HAL open science

Approchabilité, Calibration et Regret dans les Jeux à Observations Partielles

Vianney Perchet

To cite this version:

Vianney Perchet. Approchabilité, Calibration et Regret dans les Jeux à Observations Partielles. Mathématiques [math]. Université Pierre et Marie Curie - Paris VI, 2010. Français. NNT : . tel-00567079

HAL Id: tel-00567079

https://theses.hal.science/tel-00567079v1

Submitted on 18 Feb 2011

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.

Thèse de doctorat de l'Université Pierre et Marie Curie

Spécialité
Mathématiques

Présentée par

Vianney PERCHET

Pour obtenir le grade de

Docteur de l'Université Pierre et Marie Curie

Approchabilité, calibration et regret dans les jeux à observations partielles

---

Soutenue le 25 Juin 2010 devant le jury composé de MM.

Bernard	**de Meyer**	Université Paris I	*Rapporteur*
Nicolas	**Vayatis**	ENS Cachan	*Rapporteur*
Gérard	**Biau**	Université Paris VI	*Président du jury*
Jérôme	**Renault**	Université Toulouse 1	*Examinateur*
Sylvain	**Sorin**	Université Paris VI	*Directeur de thèse*
Gilles	**Stoltz**	École Normale Supérieure	*Examinateur*

Équipe Combinatoire et Optimisation - FRE 3232

à Mamette, Mamie, Papé et Papi

Remerciements

Je voudrais tant remercier mon directeur de thèse Sylvain Sorin pour avoir accepté de m'encadrer et pour toute l'attention qu'il m'a portée. J'ai pu découvrir sous sa guidance les nombreuses et différentes facettes de la théorie des jeux, en découvrant jour après jour l'étendue de ses goûts et de son impressionnante culture mathématiques. Il a d'ailleurs, tant bien que mal et au cours d'échanges parfois houleux, essayé de me les transmettre en même temps que l'importance de la rigueur et de la précision. Je tiens également à le remercier pour le cadre de travail agréable qu'il a su construire ; cela va être bien difficile de quitter son équipe.

Je souhaite aussi exprimer mes remerciements à Gilles Stoltz qui a été présent tout au long de ma thèse. Il n'a eu cesse de me prodiguer conseils et idées durant ces quatre années, tant sur un plan strictement scientifique que pour aiguillonner mes différents choix. Pour être complètement honnête, il était là avant même le commencement de cette thèse puisque l'on a rédigé ensemble mon projet de thèse ; je suis donc très heureux de le voir à la toute fin parmi les membres du jury.

C'est un honneur que m'ont fait Bernard de Meyer et Nicolas Vayatis d'accepter d'être les rapporteurs de cette thèse. J'ai eu l'occasion de partager de nombreuses et longues conversations avec Bernard lors de diverses conférences. Je suis très fier de faire l'année prochaine mon post-doctorat avec Nicolas, rencontré lors de cette sympathique conférence à Porto et j'aurais aimé que les contraintes de calendrier ne l'empêchent pas d'être présent lors de la soutenance.

Je suis aussi très reconnaissant envers les autres membres du jury. Merci à Jérôme Renault d'être venu depuis Toulouse pour cette occasion ; certains résultats de cette thèse sont issus de nos discussions. Je veux également exprimer ma gratitude envers Gérard Biau qui m'a fait le privilège d'accepter de présider ce jury.

Si j'ai pu réaliser cette thèse c'est en partie grâce à Rida Laraki. C'est toi le premier qui m'a donné goût à la recherche — en particulier pour la théorie des jeux — et qui m'a ensuite présenté à Sylvain. Je remercie également Marc Quincampoix de m'avoir invité quelques jours à Brest. Ce fut un réel plaisir de commencer notre collaboration qui, je l'espère, sera encore plus fructueuse. J'en profite pour aussi remercier tous

