delayed.

Theorem 2.3 *There exists strategies without Φ-regret such that, for every $n \in \mathbb{N}$*

$$\mathbb{E}_{\sigma,\tau}\left[\max_{\phi\in\Phi} \frac{\sum_{m=1}^n \rho(\phi(a_m), b_m)}{n} - \overline{\rho}_n\right] = \mathbb{E}_{\sigma,\tau}\Big[\left\|\overline{R^\Phi}_n^+\right\|_\infty\Big] \leq 3\sqrt{\frac{\log(|\Phi|)}{n}}.$$

The notion of Φ-regret is a refinement of respectively external and internal regret, because of the specific choices of families $\Phi_e := \Big\{\phi_{a^*}; \forall a^* \in \mathcal{A}, \phi_{a^*}(a) = a^*, \forall a \in \mathcal{A}\Big\}$ or $\Phi_i := \Big\{\phi_{a',a^*};\ \forall a', a^* \in \mathcal{A}, \phi_{a',a^*}(a') = a^* \text{ and } \phi_{a',a^*}(a) = a \text{ if } a \neq a'\Big\}$. Proposition 2.2 links the different aforementioned quantities, and shows that minimizing internal regret is, in some sense, enough to minimize each one of them (up to the cost of a factor A). We will need the following notation.

Given a family $\Phi \subset \{\phi : \mathcal{A} \to \mathcal{A}\}$, we define the matrix H_Φ of size $|\Phi| \times A^2$ by

$$H_\Phi^{\phi,(a,a')} = 1 \text{ if } \phi(a) = a' \text{ and } 0 \text{ otherwise.}$$

Proposition 2.2 *Given any family $\Phi \subset \{\phi : \mathcal{A} \to \mathcal{A}\}$, one has $\overline{R^\Phi}_n = H_\Phi \overline{R}_n$ where $\overline{R}_n$ is seen as a vector of size A^2. As a consequence, $\left\|\overline{R^\Phi}_n^+\right\|_\infty \leq A\left\|\overline{R}_n^+\right\|_\infty$.*

For the specific case of external regret, one also has $\overline{r}_n = \overline{R}_n\mathbf{1}$ (where $\overline{R}_n$ is seen as a matrix and $\mathbf{1}$ is a vector with only ones). The converse is not true as there exist externally consistent strategies with linear internal regret.

Proof: The proof of the first part follows directly from the definitions of internal and Φ-regret. For the existence of externally consistent strategies with linear internal regret, we refer to Stoltz & Lugosi [75]. ■

Another refinements of these concepts can be made, following this time Fudenberg & Levine [28] and Lehrer [43], in two different directions. The first one is to assume

that regret is computed not at every stages, but only on a restricted subset of stages (that might depend on the history) and the second direction is to consider time varying switch-mapping ϕ. Formally, let $\mathcal{X}$ be an activation function, i.e., $\mathcal{X} : H \times \mathcal{A} \to \{0,1\}$ and $\mathcal{X}(h^n, a_{n+1}) = 1$ indicated that the stage $n+1$ is active. We recall that H stands for the set of all finite histories.. A switch function $\phi : H \times \mathcal{A} \to \mathcal{A}$ indicates that, after the finite history h^n, $\rho(a_{n+1}, b_{n+1})$ will be compared to $\rho(\phi(h^n, a_{n+1}), b_{n+1})$.

Definition 2.4 *Given an activation mapping $\mathcal{X}$ and a switch mapping ϕ, a strategy σ has no $(\mathcal{X}, \phi)$-regret if, no matter the strategy τ of Nature, $\mathbb{P}_{\sigma,\tau}$-almost surely*

$$\limsup_{n\to\infty} \frac{\sum_{m=1}^{n} \mathcal{X}(h^{m-1}, a_m) \left[\rho(\phi(h^{m-1}, a_m), b_m) - \rho_m\right]}{\sum_{m=1}^{n} \mathcal{X}(h^{m-1}, a_m)} \leq 0,$$

as soon as $\sum_{m=1}^{n} \mathcal{X}(h^{m-1}, a_m)$ converges to $+\infty$.

Lehrer [43] has proved that, given a probability λ on the whole set of pairs of activations-switch mappings (embedded with the product topology), there exists a strategy without $(\mathcal{X}, \phi)$-regret, for λ-almost all pairs. However, rates of convergence are not explicit, in part because we divide the score by the number of actives stages $\sum_{m=1}^{n} \mathcal{X}(h^{m-1}, a_m)$ and not by n.

2.1.3 Reductions : form external to Φ-regret

In this section, we show how to construct a strategy with no Φ-regret based on an algorithm that only outputs externally consistent strategies, developing an idea of Stoltz & Lugosi [75] and recovering the more general result of Blum & Mansour [11]. Indeed, consider the following auxiliary game where action sets of the player and Nature are respectively Φ and a compact subset $\mathcal{U} \subset [0;1]^A$. Given an exogenous sequences $p_n \in \Delta(\mathcal{A})$ we define the payoff at stage n of the player generated by the choices of $\phi \in \Phi$ and $U_n \in \mathcal{U}$ by

$$V_n^{\phi} = \sum_{a\in\mathcal{A}} p_n[a] U_n^{\phi(a)} := \sum_{a\in\mathcal{A}} p_n \circ \phi^{-1}[a] U_n^a \ .$$

Let θ be an externally consistent strategy and θ_n^{ϕ} denote the weight put by θ on ϕ at stage n. Then the expected external regret at this stage is written as

$$\begin{aligned} r_n^{\phi'} = V_n^{\phi'} - \mathbb{E}\left[V_n^{\phi_n}\right] &= \sum_{a\in\mathcal{A}} p_n[a] U_n^{\phi'(a)} - \sum_{\phi\in\Phi} \theta_n^{\phi} \sum_{a\in\mathcal{A}} p_n[a] U_n^{\phi(a)} \\ &= \sum_{a\in\mathcal{A}} p_n[a] U_n^{\phi'(a)} - \sum_{a\in\mathcal{A}} \left(\sum_{\phi\in\Phi} \theta_n^{\phi} p_n \circ \phi^{-1}[a] \right) U_n^a \end{aligned}$$

On the other hand, the strategy that dictates to play p_n at stage n in the original game suffers an expected Φ-regret defined by

$$\mathbb{E}\left[R_n^{\phi'}\right] = \sum_{a\in\mathcal{A}} p_n[a] U_n^{\phi'(a)} - \sum_{a\in\mathcal{A}} p_n[a] U_n^a \ .$$

So, as soon as $p_n[a] = \sum_{\phi\in\Phi} \theta_n^\phi p_n \circ \phi^{-1}[a]$ for every $a \in \mathcal{A}$, Φ-regret in the original game and external regret in the auxiliary game coincide exactly (in expectation). And the latter converges to zero, at the same speed of the former, i.e., at rates indicated by Theorem 2.1 and Corollary 2.1.

The existence of such a p_n is a simple consequence of Brouwer fixed point theorem. Indeed, first, notice that θ_n depends only on the past observations, thus is independent of p_n. As a consequence, p_n can be taken as any fixed point of the continuous mapping $p \mapsto \sum_{\phi\in\Phi} \theta_n^\phi p \circ \phi^{-1}$ from the simplex $\Delta(\mathcal{A})$ to itself.

We only have proved the convergence of Φ-regret in expectation; as usual, almost sure convergence is a consequence of concentration inequalities (or see Theorem 2.7, page 47 and Example 1, page 19, in Hall & Heyde [29]).

2.2 Compact action spaces, generalizations and examples

2.2.1 Compact action spaces

Although Φ-regret can be seen as a consequence of external or internal regret in the finite case (when $\mathcal{A}$ is finite), its introduction is more useful in the following compact case.

Assume that $\mathcal{A}$, action space of the player, is no longer finite but a compact subset of some Euclidian space. On the other side, $\mathcal{U}$, action space of Nature, is a subset of mappings from $\mathcal{A}$ to $\mathbb{R}$. Choices of $a_n \in \mathcal{A}$ and $U_n \in \mathcal{U}$ generate, at stage n, a payoff of $\rho_n := U_n(a_n)$.

External regret is defined almost exactly as before, i.e., $r_n : \mathcal{A} \to \mathbb{R}$ is a continuous mapping defined by $r_n(a) = U_n(a) - \rho_n$. In the compact case, we must however be careful in the order of quantifiers when passing to limits: a strategy σ is externally consistent if, for all strategy τ of Nature, $\mathbb{P}_{\sigma,\tau}$-almost surely,

$$\sup_{a^*\in\mathcal{A}} \limsup_{n\to\infty} \overline{U}_n(a^*) - \overline{\rho}_n \leq 0, \text{ or equivalently, } \left\| \limsup_{n\to\infty} \overline{r}_n^+ \right\|_\infty \leq 0.$$

Remark 2.2 *We claimed that order of quantifiers has some importance. Assume that $\mathcal{A} = [0,1]$ and that for every $n \in \mathbb{N}$ and $a \in [0;1]$, $U_n(a) = \mathbb{1}_{a\in(0,1/n)}$. Choosing always the same fixed action a^* gives zero as an asymptotic average payoff, therefore the strategy that plays $a_n = 0$ should not have any regret (neither external, internal, or Φ for that matter).*

On the other hand, for every $N \in \mathbb{N}$, the choice of $a^* = 1/2N$ gives $\overline{U}_N(a^*) = 1$, thus $\limsup_{n\to\infty} \sup_{a^*\in\mathcal{A}} \overline{U}_n(a^*) - \overline{\rho}_n = 1$. This explains the choices in the order of quantifiers in the definition.

Difficulties arise to define internal regret, because scores are multiplied by frequencies of actions in the finite case. We shall instead only focus on Φ-regret, whose definition is also identical: $R_n^\Phi : \Phi \to \mathbb{R}$ is a mapping defined by $R_n^\Phi(\phi) = U_n(\phi(a_n)) - U_n(a_n)$. And a strategy σ has no Φ-regret if, for all strategy τ of Nature, $\mathbb{P}_{\sigma,\tau}$-almost surely,

$$\sup_{\phi\in\Phi} \limsup_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} U_n(\phi(a_n)) - U_n(a_n) \leq 0, \text{ or equivalently, } \left\| \limsup_{n\to\infty} \overline{R^\Phi}_n^+ \right\|_\infty \leq 0.$$

If $\mathcal{A}$ is not compact but $(\mathcal{A}, \mathcal{F}, \mu)$ is a probability space, then external and Φ-regret can also be defined μ-almost surely. The supremum over Φ is simply replaced by for μ-almost every mappings $\phi \in \Phi$.

2.2.2 Generalizations

The whole concept of regret minimization can be extended beyond the comparison of averages of scalar payoffs. Let $g : \mathcal{A} \times \mathcal{B} \to \mathbb{R}^d$ be a vector valued payoff mapping and define a sequence of evaluation mapping $B_n : \left(\mathbb{R}^d\right)^n \to \mathbb{R}$ and a class Ξ of departure sequence $\xi[g] : \mathcal{A} \times \mathcal{B} \to \mathbb{R}^d$. Then a strategy σ has no generalized regret if

$$\sup_{\xi \in \Xi} \limsup_{n \to \infty} B_n\Big(\xi[g](a_1, b_1), \dots, \xi[g](a_n, b_n)\Big) - B_n\Big(g(a_1, b_1), \dots, g(a_n, b_n)\Big) \leq 0 \ .$$

almost surely, no matter the strategy of Nature.

Of course, without additional assumptions on the sequences B_n and Φ, generalized regret cannot be minimized. Rakhlin, Sridharan & Tewari [66] have used the min-max techniques to infer the existence of such strategies (associated with rates of convergences) on specific cases:

i) External, internal and Φ-regret are obtained if $g = \rho$, $B_n(z_1, \dots, Z_n) = \frac{1}{n} \sum_{m=1}^n z_m$ and, for every $\phi \in \Phi$, there exists $\xi \in \Xi$ such that $\xi[g](a, b) = \rho(\phi(a), b)$.

ii) Approachability of a convex $\mathcal{C}$ if $B_n(z_1, \dots, z_n) = -d_{\mathcal{C}}\left(\frac{1}{n} \sum_{m=1}^n z_m\right)$ and the departure mappings are $\xi[g](a, b) \in \mathcal{C}$ for any $a \in \mathcal{A}$ and $b \in \mathcal{B}$.

iii) When B is a function of the average, i.e., $B_n(z_1, \dots, z_n) = G\left(\frac{1}{n} \sum_{m=1}^n z_m\right)$, an interesting (yet maybe counterintuitive) property arises even in the finite case. There might exist strategies that are not externally consistent yet internally consistent, in the sense that,

$$\limsup_{n \to \infty} \max_{x^* \in \Delta(\mathcal{A})} G\Big(\rho(x^*, \overline{b}_n)\Big) - G\Big(\overline{\rho}_n\Big) > 0$$

but for every $a \in \mathcal{A}$

$$\limsup_{n \to \infty} \frac{|\mathbb{N}_n[a]|}{n} \left(\sup_{x^* \in \Delta(\mathcal{A})} G\Big(\rho(x^*, \overline{b}_n\Big) - G\Big(\overline{\rho}_n[a]\Big)\right) \leq 0 \ .$$

2.2.3 Experts

An interpretation – which is actually also a generalization – of these results concerns games of predictions with expert advices, studied (almost exhaustively) by Cesa-Bianchi & Lugosi [14]. At each stage $n \in \mathbb{N}$, an agent must take a decision d_n in some topological convex and compact set $\mathcal{D}$. He is advised by a pool $\mathcal{E}$ of experts, i.e., expert e suggests to choose the decision d_n^e at this stage. Once his choice his made, Nature reveals the state of the world $s_n \in \mathcal{S}$ (where $\mathcal{S}$ is some arbitrary space) which generate a loss $L_n := L(d_n, s_n)$.

After n stages, the agent has suffered an average loss of $\overline{L}_n = \frac{1}{n}\sum_{m=1}^{n} L(d_m, s_m)$ while the best expert had incurred an average loss of $\overline{L}_n^\star = \frac{1}{n}\min_{e\in\mathcal{E}}\sum_{m=1}^{n} L(d_m^a, s_m)$. An evaluation criteria of a strategy of an agent compare these two quantities, as was done by Auer, Cesa-Bianchi & Gentile [3].

Corollary 2.3 *If L is convex and has value in $[0;1]$, then there exists an algorithm such that*

$$\overline{L}_n - \overline{L}_n^\star \leq 2\sqrt{\frac{\log(|\mathcal{E}|)}{n}}.$$

Proof: Consider an externally consistent strategy σ given by Theorem 2.1, where the action set is the set of experts and the payoff at stage n of choosing expert e is $\rho(e, s_n) = -L\left(d_n^e, s_n\right)$. Denote by $x_{n+1} \in \Delta(\mathcal{E})$ the mixed action dictates by σ at stage $n+1$. It induces the decision $d_{n+1} = \sum_{e\in\mathcal{E}} x_{n+1}[e]d_{n+1}^e$ which satisfies, by convexity of L:

$$L(d_{n+1}, s_{n+1}) \leq \sum_{e\in\mathcal{E}} x_{n+1}[e]L(d_{n+1}^e, s_{n+1}) = \mathbb{E}_{\sigma,s_{n+1}}\left[\rho(e, s_{n+1})\right].$$

Therefore $\overline{L}_n - \overline{L}_n^\star$ is smaller than the expected regret of σ, hence the result. ■

2.3 Links with game theory

2.3.1 Regret and sets of equilibria

Existence of consistent strategies can be used to prove classical game theory results: non-emptiness of Hannan (or correlated) sets and min-max theorems, as noticed by Blum & Mansour [11] and Cesa-Bianchi & Lugosi [14].

Consider a game between a set of players $\mathcal{I}$ of size I, where $\mathcal{A}_i$ denotes the finite action space of player i and $\rho_i : \prod_{i\in\mathcal{I}} \mathcal{A}_i \to \mathbb{R}$ his payoff function (extended multi-linearly as usual). Hannan set of player i is the subset of joint distributions of actions defined by

$$\begin{aligned}\mathcal{H}_i &= \left\{ q \in \Delta\left(\prod_{i\in\mathcal{I}} \mathcal{A}_i\right) \;;\; \rho_i\big(a, q_{-i}\big) \leq \rho_i(q), \forall a \in \mathcal{A}_i \right\} \\ &= \left\{ q \in \Delta\left(\prod_{i\in\mathcal{I}} \mathcal{A}_i\right) \;;\; \max_{a^*\in\mathcal{A}_i} \rho_i\big(a^*, q_{-i}\big) - \rho_i(q) \leq 0 \right\},\end{aligned}$$

where $\rho_i(q) = \mathbb{E}_q[\rho_i]$ and q_{-n} is the marginal of q on $\prod_{j\neq i}\mathcal{A}_j$, i.e., the empirical joint distribution of actions played by the opponents of player i. Informally, a joint distribution q belongs to $\mathcal{H}_i$ if player i has no interest to always play a fixed action $a^* \in \mathcal{A}_i$ if his opponents coordinate to play accordingly to q_{-i}.

By linearity of ρ_i, if a strategy of player i is externally consistent (independently of the behavior of its opponents), then necessarily the empirical joint distribution of

actions converges to $\mathcal{H}_i$. We qualify this property as *unilateral*, as it does not make any assumption on opponents' strategies.

If every player follows unilaterally an externally consistent strategy (but not necessary output by the same algorithm), then empirical distributions of actions will converge to the Hannan set of the game, $\overline{\mathcal{H}} = \cap_{i\in\mathcal{I}}\mathcal{H}_i$, which is therefore guaranteed to be non empty.

The main difference between elements of Hannan set and Nash equilibria is that in the latter the distribution must be a product distribution. So set of Nash equilibria is always contained, but might be in some arbitrary game, much smaller than $\overline{\mathcal{H}}$.

On the other hand, in zero sum game, elements of Hannan set satisfy the following property. If $q \in \Delta(\mathcal{A}\times\mathcal{B})$ belongs to $\overline{\mathcal{H}}$, then if we denote by $q_1 \in \Delta(\mathcal{A})$ and $q_2 \in \Delta(\mathcal{B})$ its marginals, necessarily

$$\min_{y\in\Delta(\mathcal{B})}\max_{x\in\Delta(\mathcal{A})}\rho(x,y) \le \max_{a\in\mathcal{A}}\rho(a,q_2) \le \rho(q) \le \min_{b\in\mathcal{B}}\rho(q_1,b) \le \max_{x\in\Delta(\mathcal{A})}\min_{y\in\Delta(\mathcal{B})}\rho(x,y)\ .$$

Since $\max_{x\in\Delta(\mathcal{A})}\min_{y\in\Delta(\mathcal{B})}\rho(x,y) \le \min_{y\in\Delta(\mathcal{B})}\max_{x\in\Delta(\mathcal{A})}\rho(x,y)$ always holds, both quantities must coincide and, by definition, are equal to the value v of the game. More importantly, the first and last inequality above imply that

$$\max_{a\in\mathcal{A}}\rho(a,q_2) = \max_{x\in\Delta(\mathcal{A})}\rho(x,q_2) = v \text{ and } \min_{b\in\mathcal{B}}\rho(q_1,b) = \min_{y\in\Delta(\mathcal{B})}\rho(q_1,y) = v,$$

thus (q_1,q_2) is a pair of *optimal mixed actions*.

As a consequence, in a zero sum game, if players follows unilaterally consistent strategies, they will obtain asymptotically at least the value. And if both players have consistent strategies, their empirical mixed action converges to their set of optimal mixed actions.

This property has been somehow generalized by Hart and Mas-Colell [33] in potential games, see also Viossat & Zapechelnyuk [78]. They have constructed a specific externally consistent strategy such that, if both players follows it, the product of empirical distributions of actions converges to the set of Nash equilibria (and more precisely to a subset of it whose payoff are identical). However, this is only a *global* property (as opposed to unilateral properties) as both players must follow this specific strategy. Moreover, the result does not extend to any game, even those with an unique Nash equilibria.

We proved, following Cesa-Bianchi & Lugosi [14] and Sorin [70], a min-max theorem due to von Neumann using externally consistent strategies. It is actually possible to get the following generalized version of Fan [18]. We first recall that a mapping ρ on $\mathcal{A}\times\mathcal{B}$ is said to be concave-like if for every $a, a' \in \mathcal{A}$ and $\alpha \in [0,1]$, there exists $a^* \in \mathcal{A}$ such that $\rho(a^*,\cdot) \ge \alpha\rho(a,\cdot) + (1-\alpha)\rho(a',\cdot)$. Convexity-like is defined similarly.