les organisateurs et participants du séminaire du lundi matin. Toutes ces années de thèse ont été rendues plus agréables par l'ensemble des membres de l'Équipe Combinatoire et Optimisation. Je pense notamment à Jérôme Bolte avec qui j'ai partagé mathématiques, films, séries, batteries de voitures, un nombre incalculable de cafés (fumés pour la plupart), ainsi que tant de bons moments ; j'ai également reçu le soutien et l'aide d'Hélène Frankowska, Michel Pocchiola, Eric Balandraud, mais aussi d'autres personnes proches du laboratoire comme Aris Daniilidis ou Patrick Louis Combettes, à qui je fais probablement la surprise d'arriver déjà à la soutenance. Merci aussi à tous ceux qui ont grandement facilité les petits problèmes quotidiens, comme Michèle Trouvé, Germain Gomes, Vincent Duquenne ou Jean Fonlupt, directeur du laboratoire durant les premières années de thèse et ceux qui ont accepté de partager un bout de bureau avec moi, Benjamin Girard, Yahya Hamidoune et Vincent Pilaud — ainsi que pour sa science de LaTeX. Mon année d'ATER au CEREMADE a été rendue beaucoup plus intéressante par Yannick Viossat et Filippo Santambrogio et je vous en suis reconnaissant.

Je n'oublie pas le groupe soudé et fort sympathique de thésards parisiens (ou non) en théorie des jeux : Anne, Antoine, Cheng, Fabien, Juan, Luis, Mael, Marie — qui a en plus relu la plus grande partie du manuscript, Mario, Mathieu, Miquel, Pauline, Xavier, etc. Je voulais tout spécialement adresser mes remerciements à Guillaume avec qui j'ai partagé énormément de mémorables instants tant en France qu'en Espagne ou en Italie. J'espère que l'on aura l'occasion de compléter la liste.

Je pense également à tous mes amis qui se reconnaîtront ; je ne citerais que, par ordre de taille, Joe, Laurent et Arvind ainsi que leurs compagnes Manue, Julie et Anne-Laure (dont étrangement l'ordre est, il me semble, préservé). J'embrasse en cette fin de remerciements tout mon entourage, la J.P. family, ainsi que toute ma grande famille.

Mes toutes dernières pensées vont à Anaïs qui réussit, je ne sais comment, à me supporter et même, dans tous les moments difficiles, à me porter. J'espère être capable un jour de lui rendre tout ce qu'elle m'a, et continue de me donner.

Résumé

Cette thèse s'intéresse aux jeux statistiques avec observations partielles. Ces jeux ne sont pas la formalisation d'une intéraction stratégique entre deux joueurs parfaitement rationnels, mais entre un joueur et la nature (ou l'environnement). On donne ce nom au second joueur car aucune hypothèse n'est faite sur ses paiements, ses objectifs ou sa rationalité.

Les observations du joueur sont dites complètes s'il observe les choix de la nature, *i.e.* si il apprend a posteriori soit quelle est, à chaque étape, l'action choisie par cette dernière soit au moins son propre paiement. On s'intéressera au cadre où cette hypothèse est affaiblie et où l'on suppose que le joueur n'a que des observations partielles : il ne reçoit à chaque étape qu'un signal aléatoire dont la loi dépend de l'action de la nature.

L'objectif principal de cette thèse est de généraliser des notions largement utilisées dans les jeux avec observations complètes au cadre des jeux avec observations partielles. Nous allons en effet, dans un premier temps, construire des stratégies qui n'ont pas de regret interne et dans un deuxième temps nous allons caractériser les ensembles approchables.

Mots-clés : Jeux répétés statistiques, Jeux à observations partielles, Apprentissage en ligne, Approchabilité, Calibration, Regret

Abstract

This thesis explores statistical games with partial observations. Those games are not the formalization of a strategic interaction between two rational players, but between a player and Nature (or the environment). We give this name to the second player since no assumption is made on her payoffs, objectives or rationality.

The observations of the player are said to be full if he observes the choices of Nature, *i.e.* if he learns, after each stage, either which action she chose or at least his own payoff. We will investigate the framework where this assumption is weaken and where we assume that the player has only partial observations : he only receives at each stage a random signal whose law depends on Nature's action.

The main objective of this thesis is to generalize notions widely used in games with full observations to the framework of games with partial observations. Indeed, we first construct strategies without internal regret and then we characterize approachable sets.

Keywords : Repeated statistical games, Games with partial monitoring, Online learning, Approachability, Calibration, Regret

Table des matières

1 Introduction, Plan et Principaux Résultats **1**

A Jeux à Observations Complètes **7**

2 Approchabilité **9**
2.1 En dimension finie 10
Approchabilité classique 10
Approchabilité faible 14
2.2 En dimension infinie 15
2.3 Extensions 17
Approches continues 17
Durées d'étape variables 18
Paiements non bornés et loi forte des grands nombres 20
Approchabilité avec activations 21
Stratégies à mémoire bornée 23

3 Tests de Théories et Calibration **25**
3.1 Apprentissage non-Bayésien - Test de théories 26
Manipulabilité des tests 26
Tests non manipulables 27
Plusieurs prédicteurs 28
Tests de calibration 29
3.2 Apprentissage Bayésien - Fusion 32
Jeu entre l'inspecteur et un prédicteur 34
3.3 Calibration par rapport à une grille et ε-calibration 35
Presque-calibration 37

4 Non-Regret **39**
4.1 Non-regret externe 40

Prédictions avec conseils d'experts 41
Regret externe et jeux 42
4.2 Non-regret interne 43
Équilibres corrélés 45
4.3 Extensions 45
Du regret externe au regret *swap* 45
Notions plus fines de regret 48
Approches continues 49
Autres utilisations du regret 51

5 Liens entre Approchabilité, Calibration et non-Regret 55
5.1 Approchabilité et non-regret 56
Espace d'actions du joueur 1 fini 56
Espaces d'actions infinies et regret généralisé 59
5.2 Non-regret et calibration 61
Calibration par rapport à un graphe 62
5.3 Calibration et approchabilité 65
De l'approchabilité à la calibration 65
De la calibration à l'approchabilité 69

B Jeux à Observations Partielles 75

6 Internal Consistency with Partial Monitoring 77
6.1 Full monitoring case : from approachability to calibration 79
From approachability to internal no-regret 84
From internal regret to calibration 85
From calibration to approachability 86
6.2 Internal regret in the partial monitoring framework 90
External regret 91
Internal regret 91
On the strategy space 96
6.3 Back on payoff space 96
The worst case fulfills Assumption 6.15 97
Compact case 98
Regret in terms of actual payoffs 99
External and internal consistency 101

7 Calibration-Based Optimal Algorithms 103
7.1 Full monitoring 105
Model and definitions 105
A naïve algorithm, based on calibration 107
Calibration and Laguerre diagram 111
Optimal algorithm with full monitoring 114

7.2 Partial monitoring . . . 116
Definitions . . . 116
A naïve algorithm . . . 119
Optimal algorithms . . . 119
7.3 Concluding remarks . . . 124
Second algorithm : calibration and polytopial complex. . . . 124
Extension to compact case . . . 125
Strengthening of the constants . . . 126
7.4 Proofs of technical results . . . 128
Proof of proposition 7.18 . . . 128
Proof of Lemma 7.11 . . . 131

8 Approachability of Convex Sets 135
8.1 Approachability . . . 136
Full monitoring case . . . 137
Partial monitoring case . . . 139
8.2 Internal regret with partial monitoring . . . 140
8.3 Proofs of the main results . . . 142
Proof of Theorem 8.8 . . . 142
Proof of proposition 8.9 . . . 145
Remarks on the counterexample . . . 147
8.4 Repeated game with incomplete information . . . 148

9 Purely Informative Game. 151
9.1 Approachability in the purely informative game . . . 152
Game with full monitoring . . . 152
Game with partial monitoring . . . 153
Purely informative game . . . 154
Links between approachability in these games . . . 155
9.2 Characterization of approachable sets . . . 160
Preliminaries on the space of probability measures . . . 161
Approachability in the purely informative game . . . 164
9.3 Characterization of convex approachable sets . . . 168
9.4 Convex games . . . 170