Theorem 2.4 *Let $\mathcal{A}$ be a compact set, $\mathcal{B}$ any set and ρ a concave-like convex-like mapping on $\mathcal{A}\times\mathcal{B}$ bounded from below and such that $g(\cdot,b)$ is upper-semicontinuous for every $b\in\mathcal{B}$. Then the zero-sum game on $\mathcal{A}$ and $\mathcal{B}$ has a value.*

Proof: Let $\mathcal{B}'$ be any finite subset of $\mathcal{B}$ and consider an externally consistent strategy of the first player; its existence is ensures by the following Corollary 4.4. It also implies that, for every $\varepsilon > 0$, there exists a sequence $\delta_n \geq 0$ going to zero such that, at stage n,

$$\inf_{b\in\mathcal{B}} \max_{a\in\mathcal{A}} \rho(a,b) \leq \max_{a\in\mathcal{A}} \rho(a,b_n^*) \leq \max_{a\in\mathcal{A}} \frac{1}{n}\sum_{m=1}^{n} \rho(a,b_m) \leq \overline{\rho}_n + \varepsilon + \delta_n \ ,$$

where b_n^* is given by the definition of convexity-like applied to $\sum_{m=1}^{n} b_m/n$. On Nature's side, we can assume that her strategy is such that, at stage n, b_n is an action realizing $\inf_{b\in\mathcal{B}'} \rho(a_n,b)$ up to $1/2^n$. As a consequence,

$$\overline{\rho}_n \leq \frac{1}{n}\sum_{m=1}^{n} \min_{b\in\mathcal{B}} \rho(a_m,b) + \frac{1}{2^m} \leq \min_{b\in\mathcal{B}} \frac{1}{n}\sum_{m=1}^{n} \rho(a_m,b) + \frac{1}{n} \leq \min_{b\in\mathcal{B}'} \rho(a_n^*,b) + \frac{1}{n}$$

where a_n^* is given by the definition of concavity-like. As a consequence, taking n and ε to their limits yields that, for any finite subset $\mathcal{B}'$,

$$\inf_{b\in\mathcal{B}} \max_{a\in\mathcal{A}} \rho(a,b) \leq \sup_{a\in\mathcal{A}} \inf_{b\in\mathcal{B}'} \rho(a,b).$$

Since $\mathcal{A}$ is compact and $\rho(\cdot,b)$ is upper-semicontinuous, for every ε and $\mathcal{B}'$ the set

$$\mathcal{A}_\varepsilon[\mathcal{B}'] = \left\{ a \in \mathcal{A} \text{ s.t } \rho(a,b) \geq \inf_{b\in\mathcal{B}} \max_{a\in\mathcal{A}} \rho(a,b) - \varepsilon, \forall b \in \mathcal{B}' \right\}$$

is a compact non-empty set, and this remains true for any finite intersection over different subsets. As a consequence, the whole intersection (over every ε and $\mathcal{B}'$) remains compact and non-empty, and any point a in it must satisfy that $\rho(a,b) \geq \inf_{b\in\mathcal{B}} \max_{a\in\mathcal{A}} \rho(a,b)$, for every $b \in \mathcal{B}$. Stated otherwise,

$$\inf_{b\in\mathcal{B}} \max_{a\in\mathcal{A}} \rho(a,b) \leq \max_{a\in\mathcal{A}} \inf_{b\in\mathcal{B}} \rho(a,b)$$

and the game has a value. ■

Stronger results can be proved using internally consistent strategies. Aumann [4] defined correlated equilibria in a game as a distribution on the set of profiles of action $q \in \Delta\left(\prod_{i\in\mathcal{I}} \mathcal{A}_i\right)$ such that, for every player $i \in \mathcal{I}$ and every action $a \in \mathcal{A}_i$:

$$\rho_i\big(a, q_{-i}[a]\big) \geq \max_{a^*\in\mathcal{A}_i} \rho_i\big(a^*, q_{-i}[a]\big), \text{ or } q_i[a]\Big(\max_{a^*\in\mathcal{A}_i} \rho_i\big(a^*, q_{-i}[a]\big) - \rho_i\big(a, q_{-i}[a]\big)\Big) \leq 0$$

where $q_{-i}[a] \in \Delta\left(\prod_{j\neq i} \mathcal{A}_j\right)$ is the probability induced by q knowing that $a_i = a$ and $q_i[a]$ is the probability put on $a \in \mathcal{A}_i$ by q (or the relative frequency of action $a \in \mathcal{A}_i$). In words, assume that a referee draws a lottery accordingly to q and only tells player i an action he should play. Then, a correlated equilibrium is a joint distribution such that

every player, when he is told to play action a (and assuming that the others follows their recommendation), cannot gain strictly more by playing a^* instead of a.

It is quite clear (from their very definition) that if every player follows unilaterally an internally consistent strategy then the empirical distribution of actions converges to the set of correlated equilibria (but maybe not to one specific correlated equilibrium), see Foster & Vohra [23].

2.3.2 Regret, (smooth) fictitious play and follow the perturbed leader

Fictitious play is a classic unilateral discrete time dynamic in game theory. At stage n, each player computes empirical (either joint or product) distributions of actions of his opponents and plays a best response to it. Although quite natural, this strategy is not externally consistent. On the contrary, Fudenberg & Levine [28] have introduced a slight modification, called *smooth fictitious play* that has asymptotically a regret smaller than ε (where $\epsilon > 0$ is fixed), see also Hofbauer, Sorin & Viossat [36].

Let ρ_ε denotes an ε-perturbation of ρ (induced by $\psi : \Delta(\mathcal{A}) \to \mathbb{R}$) defined by

$$\rho_\varepsilon(x, y) = \rho(x, y) + \varepsilon\psi(x), \ \forall y \in \Delta(\mathcal{B}).$$

Since we are interested in unilateral procedure, we might as well make a change of variable by defining $U = \Big(\rho(a, y)\Big)_{a\in\mathcal{A}} \in [0; 1]^A$ so that $\rho(x, U) = \langle x, U\rangle$. As a consequence, the mapping ρ_ε can be rewritten as

$$\rho_\varepsilon(x, U) = \langle x, U\rangle + \varepsilon\psi(x).$$

We also define the ε-best response mapping by $\mathrm{BR}_\varepsilon(U) = \mathrm{argmax}_{x\in\Delta(cA)}\langle x, U\rangle + \psi(x)$.

We assume that the mapping $\psi : \Delta(\mathcal{A}) \to \mathbb{R}$ is chosen so that

i) ψ is a continuously differentiable mapping and $\|\psi\|_\infty \leq 1$;

ii) The ε-best response mapping BR_ε is univoque and continuous;

iii) $\mathrm{BR}_\varepsilon(U)$ does not belong to the boundary of $\Delta(\mathcal{A})$.

Actually, point iii) ensures that ρ_ε attains its maximum at a point where its first derivative vanishes. It can therefore be weaken into one of the following

iii') for every $U \in [0; 1]^A$, $D_1\rho_\varepsilon(\cdot, U)$ is equal to zero at $x = \mathrm{BR}_\varepsilon(U)$

iii") $D_1\rho_\varepsilon(\cdot, U)$ is orthogonal to the gradient of BR_ε at U.

Study of $\sigma(h^n) = \mathrm{BR}_\varepsilon(\overline{U}_n)$, the strategy associated with this perturbation, might be simpler in continuous time. First, we introduce the mapping $W : [0; 1]^A \to \mathbb{R}$ defined by

$$W_\varepsilon(U) = \sup_{x\in\Delta(\mathcal{A})} \rho_\varepsilon(x, U) = \langle \mathrm{BR}_\varepsilon(U), U\rangle + \varepsilon\psi(\mathrm{BR}_\varepsilon(U)) \ .$$

In particular, because of point i), regret is asymptotically smaller than 2ε as soon as $\limsup_{n\to\infty} W_\varepsilon(\overline{U}_n) - \overline{\rho}_n \leq \varepsilon$. As in Section 1.4.1, the continuous-time dynamic associated with the discrete-time dynamic of $(\overline{U}_n, \overline{\rho}_n)$ is

$$\Big(\dot{\mathbf{U}}, \dot{\boldsymbol{\rho}}\Big) \in \Big\{(V, \langle \mathrm{BR}_\varepsilon(\mathbf{U}), V\rangle \ ; \ V \in [0;1]^A\Big\} - \Big(\mathbf{U}, \boldsymbol{\rho}\Big).$$

Define $\lambda(t) = W_\varepsilon(\mathbf{U}(t)) - \boldsymbol{\rho}(t)$ then one has $\dot{\lambda} + \lambda \leq \varepsilon$ thus $\lambda(t) \leq \varepsilon + Me^{-t}$ for some constant M. As a consequence, λ is a Lyapounov function with respect to the set

$$\Big\{(U, \rho) \in \mathbb{R}^A \times \mathbb{R} \ ; \ W_\varepsilon(U) - \rho \leq \varepsilon\Big\}$$

which is thus a global attractor of the dynamic (see Benaïm, Hofbauer & Sorin [7]). So $(\overline{U}_n, \overline{\rho}_n)$ converges almost surely to it and the strategy is ε-externally consistent. Benaïm & Faure [6] proved recently that external consistency can be achieve (without requiring a doubling trick argument) with a smooth fictitious play with a vanishing step size; indeed, they showed that if ε is not fixed but depends on $n \in \mathbb{N}$ as $\varepsilon_n = n^\gamma$, with $\gamma < 1$, then asymptotically the regret converges to zero.

Smooth fictitious play (also known as *follow the regularized leader*) is a generalization of two classes of algorithms, *exponential weight algorithms* or its even more general version called *follow the perturbed leader* (see Cesa-Bianchi & Lugosi [14], Sections 4.2 and 4.3). To recover the first class of algorithms, entropy must be used as regularization, i.e., $\psi(x) = -\sum_{a\in\mathcal{A}} x[a]\log(x[a])$ to get

$$\mathrm{BR}_{1/\eta}(U)[a] = \frac{\exp(\eta\, U^a)}{\sum_{a'\in\mathcal{A}} \exp(\eta\, U^a)}$$

which is, by definition, the exponential weight algorithm.

Links with *follow the perturbed leader* (or Stochastic Fictitious Play accordingly to Fudenberg & Kreps [26]) might be a bit more tedious. This algorithm does not choose a deterministic regularization $\varepsilon\psi$ but perturbs each component of $\overline{U}_n$ by a random quantity ε_n^a, such that the joint density $f : \mathbb{R}^A \to \mathbb{R}$ of the vector $\big(\varepsilon_n^a\big)_{a\in\mathcal{A}}$ is independent of $\overline{U}_n$ and n. Action played at stage $n+1$ is any maximizer of $\overline{U}_n^a + \varepsilon_n^a$. In particular, a given action a is chosen at this stage with probability $X^a(\overline{U}_n)$ where $X^a(\cdot)$ is defined by

$$X^a(U) = \mathbb{P}\big\{\, \mathrm{argmax}_{a'\in\mathcal{A}}\, U^{a'} + \varepsilon^{a'} = a\big\}\ .$$

Follow the Perturbed Leader generates a discrete stochastic process $(\overline{U}_n, \overline{\rho}_n)$ which is an A.S.D. of the following differential inclusion

$$\Big(\dot{\mathbf{U}}, \dot{\boldsymbol{\rho}}\Big) \in \Big\{\ (V, \langle X(\mathbf{U}), V\rangle)\ \ ;\ V \in [0;1]^A\Big\} - \Big(\mathbf{U}, \boldsymbol{\rho}\Big)\ .$$

This is a special case of Smooth Fictitious Play since, as soon as f is positive and X is continuously differentiable, Hofbauer & Sandholm [35] have shown that there exists a deterministic regularization $\varepsilon\psi$ such that $X(U) = \mathrm{BR}_\varepsilon(U)$. For example, in the case

where ε^a are i.i.d. with cumulative distribution $F(x) = \exp\big(-\exp(-\eta x - \gamma)\big)$ (where γ the Euler constant), follow the perturbed leader coincides exactly with exponential weight algorithm (see e.g., Lemma 1 in McFadden [54]).

As mentioned before, proofs based on A.S.D. do not exhibit rates of convergences (and this might be seen a major drawback of these techniques). However, we only considered here strategies that do not depend on the past sequence of player's actions (but only on the sequence of Nature's choices). So the discrete process is very closed to the one induced by procedures in law (this is not the case for approachability, see Section 1.3.1) which is in turn close to the continuous-time process. And it is actually possible to quantify explicitly these relative differences, see e.g., Sorin [72] or Kwon [40], in order to recover exact rates of convergence.

3 Calibration

We recall that calibration is a criterion introduced by Dawid [16] in the following repeated games between a player and Nature. At each stage $n \in \mathbb{N}$, Nature chooses a state of the world ω_n in some finite set Ω and the player makes a prediction upon its law by choosing a probability distribution $p_n \in \Delta(\Omega)$. Strategies of the player and Nature are mappings from the set of finite histories $\cup_{n\in\mathbb{N}}(\Omega \times \Delta(\Omega)^n$ into, respectively, $\Delta(\Delta(\Omega))$ and $\Delta(\Omega)$.

The usual example consisting of a meteorologist that predicts each day the probability of rain corresponds to $\Omega = \{0, 1\}$, with $\omega = 1$ if it rains. This binary case is in fact much easier than the general case, as discussed in Section 3.1.2.

3.1 Finite (ε and grid) calibration

We will need the following notations. For every $p \in \Delta(\Omega)$ – seen as a subset of $\mathbb{R}^{\Omega-1}$ – and $\varepsilon > 0$, let $\mathbb{N}_n[p, \varepsilon]$ be the set of stages where the prediction was ϵ-close to p, i.e.,

$$\mathbb{N}_n[p, \varepsilon] = \Big\{m \in \{1, \dots, n\} \text{ s.t. } \|p_m - p\| \leq \varepsilon\Big\},$$

where $\|\cdot\|$ is an Euclidian norm of $\mathbb{R}^{\Omega-1}$. We denote by $\overline{\omega}_n[p, \varepsilon] \in \Delta(\Omega)$ the empirical distribution of states on $\mathbb{N}_n[p, \varepsilon]$ and by $\overline{p}_n[p, \varepsilon]$ the average prediction on it.

Definition 3.1 *A strategy σ of the player is ε-calibrated if for every strategy τ of Nature, and for every $p \in \Delta(\Omega)$,*

$$\limsup_{n\to\infty} \frac{|\mathbb{N}_n[p, \varepsilon]|}{n} \left(\Big\| \overline{p}_n[p, \varepsilon] - \overline{\omega}_n[p, \varepsilon] \Big\| - \varepsilon \right) \leq 0, \quad \mathbb{P}_{\sigma,\tau}\text{-as.}$$

A strategy is calibrated if it is ε-calibrated, for every $\varepsilon > 0$.

Intuitively, a strategy is ε-calibrated if on the set of stages (assuming that it is big enough) where the prediction was ε-close to some $p \in \Delta(\Omega)$, the empirical distribution of states is close to this specific p. Although not stated explicitly in Definition 3.1, it

is possible to require that rates of convergence are independent of Nature's strategy see Section 4.2 below. With a careful concatenation of ε-calibrated strategies, following the doubling trick, one can easily obtain a calibrated strategies, as did Foster & Vohra [24] or Fudenberg & Levine [27]. It remains to construct such strategies, which can be done using the slightly weaker concept of calibrated strategies with respect to an ε-grid of $\Delta(\Omega)$ defined below.

We recall that a finite subset $\big\{x[\ell]\ ;\ \ell \in \mathcal{L}\big\}$ of $\mathcal{K} \subset \mathbb{R}^d$ is an ε-grid of $\mathcal{K}$ if for every $x \in \mathcal{K}$, there exists $\ell \in \mathcal{L}$ such that $\|x - x[\ell]\| \leq \varepsilon$. Moreover, such a grid is regular if there exists $\{e_1, \ldots, e_d\}$, d linearly independent vectors, such that

$$\Big\{x[\ell]\ ;\ \ell \in \mathcal{L}\Big\} = \left\{\sum_{k=1}^{d} n_k e_k\ ;\ n_k \in \mathbb{Z}\right\} \cap \mathcal{K}\ .$$

Assume that the player can only make predictions on a grid $\big\{p[\ell]\ ;\ \ell \in \mathcal{L}\big\}$ of $\Delta(\Omega)$, so that a strategy is a mapping from the finite histories into $\Delta(\mathcal{L})$. Empirical distribution of states on $\mathbb{N}_n(\ell) := \big\{m \in \{1, \ldots, n \text{ s.t. } p_m = p[\ell]\big\}$ is denoted by $\overline{q}_n[\ell]$.

Definition 3.2 *A strategy σ of the player is calibrated with respect to $\big\{p[\ell]\ ;\ \ell \in \mathcal{L}\big\}$ if for every strategy τ of Nature, for every $\ell \in \mathcal{L}$,*

$$\limsup_{n\to\infty} \frac{|\mathbb{N}_n[\ell]|}{n}\left(\big\|\overline{\omega}_n[\ell] - p[\ell]\big\| - \min_{k\in\mathcal{L}} \big\|\overline{\omega}_n[\ell] - p[k]\big\|\right) \leq 0, \quad \mathbb{P}_{\sigma,\tau}\text{-as.}$$

In words, a strategy is calibrated with respect to a grid if on the set of stages where $p[\ell]$ is predicted, the empirical distributions of states is closer to $p[\ell]$ than to any other $p[k]$.

Remark 3.1 *Given a finite grid, the Voronoï cell associated with $p[\ell]$ is the set of points closer to $p[\ell]$ than to any other $p[k]$, i.e.,*

$$V[\ell] := \Big\{p \in \Delta(\Omega)\ \text{s.t.}\ \big\|p - p[\ell]\big\| \leq \min_{k\in\mathcal{L}} \big\|p - p[k]\big\|\Big\}.$$

Each Voronoï is a polytope since they are defined by a finite number of linear inequalities, their union covers $\Delta(\Omega)$ and any intersection has empty interior. The fact that the calibration score

$$\frac{|\mathbb{N}_n[\ell]|}{n}\left(\big\|\overline{\omega}_n[\ell] - p[\ell]\big\| - \min_{k\in\mathcal{L}} \big\|\overline{\omega}_n[\ell] - p[k]\big\|\right)$$

is non positive means that $\overline{\omega}_n[\ell]$ belongs to (or converges to) the Voronoï cell $V[\ell]$.

Dawid [17] and Oakes [57] proved that there does not exist deterministic ε-calibrated strategies, based on a counter example given in the following section. On the other hand, there exists random ε-calibrated strategies, as proved by Foster and Vohra [23] by exhibiting an algorithm to construct makes the so-called Brier score decrease to zero.

Theorem 3.1 *For every grid, there exists a calibrated strategy with respect to it. As a consequence, for every $\varepsilon > 0$, there exist ε-calibrated strategies, and thus calibrated strategies.*

To end this section, we note that finite calibration can also be defined with respect to some weights $\{\nu[\ell] \in \mathbb{R}; \ell \in \mathcal{L}\}$. A strategy σ of the player is weighted-calibrated with respect to $\big\{p[\ell], \nu[\ell]\ ;\ \ell \in \mathcal{L}\big\}$ if for every strategy τ of Nature, for every $\ell \in \mathcal{L}$,

$$\limsup_{n\to\infty} \frac{|\mathbb{N}_n[\ell]|}{n}\Big(\Big(\big\|\overline{\omega}_n[\ell]-p[\ell]\big\|^2-\nu[\ell]\Big)-\Big(\min_{k\in\mathcal{L}}\big\|\overline{\omega}_n[\ell]-p[k]\big\|^2-\nu[k]\Big)\Big)\leq 0,\quad \mathbb{P}_{\sigma,\tau}\text{-as.}$$

Corollary 3.1 *For every grid and weights, there exists a calibrated strategy with respect to them.*

Given $\big\{p[\ell], \nu[\ell]\ ;\ \ell \in \mathcal{L}\big\}$, the Laguerre cell (or Power cell) associated with $p[\ell]$ and $\nu[\ell]$ is

$$P[\ell] := \Big\{p \in \Delta(\Omega) \text{ s.t. } \big\|p-p[\ell]\big\|^2-\nu[\ell] \leq \min_{k\in\mathcal{L}}\big\|p-p[k]\big\|^2-\nu[k]\Big\};$$

as in Remark 3.1, a weighted-calibrated strategy ensures that $\overline{\omega}_n[\ell]$ converges, as soon as the frequency of ℓ is not zero, to $P[\ell]$. Because of the squared norms, this set is also a polytope.

3.1.1 Discussion on the impossibility of deterministic calibration

When $\Omega = \{0,1\}$, Oakes [57] and Dawid [17] output an example of Nature's strategy ensuring that no ε-deterministic calibrated strategies exist. Their idea is actually quite simple yet highly unstable. Define the strategy as follows: given the past history h^n,

$$\text{if } p_{n+1} \geq \frac{1}{2} \text{ then } \omega_{n+1} = 0 \text{ and if } p_{n+1} < \frac{1}{2} \text{ then } \omega_{n+1} = 1;$$

In words, if the forecaster claims that it will rain with high probability then Nature does not make it rain and if it claims that it will not rain, Nature makes it rain.

This prevents any deterministic strategies from being ε-calibrated, but this is not immediate (and the proof, although quite simple will shed lights on the following discussion). We distinguish two cases, either the predictions of $1/2$ have an asymptotic positive frequency or a null frequency, i.e., if

$$\text{either } \limsup_{n\to\infty} \frac{\Big|\big\{m \leq n \text{ s.t. } p_m = 1/2\big\}\Big|}{n} > 0 \quad \text{or} \quad \lim_{n\to\infty} \frac{\Big|\big\{m \leq n \text{ s.t. } p_m = 1/2\big\}\Big|}{n} = 0.$$

In the first case, $\overline{p}[1/2+\varepsilon,\varepsilon] \geq 1/2$ while $\overline{q}[1/2+\varepsilon,\varepsilon] = 0$ thus such a strategy is not ε-calibrated.

In the second case, we can assume that no prediction falls exactly at $1/2$ (since their frequency goes to zero). If the predictions bigger than $1/2$ have an asymptotic positive frequency, then necessarily, there must exist p^* such that the set of stages where predictions belong to $[p^\star-\varepsilon, p^*+\varepsilon] \subset [1/2;1]$ also has a positive frequency. And since $\overline{p}_n[p^*,\varepsilon] \geq 1/2$ and $\overline{\omega}_n[p^*,\varepsilon] = 0$, the strategy is not ε-calibrated.

If the predictions bigger than 1/2 have an asymptotic null frequency, then necessarily the predictions smaller than 1/2 have an asymptotic positive frequency, and the same arguments hold (because we assumed that no predictions were equal to exactly 1/2). So no deterministic strategy can be ε-calibrated.