Bibliographie 177

1

Introduction, Plan et Principaux Résultats

Cette thèse s'intéresse aux *jeux statistiques* avec observations partielles. Elle s'inscrit dans la lignée des travaux de Blackwell et Girshick [20] et récemment ceux de Cesa-Bianchi et Lugosi [29], par exemple. Ces jeux ne sont pas la formalisation d'une intéraction stratégique entre deux joueurs parfaitement rationnels, mais entre un joueur et *la nature* (ou l'environnement). On donne ce nom au second joueur car aucune hypothèse n'est faite sur ses paiements, ses objectifs ou sa rationalité.

Une tentative de résolution, que l'on pourrait qualifier de naïve, des jeux statistiques serait de les considérer comme à somme nulle et de chercher ensuite des stratégies optimales — on ne rentrera pas dans les détails des jeux à somme nulle, pour ceux-ci il vaut mieux consulter Sorin [107]. Dans les jeux statistiques, jouer une telle stratégie peut conduire à de mauvais résultats et il est donc nécessaire de définir d'autres critères Par exemple, considérons le jeu répété où les paiements sont définis par la matrice de gauche dans la figure 1.

	G	D
H	(0,0)	(-1,1)
B	(1,-1)	(-2,2)

	G	D
H	(0,0,0)	(-1,-2,-3)
B	(1,0,-1)	(-2,2,100)

Figure 1.1: À gauche, un exemple de jeux à somme nulle. À droite, un jeu à paiements vectoriels.

À chaque étape, la nature choisit une colonne G ou D et simultanément (et indépendamment) le joueur choisit une ligne H ou B. Le paiement du joueur est alors la première coordonnée de la case choisie, et celui de la nature le second (par exemple si la nature choisit G et le joueur B, ce dernier obtient 1 et la nature -1). Si l'on suppose que la nature est rationnelle, elle devrait jouer à chaque étape D et donc le joueur devrait toujours jouer H pour obtenir un paiement de -1 ; il s'agit d'ailleurs de sa stratégie optimale. Si l'on ne suppose pas de rationalité de la part de la nature,

il est possible qu'elle joue à chaque étape G. La *bonne stratégie* deviendrait alors de toujours jouer B pour obtenir un paiement égal à 1 au lieu de 0.

La première difficulté (et peut être la plus grande) lorsque l'on s'intéresse aux jeux statistiques est tout simplement de définir la (ou du moins une) notion de *bonne stratégie*. Celle que Hannan [56] proposa a une intuition simple ; il faudrait qu'un joueur ne puisse pas se dire après suffisamment d'étapes : *Si j'avais su comment la nature allait jouer, j'aurais mieux fait de toujours choisir la même action.* La différence entre la moyenne de ce que le joueur aurait pu obtenir et ce qu'il a effectivement obtenu s'appelle le regret externe.

L'utilisation du regret est valable car les objectifs du joueur — à défaut de ceux de la nature — sont bien définis : il cherche à maximiser son paiement. Blackwell [17] a, quant à lui, étudié les jeux où les paiements sont des vecteurs de $\mathbb{R}^d$. Par exemple, dans le jeu de droite de la figure 1, si la nature joue D et le joueur H, son paiement est $(-1, -2, -3)$ alors qu'il aurait obtenu $(-2, 2, 100)$ en jouant B. Dans ce cadre d'optimisation multi-critères, les objectifs du joueur ne sont pas intrinsèques. En effet, il pourrait maximiser les coordonnées suivant un ordre (par exemple la première, puis la deuxième, puis la troisième) ou alors maximiser la plus grande (ou la plus petite) des coordonnées, leur somme, etc.

Supposons que les objectifs du joueur soient représentés par un ensemble $E \subset \mathbb{R}^d$ (*i.e.* un sous-ensemble de l'espace des paiements). Une *bonne stratégie* pour Blackwell est une stratégie qui approche E, c'est-à-dire telle que, quoi que fasse la nature, la moyenne des paiements converge vers E.

On dit que le jeu est à observations complètes si le joueur observe les choix de la nature, *i.e.* si il sait a posteriori quelle est la colonne choisie à chaque étape — ou du moins son paiement. Une seconde difficulté s'ajoute lorsque cette hypothèse est affaiblie et que l'on suppose que le joueur n'a que des observations partielles : il ne reçoit qu'un signal aléatoire dont la loi dépend de l'action de la nature.

Par exemple, dans les jeux de la figure 1, on peut supposer qu'à chaque étape le joueur voit quelle était la colonne choisie avec probabilité 0.5. Ainsi en moyenne une fois sur deux , il a connaissance de la case choisie, sinon il ne connaît que la ligne. Sous ces conditions, construire des stratégies avec les propriétés précédentes peut paraître plus complexe.

Plan et résultats principaux

Les objectifs de cette thèse sont doubles :
1) Construire des stratégies qui n'ont pas de regret interne avec observations partielles en suivant l'approche de Lehrer et Solan [74]).
2) Caractériser les ensembles approchables (en étendant les résultats de Blackwell [17]) avec observations partielles.

Observations complètes

La première partie de la thèse est consacrée aux jeux avec observations complètes. Dans les chapitres 2, 3 et 4, on rappelle les définitions et des résultats sur l'approchabilité et le non-regret. On introduit également la notion de calibration qui est l'outil statistique à la base de nos algorithmes, avant de mettre en évidence les liens entre ces trois concepts. Les démonstrations sont omises, sauf lorsque l'on propose des modifications des preuves originales (ou lorsque l'on applique des théorèmes généraux à des cas particuliers). Le cas échéant, on précisera quel est le nouvel argument et son intérêt.

Le cinquième chapitre est consacré aux équivalences entre ces trois notions, ce qui se traduit par le fait que :
i) une stratégie sans regret peut être déduite d'une stratégie d'approchabilité (Hart et Mas-Colell [57]) ;
ii) une stratégie calibrée peut être déduite d'une stratégie sans regret (Foster et Vohra [43] et Sorin [108]) ;
iii) une stratégie d'approchabilité d'un convexe peut être déduite d'une stratégie calibrée (Perchet [92]).

Il existe également d'autres implications intermédiaires, comme la construction d'une stratégie calibrée directement à partir d'une stratégie d'approchabilité (Mannor et Stoltz [80] et Foster [41]) ou d'une stratégie sans regret à partir d'une stratégie calibrée (Foster et Vohra [42]).

On fournit les preuves de ces résultats car elles fournissent des intuitions pour le cadre des jeux avec observations partielles. Cette première partie s'inspire fortement du cours *Lectures on Dynamics in Games* de Sorin [108].

Observations partielles

La seconde partie de la thèse est consacrée aux jeux avec observations partielles et contient les contributions originales (le point iii) est repris dans le chapitre 6). Les principaux résultats sont :

Dans le sixième chapitre : le théorème 6.16 donne l'existence et la construction de stratégies sans regret interne (sous des conditions plus fortes, leur existence a été montrée par Lehrer et Solan [74]).

Dans le septième chapitre : le théorème 7.20 donne un algorithme qui construit une stratégie sans regret interne (avec l'évaluation au *pire cas*) dont la vitesse de convergence est optimale, *i.e.* en $O\left(n^{-1/3}\right)$ (sous ces hypothèse la vitesse de Lugosi, Mannor et Stoltz [78] était de $O\left(n^{-1/5}\right)$).

Dans le huitième chapitre : le théorème 8.8 caractérise, de manière géométrique, les ensembles convexes approchables.

Dans le neuvième chapitre : le théorème 9.20 caractérise les ensembles quelconques approchables, en introduisant un nouveau jeu abstrait appelé *purement informatif* où le paiement d'une étape est l'information maximale disponible.