On the other hand, consider the deterministic strategy of the player that predicts at odd stages $p_n = 1/2$ and at even stages $p_n = 1/2 - 1/n$. The only accumulation point of the sequence of predictions if 1/2, so for every $p \neq 1/2$ there are a finite number of prediction ε-close to p, for every ε smaller than some $\varepsilon_p > 0$. And on the other hand, for $p = 1/2$, no matter ε, if n is big enough, $\mathbb{N}_n[0.5, \varepsilon]$ contains approximatively half predictions below 1/2 and half above, so the empirical distributions is asymptotically equal to 1/2. As a consequence, no matter $\varepsilon > 0$ and $p \neq 0.5$,

$$\limsup_{n\to\infty} \frac{|\mathbb{N}_n[0.5,\varepsilon]|}{n} \Big\| \overline{p}_n[0.5,\varepsilon] - \overline{\omega}_n[0.5,\varepsilon] \Big\| = 0 \text{ and } \lim_{\varepsilon'\to 0} \limsup_{n\to\infty} \frac{|\mathbb{N}_n[p,\varepsilon']|}{n} = 0.$$

Obviously, this does not contradict Oakes [57] and Dawid [17] counter-example. The reason is that, on the stages when the prediction is *ε-close to* $p^* = 1/2 + \varepsilon$, the average prediction is 1/2 while the empirical state is 0. But one might argue that predictions are actually *never close* to p^* (but ε-away), so Oakes and Dawid argument fails if calibration was defined only with respect only to those points p that are accumulation points of the sequence of predictions (i.e., there are predictions arbitrarily close to them).

This argument can be generalized to any stationary strategy of Nature (i.e., if $\omega_n = f(p_n)$ for some fixed but possibly random mapping f). Unfortunately, we are unable to claim that there exists deterministic (ε-)calibrated strategies with respect to accumulation points, but this shows how the very concept of calibration is unstable with respect to small variations in definition or objectives. This subject is somehow once again developed in Section 3.3.

3.1.2 Efficient calibration in the binary case

Foster [21] has designed an algorithm that computes efficiently an ε-strategy in the binary case (although it seems that it was Abernethy, Bartlett & Hazan [1] that noticed its efficiency). The idea is to consider a calibrated strategy with respect to the regular grid $\big\{p[\ell] := \varepsilon + 2\ell\varepsilon \ ; \ \ell \in \mathcal{L}\big\}$ where $\mathcal{L} := \big\{0, 1, \ldots, (\lfloor \varepsilon^{-1} \rfloor - 1)/2\big\}$

Following Foster's notation, we define, for every $\ell \in \mathcal{L}$,

$$e_n^\ell = \frac{|\mathbb{N}_n[\ell]|}{n}\Big(\overline{\omega}_n[\ell] - (p[\ell] + \varepsilon)\Big) \text{ and } d_n^\ell = \frac{|\mathbb{N}_n[\ell]|}{n}\Big(\overline{\omega}_n[\ell] - (p[\ell] - \varepsilon)\Big)$$

so that a strategy is calibrated if, asymptotically, every e_n^ℓ and d_n^ℓ are smaller than zero. Foster's algorithm consists in finding at stage n an element $\ell^* \in \mathcal{L}$ such that

- either both $e_n^{\ell^*} \leq 0$ and $d_n^{\ell^*} \leq 0$; in that case, predict $p[\ell^*]$
- or $e_n^{\ell^*-1} > 0$ and $d_n^{\ell^*} > 0$; in that case play $p[\ell^*]$ or $p[\ell^* - 1]$ with a respective probability proportional to $d_n^{\ell^*}$ and $e_n^{\ell^*-1}$.

Existence of such a ℓ^* is ensured by the fact that the first d_n^1 and the last e_n^L are always non positive. Computations show that the error converges to zero.

So the tricky remaining part consists in finding efficiently this ℓ^*. To this purpose, Abernethy, Bartlett & Hazan [1] introduced, for every $\ell \in \mathcal{L}$, the quantity

$$\theta_n^\ell = e_n^\ell \text{ if } e_n^\ell > 0, \quad \theta_n^\ell = -d_n^\ell \text{ if } d_n^\ell > 0 \quad \text{and} \quad \theta_n^\ell = 0 \quad \text{otherwise},$$

which is well defined since e_n^ℓ and d_n^ℓ cannot be simultaneously positive. Specifically, it always holds that $\theta_n^1 \geq 0$ and $\theta_n^L \leq 0$ so if any of them is equal to zero, Foster's strategy dictates to predict it. Otherwise, one must find ℓ^* such that $\theta_n^{\ell^*-1} > 0$ and $\theta_n^{\ell^*} < 0$ and the main argument is that it can be done through a binary search, thus in $O\left(\log(1/\varepsilon)\right)$ steps.

Foster's strategy can be somehow generalized with more than two outputs (see e.g., Mannor & Stoltz [50]) although, unfortunately, at the cost of efficiency since the binary search trick does not extend.

3.2 Generalization

Recall that, roughly speaking, a strategy is calibrated if on the set of stages where the prediction was *close* to p, the average prediction and the empirical distribution of outcome asymptotically coincide. General concepts of calibration are induced by a different definition of *closeness*.

Let $\mathcal{F}$ be a family of Borel measurable subsets of $\Delta(\Omega)$ and denote, for every $F \in \mathcal{F}$,

$$\mathbb{N}_n[F] = \Big\{ m \leq n \text{ s.t. } p_m \in F \Big\}, \ \overline{\omega}_n[F] = \frac{\sum_{m \in \mathbb{N}_n[F]} \omega_m}{|\mathbb{N}_n[F]|} \text{ and } \overline{p}_n[F] = \frac{\sum_{m \in \mathbb{N}_n[F]} p_m}{|\mathbb{N}_n[F]|}$$

respectively the set of stages where the prediction was in F (before the n-th), the empirical distribution of outcomes and the average prediction on it.

Definition 3.3 *A strategy σ of the player is $\mathcal{F}$-calibrated if for every strategy τ of Nature,*

$$\limsup_{n\to\infty} \sup_{F \in \mathcal{F}} \frac{|\mathbb{N}_n[F]|}{n} \Big\| \overline{\omega}_n[F] - \overline{p}_n[F] \Big\| \leq 0, \quad \mathbb{P}_{\sigma,\tau}\text{-}as\ .$$

Several types of families have been considered by various authors. For instance, from the most to the least complicated

i) Mannor & Stoltz [50] treated the most difficult problem where $\mathcal{F}$ is the family of all Borel measurable subsets of $\Delta(\Omega)$;

ii) Rakhlin, Sridharan & Tewari [66] considered the family of every possible ℓ_1 balls;

ii) Perchet [60] defined $\mathcal{F}$ to be some neighborhood basis of $\Delta(\Omega)$.

In the first case, the minmax techniques of Rakhlin, Sridharan and Tewari [65, 66] upper-bound the calibration error at stage n (but with a strategy that depends on n) in $O\left(n^{-1/(|\Omega|+1)}\right)$ while for the two last cases the bound shrinks to $O\left(n^{-1/2}\right)$. On the other hand, Mannor & Stoltz [50] and Perchet [60] obtained (actually before) the same results, yet in a constructive way. They are developed in Section 4.2.

Drawbacks of these definitions of calibration (which will lead to another type of generalization) are illustrated by the following examples.

Assume that $\Omega = \{0, 1\}$ and that the sequence of outcomes is $0, 1, 0, 1, 0...$ (i.e., $\omega_n = 1$ iff n is even). Consider a player that predicts, at every stage, that the probability of 1 is exactly $1/2$. Then this strategy is calibrated accordingly to any of the previous definitions of calibration. On the other hand, on the set of even stages, empirical distribution is 1 while average prediction is $1/2$ which contradict precepts of calibration.

Even more intricate: assume that $\Omega = \{0, 1, 2\}$ that $\omega_n = 0$ with probability $1/3$ and that 1 and 2 alternates on the remaining set of stages. The sequence of outcomes on any fixed subset of $\mathbb{N}$ contains asymptotically as many 0 than 1 and 2 so predicting $1/3, 1/3, 1/3$ at every stage is not contradicting. On the other hand, if we consider only the set of stages where the outcome was 1 or 2 then the prediction is always $1/2, 1/2$ while 1 and 2 alternate.

We introduce the following concepts of checking rules. Let $\mathcal{U}$ and $\mathcal{T}$ be respectively an *active universe mapping* and a *testing mapping*, i.e.,

$$\mathcal{U} : \bigcup_{n\in\mathbb{N}} (\Delta(\Omega) \times \Omega)^n \rightrightarrows \Delta(\Omega) \times \Omega \quad \text{and} \quad \mathcal{T} : \bigcup_{n\in\mathbb{N}} (\Delta(\Omega) \times \Omega)^n \rightrightarrows \Delta(\Omega) \times \Omega$$

such that $\mathcal{T}(h^n) \subset \mathcal{U}(h^n)$. The interpretation is that stage $n+1$ is *active* if (p_{n+1}, ω_{n+1}) belongs to the active universe $U(h^n)$; given a set of active stages, calibration compares the empirical frequency of the *tested* event with the average prediction of this event.

Such a pair $(\mathcal{U}, \mathcal{T})$ forms a *checking rule* and we define as before the set of active stages

$$\mathbb{N}_n[\mathcal{U}, \mathcal{T}] = \left\{ m \leq n \text{ s.t. } (p_m, \omega_m) \in \mathcal{U}(h^{m-1}) \right\} ,$$

the empirical probability of tested events

$$\overline{\omega}_n[\mathcal{U}, \mathcal{T}] = \frac{\sum_{m\in\mathbb{N}_n[\mathcal{U},\mathcal{T}]} \mathbb{1}\{(p_m, \omega_m) \in \mathcal{T}(h^{m-1})\}}{|\mathbb{N}_n[\mathcal{U}, \mathcal{T}]|} ,$$

and the average predicted conditional probability of tested events

$$\overline{p}_n[\mathcal{U}, \mathcal{T}] = \frac{\sum_{m\in\mathbb{N}_n[\mathcal{U},\mathcal{T}]} p_m \left\{ (p_m, \omega_m) \in \mathcal{T}(h^{m-1}) \,\middle|\, \mathcal{U}(h^{m-1}) \right\}}{|\mathbb{N}_n[\mathcal{U}, \mathcal{T}]|} .$$

Definition 3.4 *A strategy σ is calibrated with respect to some given checking rule $(\mathcal{U}, \mathcal{T})$ if, for every strategy τ of Nature,*

$$\limsup_{n\to\infty} \frac{|\mathbb{N}_n[\mathcal{U}, \mathcal{T}]|}{n} \left| \overline{\omega}_n[\mathcal{U}, \mathcal{T}] - \overline{p}_n[\mathcal{U}, \mathcal{T}] \right| \leq 0, \quad \mathbb{P}_{\sigma,\tau}\text{-}as ,$$

with the assumption that $p\{A|B\} = +\infty$ *if* $p\{B\} = 0$.

The following theorem (a weaker version first appeared in Lehrer [41]) continues the discussion of Section 3.1.1 and weakens furthermore the range of the counterexample of Oakes and Dawid. It shows that if checking rules do not depend on current predictions (but possibly on past predictions), then deterministic calibration does exist; this is quite obvious if one faces only one checking rule, but the result actually holds with an infinite number of them.

To be formal, we embed the set of checking-rules independent of current prediction (i.e. pairs of mappings from $\bigcup_{n\in\mathbb{N}}(\Delta(\Omega)\times\Omega)^n$ into Ω) with the cylinder topology.

Theorem 3.2 *Let* λ *be a probability distribution on the set of checking-rules independent of current predictions. Then there exists a deterministic strategy* σ *that is calibrated with* λ*-almost every checking rules.*

Actually, the result that we shall prove is stronger as we will show that, as soon as $|\mathbb{N}_n[\mathcal{U},\mathcal{T}]|$ goes to infinity, $\limsup_{n\to\infty}\left|\overline{q}_n[\mathcal{U},\mathcal{T}]-\overline{p}_n[\mathcal{U},\mathcal{T}]\right|\leq 0$, $\mathbb{P}_{\sigma,\tau}$-as.

A similar result (that extends Foster & Vohra [24]) due to Sandroni, Smorodinsky & Vohra [68] deals with checking rules depending on current predictions, under the following extra assumptions. We assume that the calibration test compares the empirical distribution of outcomes with the average prediction on the set of *active* stages where predictions were in some given set $F\subset\Delta(\Omega)$; activeness of stages might depend on past histories. Formally, $\mathcal{U}(h^n)$ is either empty (so the stage $n+1$ is not active) or $\mathcal{U}(h^n)=F\times\Omega$. Mapping $\mathcal{T}$ is, on the other side, constant, i.e., $\mathcal{T}(h^n)=F\times\{\omega\}$ for some $\omega\in\Omega$ (at least on active stages).

Proposition 3.2 *Consider a countable number of such checking rules. Then there exists a strategy of the player that is calibrated with every one of them.*

3.3 Smooth calibration

Smooth calibration is another criterion (close to usual calibration) that can be satisfied with a deterministic strategy, as proved by Foster and Kakade. Even more surprisingly, it can be used to output a calibrated strategy showing again the instability of Oakes and Dawid's result.

The idea is to *smooth* definitions of calibrations. Indeed, notice that given $F\subset\Delta(\Omega)$, the calibration score can be written as

$$\frac{|\mathbb{N}_n[F]|}{n}\left\|\overline{\omega}_n[F]-\overline{p}_n[F]\right\|=\frac{1}{n}\left\|\sum_{m=1}^{n}\mathbb{1}\{p_m\in F\}(\omega_m-p_m)\right\|$$

and the mapping $p\mapsto\mathbb{1}\{p\in F\}$ is not continuous. Instead, given some continuous mapping $g:\Delta(\Omega)\to[0,1]$, consider the following smoothen version of the score

$$\frac{1}{n}\left\|\sum_{m=1}^{n}g(p_m)(\omega_m-p_m)\right\|;$$

with respect to some checking rule $(\mathcal{U}, \mathcal{T})$ independent of the current predictions (so that $\mathcal{U}(h^n)$ and $\mathcal{T}(h^n)$ can be seen as subset of Ω), this score becomes

$$\frac{1}{n}\left\|\sum_{m=1}^{n} g(p_m)\mathbb{1}\{\omega_m \in \mathcal{U}(h^{m-1})\Big(\mathbb{1}\{\omega_m \in \mathcal{T}(h^{m-1}) - p_m\{\mathcal{T}(h^{m-1})|\mathcal{U}(h^{m-1})\}\Big)\right\| .$$

A weaker version of the following Proposition has been proved independently by Kakade & Foster [37] and Vovk, Nouretdinov, Takemura & Shafer [80]; the former named this property *weak calibration*, but we used the term weak in another meaning (i.e., when horizon of the game is fixed and known).

Proposition 3.3 *There exists a deterministic strategy σ of the player such that, no matter Nature's strategy, for every continuous mapping $g : \Delta(\Omega) \to \mathbb{R}_+$,*

$$\limsup_{n\to\infty} \frac{1}{n}\left\|\sum_{m=1}^{n} g(p_m)(\omega_m - p_m)\right\| \leq 0 .$$

If μ is a probability distribution on the set of checking rules independent of current predictions, then there exists a deterministic σ such that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{m=1}^{n} g(p_m)\mathbb{1}\{\omega_m \in \mathcal{U}(h^{m-1})\Big(\mathbb{1}\{\omega_m \in \mathcal{T}(h^{m-1}) - p_m\{\mathcal{T}(h^{m-1})|\mathcal{U}(h^{m-1})\}\Big) \leq 0,$$

for μ-ae checking rule and every continuous mapping g, no matter Nature's strategy.

As noticed by Foster and Kakade, the convergence in first part of the result can be made uniform with respect to Nature's strategy.

Actually, the most surprising and interesting property of smooth calibration is not so much that there exist deterministic smooth calibrated algorithms, but that they can be used to construct an *almost deterministic* ε-calibrated strategy as follows, see Kakade & Foster [37] for more details

Let ε be fixed and consider a finite ε-triangulation of $\Delta(\Omega)$ whose vertices are $\mathcal{V} := \{v_1, \ldots, v_V\}$. Any $p \in \Delta(\Omega)$ belongs to one simplex of the triangulation and we denote by $V(p)$ its vertices (if there are more than one simplices, then choose one arbitrarily). The point p can be written as a convex combination of vertices in $V(p)$, i.e. $p = \sum_{v\in V(p)} \mu_v(p)v$ and it is even possible to decompose $p = \sum_{v\in\mathcal{V}} \mu_v(p)v$ by assuming that $\mu_v(p) = 0$ for any p that does not belong to the same simplex. All those mappings μ_v are continuous and Lipschitz.

We construct an ε-calibrated strategy σ using a fixed deterministic smooth calibrated strategy σ_d in the following way. Whenever σ_d dictates to predict $p \in \Delta(\Omega)$, σ predicts $v \in \mathcal{V}$ with probability $\mu_v(p)$. Immediate calculations show that, for every $v \in \mathcal{V}$,

$$\mathbb{E}_{\sigma,\tau}\left[\frac{1}{n}\sum_{m=1}^{n} \mathbb{1}\{p_m = v\}(\omega_m - v)\right] = \frac{1}{n}\sum_{m=1}^{n} \mu_v(p_m)(\omega_m - p_m) + \frac{1}{n}\sum_{m=1}^{n} \mu_v(p_m)(p_m - v).$$

Since $\mu(p_m)\|p_m - v\| \leq \varepsilon$, expected calibration score (and the actual score, thanks to concentration inequalities) are ε-close to the smooth calibration score, hence the result.

Key features of this construction are that, although it is impossible to construct an ε-calibrated strategy deterministically (as proved by Oakes and Dawid), it is possible by using randomizations on arbitrarily small balls. This is why we used the term of *almost deterministic* strategies.

Concerning the complexity of (weak) calibration, a recent result of Hazan & Kakade [34], based on an idea of Kakade & Foster [37], shows that it is a hard criterion to satisfy. Indeed, an almost deterministic strategy σ (based on some triangulation of $\Delta(\Omega)$) can be used to find ε-Nash equilibria of games. We sketch the proof in the following.

Consider a game between a set of players $\mathcal{I}$ with actions sets $\mathcal{A}_i$ and payoff mappings ρ_i. Define $\Omega = \prod_{i\in\mathcal{I}} \mathcal{A}_i$ and let $X_i : \Delta(\Omega) \to \Delta(\mathcal{A}_i)$ be a smooth ε-best response of player i (i.e., given any $p \in \Delta(\Omega)$, if p^{-i} denotes the i-th marginal of p, then $\rho_i(X_i(p), p^{-i}) \geq \max_{x\in\Delta(\mathcal{A}_i)} \rho_i(x, p^{-i}) - \varepsilon$).

We denote by $p_n \in \Delta(\Omega)$ the prediction output at stage n by the strategy σ and we assume that player i plays accordingly to $X_i(p_n)$. The profile of actions actually played is $\omega_n \in \Omega$ and one has $\mathbb{E}[\omega_n] = \Big(X_1(p_n), \ldots, X_I(p_n)\Big) =: X(p_n)$. Since σ is ε-calibrated, for every vertex v and with probability one,

$$\limsup_{n\to\infty} \frac{\sum_{m=1}^n \mathbb{1}\{p_m = v\}}{n} \left\| \frac{\sum_{m=1}^n \mathbb{1}\{p_m = v\}(v - \omega_m)}{\sum_{m=1}^n \mathbb{1}\{p_m = v\}} \right\| - \frac{\sum_{m=1}^n \mu_v(p_m)}{n}\varepsilon \leq 0.$$

Concentration inequalities, and the fact that $X(p_n) - \omega_n$ and $\mathbb{1}\{p_n = v\} - \mu_v(p_n)$ are martingale differences imply that, with probability one,

$$\lim_{n\to\infty} \frac{\sum_{m=1}^n \mathbb{1}\{p_m = v\}(X(v) - \omega_m)}{n} = 0 \text{ and } \lim_{n\to\infty} \frac{\sum_{m=1}^n \mathbb{1}\{p_m = v\} - \mu_v(p_m)}{n} = 0 \ .$$

As a consequence, summing terms, for every vertex v that is predicted with a positive density (i.e., such that $\limsup_{n\to\infty} \sum_{m=1}^n \mathbb{1}\{p_m = v\}/n > 0$), one must have $\|v - X(v)\| \leq \varepsilon$. And so, by the very definition of $X(\cdot)$, v must be a 2ε-Nash equilibrium.

Therefore, not only does the empirical profile of action converge to the convex hull of ε-Nash equilibria, but also if a stage n chosen at random then, with arbitrarily great probability, $X(v_n)$ is an 2ε-Nash equilibrium.

4 Equivalences between approachability, regret and calibration

This part is devoted mainly to describe how approachability can be used to construct consistent and calibrated strategies. We also show why calibration is an important and useful tool, as it can be used to construct general (and even approachability) consistent strategies. Since we can also reduce approachability to regret, this complete the circle and this is the reason why we called these notions *equivalent*.

4.1 Using approachability to get regret

4.1.1 From approachability to regret; the finite case

Although Blackwell [10] was the first to notice that consistent strategies can be constructed using approachability theory, we first treat Hart & Mas-Colell [32] idea in finite dimension.

We recall that choices of actions $a_n \in \mathcal{A}$ and $b_n \in \mathcal{B}$ generate at stage $n \in \mathbb{N}$ an external regret r_n defined by

$$r_n = r(a_n, b_n) := \Big(\rho(1, b_n) - \rho(a_n, b_n), \ldots, \rho(A, b_n) - \rho(a_n, b_n)\Big) \in \mathbb{R}^A$$

and that a strategy is externally consistent if $\|\overline{r}_n^+\|_\infty$ goes to 0 almost surely. Actually, using approachability theory, Hart & Mas-Colell [32] proved the following

Proposition 4.1 *The strategy σ defined by playing, at stage $n+1$, proportionally to $\overline{r}_n^+$ (and arbitrarily if every component is non-positive) is externally consistent. Moreover, for every strategy τ of Nature and $n \in \mathbb{N}$,*

$$\mathbb{E}_{\sigma,\tau}\Big[\|\overline{r}_n^+\|_\infty\Big] \leq \mathbb{E}_{\sigma,\tau}\Big[\|\overline{r}_n^+\|_2\Big] \leq \sqrt{\frac{A}{n}}\ ,$$

and, for every $\eta > 0$, $\mathbb{P}_{\sigma,\tau}\left\{\sup_{N \geq n} \|\overline{r}_n^+\| \geq \eta\right\} \leq 3\exp\left(-\frac{\eta^2 n}{64A}\right)$ as soon as $\frac{\eta^2 n}{32A} \geq 1$

Proof: We simply have to prove that σ is exactly Blackwell's approachability strategy of the negative orthant $\mathbb{R}_-^A$ (which is a cone) in the game where the vector payoff is $r(a, b)$. This is a consequence of the following geometric property

$$\text{No matter the choice of } x \in \Delta(\mathcal{A}),\ \Big\langle\, x\,, \mathbb{E}_x[r(a,b)]\,\Big\rangle = 0, \text{ for all } b \in \mathcal{B}.$$

Indeed, the k-th component of $\mathbb{E}_x[r(a,b)]$ is, by linearity, $r(k,b) - r(x,b)$, thus the inner product is equal to $\sum_{a\in\mathcal{A}} x[a]\big(r(a,b) - r(x,b)\big) = r(x,b) - r(x,b) = 0$.

Since $x_{n+1} = \sigma(h^n)$ is proportional to $\overline{r}_n^+$, the geometric property implies that

$$\Big\langle\, \overline{r}_n^+, \mathbb{E}_{\sigma,\tau}[r_{n+1}]\,\Big\rangle = 0 \quad \text{thus} \quad \Big\langle\, \overline{r}_n - \overline{r}_n^-\,, \mathbb{E}_{\sigma,\tau}[r_{n+1}] - \overline{r}_n^-\,\Big\rangle = 0$$

because one always has $\langle z^+, z^-\rangle = 0$. Since $\overline{r}_n^-$ is the projection of $\overline{r}_n$ on the negative orthant, this proves that σ satisfies Blackwell property, hence is an approachability strategy. Bonds follow from Corollary 1.15. ■

Once the reduction from external regret minimization to approachability of $\mathbb{R}_-^A$ has been made, the existence of externally consistent strategies is immediate because $\mathbb{R}_-^A$ is obviously a convex approachable set. Indeed, for every $y \in \Delta(\mathcal{B})$, there exists $x \in \Delta(\mathcal{A})$ such that $r(x,y) \in \mathbb{R}_-^A$: it suffices to take for x any best response to y. The most

interesting feature of Proposition 4.1 is that the strategy is very simple and natural: the more regret a specific action induces, the more it should be played (and with a weight exactly proportional to this regret generated).

Generalizations to the compact case (when Nature chooses at stage $n \in \mathbb{N}$ an outcome vector $U_n \in [0,1]^A$) are immediate and omitted.

Remark 4.1 *One might argue that with exponential weight algorithm, the dependency in A in rates of convergence shrinks to $\sqrt{\log(A)}$ instead of $\sqrt{A}$, so the strategy we output might not be optimal. Actually this argument is flawed, rates of convergence are indeed optimal since we minimized the ℓ_2-norm of the regret. It is only possible to upperbound with $\sqrt{\log(A)/n}$ the ℓ_∞-norm of the regret.*

Actually, Hart & Mas-Colell strategy is an approachability strategy of $\mathbb{R}^A_-$ driven by the potential $\Phi(z) = \|z^+\|^2$ (that represents the ℓ_2-norm of the regret) while exponential weights are driven by the soft-max potential $\Phi(z) = \frac{1}{\eta}\log\left(\sum_{a\in\mathcal{A}} e^{\eta z_a}\right)$ which is a twice differentiable surrogate of $\|z^+\|_\infty$. However, minimization of the infinite norm of regret can also be reduced to approachability, see Proposition 4.2 below (following actually an idea of Blackwell [10]).

Proposition 4.2 *Assume that Nature chooses outcome vectors $U \in [0,1]^A$ and define the game with vector payoffs and target set defined as follows*

$$g(a,U) = \Big(U^a, U\Big) \in [0,1]\times[0,1]^A \quad \textit{and} \quad \mathcal{C} = \Big\{(z,V) \in [0,1]\times[0,1]^A \ \textit{s.t.}\ z \geq \max_{a\in\mathcal{A}} V^a\Big\}.$$

Then any approachability strategy of $\mathcal{C}$ (which is a convex approachable set) minimizes the ℓ_∞ norm of the external regret since

$$d_{\mathcal{C}}\big(\overline{g}_n\big) \leq \big\|\overline{r}_n^+\big\|_\infty \leq \sqrt{2}\, d_{\mathcal{C}}\big(\overline{g}_n\big).$$

Proof: Convexity of $\mathcal{C}$ (which is actually a polytope, i.e., the intersection of a finite number of half-spaces and a compact set) is a direct consequence of its definition since

$$\mathcal{C} = \bigcap_{a\in\mathcal{A}} \Big\{(z,V) \text{ s.t } z \geq V^a\Big\} \bigcap [0,1]\times[0,1]^A \ .$$

Approachability of $\mathcal{C}$ is immediate: for every $U \in \mathcal{U}$, choosing a to be one of the highest component of U ensures that $g(a,U) = (U^a, U)$ belongs to $\mathcal{C}$. It remains to prove the inequalities.

Notice that if we denote by $(\overline{z}_n, \overline{U}_n)$ the average vector payoff at stage n, then $\overline{U}_n$ is the average outcome vector and $\overline{z}_n$ is the average actual payoff. As a consequence, the ℓ_∞ norm of the regret, $\|\overline{r}_n\|_\infty = \max_{a\in\mathcal{A}} \overline{U}_n^a - \overline{z}_n$, is exactly equals to the distance between $\big(\overline{z}_n, \overline{U}_n\big)$ and $\big(\max_{a\in\mathcal{A}} \overline{U}_n^a, \overline{U}_n\big)$. By definition, the latter belongs to $\mathcal{C}$, therefore one has $d_{\mathcal{C}}\big(\overline{g}_n\big) \leq \big\|\overline{r}_n^+\big\|_\infty$.

Let $a^* \in \operatorname{argmax}_{a\in\mathcal{A}} \overline{U}_n^a$ and $\big(\overline{z}_n^c, \overline{U}_n^c\big) = \Pi_{\mathcal{C}}\big(\overline{z}_n, \overline{U}_n\big)$, then

$$\big\|\overline{r}_n^+\big\|_\infty = \overline{U}_n^{a^*} - \overline{z}_n = \overline{U}_n^{a^*} - \overline{z}_n^c + \overline{z}_n^c - \overline{z}_n \leq \Big\|\overline{U}_n - \overline{U}_n^c\Big\| + \Big|\overline{z}_n^c - \overline{z}_n\Big| \leq \sqrt{2} d_{\mathcal{C}}(\overline{g}_n)$$

where we used the fact that $U \mapsto \max_{a\in\mathcal{A}} U^a$ is 1-Lipschitz. ■

Extensions to the case where $\mathcal{A} \subset \mathbb{R}^A$ is a compact convex set are immediate as the finiteness of $\mathcal{A}$ is not used in the proof.

Actually, Blackwell proved this result in the finite case, where Nature chooses action in $\mathcal{B}$; in that case, stage payoffs are $g'(a,b) = (\rho(a,b), \delta_b) \in \mathbb{R} \times \Delta(\mathcal{B})$ where, as usual, $\Delta(\mathcal{B})$ is seen as a subset of $\mathbb{R}^B$. The target set is

$$\mathcal{C}' = \left\{(z,y) \in \mathbb{R} \times \Delta(\mathcal{B}) \text{ s.t. } z \geq \max_{a\in\mathcal{A}} \rho(a,y)\right\}$$

and since $\|g'(a,b)\| \leq \sqrt{2}$, approachability results imply that $\overline{g}'_n$ converges to $\mathcal{C}'$ at the rate of $\sqrt{2/n}$ thus expected regret is bounded in the order of $\sqrt{B/n}$ (because in this framework, $y \mapsto \rho(a,y)$ is $\sqrt{B}$-Lipschitz and not 1-Lipschitz).

This shows that regret can be bounded, not only with respect to the number of player's actions (i.e. in $\sqrt{\log(A)/n}$), but also with respect to Nature's one (in $\sqrt{B/n}$). This might lead to some improvement if the former is exponentially larger than the latter.

Remark 4.2 *In the compact case, usual proofs show that Blackwell's approachability strategy ensures that $d_{\mathcal{C}}(\overline{g}_n) \leq \sqrt{\|g\|_\infty/n} = \sqrt{(A+1)/n}$. However, there exist a consistent strategy such that $\|\overline{r}_n^+\|_\infty \leq 3\sqrt{\log(A)/n}$. So this is an example where the optimal dimension dependency of rates of approachability is not $\sqrt{\|g\|_\infty}$, but much smaller.*

There are two possible explanations: either minimizing step by step the ℓ_2 distance (i.e. following Blackwell's strategy) is not optimal, or some important facts are hidden within proofs. In Remarks 4.1 we claimed that the answer was the first possibility: indeed, the final objective was to minimize the ℓ_∞-distance, so minimizing the ℓ_2 norm must induce an additional dimension-dependent constant. This is not the case here, because the final objective is within constant of the ℓ_2-distance.

An open and fairly question is wether the dimension dependent term should depend on the specific target set $\mathcal{C}$ or not. In these examples, respective sizes of the target sets $\mathcal{C}$ within the set of feasible payoff vectors are rather intriguing. For instance, in the framework of Proposition 4.1, the volume of $\mathrm{co}\{g(a,b)\}$ is 2^A times the volume of $\mathcal{C}$ while it is only $A+1$ times the volume of $\mathcal{C}$ in the framework of Proposition 4.2. This has to be compared with the respective size of dimension dependent constants which were $\sqrt{A}$ and $\sqrt{\log(A)}$.

We now turn to the minimization of internal regret $R_n = R(a_n, b_n)$. We recall that it is a $A \times A$-matrix whose (a,a') component is $\rho(a',b) - \rho(a,b)$ if $a = a_n$ and 0 otherwise. The generalization of Hart & Mas-Colell strategy will appeal to the concept of *invariant measures* of matrices.

A probability distribution $\lambda \in \Delta(\{1,\ldots,d\})$ is an invariant measure of a some $d \times d$-matrix M with non-negative coefficient, if

$$\sum_{k=1}^{d} \lambda^k M^{k,i} = \lambda^i \sum_{k=1}^{d} M^{i,k}, \quad \forall i \in \{1,\ldots,d\} \ ,$$

and their existence is a consequence of Perron-Frobenius theorem (this also generalizes usual invariant measure of Markov chains (see e.g. Seneta [69]).

Sorin [71], but also Hart & Mas-Colell [31] and Foster & Vohra [25] used the existence of invariant measure to output a simple internally consistent strategy.

Proposition 4.3 *The strategy σ that dictates to play at stage $n+1$ an invariant measure of $\overline{R}_n^+$ (and arbitrarily if every component is non-positive) is internally consistent. Moreover, for every strategy τ of Nature and $n \in \mathbb{N}$,*

$$\mathbb{E}_{\sigma,\tau}\Big[\left\|\overline{R}_n^+\right\|_\infty\Big] \leq \mathbb{E}_{\sigma,\tau}\Big[\left\|\overline{R}_n^+\right\|_2\Big] \leq \sqrt{\frac{A}{n}}\ ,$$

and, for every $\eta > 0$, $\mathbb{P}_{\sigma,\tau}\left\{\sup_{N\geq n}\left\|\overline{R}_n^+\right\| \geq \eta\right\} \leq 3\exp\left(-\frac{\eta^2 n}{64A}\right)$ as soon as $\frac{\eta^2 n}{32A} \geq 1$.

Proof: As for external regret, we just need to prove that σ is exactly Blackwell's approachability strategy of the negative orthant. And again, this is a consequence of a geometric property:

Any invariant measure λ of any matrix M with non-negative coefficient satisfies, no matter the choice of $b \in \mathcal{B}$, $\Big\langle M\ , \mathbb{E}_\lambda[R(a,b)]\Big\rangle = 0$.

Let $U^a := \rho(a,b)$, then the (i,k)-component of $\mathbb{E}_\lambda[R(a,b)]$ is $\lambda^i\Big(U^k - U^i\Big)$. So the inner product is equal to $\sum_{i,k} M^{i,k}\lambda^i\Big(U^k - U^i\Big)$ and the coefficient before U^i in this sum is

$$\sum_k \lambda^k M^{k,i} - \lambda^i \sum_k M^{i,k} = 0$$

since λ is an invariant measure of M.

Since $x_{n+1} = \sigma(h^n)$ is an invariant measure of $\overline{R}_n^+$, geometric properties implies that

$$\Big\langle \overline{R}_n^+, , \mathbb{E}_{\sigma,\tau}[R_{n+1}]\Big\rangle = 0 \quad \text{thus} \quad \Big\langle \overline{R}_n - \overline{R}_n^-\ , \mathbb{E}_{\sigma,\tau}[R_{n+1}] - \overline{R}_n^-\Big\rangle = 0\ .$$

This proves that σ satisfies Blackwell property, hence is an approachability strategy and bonds follows from Corollary 1.15. ■

Once again, using approachability theory to prove existence of internally consistent strategies is immediate: the negative orthant satisfies Blackwell's property. An interesting feature of this algorithm is the simple characterization of this optimal (for the minimization of the ℓ_2 norm) strategy.

Interestingly, the reduction from external to internal consistent strategies (see Section 2.1.3 or Stoltz & Lugosi [75]) run with the algorithm of Proposition 4.1 constructs exactly the strategy of Proposition 4.3.

So both Propositions 4.1 and 4.3 can be unified into the the following theorem that deals more generally with Φ-regret. It exhibits a strategy with the same complexity as the previous internally consistent strategy, dictating to play at each stage an invariant measure of some matrix. Given a family Φ, we recall that Φ-regret at stage n is denoted by $R_n^\Phi \in \mathbb{R}^{|\Phi|}$ and defined by

$$R_n^\Phi = R^\Phi(a_n, b_n) := \Big(\rho(\phi(a_n), b_n) - \rho(a_n, b_n)\Big)_{\phi\in\Phi}.$$

Finally, given $M \in \mathbb{R}^{|\Phi|}$, let $\Theta^\Phi(M)$ be the $A \times A$-matrix whose (a, a') component is $\Theta^\Phi(M)^{a,a'} = \sum_{\phi:\phi(a)=a'} M^\phi$

Theorem 4.1 *Let Φ be a family of swap mappings. The strategy playing at stage $n+1$ accordingly to any invariant measure of $\Theta^\Phi(\overline{R_n^\Phi}^+)$ has no Φ-regret. Moreover, for every strategy τ of Nature and $n \in \mathbb{N}$,*

$$\mathbb{E}_{\sigma,\tau}\Big[\Big\|\overline{R_n^\Phi}^+\Big\|_\infty\Big] \le \mathbb{E}_{\sigma,\tau}\Big[\Big\|\overline{R_n^\Phi}^+\Big\|_2\Big] \le \sqrt{\frac{A_\Phi}{n}}\ , \textit{with } A_\Phi = \max_{a\in\mathcal{A}}\Big|\Big\{\phi\in\Phi \textit{ s.t. } \phi(a)\neq a\Big\}\Big|$$

and, for every $\eta > 0$, $\mathbb{P}_{\sigma,\tau}\Big\{\sup_{N\ge n}\Big\|\overline{R_n^\Phi}^+\Big\| \ge \eta\Big\} \le 3\exp\Big(-\frac{\eta^2 n}{64A_\Phi}\Big)$ as soon as $\frac{\eta^2 n}{32A_\Phi} \ge 1$.

Proof: The proof follows closely the ones of Propositions 4.1 and 4.3. Indeed, one just has to prove that this strategy is an approachability strategy of $\mathbb{R}_-^{|\Phi|}$, using the following geometric property:

Any invariant measure λ of any matrix $\Theta(M)$ with non-negative coefficient satisfies, no matter the choice of $b \in \mathcal{B}$, $\Big\langle M\,, \mathbb{E}_\lambda[R^\Phi(a,b)]\Big\rangle = 0$.

Indeed, if one denote $U = \rho(\cdot, b)$, then

$$\begin{aligned}
\Big\langle M, \mathbb{E}_\lambda[R^\Phi(a,b)]\Big\rangle &= \sum_{\phi\in\Phi} M^\phi \sum_{a\in\mathcal{A}} \lambda^a\Big(U^{\phi(a)} - U^a\Big)\\
&= \sum_{a\in\mathcal{A}} \lambda^a \sum_{a'\in\mathcal{A}} \sum_{\phi:\phi(a)=a'} M^\phi\Big(U^{a'} - U^a\Big)\\
&= \sum_{a\in\mathcal{A}} \lambda^a \sum_{a'\in\mathcal{A}} \Theta^\Phi(M)^{a,a'}\Big(U^{a'} - U^a\Big)\\
&= \sum_{a\in\mathcal{A}} \left(\sum_{a'\in\mathcal{A}} \lambda^{a'}\Theta^\Phi(M)^{a',a} - \lambda^a \sum_{a'\in\mathcal{A}} \Theta^\Phi(M)^{a,a'}\right) U^a = 0
\end{aligned}$$

since λ is an invariant measure of $\Theta^\Phi(M)$. As a consequence, this strategy is exactly Blackwell's approachability strategy of the negative orthant. The result comes from the fact that $R^\Phi(a,b)$ has at most A_Φ non-zero components, each one in $[-1, 1]$, thus $\|R^\Phi(a,b)\|^2 \le A_\Phi$. ■

Remark 4.3 *As usual, if techniques from approachability in infinite dimension are used instead of regular approachability, the term in $\sqrt{A}$ for external and internal regret or $\sqrt{A_\Phi}$ for Φ-regret can be replaced by respectively $\sqrt{\log(A)}$ or $\sqrt{A\log(A)}$, up to some constant.*

4.1.2 From approachability to regret; the infinite case

We turn in this section to the case where action set $\mathcal{A}$ is no longer finite but some convex compact metric set and at stage $n \in \mathbb{N}$, Nature chooses a mapping $U_n : \mathcal{A} \to [0,1]$ in a set $\mathcal{U}$ of equicontinuous mapping. We show how previous results can be extended to this *compact case* (indeed, Arzela-Ascoli theorem ensures that $\mathcal{U}$ is relatively compact).

Theorem 4.2 *In this compact case, there exists a strategy without Φ_c regret, where Φ_c is the set of continuous mapping from $\mathcal{A}$ to itself.*

Proof: Consider an auxiliary game where action sets of player and Nature are $\mathcal{A}$ and $\mathcal{U}$. Choices of $a \in \mathcal{A}$ and $U \in \mathcal{U}$ generates a payoff $\widetilde{U}[a] \in \mathcal{L}_2(\Phi_c, \lambda)$, where λ is some fixed probability distribution over $(\Phi_c, \|\cdot\|_\infty)$ embedded with the Borelian σ-field, defined by

$$\widetilde{U}[a](\phi) := U(\phi(a)) - U(a), \quad \forall \phi \in \Phi_c.$$

The convex set $\mathcal{C} = \mathcal{L}_2^-(\Phi_c, \lambda) := \left\{\widetilde{U} \in \mathcal{L}_2(\Phi_c, \lambda) \text{ s.t. } \widetilde{U} \leq 0\right\}$ is not excludable by Nature; indeed, for any $U \in \mathcal{U}$, there exists $a \in \mathcal{A}$ (any global maximizer of U) such that $\widetilde{U}[a]$ belongs to $\mathcal{C}$. Thus it is approachable by the player, and any approachability strategy has no ϕ-regret, for λ-almost all mapping $\phi \in \Phi_c$.

However, Φ_c is separable (see Rudin [67] or Stoltz & Lugosi [76]), so there exists $\{\phi_k;\, k \in \mathbb{N}\}$ a countable dense subset of Φ_c; the corresponding probability λ we consider is $\lambda = \sum_{k\in\mathbb{N}} 2^{-k}\delta_{\phi_k}$. Since $\mathcal{U}$ is a family of equicontinuous mappings, every mapping $U \in \mathcal{U}$ share the same modulus of continuity $\omega(\cdot)$; this means that, for every $\varepsilon > 0$ there exists $\delta := \omega(\varepsilon)$ such that if $d(a, a') \leq \delta$ then $|U(a) - U(a')| \leq \varepsilon$, for any mapping $U \in \mathcal{U}$. Given $\phi \in \Phi_c$, there exists ϕ_k such that $\|\phi - \phi_k\| \leq \delta$ thus

$$\frac{1}{n}\sum_{m=1}^{n} U_m[\phi(a_m)] - U_m[a_m] \leq \frac{1}{n}\sum_{m=1}^{n} U_m[\phi_k(a_m)] - U_m[a_m] + \varepsilon\,.$$

Since σ has no ϕ_k-regret, its ϕ-regret is asymptotically smaller than ε, for every $\varepsilon > 0$, thus it has no Φ_c-regret. ■

Corollary 4.4 *Conclusions of Theorem 4.2 hold if $\mathcal{U}$ is the convex hull of a finite set of upper-semicontinuous mappings bounded from below and Φ_c is the set of constant mappings.*

Proof: Every $U \in \mathcal{U}$ is upper-semicontinuous over a compact set, it admits a maximum. Therefore $\mathcal{U}$ is uniformly bounded and the set $\mathcal{C}$ is approachable, with respect to some probability distribution λ that remains to be defined.