Outline and main results

The objectives of this thesis are double :

1) To construct strategies without internal regret with partial monitoring (as introduced by Lehrer and Solan [74]) ;
2) To characterize approachable sets with partial monitoring (following Blackwell [17].

Full monitoring

The first part of this thesis is devoted to the full monitoring case. In chapter Z, 3 and 4, we recall definitions and results on approachability and no-regret. We also introduce the notion of calibration, which is the statistical tool on which our algorithms rely. Proofs are omitted, except if we provide modifications — in that case we will specify what is the new argument and its interest — or if we apply general theorems to specific cases.

The fifth chapter is concerned with the equivalence between these three notion, which can be seen through the following properties :

i) a strategy without regret can be deduced from an approachability strategy (Hart and Mas-Colell [57]) ;
ii) a calibrated strategy can be deduced from a strategy without regret (Foster and Vohra [43] and Sorin [108]) ;
iii) an approachability strategy of a convex set can be deduced from a calibrated strategy (Perchet [92]).

There also exist implications within this cycle : a calibrated strategy can be derived from an approachability strategy (Mannor et Stoltz [80] and Foster [41]) or a strategy with no-regret can be derived from a calibrated strategy (Foster and Vohra [42]).

We provide the proofs of those results since they gave intuitions for the partial monitoring framework. This first part is greatly inspired from the course *Lectures on Dynamics in Games* of Sorin [108].

Partial monitoring

The second part of this thesis is devoted to the partial monitoring framework and contains our original contributions (point iii) can be also found in chapter 6). The main results are :

In Chapter 6 : Theorem 6.16 gives the existence and the construction of strategies with no internal regret (under stronger assumptions, their existence was proved by Lehrer et Solan [74]).

In Chapter 7 : Theorem 7.20 provides an algorithm that constructs a strategy with no internal regret (with the *worst case* evaluation) whose rate of convergence is

optimal, *i.e.* in $O\left(n^{-1/3}\right)$ (under those assumptions the rate of Lugosi, Mannor and Stoltz [78] was $O\left(n^{-1/5}\right)$).

In Chapter 8 : Theorem 8.8 characterizes, in a purely geometrical way, convex sets that are approchable.

In Chapter 9 : Theorem 9.20 characterizes approachable sets, by introducing a new abstract game called *purely informative* where the payoff at any stage is the maximal information available.

Première partie

Jeux à Observations Complètes

2

Approchabilité

Ce chapitre est consacré aux définitions et caractérisations d'ensembles approchables, en dimension finie comme en dimension infinie. Il se conclut par plusieurs extensions (stratégies à mémoire bornée, approches continues, durées d'étape variables, etc.)

Sommaire

2.1 En dimension finie **10**
Approchabilité classique 10
Approchabilité faible 14
2.2 En dimension infinie **15**
2.3 Extensions **17**
Approches continues 17
Durées d'étape variables 18
Paiements non bornés et loi forte des grands nombres 20
Approchabilité avec activations 21
Stratégies à mémoire bornée 23

Blackwell [17] a défini l'approchabilité d'un ensemble donné E, dans un jeu répété Γ à deux joueurs avec paiements vectoriels, comme une propriété analogue à l'existence de la valeur d'un jeu à somme nulle Γ_0. Dans ce dernier, si l'on note I (resp. J) l'espace d'actions du joueur 1 (resp. du joueur 2) et ρ la fonction de paiement définie sur $I \times J$ à valeurs dans $\mathbb{R}$, alors l'objectif des deux joueurs est défini intrinsèquement : le joueur 1 maximise le paiement et le joueur 2 le minimise. On appelle $\overline{v} = \max_{i \in I} \min_{j \in J} \rho(i,j)$ le *maxmin* de Γ_0, c'est-à-dire le paiement minimal que le joueur 1 peut être assuré d'obtenir quelle que soit la stratégie du joueur 2. De manière duale, on appelle $\underline{v} = \min_{j \in J} \max_{i \in I} \rho(i,j)$ le *minmax* de Γ_0. Sous certaines conditions de régularité (voir par exemple Sorin [107], Appendice A), $\overline{v}$ et $\underline{v}$ sont égaux et leur valeur commune, notée v, est appelée valeur de Γ_0.