Denote by $U_1, \ldots, U_m$ the extreme points of $\mathcal{U}$. As they are upper-semicontinuous and bounded, their exists a countable subset $\{a_k; k \in \mathcal{N}\} \subset \mathcal{A}$ such that, for every $\varepsilon > 0$ and every $a \in \mathcal{A}$, there exists a_k satisfying $U_i(a_k) \geq U_i(a) - \varepsilon$, for every $i \in \{1, \ldots, m\}$. Define λ as any probability measure whose support is exactly this countable subset.

The rest of the proof follows the one of Theorem 4.2. ■

In the finite case, approachability theory not only provides a quick and easy proof of consistent strategies, but also exhibit explicitly some of them. In fact, playing somehow proportionally to the positive part of the regret is still externally consistent in the compact case. Let λ be any positive probability measure on $\{a_k;\ k \in \mathbb{N}\}$, a countable dense subset of $\mathcal{A}$ and denote by $\overline{r}_n^+[a_k]$ the external regret at stage n induces by action a_k.

Consider the strategy that chooses a_k at stage $n+1$ with probability $\frac{\lambda_k \overline{r}_n^+[a_k]}{\sum_\ell \lambda_\ell \overline{r}_n^+[a_\ell]}$. Then, as in the finite case, one can easily show that the geometric property holds, i.e.,

$$\langle \mathbb{E}[r_{n+1}], \overline{r}_n^+ \rangle = \sum_k \frac{\lambda_k \overline{r}_n^+[a_k]}{\sum_\ell \lambda_\ell \overline{r}_n^+[a_\ell]} \sum_\ell \Big(U_{n+1}(a_\ell) - U_{n+1}(a_k) \Big) \overline{r}_n^+[a_\ell] \lambda_\ell = 0 \ .$$

Approachability in infinite dimension (along with the density argument) ensures that this strategy has no external regret.

Concerning Φ-regret, one cannot simply play accordingly to any invariant measure of some infinite dimensional matrix, as their existence is not ensured. However, it is still possible to discretize finitely $\mathcal{A}$ to get a Φ-regret smaller than ε, with ε-arbitrary small (or even equal to 0, if ε is taken as a decreasing sequence, see Proposition 1.7).

Let $\omega_{\mathcal{U}}(\cdot)$ be the common modulus of continuity of $U \in \mathcal{U}$ and $\mathbf{A}$ a finite $\omega_{\mathcal{U}}(\varepsilon)$-grid of $\mathcal{A}$. For any $\phi \in \Phi$, we define $\underline{\phi} : \mathcal{A} \to \mathbf{A}$ by $\underline{\phi} = \operatorname{argmin}_{\mathbf{a}' \in \mathbf{A}} d(\phi(a), \mathbf{a}')$ with ties broken arbitrarily. As a consequence, for every $a \in \mathcal{A}$, $U \in \mathcal{U}$ and non negative $q \in \mathcal{L}_2(\Phi, \lambda)$,

$$\left| \int_\Phi q(\phi) U(\underline{\phi}(a)) d\lambda - \int_\Phi q(\phi) U(\phi(a)) d\lambda \right| \leq \varepsilon \ .$$

We define, for any $(\mathbf{a}, \mathbf{a}')$, $\Theta[q]^{\mathbf{a},\mathbf{a}'} := \int_{\Phi^{\mathbf{a},\mathbf{a}'}} q d\lambda$, where $\Phi^{\mathbf{a},\mathbf{a}'} := \Big\{ \phi \in \Phi \text{ s.t. } \underline{\phi}(\mathbf{a}) = \mathbf{a}' \Big\}$. Let x be any invariant measure of the matrix $\Theta[q]$ then one has

$$\Big\langle q \,, \widetilde{U}(x) \Big\rangle \leq \sum_{\mathbf{a} \in \mathbf{A}} x^{\mathbf{a}} \sum_{\mathbf{a} \in \mathbf{A}} \Theta[q]^{\mathbf{a},\mathbf{a}'} \Big(U(\mathbf{a}') - U(\mathbf{a} \Big) + \left| \int_\Phi q(\phi) \Big(U(\underline{\phi}(a)) - U(\phi(a)) \Big) d\lambda \right| \leq \varepsilon.$$

This proves that $\mathcal{L}_2^-(\Phi_c, \lambda)$ is a B-set, hence approachable. We can only claim that the strategy we exhibited has some *flavors* of invariant measures.

4.2 Using regret to get calibration

We show in this section that finite calibration can easily be understood in terms of internal regret. The first idea goes to Foster & Vohra [23] and it has been somehow clarified by Sorin [71]. Recall that, in finite calibration, Nature chooses at stage n an outcome $\omega_n \in \Omega$. The player formulates a prediction on ω_n by choosing a probability distribution $p[\ell_n] \in \Delta(\Omega)$ that must belong to a finite grid $\{p[\ell]\,;\ell \in \mathcal{L}\}$.

Theorem 4.3 *There exists a strategy σ calibrated with respect to the grid $\{p[\ell]\,;\ell \in \mathcal{L}\}$, such that, no matter the strategy τ of Nature,*

$$\mathbb{E}_{\sigma,\tau}\left[\sup_{\ell\in\mathcal{L}} \frac{|\mathbb{N}_n[\ell]|}{n}\Big(\left\|\overline{\omega}_n[\ell]-p[\ell]\right\|^2 - \min_{k\in\mathcal{L}}\left\|\overline{\omega}_n[\ell]-p[k]\right\|^2\Big)\right] \leq 6\sqrt{\frac{\log(L)}{n}},\ \textit{so}$$

$$\mathbb{E}_{\sigma,\tau}\left[\sup_{\ell\in\mathcal{L}} \frac{|\mathbb{N}_n[\ell]|}{n}\Big(\left\|\overline{\omega}_n[\ell]-p[\ell]\right\| - \min_{k\in\mathcal{L}}\left\|\overline{\omega}_n[\ell]-p[k]\right\|\Big)\right] \leq \frac{6}{\delta(\mathcal{L})}\sqrt{\frac{\log(L)}{n}},$$

where $\delta(\mathcal{L}) = \inf_{\ell\neq k\in\mathcal{L}}\left\|p[\ell]-p[k]\right\|$ is the diameter of the grid.

Proof: The proof uses the fact (simply obtained by expanding sums) that, for any sequence q_m and every $\ell, k \in \mathcal{L}$,

$$\sum_{m\in\mathbb{N}_n[\ell]} \frac{\|\omega_m - p[\ell]\|^2}{|\mathbb{N}_n[\ell]|} - \frac{\|\omega_m - p[k]\|^2}{|\mathbb{N}_n[\ell]|} = \left\|\overline{\omega}_n[\ell]-p[\ell]\right\|^2 - \left\|\overline{\omega}_n[\ell]-p[k]\right\|^2 .$$

Now consider the game with action space $\mathcal{L}$ and Ω where choices of ℓ and ω generate the payoff $\rho(\ell,\omega) = -\|\omega - p[\ell]\|^2$. An internally consistent strategy satisfies, by definition,

$$\limsup_{n\to\infty}\sup_{\ell,k} \frac{|\mathbb{N}_n[\ell]|}{n}\Big(\sum_{m\in\mathbb{N}_n[\ell]} \frac{\left\|\omega_m - p[\ell]\right\|^2}{|\mathbb{N}_n[\ell]|} - \frac{\left\|\omega_m - p[k]\right\|^2}{|\mathbb{N}_n[\ell]|}\Big) \leq 0.$$

So this, along with the basic fact, shows that any internally consistent strategy is calibrated with respect to the grid $\{p[\ell]; \ell \in \mathcal{L}\}$. Rates of convergences follows from those of internal consistency. ■

We stress out that we proved a stronger result than require; the calibration score converges almost surely to zero, at a rate independent of Nature's strategy.

Remark 4.4 *This proof of calibration highlights the following fact. It does not really matter that ω_m belongs to a finite set Ω and that p_n are probability distributions over Ω. Indeed, one can just assume that sequences ω_n and p_n belong to some compact set of an Euclidian space $\mathbb{R}^d$. Similarly, given two finite families of predictions $\{p[\ell] \in \mathbb{R}^d; \ell \in \mathcal{L}\}$ and weights $\{\nu[\ell] \in \mathbb{R}; \ell \in \mathcal{L}\}$, we recall that weighted calibration sis defined as*

$$\sup_{\ell\in\mathcal{L}} \frac{|\mathbb{N}_n[\ell]|}{n}\Big(\Big(\left\|\overline{\omega}_n[\ell]-p[\ell]\right\|^2 - \nu[\ell]\Big) - \Big(\min_{k\in\mathcal{L}}\left\|\overline{\omega}_n[\ell]-p[k]\right\|^2 - \nu[k]\Big)\Big),$$

and the exact same proof (yet with $\rho(\ell,\omega) = -\|\omega - p[\ell]\|^2 + \nu[\ell]$) gives the existence of weighted calibrated strategies. Rates of convergence are identical except that the constant 6 *is replaced with* $6 + 3\max_{\ell\in\mathcal{L}} |\nu[\ell]|$.

Notice that we defined finite calibration with respect to the uniform norm of the positive part of $\left(\left\|\overline{\omega}_n[\ell] - p[\ell]\right\| - \min_{k\in\mathcal{L}} \left\|\overline{\omega}_n[\ell] - p[k]\right\|\right)_{\ell,k}$. And this quantity is upper-bounded optimally by the exponential weight algorithm. We could as well have defined calibration in terms of the ℓ^2 norm of this vector and as in regret minimization, playing an invariant measure could then improve bounds.

Next proposition states that, quite surprisingly, there exist ε-calibrated strategies with rates of convergence independent of ε (and even of Ω, for a slightly weaker notion).

Proposition 4.5 *For every $\varepsilon > 0$, there exists a grid $\{p[\ell]; \ell \in \mathcal{L}\}$ and a strategy σ such that, no matter the strategy τ of Nature and for every $n \in \mathbb{N}$,*

$$\mathbb{E}_{\sigma,\tau}\left[\sup_{\ell\in\mathcal{L}} \frac{|\mathbb{N}_n[\ell]|}{n}\left(\left\|\overline{p}_n[\ell] - \overline{\omega}_n[\ell]\right\| - \varepsilon\right)\right] \leq \sqrt{\frac{1}{n}}\ .$$

Moreover, this strategy is ε-calibrated, with a rate of convergence independent of ε, since one also has, for every $n \in \mathbb{N}$,

$$\mathbb{E}_{\sigma,\tau}\left[\sup_{p\in\Delta(\Omega)} \frac{|\mathbb{N}_n[p,\varepsilon]|}{n}\left(\left\|\overline{p}_n[p,\varepsilon] - \overline{\omega}_n[p,\varepsilon]\right\| - \varepsilon\right)\right] \leq \sqrt{\frac{\gamma(\Omega)}{n}},\ \textit{with } \gamma(\Omega) \leq (2\Omega^{\Omega}).$$

Proof: Let ε be fixed; the strategy considered is simply a calibrated strategy with respect to some well chosen grid of $\Delta(\Omega)$. Recall that $\Delta(\Omega)$ is written as the following subset of $\mathbb{R}^d$ with $d = \Omega - 1$:

$$\Delta(\Omega) := \left\{q = (q_1,\ldots,q_d) \in \mathbb{R}^d \text{ s.t. } q_1,\ldots,q_d \geq 0 \text{ and } \sum_{k=1}^{d} q_k \leq 1\right\}.$$

Denote by $\mathbf{e}_k$ the unit vector of $\mathbb{R}^d$ whose components are all zero except the k-th which is one. The regular grid considered is indexed by $\mathcal{L}_\varepsilon$ and defined by

$$\left\{\sum_{k=1}^{d} \frac{2n_k}{\sqrt{d}}\varepsilon.\mathbf{e}_k \in \Delta(\Omega); n_k \in \mathbb{N}\right\} =: \left\{p[\ell] = \sum_{k=1}^{d} \frac{2n_k[\ell]}{\sqrt{d}}\varepsilon.\mathbf{e}_k\ ;\ \ell \in \mathcal{L}_\varepsilon\right\}$$

Given a point $p[\ell]$ of the grid, its neighbors are points $p[\ell']$ such that $n_k[\ell] = n_k[\ell']$ for every $k \in \{1,\ldots,d\}$ except for exactly one k_0 which is such that $\left|n_{k_0}[\ell] - n_{k_0}[\ell']\right| = 1$. So if we denote by $\mathcal{N}[\ell] \subset \mathcal{L}_\varepsilon$ the neighbors of $p[\ell]$, it contains at most $2d$ elements.

The basic idea behind the specific geometry of this grid is that

any point $q \in \Delta(\Omega)$ is closer to $p[\ell]$ than to any other $p[\ell']$ if and only if it is closer to $p[\ell]$ than to any of its neighbors.

Consider the game introduced in the proof of Theorem 4.3, except that choices of $\ell_n \in \mathcal{L}_\varepsilon$ and ω_n generate an internal regret R'_n whose (ℓ,ℓ')-th component is

$$(R'_n)^{\ell,\ell'} = \begin{cases} \left\|\omega_n - p[\ell]\right\|^2 - \left\|\omega_n - p[\ell']\right\|^2 & \text{if } \ell = \ell_n \text{ and } \ell' \in \mathcal{N}[\ell] \\ 0 & \text{otherwise} \end{cases}.$$

As a consequence, using the simple fact concerning averages of norms,

$$\|R'_n\| \le 4\frac{\varepsilon}{\sqrt{d}} \text{ and } \overline{R}'_n = \frac{|\mathbb{N}_n[\ell]|}{n}\left(\left\|\overline{\omega}_n[\ell] - p[\ell]\right\|^2 - \left\|\overline{\omega}_n[\ell] - p[\ell']\right\|^2\right)_{l\in\mathcal{L}_\varepsilon, k\in\mathcal{N}[\ell]}.$$

Same arguments as in the proof of Proposition 4.3 yield that playing, at stage $n+1$, any invariant measure of $(\overline{R}'_n)^+$ ensures that $\mathbb{E}_{\sigma,\tau}\left[\left\|(\overline{R}'_n)^+\right\|^2\right] \le 16\varepsilon^2/(dn)$.

It remains to relate $\left\|\overline{\omega}_n[\ell] - p[\ell]\right\|$ to $\|\overline{R}'_n\|$. First, we write $\overline{\omega}_n[\ell] = p[\ell] + \sum_{k=1}^d x_k \mathbf{e}_k$ where we can assume (up to a change of signe) that every x_k is positive and even $x_k \ge \varepsilon/\sqrt{d}$ (otherwise $\overline{\omega}_n[\ell]$ is even closer to $p[\ell]$).

We denote by $p[\ell_k] = p[\ell] + 2\varepsilon\mathbf{e}_k/\sqrt{d}$ the neighbor of $p[\ell]$ in the direction of $\mathbf{e}_k$, so that $\mathbf{e}_k = (p[\ell_k] - p[\ell])/\|p[\ell_k] - p[\ell]\|$. Triangle inequality implies that

$$\begin{aligned}
\left\|p[\ell] - \overline{\omega}_n[\ell]\right\| &\le \left\|p[\ell] - \left(p[\ell] + \sum_{k=1}^d \frac{\varepsilon}{\sqrt{d}}\mathbf{e}_k\right)\right\| + \left\|\left(p[\ell] + \sum_{k=1}^d \frac{\varepsilon}{\sqrt{d}}\mathbf{e}_k\right) - \overline{\omega}_n[\ell]\right\| \\
&= \varepsilon + \left\|p[\ell] + \sum_{k=1}^d \frac{p[\ell_k] - p[\ell]}{2} - \overline{\omega}_n[\ell]\right\| \\
&= \varepsilon + \sqrt{\sum_{k=1}^d \left\langle p[\ell] + \sum_{k=1}^d \frac{p[\ell_k] - p[\ell]}{2} - \overline{\omega}_n[\ell], \mathbf{e}_k\right\rangle^2} \\
&= \varepsilon + \sqrt{\sum_{k=1}^d \left\langle \frac{p[\ell_k] + p[\ell]}{2} - \overline{\omega}_n[\ell], \frac{p[\ell_k] - p[\ell]}{\|p[\ell_k] - p[\ell]\|}\right\rangle^2} \\
&= \varepsilon + \sqrt{\sum_{k=1}^d \frac{\left(\left\|\overline{\omega}_n[\ell] - p[\ell]\right\|^2 - \left\|\overline{\omega}_n[\ell] - p[\ell_k]\right\|^2\right)^2}{4\left\|p[\ell_k] - p[\ell]\right\|^2}} \\
&= \varepsilon + \frac{\sqrt{d}}{4\varepsilon}\sqrt{\sum_{k=1}^d \left(\left\|\overline{\omega}_n[\ell] - p[\ell]\right\|^2 - \left\|\overline{\omega}_n[\ell] - p[\ell_k]\right\|^2\right)^2}.
\end{aligned}$$

To sum up, we have prove that, for every $\ell \in \mathcal{L}_\varepsilon$,

$$\left\|p[\ell] - \overline{\omega}_n[\ell]\right\| - \varepsilon \le \frac{\sqrt{d}}{4\varepsilon}\sqrt{\sum_{k\in\mathcal{N}[\ell]} \left[\left(\left\|\overline{\omega}_n[\ell] - p[\ell]\right\|^2 - \left\|\overline{\omega}_n[\ell] - p[\ell_k]\right\|^2\right)^+\right]^2}.$$

Multiplying both sides of this inequality by $\frac{|\mathbb{N}_n[\ell]|}{n}$, taking the square and summing over $\ell \in \mathcal{L}_\varepsilon$, one obtains

$$\left\|\widetilde{R}_n\right\|_2^2 := \sum_{\ell\in\mathcal{L}_\varepsilon}\left[\frac{|\mathbb{N}_n[\ell]|}{n}\Big(\left\|p[\ell]-\overline{\omega}_n[\ell]\right\|-\varepsilon\Big)\right]^2 \le \frac{d}{16\varepsilon^2}\left\|\left(\overline{R}'_n\right)^+\right\|_2^2 ,$$

therefore the strategy σ ensures that

$$\mathbb{E}_{\sigma,\tau}\left[\sup_{\ell\in\mathcal{L}_\varepsilon}\frac{|\mathbb{N}_n[\ell]|}{n}\Big(\left\|p[\ell]-\overline{\omega}_n[\ell]\right\|-\varepsilon\Big)\right] \le \mathbb{E}_{\sigma,\tau}\left[\left\|\widetilde{R}_n\right\|_2\right] \le \sqrt{\frac{1}{n}} .$$

This gives the first part of the proof. The last part is due to the fact that there are less than $(2\sqrt{d})^d$ points in the ε-ball centered at some $p \in \Delta(\Omega)$. ■

Proposition 4.5 also allows to recover the following result of Mannor & Stoltz [50]

Theorem 4.4 *When $\mathcal{F}$ is the family of all Borel subsets of $\Delta(\Omega)$, there exists a $\mathcal{F}$-calibrated strategy σ such that, for every strategy τ of Nature,*

$$\mathbb{E}_{\sigma,\tau}\left[\frac{|N_n[F]|}{n}\left\|\overline{p}_n[F]-\overline{\omega}_n[F]\right\|\right] \le 7n^{-\frac{1}{\Omega+1}}, \quad \mathbb{P}_{\sigma,\tau}\text{-a.s.},$$

and, for every $\delta > 0$, with probability at least $1-\delta$, one also has

$$\frac{|N_n[F]|}{n}\left\|\overline{p}_n[F]-\overline{\omega}_n[F]\right\| \le \frac{9}{n^{\frac{1}{\Omega+1}}} + 2\sqrt{\frac{\log\left(\frac{2}{\delta}\right)}{n}}, \quad \mathbb{P}_{\sigma,\tau}\text{-a.s.}.$$

Proof: The result is a consequence of a doubling trick applied to strategies constructed in Proposition 4.5. Assume that the strategy adapted to some ε is played during N stages. On those stages, one has

$$\frac{|\mathbb{N}_n[F]|}{n}\left\|\overline{p}_n[F]-\overline{\omega}_n[F]\right\| - \varepsilon \le \sum_{\ell\in\mathcal{L}_{\varepsilon_k}}\frac{|\mathbb{N}_n[\ell]|}{n}\left(\left\|p[\ell]-\overline{\omega}_n[\ell]\right\|-\varepsilon\right) \le \sqrt{L_\varepsilon}\left\|\widetilde{R}_n\right\|_2 .$$

Taking expectation and using the fact that $L_\varepsilon \le \varepsilon^{-d}$ yield that

$$\mathbb{E}_{\sigma,\tau}\left[\frac{|\mathbb{N}_n[F]|}{n}\left\|\overline{p}_n[F]-\overline{\omega}_n[F]\right\|\right] \le \varepsilon + \sqrt{\frac{1}{\varepsilon^d}}\sqrt{\frac{1}{n}}$$

Hence, the doubling trick adapted to the sequences $\varepsilon_k = \left(\frac{d}{2^k}\right)^{\frac{1}{d+2}}$, played during 2^k stages ensures that, denoting $n = 2^{k_0} + m < 2^{k_0+1}$,

$$\begin{aligned}\mathbb{E}_{\sigma,\tau}\left[\frac{|\mathbb{N}_n[F]|}{n}\left\|\overline{p}_n[F]-\overline{\omega}_n[F]\right\|\right] &\le \frac{1}{n}\left(\sum_{k=0}^{k_0-1}2^k 2\varepsilon_k + m\varepsilon_{k_0} + \sqrt{\frac{1}{\varepsilon_{k_0}^d}m}\right)\\ &\le 2\frac{1}{2^{k_0}}\sum_{k=0}^{k_0}2^{k\frac{d+1}{d+2}} \le \frac{4}{2^{\frac{d+1}{d+2}}-1}\frac{1}{2^{\frac{k_0+1}{d+2}}} \le 7\frac{1}{n^{\frac{1}{d+2}}}.\end{aligned}$$

High probability bounds are classics consequences of concentration inequalities, since $|\mathbb{N}_n[F]|\left\|\overline{\omega}_n[F]-\overline{p}_n[F]\right\|/n = \left\|\overline{Y}_n\right\|$ where the sequence $Y_n = (\omega_n - p_n)\,\mathbb{1}\{p_n \in F\}$ is such that $\|Y_n\|^2 \leq 2$ and, by Jensen inequality, $\|\overline{Y}_n\| \leq \|\overline{Y}_n - \mathbb{E}[\overline{Y}_n]\| + E[\|\overline{Y}_n\|]$. ■

In fact, Theorem 4.4 slightly improves the result of Mannor & Stoltz [50] since it implies that

$$\limsup_{n\to\infty} n^{\frac{1}{\Omega+1}} \frac{|N_n[F]|}{n}\left\|\overline{\omega}_n[F]-\overline{p}_n[F]\right\| \leq 7, \quad \mathbb{P}_{\sigma,\tau}\text{-as.}$$

Rakhlin, Sridharan & Tewari [66] wrote the calibration problem in terms of a generalized regret, see Section 2.2.2. Formally, assume that actions spaces are respectively $\Delta(\Omega)$ and Ω and that the stage game payoff is null, i.e. $g(p,\omega) = 0$. The class of departure function considered are $\left\{\xi_{p,\lambda};\ p \in \Delta(\Omega), \lambda > 0\right\}$ where $\xi_{p,\lambda} : \Delta(\Omega) \times \Omega \to \mathbb{R}^\Omega$ and the evaluation mappings $B_n : \left(\mathbb{R}^\Omega\right)^n \to \mathbb{R}$ are defined by, for every $n \in \mathbb{N}$,

$$\xi_{p,\lambda}[g](p_n,\omega_n) = \mathbb{1}\{\|p_n - p\|_1 \leq \lambda\}(p_n - \delta_{\omega_n}) \text{ and } B_n(Z_1,\ldots,Z_n) = \left\|\frac{1}{n}\sum_{m=1}^{n} Z_m\right\|.$$

As a consequence, one easily has that regret is upper bounded by calibration score, as

$$\begin{aligned}&\sup_{p,\lambda} B_n\Big(\xi_{p,\lambda}[g](p_1,\omega_1),\ldots,\xi_{p,\lambda}[g](p_n,\omega_n)\Big) - B_n\Big(g(p_1,\omega_1),\ldots,g(p_n,\omega_n)\Big)\\ =&\sup_{p,\lambda} \frac{\mathbb{N}_n[p,\lambda]}{n}\left\|\overline{\omega}_n[p,\lambda]-\overline{p}_n[p,\lambda]\right\|.\end{aligned}$$

The max-min formulation of the regret minimization problem (see Section 2.2.2) proves that $\sup_{p,\lambda}\left\|\overline{\omega}_N[p,\lambda]-\overline{p}_N[p,\lambda]\right\|\mathbb{N}_N[p,\lambda]/N$ is upper bounded at the final stage N by $c\Omega^2\sqrt{\log(N)/N}$ where c is a universal constant. An alternative (and actually more general) proof is given in the next section.

4.3 Using Approachability to get (smooth and generalized) Calibration

In this section, we show that recent results in calibration can be rewritten solely as the existence or construction of some approachability strategy. The first result we exhibit is a generalization of both a previous one of Perchet [60] (since the strategy is calibrated with respect to much larger families) and Rakhlin, Sridharan & Tewari [66] (because the proof is constructive and not horizon dependent).

Theorem 4.5 *Let $\mathcal{F} := \left\{B[p,\lambda]_\infty; p \in \Delta(\Omega), \lambda > 0\right\}$ be the family of ℓ_∞-balls. Then there exists a calibrated strategy σ such that, no matter the strategy τ of Nature and for*

every $n \in \mathbb{N}$,

$$\mathbb{E}_{\sigma,\tau}\left[\sup_{p\in\Delta(\Omega),\lambda>0} \frac{|\mathbb{N}_n[p,\lambda]|}{n}\left\|\omega_n[p,\lambda]-\overline{p}_n[p,\lambda]\right\|_\infty\right] \leq 12\sqrt{2\frac{\Omega}{n}\log\left(\frac{2n}{\Omega}\right)}.$$

Proof: Let $\varepsilon > 0$ be fixed. As in the proof of Proposition 4.5, the set $\Delta(\Omega)$ is represented as a subset of $\mathbb{R}^d$ (with $d = \Omega - 1$),

$$\Delta(\Omega) = \left\{p = (p_1,\ldots,p_d) \in \mathbb{R}^d \text{ s.t. } p_1,\ldots,p_d \geq 0 \text{ and } \sum_{k=1}^d p_k = 1\right\}$$

and we consider the regular ε-grid $\mathcal{L}_\varepsilon$ defined by

$$\left\{\sum_{k=1}^d 2n_k\varepsilon.\mathbf{e}_k \in \Delta(\Omega); n_k \in \mathbb{N}\right\} =: \left\{p[\ell] = \sum_{k=1}^d 2n_k[\ell]\varepsilon.\mathbf{e}_k\,;\, \ell \in \mathcal{L}_\varepsilon\right\}.$$

Although the family of ℓ_∞-balls is infinite, the number of different possible intersections of such a ball with the grid $\mathcal{L}_\varepsilon$ is obviously finite (it is trivially bounded by its number of subsets, 2^{L_ε}). However, an ℓ_∞-ball $B_\infty(p,\lambda)$ is rectangular and can be described by two extreme points: the lowest corner $p - \sum_{k=1}^d \lambda\mathbf{e}_k$ and the highest corner (in every direction) $p + \sum_{k=1}^d \lambda\mathbf{e}_k$

The grid $\mathcal{L}_\varepsilon$ is regular, so this characterization holds for intersections with ℓ_∞ balls: they are characterized by two extreme points. As a consequence, they are at most $\mathcal{L}_\varepsilon \leq \varepsilon^{-2d}$ different possible intersections. Consider a fixed family of ℓ_∞-balls that induce exactly these different intersections, and denote it $\left\{B_\infty(\mathbf{p}[k],\lambda_k);\ k \in \mathcal{K}\right\}$.

We introduce an auxiliary game with action space $\mathcal{L}_\varepsilon$ and Ω, payoff mapping

$$g(\ell,\omega) = \left(\mathbb{1}\left\{\|p[\ell]-\mathbf{p}[k]\| \leq \lambda_k\right\}(p[\ell]-\delta_\omega)\right)_{k\in\mathcal{K}}$$

and consider the closed and convex target set $\mathcal{C} := B_\infty(0,\varepsilon) \subset \left(\mathbb{R}^d\right)^K$.

Given $q \in \Delta(\Omega)$, the pure action ℓ corresponding to a point of the grid $p[\ell]$ such that $\|p[\ell]-q\|_\infty \leq \varepsilon$ ensures that $g(\ell,q)$ belongs to $\mathcal{C}$ which is therefore approachable. Moreover, since $\mathcal{C}$ is rectangular, the approachability strategy of Corollary 1.16, adapted to the potential $\Phi(z) = \frac{1}{\eta}\log\left(\sum_{k\in\mathcal{K}}\sum_{i=1}^d e^{\eta(z_i^k-\varepsilon)} + e^{-\eta(z_i^k-\varepsilon)}\right)$, ensures that

$$\mathbb{E}_{\sigma,\tau}\left[\|\overline{g}_n\|_\infty\right] \leq \varepsilon + 4\eta + \frac{\log(2dK)}{\eta n} \leq \varepsilon + 4\eta + \frac{\log(2d)+2d\log(1/\varepsilon)}{\eta n}.$$

Therefore, given $N \in \mathbb{N}$ such that $N \geq 2ed$, the choice of $\varepsilon/4 = \eta = \sqrt{\frac{d}{8N}\log\left(\frac{N}{2d}\right)}$ ensures in particular that

$$\mathbb{E}_{\sigma,\tau}\left[\|\overline{g}_N\|_\infty\right] \leq 6\sqrt{\frac{d}{N}\log\left(\frac{N}{2d}\right)}.$$

As usual, when playing by blocks of increasing size 2^m (starting at $\underline{m}$ such that $2^{\underline{m}} \geq ed$), the last two displays ensure that, for every $n \in \mathbb{N}$,

$$\mathbb{E}_{\sigma,\tau}\left[\|\overline{g}_n\|_\infty\right] \leq 12\sqrt{2\frac{d}{n}\log\left(\frac{2n}{d}\right)}.$$

The result comes from the fact that, by construction, for every $n \in \mathbb{N}$,

$$\sup_{p\in\Delta(\Omega),\lambda>0} \frac{|\mathbb{N}_n[p,\lambda]|}{n}\left\|\omega_n[p,\lambda] - \overline{p}_n[p,\lambda]\right\|_\infty = \|\overline{g}_n\|_\infty.$$

■

If $d \geq 3$, since $\mathcal{L}_\varepsilon \leq \varepsilon^{2d}/d!$, constants in Theorem 4.5 can be lowered if one is only interested in the asymptotic behavior. This result holds almost surely since, using concentration inequalities, with $\mathbb{P}_{\sigma,\tau}$ probability at least $1-\delta$,

$$\left\|\overline{g}_n\right\|_\infty \leq 12\sqrt{2\frac{d}{n}\log\left(\frac{2n}{d}\right)} + 2\sqrt{\frac{2d}{n}\log\left(\frac{2n}{d}\right) + \frac{1}{n}\log\left(\frac{1}{\delta}\right)}.$$

Statement concern ℓ_∞ balls; however, it is also possible to show that for other ℓ_p-balls, the number of possible intersection with the grid is bounded by $O\left(\frac{1}{\varepsilon}\right)^{O(\Omega^2)}$ (see e.g. Rakhlin, Sridharan & Tewari [66]). Thus the results holds, up to some polynomial term in Ω, for any other ℓ_p-norm.

This technique could actually have been used to proved Theorem 4.4, a similar result with respect to the family of Borel sets. The difference is that the number of possible intersection between Borel sets and our grid would have been in the order of $2^{1/\varepsilon^d}$. After taking the logarithm, equalizing the three remaining terms in regret ε, η and $1/(\varepsilon^d\eta n)$ yields that $\varepsilon = \eta = n^{-1/(d+2)}$. This would have been the bound on expected regret.

We now turn to calibration with checking rules and smooth calibration, and we show that they can be reduced to approachability problems. We recall that given a pair of mappings $\mathcal{U}$ and $\mathcal{T}$, we defined

$$\mathbb{N}_n[\mathcal{U},\mathcal{T}] = \left\{m \leq n \text{ s.t. } (p_m,\omega_m) \in \mathcal{U}(h^{m-1})\right\},$$

the empirical probability of tested events

$$\overline{\omega}_n[\mathcal{U},\mathcal{T}] = \frac{\sum_{m\in\mathbb{N}_n[\mathcal{U},\mathcal{T}]} \mathbb{1}\{(p_m,\omega_m) \in \mathcal{T}(h^{m-1})\}}{|\mathbb{N}_n[\mathcal{U},\mathcal{T}]|},$$

and the average predicted conditional probability of tested events

$$\overline{p}_n[\mathcal{U},\mathcal{T}] = \frac{\sum_{m\in\mathbb{N}_n[\mathcal{U},\mathcal{T}]} p_m\left\{(p_m,\omega_m) \in \mathcal{T}(h^{m-1}) \,\middle|\, \mathcal{U}(h^{m-1})\right\}}{|\mathbb{N}_n[\mathcal{U},\mathcal{T}]|}.$$

If a checking rule is independent of current predictions, then the same definition hold with $(p_m,\omega_m) \in \mathcal{U}(h^n)$ (resp. in $\mathcal{T}(h^n)$) replaced by $\omega_m \in \mathcal{U}(h^n)$ (resp. in $\mathcal{T}(h^n)$).

Theorem 4.6 *Let λ be a probability distribution on the set of checking-rules independent of current predictions. Then there exists a deterministic strategy σ that is calibrated with λ-almost every checking rules such that, $\mathbb{P}_{\sigma,\tau}$-almost surely,*

$$\limsup_{n\to\infty} \left|\overline{\omega}_n[\mathcal{U},\mathcal{T}] - \overline{p}_n[\mathcal{U},\mathcal{T}]\right| \leq 0,$$

as soon as $|\mathbb{N}_n[\mathcal{U},\mathcal{T}]|$ increases to infinity.

Proof: Proof relies essentially on approachability with activation in infinite dimension. We define an auxiliary game where payoff is a random variable over the set of checking rules independent of current predictions. Action set of the player is reduced to $\Delta(\Omega)_0$, the interior of $\Delta(\Omega)$ – so that conditional probabilities are well defined – and payoff at stage n is $\mathbb{1}\{\omega_n \in \mathcal{T}(h^{n-1})\} - p\{\mathcal{T}(h^{n-1})|\mathcal{U}(h^{n-1})\}$ if the coordinates $(\mathcal{U},\mathcal{T})$ is active, i.e., if $n \in \mathbb{N}_n[\mathcal{U},\mathcal{T}]$.

By definition, average payoff at stage n is exactly $\overline{\omega}_n[\mathcal{U},\mathcal{T}] - \overline{p}_n[\mathcal{U},\mathcal{T}]$ and we shall construct a strategy σ that approaches the convex set $\{0\}$, that is, using Theorem 1.6, find $p_{n+1} \in \Delta(\Omega)_0$ such that, for every $\omega \in \Omega$,

$$\int \Big(\overline{\omega}_n[\mathcal{U},\mathcal{T}] - \overline{p}_n[\mathcal{U},\mathcal{T}]\Big).\frac{\Big(\mathbb{1}\{\omega \in \mathcal{U}(h^n)\}\Big(\mathbb{1}\{\omega \in \mathcal{T}(h^n)\} - p_{n+1}\{\mathcal{T}(h^n)|\mathcal{U}(h^n)\}\Big)}{|\mathbb{N}_{n+1}[\mathcal{U},\mathcal{T}]|}d\lambda$$

is less or equal to zero (or at least smaller than $\varepsilon_n = 1/n^2$).

To construct this p_{n+1}, we consider the game with payoff defined on $\Delta(\Omega)_0$ and Ω by

$$g(p,\omega) = \int \frac{\overline{\omega}_n[\mathcal{U},\mathcal{T}] - \overline{p}_n[\mathcal{U},\mathcal{T}]}{1+|\mathbb{N}_n[\mathcal{U},\mathcal{T}]|}\mathbb{1}\{\omega \in \mathcal{U}(h^n)\}\Big(\mathbb{1}\{\omega \in \mathcal{T}(h^n)\} - p\{\mathcal{T}(h^n)|\mathcal{U}(h^n)\}\Big)d\lambda$$

and g is extended linearly in its second variable on $\Delta(\Omega)$. Since one always has

$$\mathbb{1}\{\omega \in \mathcal{U}(h^n)\}\frac{\overline{\omega}_n[\mathcal{U},\mathcal{T}] - \overline{p}_n[\mathcal{U},\mathcal{T}]}{1+|\mathbb{N}_n[\mathcal{U},\mathcal{T}]|} = \mathbb{1}\{\omega \in \mathcal{U}(h^n)\}\frac{\overline{\omega}_n[\mathcal{U},\mathcal{T}] - \overline{p}_n[\mathcal{U},\mathcal{T}]}{|\mathbb{N}_{n+1}[\mathcal{U},\mathcal{T}]|}$$

the integrals in the last two displayed equations coincide, so we just need to prove that there exists $p_{n+1} \in \Delta(\Omega)_0$ such that $g(p_{n+1},\omega) \leq \varepsilon_n$, for every $\omega \in \Omega$ or, more generally, that

$$\inf_{p\in\Delta(\Omega)_0} \sup_{\omega\in\Omega} g(p,\omega) \leq 0.$$

And this is a consequence of Lemma 5.1, since $g(p,p) = 0$ for every $p \in \Delta(\Omega)_0$, $g(p,\cdot)$ is affine and $g(\cdot,p)$ is continuous on $\Delta(\Omega)_0$. ■

When checking rules might depend on current predictions (see Sandroni, Smorodinsky & Vohra [68] or Foster, Rakhlin, Sridharan & Tewari [22]), the result and proof are almost identical.

Proposition 4.6 *Let λ be a probability distribution on the set of checking-rules. Then there exists a strategy σ that is calibrated with λ-almost every checking rules such that, $\mathbb{P}_{\sigma,\tau}$-almost surely,*

$$\limsup_{n\to\infty} \left|\overline{\omega}_n[\mathcal{U},\mathcal{T}] - \overline{p}_n[\mathcal{U},\mathcal{T}]\right| \leq 0,$$

as soon as $|\mathbb{N}_n[\mathcal{U},\mathcal{T}]|$ increases to infinity.

Proof: the proof is almost identical to the case of checking rule independent of predictions. The only difference lies in the definition of the payoff $g(p,\omega)$ which is

$$\int \frac{\overline{\omega}_n[\mathcal{U},\mathcal{T}] - \overline{p}_n[\mathcal{U},\mathcal{T}]}{1+|\mathbb{N}_n[\mathcal{U},\mathcal{T}]|} \mathbb{1}\{(\omega,p)\in\mathcal{U}(h^n)\}\Big(\mathbb{1}\{(\omega,p)\in\mathcal{T}(h^n)\} - p\left\{\mathcal{T}(h^n)|\mathcal{U}(h^n)\right\}\Big)d\lambda.$$

Since $g(\cdot,\omega)$ might not be continuous, Lemma 5.1 does not apply. However, g is bounded and defined over $\Delta(\Omega)_0$ and Ω, the former being measurable and the latter finite. Therefore, see Sorin [70] Theorem A.9, this game has a value in mixed action. And this value has to be smaller than 0 since $g(p,p)=0$ for every $p\in\Delta(\Omega)_0$. ■

The last similar reduction to approachability concerns smooth calibration.

Theorem 4.7 *There exists a deterministic strategy σ of the player such that, no matter Nature's strategy, for every continuous mapping $g:\Delta(\Omega)\to\mathbb{R}_+$,*

$$\limsup_{n\to\infty} \frac{1}{n}\left\|\sum_{m=1}^{n} g(p_m)(\omega_m - p_m)\right\| \leq 0\ .$$

The same result holds if one adds checking rules independent of current predictions.

Proof: The set of continuous mappings from $\Delta(\Omega)$ to $\mathbb{R}_+$ is separable and we denote by λ a probability distribution with support $\{g_k; k\in\mathbb{N}\}$, a dense countable family.

Following the lines of the proof of Therorem 4.6, we define

$$\overline{\omega}_n[g_k] = \frac{\sum_{m=1}^{n} g_k(p_m)\omega_m}{n} \text{ and } \overline{p}_n[g_k] = \frac{\sum_{m=1}^{n} g_k(p_m)p_m}{n}.$$

Then, Corollary 1.4 ensures the existence of an approachability strategy such that, for every $k\in\mathbb{N}$, $\left\|\overline{\omega}_n[g_k] - \overline{p}_n[g_k]\right\|$ converges to zero. Indeed, one just has to prove that $\{0\}\subset\mathcal{L}_2$ is approachable, thus that for every $n\in\mathbb{N}$, there exists $p_{n+1}\in\Delta(\Omega)$ such that, no matter $\omega\in\Omega$,

$$\int \Big(\overline{\omega}_n[g_k] - \overline{p}_n[g_k]\Big)\Big(g_k[p_{n+1}](\omega - p_{n+1})\Big)d\lambda \leq 0,$$

where we assumed that $0/0=0$. The existence of such $p\in\Delta(\Omega)$ is again a consequence of Ky Fan's inequality generalized in Lemma 5.4.

Since $\{g_k; k \in \mathbb{N}\}$ is a dense family, necessarily $\left\|\overline{\omega}_n[g] - \overline{p}_n[g]\right\|$ must converges to zero, for every continuous mapping g. ■

A close look to the first proof of existence of deterministic smooth calibrated strategies, due to Kakade & Foster [37], shows that they also have constructed an ε-approachability strategy (and then used a doubling trick). We proposed here a direct (and maybe more intuitive) proof.

4.4 Using calibration to get regret and approachability

Calibration in some auxiliary game can be seen as a useful tool to construct strategies that satisfies another criterion as approachability, no internal regret and so on. This idea goes back to Foster & Vohra [23] and was used, recently, by Perchet [59, 61, 62]; in particular, it is useful in a specific case of general regret (see Section 2.2.2) defined below.

But first, we focus on usual internal regret in the finite case (although it can be generalized immediately when $\mathcal{B}$ is any compact set). Recall that a strategy is internally consistent if the supremum limit of

$$\frac{|\mathbb{N}_n[a]|}{n}\left(\max_{a^*\in\mathcal{A}} \rho\big(a^*, \overline{b}_n[a]\big) - \rho\big(a, \overline{b}_n[a]\big)\right)$$

is non positive. By linearity of $\rho(a,\cdot)$, this quantity can be immediately rewritten into

$$\frac{|\mathbb{N}_n[a]|}{n}\left(\left(\left\|\rho\big(a,\cdot\big) - \overline{b}_n[a]\right\|^2 - \|\rho(a,\cdot)\|^2\right) - \min_{a^*\in\mathcal{A}}\left(\left\|\rho\big(a^*,\cdot\big) - \overline{b}_n[a]\right\|^2 - \|\rho(a^*,\cdot)\|^2\right)\right),$$

up to a factor 2. As a consequence, any weighted-calibrated strategy with respect to $\{\rho(a,\cdot), \|\rho(a,\cdot)\|^2; a \in \mathcal{A}\}$ is internally consistent. Since scores are actually exactly the same, rates of convergence of weighted calibration give rates for regret minimization.

We now turn to generalized regret. Assume that $\mathcal{A}$ and $\mathcal{B}$ are two compact and convex sets and let $G : \mathcal{A}\times\mathcal{B} \to \mathbb{R}$ be any fixed *evaluation* mapping that might not be linear in any of its coordinates. In this framework, a strategy has no G-external regret if

$$\limsup_{n\to\infty} \sup_{a^*\in\mathcal{A}} G(a^*, \overline{b}_n) - G(\overline{a}_n, \overline{b}_n) \le 0.$$

To define internal regret, assume that a strategy only uses a finite number of actions in $\mathcal{A}_{\mathcal{L}} = \{a[\ell]; \ell \in \mathcal{L}\}$, so that σ is actually a mapping from the set of finite histories into $\mathcal{L}$, and $\ell_n = \ell$ means that action $a[\ell]$ is played at stage n. Define

$$\mathbb{N}_n[\ell] = \{m \le n \text{ s.t. } \ell_m = \ell\}, \text{ and } \overline{b}_n[\ell] = \frac{1}{|\mathbb{N}_n[\ell]|}\sum_{m\in\mathbb{N}_n[\ell]} b_m.$$

A strategy has no $(\mathcal{L},\varepsilon)$-internal regret if, no matter the strategy τ of Nature, $\mathbb{P}_{\sigma,\tau}$-almost surely,

$$\limsup_{n\to\infty}\frac{|\mathbb{N}_n[\ell]|}{n}\Big(\sup_{a^*\in\mathcal{A}}G\big(a^*,\overline{b}_n[\ell]\big)-G\big(a[\ell],\overline{b}_n[\ell]\big)-\varepsilon\Big)\leq 0,\quad \forall\ell\in\mathcal{L}.$$

Proposition 4.7 *If G is continuous, then for every $\varepsilon>0$ there exists a $(\mathcal{L},\varepsilon)$-internally consistent strategy. However, their might not exist any (ε-)externally consistent strategies.*

Proof: Since G is continuous, for every $\varepsilon>0$, there exists some δ such that $\|(a,b)-(a',b')\|\leq\delta$ implies that $|G(a,b)-G(a',b')|\leq\varepsilon/2$. Consider σ' any calibrated strategy with respect to $\Big\{b[\ell];\ell\in\mathcal{L}\Big\}$, a $\delta/2$ grid of $\mathcal{B}$. Assume that when σ' predicts $b[\ell]$, then σ dictates to play $a[\ell]\in\operatorname{argmax}_{a\in\mathcal{A}}G(a,b[\ell])$.

Since σ' is calibrated, for every $\eta>0$, one has that, $\mathbb{P}_{\sigma',\tau}$-as after some stage N,

$$\sup_{\ell\in\mathcal{L}}\frac{|\mathbb{N}_n[\ell]|}{n}\left(\Big\|b[\ell]-\overline{b}_n[\ell]\Big\|^2-\frac{\delta^2}{4}\right)\leq\eta.$$

In particular, as soon as $\eta\leq\delta^2/4$, either $\eta\frac{n}{|\mathbb{N}_N[\ell]|}$ is smaller than $\delta^2/4$ and then $\|b[\ell]-\overline{b}_n[\ell]\Big\|^2\leq\delta^2$, or $\frac{|\mathbb{N}_N[\ell]|}{n}$ is smaller than $4\eta/\delta^2$.

The first case implies that $G(a,b[\ell])-G(a,\overline{b}_n[\ell])\leq\varepsilon/2$ for every $a\in\mathcal{A}$, thus in both cases one has that, after stage N,

$$\sup_{\ell\in\mathcal{L}}\frac{|\mathbb{N}_n[\ell]|}{n}\Big(\sup_{a^*\in\mathcal{A}}G(a^*,b[\ell])-G(a[\ell],\overline{b}_n[\ell])-\varepsilon\Big)\leq\frac{8\|G\|_\infty}{\delta^2}\eta,$$

which characterizes a $(\mathcal{L},\varepsilon)$-calibrated strategy.

It remains to prove that there might not exist externally consistent strategies. Define $G(a,b)=(1-4b)a$, for every $a\in[0,1]$ and $b\in[0,1]$ and assume that during the first N stages (with N is large enough) $b_n=0$. Necessarily $\overline{a}_N$ is arbitrarily close to 1. During the next N stage, define $b_n=1$ then $\overline{a}_{2N}$ is at most $1/2$ thus the external regret is of at least $1/2$. ■

We now prove how to construct an ε-approachability strategy via calibration. Given a closed and compact set $\mathcal{C}\subset\mathbb{R}^d$ and a vector payoff mapping $g:\Delta(\mathcal{A})\times\Delta(\mathcal{B})\to\mathbb{R}^d$, define $G(x,y)=-d_{\mathcal{C}}(g(x,y))$ for every $x\in\Delta(\mathcal{A})$ and $y\in\Delta(\mathcal{B})$. If $\mathcal{C}$ is approachable, then Blackwell's condition ensures that $\sup_{x^*\in\Delta(\mathcal{A})}G(x^*,y)=0$ for every $y\in\Delta(\mathcal{B})$. By convexity of $d_{\mathcal{C}}$ and the triangle inequality,

$$\begin{aligned}d_{\mathcal{C}}(\overline{g}_n)&\leq\sum_{\ell\in\mathcal{L}}\frac{|\mathbb{N}_n[\ell]|}{n}d_{\mathcal{C}}(\overline{g}_n[\ell])\\&\leq\varepsilon+\sum_{\ell\in\mathcal{L}}\frac{|\mathbb{N}_n[\ell]|}{n}\Big(d_{\mathcal{C}}\big(g(a[\ell],\overline{b}_n[\ell])\big)-\varepsilon\Big)+\sum_{\ell\in\mathcal{L}}\frac{|\mathbb{N}_n[\ell]|}{n}\Big\|g(x[\ell],\overline{b}_n[\ell])-\overline{g}_n[\ell]\Big\|.\end{aligned}$$

Both sums converges almost surely to zero, respectively because σ has no internal regret (with respect to G) and because of concentration inequalities since $g(x[\ell], b_n) = \mathbb{E}[g(a_n, b_n)]$. One can resort to the doubling trick (since we can easily derive uniform speed of convergence) to get an approachability strategy.

4.5 Using regret to get approachability

We proved in the last section how calibration and generalized regret can be used to construct approachability strategy, as noticed by Perchet [59] or Rakhlin, Sridharan & Tewari [66]. A completely different link can also be formulated between regret and approachability, as discovered recently by Abernathy, Bartlett & Hazan [1]. We recall that Blackwell's strategy consists in playing, at stage $n+1$, optimally in the zero-sum projected game $\langle g(x,y) - \pi_{\mathcal{C}}(\overline{g}_n), \overline{g}_n - \pi_{\mathcal{C}}(\overline{g}_n)\rangle$. Abernathy, Bartlett & Hazan [1] proposed to use a regret minimization scheme to determine, stage by stage, in which projected game to play (i.e., not necessarily along the direction $\overline{g}_n - \pi_{\mathcal{C}}(\overline{g}_n)$).

The formulation is rather simple when $\mathcal{C} = \{0\} \subset \mathbb{R}^d$, so we will focus only on this case. It can however be generalized to any convex cone and therefore to any convex set in $\mathbb{R}^d$ (seen as a section of a convex cone in $\mathbb{R}^{d+1}$). The basic idea is to notice that, for $\mathcal{C} = \{0\}$ and every $n \in \mathbb{N}$,

$$d_{\mathcal{C}}(\overline{g}_n) = \|\overline{g}_n\| = \sup_{\theta \in B(0,1)} \langle \theta, \overline{g}_n \rangle, \text{ where } B(0,1) = \Big\{\theta \in \mathbb{R}^d, \|\theta\|_2 \leq 1\Big\}.$$

Assume that at stage m, the player played optimality in the projected game along the direction θ_{m-1}. Since $\mathcal{C}$ is approachable, this zero-sum game has a negative value, hence $\langle \theta_{m-1}, \mathbb{E}[g_m]\rangle \leq 0$. As a consequence,

$$\mathbb{E}\Big[d_{\mathcal{C}}(\overline{g}_n)\Big] = \mathbb{E}\Big[\|\overline{g}_n\|\Big] \leq \mathbb{E}\left[\sup_{\theta \in B(0,1)} \langle \theta, \overline{g}_n \rangle - \frac{1}{n}\sum_{m=1}^{n} \langle \theta_{m-1}, g_m \rangle\right].$$

The term inside the expectation can be written as the external regret if player and Nature's action set are respectively $B(0,1)$ and $\Big\{g(a,b); (a,b) \in \mathcal{A} \times \mathcal{B}\Big\}$. As a consequence, an approachability strategy can indeed be described as a two step procedure. At any stage n, choose, in a first step, a direction $\theta_n \in B(0,1)$ following any regret minimization algorithm. Then, in a second step, play optimally in the projected zero-sum game on θ_n.

Blackwell's strategy dictates to choose (in the first step) the direction θ_n that maximizes $\langle \theta, \overline{g}_n \rangle$; in other words, this is precisely the *follow the leader* algorithm that does not guarantee a shrinking regret (in full generality). The key point to understand this feature is that, by definition of the second step, $\langle \theta_m, \mathbb{E}[g_{m+1}]\rangle$ is always non-positive (no-matter the choice of θ_m) ; so, in this auxiliary game, Nature is in fact very restricted on her choice of actions and what is even more intricate, these restrictions depend on the player's move.

5 Appendix

5.1 Game Theory lemma

The following Lemma generalizes Ky Fan's inequality [19] recalled below:

Let $\mathcal{K}$ be a convex compact set of some Euclidian space and $g : \mathcal{K}\times\mathcal{K} \to \mathbb{R}$ such that, for every $y \in \mathcal{K}$, $g(\cdot, y)$ is concave over $\mathcal{K}$ and for every $x \in \mathcal{K}$, $g(x, \cdot)$ is continuous over K. If $g(x, x) = 0$ for every $x \in \mathcal{K}$, then there exists $x_0 \in \mathcal{K}$ such that $\sup_{x\in\mathcal{K}} g(x, x_0) \leq 0$.

Lemma 5.1 *Let g be a mapping on some compact and convex set $\mathcal{X} \subset \mathbb{R}^d$ such that $g(x, x) = 0$ for every $x \in \mathcal{X}_0$ the interior of $\mathcal{X}$ (such a mapping is called anti-symmetric).*

If for every $x \in \mathcal{X}$, $g(\cdot, x)$ is concave and $g(x, \cdot)$ is continuous and uniformly bounded by some $M > 0$ on X_0, then

$$\inf_{x\in\mathcal{X}_0} \sup_{x'\in\mathcal{X}} g(x', x) \leq 0.$$

Proof: Without loss of generality, we assume that 0 belongs to $\mathcal{X}_0$ and we denote, for every $\varepsilon > 0$ small enough, the convex compact set $\mathcal{X}_\varepsilon := \big\{(1-\varepsilon)x; x \in \mathcal{X}\big\}$. Then g and $\mathcal{X}_\varepsilon$ satisfy assumptions of Ky Fan's inequality. Thus, there exists x_ε such that $g(x, x_\varepsilon) \leq 0$ for every $x \in \mathcal{X}_\varepsilon$.

Given $x \in \mathcal{X}$, we denote by x_- the point on the boundary of $\mathcal{X}$ on the opposite direction of x, i.e., such that $\frac{x_-}{\|x_-\|} = -\frac{x}{\|x\|}$. We also define $\|\mathcal{X}_-\| = \inf_{x\in\mathcal{X}} \|x\|$.

Since $g(\cdot, x_\varepsilon)$ is concave, for every x in $\mathcal{X}$ that is not in $\mathcal{X}_\varepsilon$, one has

$$\frac{g\big(x, x_\varepsilon\big) - g\big((1-\varepsilon)x, x_\varepsilon\big)}{\varepsilon\|x\|} \leq \frac{g\big((1-\varepsilon)x, x_\varepsilon\big) - g\big(x_-, x_\varepsilon\big)}{(1-\varepsilon)\|x\| + \|x_-\|},$$

therefore, since $(1-\varepsilon)x \in \mathcal{X}_\varepsilon$ and $g(\cdot, x_\varepsilon) \leq 0$ on $\mathcal{X}_\varepsilon$, one has

$$g\big(x, x_\varepsilon\big) \leq -\frac{\varepsilon\|x\|}{(1-\varepsilon)\|x\| + \|x_-\|} g(x_-, x_\varepsilon) \leq \varepsilon \frac{M\|\mathcal{X}\|}{(2-\varepsilon)\|\mathcal{X}_-\|}.$$

Hence the result, since the right hand term goes to 0 as ε decreases to 0. ■

5.1.1 Uniform concentration inequalities

The following lemmas are central in different proofs. We recall that a process $Z_t \in \mathbb{R}^d$ is a martingale difference sequence if $\mathbb{E}\big[Z_{t+1}\big|Z_1, \ldots, Z_t\big] = 0$. Moreover, if $\|Z_t\|_2 \leq K$ then Hoeffding-Azuma's inequality in Euclidian spaces (see Corollary 3.5 in Kallenberg & Sztencel [38]) yields that, for every integer $T \geq 1$,

$$\mathbb{P}\left\{\|\overline{Z}_T\| \geq \varepsilon\right\} \leq \left(1 + \sqrt{\frac{T}{K^2}\varepsilon}\right) \exp\left(-\frac{1}{2}\frac{T}{K^2}\varepsilon^2\right) \leq 2\exp\left(-\frac{1}{4}\frac{T}{K^2}\varepsilon^2\right), \qquad (14)$$

or, $\mathbb{P}\left\{\|\overline{Z}_T\| \geq \phi^{-1}(\delta)/\sqrt{T}\right\} \leq \delta$ with $\phi(x) := (1+x/K)\exp\left(-x^2/2K^2\right)$. Actually, a weak maximal version of this inequality holds:

$$\mathbb{P}\left\{\exists\, t \leq T\,,\ \|\overline{Z}_t\| \geq \frac{T}{t}\varepsilon\right\} \leq \phi\left(\sqrt{T}\varepsilon\right) \quad \text{or} \quad \mathbb{P}\left\{\exists\, t \leq T\,,\ \|\overline{Z}_t\| \geq \frac{\sqrt{T}}{t}\phi^{-1}(\delta)\right\} \leq \delta.$$

For $d = 1$, one can define $\phi(x) = 2\exp\left(-x^2/2\right)$ and $\phi(x) = 2\exp\left(-x^2/4\right)$ otherwise.

Stronger maximal inequalities for averages of martingale differences exist:

Lemma 5.2 *Let Z_t be a martingale difference sequence with $\|Z_t\| \leq K$ then, for every $\delta > 0$ and every integer $T \geq 1$,*

$$\mathbb{P}\left\{\exists\ t \leq T,\ \|\overline{Z}_t\| \geq \frac{2}{\sqrt{t}}\phi^{-1}\left(\frac{\delta}{4}\frac{t}{T}\right)\right\} \leq \delta.$$

Proof: Define $\varepsilon_t = 2\phi^{-1}\left(\delta t/4T\right)/\sqrt{t}$. Using a peeling argument, one obtains

$$\begin{aligned}
\mathbb{P}\left\{\exists\ t \leq T,\ \|\overline{Z}_t\| \geq \varepsilon_t\right\} &\leq \sum_{m=1}^{\lfloor \log_2(T)\rfloor} \mathbb{P}\Big\{\bigcup_{t=2^m}^{2^{m+1}-1}\{\|\overline{Z}_t\| \geq \varepsilon_t\}\Big\} \\
&\leq \sum_{m=1}^{\lfloor \log_2(T)\rfloor} \mathbb{P}\Big\{\bigcup_{t=2^m}^{2^{m+1}}\{\|\overline{Z}_t\| \geq \varepsilon_{2^{m+1}}\}\Big\} \\
&\leq \sum_{m=1}^{\lfloor \log_2(T)\rfloor} \mathbb{P}\Big\{\bigcup_{t=2^m}^{2^{m+1}}\{t\|\overline{Z}_t\| \geq 2^m\varepsilon_{2^{m+1}}\}\Big\} \\
&\leq \sum_{m=1}^{\lfloor \log_2(T)\rfloor} \mathbb{P}\Big\{\bigcup_{t=2^m}^{2^{m+1}}\{t\|\overline{Z}_t\| \geq \sqrt{2^{m+1}}\phi^{-1}\left(\frac{\delta}{4}\frac{2^{m+1}}{T}\right)\}\Big\} \\
&\leq \sum_{m=1}^{\lfloor \log(T)\rfloor} \frac{2^{m+1}}{T}\frac{\delta}{4} \leq \frac{2^{\log_2(T)+2}}{T}\frac{\delta}{4} \leq \delta.
\end{aligned}$$

Hence the result. ■

Similarly, maximal inequalities can be derived for tail events:

Lemma 5.3 *Let $Z_t \in \mathbb{R}^d$ be a martingale difference sequence with $\|Z_t\| \leq K$ then, for every $\varepsilon > 0$ and every integer $T \geq 1$,*

$$\mathbb{P}\Big\{\exists\ t \geq T,\ \|\overline{Z}_t\| \geq \varepsilon\Big\} \leq 4\exp\left(-\frac{T\varepsilon^2}{8K^2}\right).$$

The exponential dependency in T can be reduced since one has, as soon as $\frac{T\varepsilon^2}{2K^2} \geq 1$,

$$\mathbb{P}\Big\{\exists\ t \geq T+1,\ \|\overline{Z}_t\| \geq \varepsilon\Big\} \leq \frac{2K^2}{\varepsilon^2}\left(1 + \sqrt{\frac{T\varepsilon^2}{2K^2}} + \frac{1}{\sqrt{\frac{T\varepsilon^2}{2K^2}}}\right)\exp\left(-\frac{T\varepsilon^2}{2K^2}\right).$$

Proof: Again, using a peeling argument, one obtains

$$\begin{aligned}
&\mathbb{P}\left\{\exists\, t \geq T,\ \|\overline{Z}_t\| \geq \varepsilon\right\} \leq \sum_{m=0}^{+\infty} \mathbb{P}\Big\{\bigcup_{t=2^mT}^{2^{m+1}T-1} \{\|\overline{Z}_t\| \geq \varepsilon\}\Big\} \\
&\leq \sum_{m=0}^{+\infty} \mathbb{P}\Big\{\bigcup_{t=2^mT}^{2^{m+1}T} \{t\|\overline{Z}_t\| \geq 2^mT\varepsilon\}\Big\} \leq \sum_{m=0}^{+\infty} \phi\left(\sqrt{2^{m+1}T}\frac{\varepsilon}{2}\right) \leq \int_0^{\infty} \phi\left(\sqrt{2^xT}\frac{\varepsilon}{2}\right) dx \\
&= \frac{1}{\log(2)} \int_{\frac{T\varepsilon^2}{8K^2}}^{\infty} \frac{\phi(K\sqrt{2v})}{v} dv = \frac{1}{\log(2)} \int_{\frac{T\varepsilon^2}{8K^2}}^{\infty} \frac{(1+\sqrt{2v})}{v} e^{-v} dv
\end{aligned}$$

hence

$$\mathbb{P}\Big\{\exists\, t \geq T,\ \|\overline{Z}_t\| \geq \varepsilon\Big\} \leq \frac{1}{\log(2)} \left(\frac{1}{\frac{T\varepsilon^2}{8K^2}} + \frac{\sqrt{2}}{\sqrt{\frac{T\varepsilon^2}{8K^2}}}\right) \exp\left(-\frac{T\varepsilon^2}{8K^2}\right)$$

and the first part of the result follows.

The second part of the proof follows from the facts that

$$\begin{aligned}
\mathbb{P}\left\{\exists\, t \geq T+1,\ \|\overline{Z}_t\| \geq \varepsilon\right\} &\leq \sum_{t=T+1}^{+\infty} \mathbb{P}\Big\{\|\overline{Z}_t\| \geq \varepsilon\Big\} \\
&\leq \sum_{t=T+1}^{+\infty} \left(1 + \sqrt{\frac{t}{K^2}}\varepsilon\right) \exp\left(-\frac{1}{2}\frac{t}{K^2}\varepsilon^2\right) \\
&\leq \int_T^{\infty} \left(1 + \sqrt{\frac{x\varepsilon^2}{K^2}}\right) \exp\left(-\frac{x\varepsilon^2}{2K^2}\right) dx \\
&\leq \frac{2K^2}{\varepsilon^2} \int_{\sqrt{\frac{T\varepsilon^2}{K^2}}}^{\infty} (u+u^2)\exp(-u^2/2) du,
\end{aligned}$$

and

$$\int_x^{\infty} (u+u^2)e^{-\frac{u^2}{2}} du = (1+x)e^{-\frac{x^2}{2}} + \int_x^{\infty} e^{-\frac{u^2}{2}} du \leq \left(1 + x + \frac{1}{x}\right) e^{-\frac{x^2}{2}}.$$

■

5.2 Probability lemmas

Lemma 5.4 *Let $f_n \in L_2(\Omega, \mu, \mathcal{F})$ such that $\sum_{n\in\mathbb{N}} \|f_n\|^2/n < \infty$ and $\|f_{n+1} - f_n\| \leq \frac{1}{n}$, then f_n converges to 0, μ-as.*

Proof: First, we prove a weaker version when $\|f_n\| < 1/\sqrt{n}$ for every $n \in \mathbb{N}$.

Let $M_n = \lceil n^{6/5} \rceil$ and k_n be the integer minimizing $\|f_k\|$ over $[M_n+1, M_{n+1}]$. Then,

$$\|f_{k_n}\|^2 \leq \frac{\sum_{k=M_n+1}^{M_{n+1}} \|f_k\|^2}{M_{n+1}-M_n} \leq \frac{M_{n+1}}{M_{n+1}-M_n} \sum_{k=M_n+1}^{M_{n+1}} \frac{\|f_k\|^2}{k}$$
$$\leq \sum_{k=M_n+1}^{M_{n+1}} 2n \frac{\|f_k\|^2}{k} \leq 2 \sum_{k=M_n+1}^{M_{n+1}} \frac{\|f_k\|^2}{k^{1/6}} \leq 2 \sum_{k=M_n+1}^{M_{n+1}} \frac{1}{k^{6/5}}.$$

Therefore, $\sum_{n\in\mathbb{N}} \|f_{k_n}\| < \infty$ and Fatou's lemma ensures that f_{k_n} converges to 0, μ-as.

Let us define $h_k = f_k - f_{k_n}$, then for evert $k > k_n$ (and similarly for $k < k_n$)

$$\|h_k\| = \left\| \sum_{j=k_n+1}^{k} (f_j - f_{j-1}) \right\| \leq \frac{k-k_n}{M_n} \leq \frac{M_{n+1}-M_n}{M_n}.$$

Summing over k, one gets

$$\sum_{k\in\mathbb{N}} \|h_k\|^2 \leq \sum_{n\in\mathbb{N}} (M_{n+1}-M_n) \left(\frac{M_{n+1}-M_n}{M_n} \right)^2 < \infty \quad \text{since} \quad \frac{M_{n+1}-M_n}{M_n} \leq O\left(\frac{1}{n}\right).$$

So both h_k and f_{k_n} converge μ-as to 0 and thus, so is $f_k = h_k + f_{k_n}$.

In the general case, one just need to notice that there exists an increasing sequence $\beta_n > 1$ such that $\sum_{n\in\mathbb{N}} \beta_n \|f_n\|^2/n < \infty$ and to define $M_{n+1} = \left\lceil \frac{\beta_{M_n}}{\beta_{M_n}-1} \right\rceil M_n + 1$. The proof follows as before, since $M_{n+1} - M_n \sim \frac{M_n}{\beta_{M_n}-1}$ and thus $M_{n+1}/(M_{n+1}-M_n) \sim \beta_{M_n}$ ■

Lemma 5.5 *Let $\mathcal{C}$ be a product set in $L_2(\Omega, \mu, \mathcal{F})$ and assume that for every $n \in \mathbb{N}$*

i) $g_n \in L_2(\Omega, \mu, \mathcal{F})$ is bounded by $B \in L_2(\Omega, \mu, \mathcal{F})$

ii) $\mathcal{X}_n \in L_2(\Omega, \mu, \mathcal{F})$ takes value in $\{0, 1\}$

iii) $\overline{g}_{\mathcal{X},n} = \sum_{k=1}^{n} \mathcal{X}_m g_m / S_n$ where $\mathcal{S}_n = \sum_{m=1}^{n} \mathcal{X}_m$

iv) $\left\langle \mathcal{X}_{n+1}(\overline{g}_{\mathcal{X},n} - \Pi_{\mathcal{C}}(\overline{g}_{\mathcal{X},n}), \frac{g_{n+1} - \Pi_{\mathcal{C}}(\overline{g}_{\mathcal{X},n})}{\mathcal{S}_{n+1}} \right\rangle \leq \varepsilon_n$, for a sequence of non-negative ε_n such that $\sum_{n\in\mathbb{N}} \varepsilon_n < \infty$.

Then $\overline{g}_{\mathcal{X},n}$ converges to $\mathcal{C}$, μ_∞-as, where $\mu_\infty(A) = \mu(A \cap \{\lim \mathcal{S}_n = \infty\})$.

Proof: Let $f_n = \overline{g}_n - \Pi_{\mathcal{C}}(\overline{g}_n)$ then point *iv)* implies that:

$$\|f_{n+1}\|^2 \leq \|f_n\|^2 - 2\left\langle \mathcal{X}_{n+1} \frac{f_n}{\overline{\mathcal{X}}_{n+1}}, f_n \right\rangle + \left\| \frac{\mathcal{X}_{n+1}}{\mathcal{S}_{n+1}} (g_{n+1} - \overline{g}_n) \right\|^2 + \varepsilon_n.$$

Since both $g_{\mathcal{X},n+1}$ and $\overline{g}_{\mathcal{X},n}$ are bounded by B and $\mathcal{X}_n \in \{0,1\}$, one has:

$$2\left\langle \mathcal{X}_{n+1}\frac{f_n}{\mathcal{S}_{n+1}}, f_n\right\rangle \leq \|f_n\|^2 - \|f_{n+1}\|^2 + 4\int_\Omega \frac{\mathcal{X}_{n+1}(\omega)}{\mathcal{S}_{n+1}^2(\omega)}B^2(\omega)d\mu(\omega) + \varepsilon.$$

Notice that, for every ω, $\sum_{n\in\mathbb{N}} \mathcal{X}_{n+1}(\omega)/\mathcal{S}_{n+1}^2(\omega) \leq \sum_{n\geq 1} 1/n^2 = \pi^2/6$. For every $n \in \mathbb{N}$ and $\omega \in \Omega$, let $j(n,\omega) := \inf\{m \in \mathbb{N}, \mathcal{S}_m(\omega) = n\}$ be the first time such that $\mathcal{S}_m(\omega)$ is bigger than n (if it exists, otherwise it is ∞). Define $\widetilde{f}_n(\omega) := f_{j(n,\omega)}(\omega)$, with $f_\infty(\omega) = 0$, so that

$$\sum_{n\in\mathbb{N}} \frac{\left\|\widetilde{f}_n\right\|^2}{n+1} = \sum_{n\in\mathbb{N}} \left\langle \mathcal{X}_{n+1}\frac{f_n}{\mathcal{S}_{n+1}}, f_n\right\rangle \leq \frac{\|f_0\|^2}{2} + \frac{\pi^2}{3}\|B\|^2 + \sum_{n\in\mathbb{N}} \varepsilon_n.$$

Since $\mathcal{C}$ is a product set, projection on $\mathcal{C}$ is a coordinate-wise projection, thus

$$\begin{aligned}\left|\widetilde{f}_{n+1}(\omega) - \widetilde{f}_n(\omega)\right| &= \left|\left[\overline{g}_{j_{n+1,\omega}}(\omega) - \overline{g}_{j_{n,\omega}}(\omega)\right] - \left[\Pi_C\left(\overline{g}_{j_{n+1,\omega}}\right)(\omega) - \Pi_C\left(\overline{g}_{j_{n,\omega}}\right)(\omega)\right]\right| \\ &\leq 2\left|\overline{g}_{j_{n+1,\omega}}(\omega) - \overline{g}_{j_{n,\omega}}(\omega)\right| = 2\left|\frac{g_{j_{n+1,\omega}}(\omega) - \overline{g}_{j_{n,\omega}}(\omega)}{n+1}\right| \leq \frac{4B(\omega)}{n+1}\end{aligned}$$

and so $\|\widetilde{f}_{n+1} - \widetilde{f}_n\|^2 = 16\|B\|^2/(n+1)^2$.

Using Lemma 5.4, with β_n an increasing sequence such that $\sum \beta_n\|\widetilde{f}\|^2/n$ converges and $M_{n+1} = \left\lceil \frac{\beta_{M_n}}{\beta_{M_n}+1}\right\rceil$, we obtain the μ-a.s. convergence of $\widetilde{f}_n$. As a consequence, after restriction to the event $\{\lim S_n = \infty\}$, f_n converges μ-as to zero. ■

Being a product set is used to bound $\left\|\widetilde{f}_{n+1} - \widetilde{f}_n\right\|$; convexity of $\mathcal{C}$ (nor actually its $\mathcal{L}_2$-boundedness, as defined by Lehrer [42]) is not enough for this proof. The reason is that if we define the mapping $\widetilde{g}_n \in L_2(\Omega,\mu,\mathcal{F})$ by $\widetilde{g}_n(\omega) = g_{j(n,\omega)}(\omega)$ and let $N := j(n,\omega)$. Then, without the product property which induces a coordinate-wise projection, $\Pi_{\mathcal{C}}\left(\widetilde{g}_n\right)(\omega)$ has no reason to be equal to $\Pi_{\mathcal{C}}\left(g_N\right)(\omega)$.

Corollary 5.6 *Same results hold if $\mathcal{X}_n \in [0,1]$ does not necessarily take values in $\{0,1\}$.*

Proof: Assume that $g_0 = 0$ and $\mathcal{X}_0 = \mathcal{S}_0 = 1$, then $\sum_{n=0}^\infty \mathcal{X}_n^2/\mathcal{S}_n^2$ is uniformly bounded. Indeed, define $k_n = \min\{m \text{ s.t. } S_m \geq n\}$, so that

$$\sum_{n=0}^\infty \frac{\chi_n^2}{\mathcal{S}_n^2} - 1 = \sum_{n=1}^\infty \sum_{m=k_n}^{k_{n+1}-1} \frac{\chi_n^2}{\mathcal{S}_n^2} \leq \sum_{n=1}^\infty \frac{1}{n^2}\sum_{m=k_n}^{k_{n+1}-1} \chi_n^2 \leq \sum_{n=1}^\infty \frac{1}{n^2}\sum_{m=k_n}^{k_{n+1}-1} \chi_n \leq \sum_{n=1}^\infty \frac{1}{n^2}.$$

Therefore, on $\{\lim S_n \to \infty\}$, $\widetilde{g}_{\chi,n} = \sum_{m=0}^n \mathcal{X}_m\, g_m/\mathcal{S}_m$ converges to $\mathcal{C}$, and $\widetilde{g}_{\chi,n} - \overline{g}_{\mathcal{X},n}$ converges to zero.

■

References

[1] J. Abernathy, P.L. Bartlett, and E. Hazan. Blackwell approachability and low-regret learning are equivalent. In *Proceedings of the 24th Annual Conference on Learning Theory*, volume 19 of *JMLR: Workshop and Conference Proceedings*, pages 27–46, 2011.

[2] S. As Soulaimani, M. Quincampoix, and S. Sorin. Repeated games and qualitative differential games: approachability and comparison of strategies. *SIAM J. Control Optim.*, 48:2461–2479, 2009.

[3] P. Auer, N. Cesa-Bianchi, and C. Gentile. Adaptive and self-confident on-line learning algorithms. *J. Comput. System Sci.*, 64:48–75, 2002. Special issue on COLT 2000 (Palo Alto, CA).

[4] R. J. Aumann. Subjectivity and correlation in randomized strategies. *J. Math. Econom.*, 1:67–96, 1974.

[5] R. J. Aumann and M. B. Maschler. *Repeated Games with Incomplete Information.* MIT Press, Cambridge, MA, 1995. With the collaboration of Richard E. Stearns (contains a reedition of chapters of Reports to the US Arms Control and Disarmament Angency ST-80, 116 and 143, Mathematica, 1966-1967-1968).

[6] M. Benaïm and M. Faure. Consistency of vanishingly smooth fictitious play. *Math. Oper. Res.*, to appear.

[7] M. Benaïm, J. Hofbauer, and S. Sorin. Stochastic approximations and differential inclusions. II. Applications. *Math. Oper. Res.*, 31:673–695, 2006.

[8] L. J. Billera and B. Sturmfels. Fiber polytopes. *The Annals of Mathematics*, 135(3):pp. 527–549, 1992.

[9] D. Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific J. Math.*, 6:1–8, 1956.

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

[11] A. Blum and Y. Mansour. From external to internal regret. In *Learning theory*, volume 3559 of *Lecture Notes in Comput. Sci.*, pages 621–636. Springer, Berlin, 2005.

[12] S. Bubeck. Introduction to online optimization. *manuscript*, 2012.

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

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

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

[16] A. P. Dawid. The well-calibrated Bayesian. *J. Amer. Statist. Assoc.*, 77:605–613, 1982.

[17] A. P. Dawid. Self-calibrating priors do not exist: Comment. *J. Amer. Statist. Assoc.*, 80:340–341, 1985.

[18] K. Fan. Minimax theorems. *Proc. Nat. Acad. Sci. U. S. A.*, 39:42–47, 1953.

[19] K. Fan. A minimax inequality and applications. In *Inequalities, III (Proc. Third Sympos., Univ. California, Los Angeles, Calif., 1969; dedicated to the memory of Theodore S. Motzkin)*, pages 103–113. Academic Press, New York, 1972.

[20] W. Feller. *An Introduction to Probability Theory and its Applications. Vol. I.* Third edition. John Wiley & Sons Inc., New York, 1968.

[21] D. Foster. A proof of calibration via blackwell's approachability theorem. *Games and Economic Behavior*, 29:73 – 78, 1999.

[22] D. P. Foster, A. Rakhlin, K. Sridharan, and A. Tewari. Complexity-based approach to calibration with checking rules. In *Proceedings of the 24th Annual Conference on Learning Theory*, volume 19 of *JMLR: Workshop and Conference Proceedings*, pages 293–314, 2011.

[23] D. P. Foster and R. V. Vohra. Calibrated learning and correlated equilibrium. *Games Econom. Behav.*, 21:40–55, 1997.

[24] D. P. Foster and R. V. Vohra. Asymptotic calibration. *Biometrika*, 85:379–390, 1998.

[25] D. P. Foster and R. V. Vohra. Regret in the on-line decision problem. *Games Econom. Behav.*, 29:7–35, 1999.

[26] D. Fudenberg and D. M. Kreps. Learning mixed equilibria. *Games Econom. Behav.*, 5:320–367, 1993.

[27] D. Fudenberg and D. Levine. An easier way to calibrate. *Games Econom. Behav.*, 29:131–137, 1999. Learning in games: a symposium in honor of David Blackwell.

[28] D. Fudenberg and D. K. Levine. Conditional universal consistency. *Games Econom. Behav.*, 29:104–130, 1999.

[29] P. Hall and C. C. Heyde. *Martingale Limit Theory and its Application.* Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1980. Probability and Mathematical Statistics.

[30] J. Hannan. Approximation to Bayes risk in repeated play. In *Contributions to the Theory of Games*, volume 3 of *Annals of Mathematics Studies*, pages 97–139. Princeton University Press, Princeton, N. J., 1957.

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

[32] S. Hart and A. Mas-Colell. A general class of adaptive strategies. *J. Econom. Theory*, 98:26–54, 2001.

[33] S. Hart and A. Mas-Colell. Regret-based continuous-time dynamics. *Games Econom. Behav.*, 45:375–394, 2003. Special issue in honor of Robert W. Rosenthal.

[34] E. Hazan and S. M. Kakade. (weak) calibration is computationaly hard. In *Proceedings of the 25th Annual Conference on Learning Theory*, volume 23, pages 3.1 – 3.10, 2012.

[35] J. Hofbauer and W. H. Sandholm. On the global convergence of stochastic fictitious play. *Econometrica*, 70:2265–2294, 2002.

[36] J Hofbauer, S. Sorin, and Y. Viossat. Time average replicator and best-reply dynamics. *Math. Oper. Res.*, 34:263–269, 2009.

[37] S. M. Kakade and D. P. Foster. Deterministic calibration and Nash equilibrium. In *Learning theory*, volume 3120 of *Lecture Notes in Comput. Sci.*, pages 33–48. Springer, Berlin, 2004.

[38] O. Kallenberg and R. Sztencel. Some dimension-free features of vector-valued martingales. *Probability Theory and Related Fields*, 88:215–247, 1991.

[39] E. Kohlberg. Optimal strategies in repeated games with incomplete information. *Internat. J. Game Theory*, 4:7–24, 1975.

[40] J. Kwon. Hilbert distance, bounded convex functions, and application to the exponential weight algorithm. Master's thesis, ENS Lyon, 2012.

[41] E. Lehrer. Any inspection is manipulable. *Econometrica*, 69:1333–1347, 2001.

[42] E. Lehrer. Approachability in infinite dimensional spaces. *Internat. J. Game Theory*, 31:253–268, 2002.

[43] E. Lehrer. A wide range no-regret theorem. *Games Econom. Behav.*, 42:101–115, 2003.

[44] E. Lehrer and E. Solan. Excludability and bounded computational capacity. *Math. Oper. Res.*, 31:637–648, 2006.

[45] E. Lehrer and E. Solan. Learning to play partially-specified equilibrium. *manuscript*, 2007.

[46] E. Lehrer and E. Solan. Approachability with bounded memory. *Games Econom. Behav.*, 66:995–1004, 2009.

[47] N. Littlestone and M. Warmuth. The weighted majority algorithm. *Information and Computation*, 108:212–261, 1994.

[48] R. D. Luce and H. Raiffa. *Games and Decisions: Introduction and Critical Survey.* John Wiley & Sons Inc., New York, N. Y., 1957.

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

[50] S. Mannor and G. Stoltz. A geometric proof of calibration. *Math. Oper. Res.*, 35:721–727, 2010.

[51] S. Mannor, G. Stoltz, and V. Perchet. Robust approachability and regret minimization in games with partial monitoring. In *Proceedings of the 24th Annual Conference on Learning Theory*, volume 19 of *JMLR: Workshop and Conference Proceedings*, pages 515–536, 2011.

[52] S. Mannor, G. Stoltz, and V. Perchet. Set-valued approachability, with applications to regret minimization in games with partial monitoring. *manuscript*, 2012.

[53] S. Mannor and J. N. Tsitsiklis. Approachability in repeated games: computational aspects and a Stackelberg variant. *Games Econom. Behav.*, 66:315–325, 2009.

[54] D. McFadden. *Conditional logit analysis of qualitative choice behavior*, pages 105–142. Frontiers in econometrics, Academic Press: New York, 1974.

[55] J.-F. Mertens, S. Sorin, and S. Zamir. *Repeated Games.* CORE discussion paper 9420–9422, 1994.

[56] J. Neveu. *Martingales à Temps Discret.* Masson et Cie, éditeurs, Paris, 1972.

[57] D. Oakes. Self-calibrating priors do not exist. *J. Amer. Statist. Assoc.*, 80:339–342, 1985. With comments by A. P. Dawid and Mark J. Schervish.

[58] W. Olszewski. Calibration and expert testing. In *Handbook of Game Theory*, volume IV. 2012.

[59] V. Perchet. Calibration and internal no-regret with random signals. *Proceedings of the 20th International Conference on Algorithmic Learning Theory*, pages 68–82, 2009.

[60] V. Perchet. *Approachability, Calibration and Regret in Games with Partial Observations (in French).* PhD thesis, Université Pierre et Marie Curie, 2010.

[61] V. Perchet. Approachability of convex sets in games with partial monitoring. *J. Optim. Theory Appl.*, 149:665–677, 2011.

[62] V. Perchet. No-regret with partial monitoring: Calibration-based optimal algorithms. *J. Mach. Learn. Res.*, 12:1893–1921, 2011.

[63] V. Perchet and M. Quincampoix. Purely informative game: Application to approachability with partial monitoring. *manuscript*, 2012.

[64] A. Rakhlin. Lecture notes on online learning. *manuscript*, 2009.

[65] A. Rakhlin, K. Sridharan, and A. Tewari. Online learning: random averages, combinatorial parameters, and learnability. In *NIPS*, 2010.

[66] A. Rakhlin, K. Sridharan, and A. Tewari. Online learning: Beyond regret. In *Proceedings of the 24th Annual Conference on Learning Theory*, volume 19 of *JMLR: Workshop and Conference Proceedings*, pages 559–594, 2011.

[67] W. Rudin. *Real and Complex Analysis*. McGraw-Hill Series in Higher Mathematics, New York, 1974.

[68] A. Sandroni, R. Smorodinsky, and R. V. Vohra. Calibration with many checking rules. *Math. Oper. Res.*, 28:141–153, 2003.

[69] E. Seneta. *Nonnegative Matrices and Markov Chains*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 1981.

[70] S. Sorin. *A First Course on Zero-Sum Repeated Games*. Springer-Verlag, 2002.

[71] S. Sorin. *Lectures on Dynamics in Games*. Unpublished Lecture Notes, 2008.

[72] S. Sorin. Exponential weight algorithm in continuous time. *Math. Program.*, 116:513–528, 2009.

[73] X. Spinat. A necessary and sufficient condition for approachability. *Math. Oper. Res.*, 27:31–44, 2002.

[74] G. Stoltz. *Incomplete information and internal regret in prediction of individual sequences*. PhD thesis, Université Paris-Sud, 2005.

[75] G. Stoltz and G. Lugosi. Internal regret in on-line portfolio selection. *Mach. Learn.*, 59:125–159, 2005.

[76] G. Stoltz and G. Lugosi. Learning correlated equilibria in games with compact sets of strategies. *Games Econom. Behav.*, 59:187–208, 2007.

[77] N. Vieille. Weak approachability. *Math. Oper. Res.*, 17:781–791, 1992.

[78] Y. Viossat and A. Zapechelnyuk. No-regret dynamics and fictitious play. *to appear in: Journal of Econ. Th.*

[79] V. Vovk. Aggregating strategies. In *Proceedings of the 3rd Annual Workshop on Computational Learning Theory*, pages 372–383, 1990.

[80] V. Vovk, I. Nouretdinov, A. Takemura, and G. Shafer. Defensive forecasting for linear protocols. In *Algorithmic learning theory*, volume 3734 of *Lecture Notes in Comput. Sci.*, pages 459–473. Springer, Berlin, 2005.

[81] D. Walkup and R. J.-B. Wets. A lipschitzian characterization of convex polyhedra. *Proceedings of the American Mathematical Society*, 23:167–173, 1969.

[82] A. Zapechelnyuk. Better-reply dynamics with bounded recall. *Math. Oper. Res.*, 33:869–879, 2008.

[83] M. Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the Twentieth International Conference on Machine Learning (ICML)*, 2003.