Une telle définition des objectifs n'est pas intrinsèque avec une fonction de paiement vectorielle (*i.e.* à valeurs dans un espace euclidien $\mathbb{R}^d$). Un joueur pourrait maximiser la première coordonnée et minimiser la seconde coordonnée de son paiement, ou alors maximiser la somme des coordonnées, etc. Les objectifs du joueur 1 dans Γ sont représentés par un sous-ensemble de $\mathbb{R}^d$, noté E. On dit que celui-ci est approchable par le joueur 1, si ce dernier a une stratégie telle qu'à partir d'une certaine étape et avec une grande probabilité, la moyenne de Cesaro des paiements reste proche de E, quelle que soit la stratégie du joueur 2. L'analogie entre approchabilité et valeur d'un jeu est visible lorsque l'on remarque que dans un jeu à somme nulle l'ensemble $[\overline{v}, +\infty[$ est approchable par le joueur 1 et l'ensemble $]-\infty, \underline{v}]$ est approchable par le joueur 2.

2.1 En dimension finie

Soit $\Gamma = (I, J, \rho)$ le jeu répété à deux joueurs où I (resp. J) est l'espace fini d'actions du joueur 1 (resp. du joueur 2) et $\rho : I \times J \to \mathbb{R}^d$ est la fonction de paiement à valeur vectorielle. On note $\rho_n = \rho(i_n, j_n)$ le paiement induit à l'étape $n \in \mathbb{N}$ par les choix des actions $i_n \in I$ et $j_n \in J$. Ceux-ci sont fonctions de l'histoire, *i.e.* des observations passées $h^{n-1} = (i_1, j_1, \ldots, i_{n-1}, j_{n-1}) \in (I \times J)^{n-1} := H_{n-1}$.

Explicitement, une stratégie σ du joueur 1 (resp. τ du joueur 2) est une application de l'ensemble des histoires finies $H := \bigcup_{n\in\mathbb{N}} H_n$ dans $\Delta(I)$ (resp. $\Delta(J)$), l'ensemble des mesures de probabilité sur I (resp. J). Un couple de stratégies (σ, τ) génère une mesure de probabilité — d'après le théorème d'extension de Kolmogorov — sur l'ensemble des histoires infinies (ou parties) du jeu muni de la tribu produit, noté $\mathcal{H} = (I \times J)^{\mathbb{N}}$.

Étant donné un ensemble fermé E de $\mathbb{R}^d$, on note $d(x, E) = \inf\{\|x - z\|_2 ; z \in E\}$ la distance de x à E, $E^\delta = \left\{x \in \mathbb{R}^d, d(x, E) < \delta\right\}$ le δ-voisinage ouvert de E et $\Pi_E(x) = \{\pi \in E; \|x - \pi\| = d(\pi, E)\}$ la projection de x sur E, en général non univoque. La fonction ρ est étendue à $\Delta(I)\times\Delta(J)$ par $\rho(x, y) = \mathbb{E}_{x,y}[\rho(i, j)]$. La moyenne d'une suite $a = \{a_m \in \mathbb{R}^d\}_{m\in\mathbb{N}}$ jusqu'à l'étape n est notée $\overline{a}_n = \sum_{m=1}^n a_m/n$.

Approchabilité classique

Définition 2.1

i) Un ensemble fermé $E \subset \mathbb{R}^d$ est approchable par le joueur 1 si pour tout $\varepsilon > 0$, il existe une stratégie σ_ε et un entier $N \in \mathbb{N}$ tels que pour toute stratégie τ du joueur 2 et tout $n \geq N$:

$$\mathbb{E}_{\sigma_\varepsilon,\tau}[d(\overline{\rho}_n, E)] \leq \varepsilon \quad \text{et} \quad \mathbb{P}_{\sigma_\varepsilon,\tau}\left(\sup_{n\geq N} d(\overline{\rho}_n, E) \geq \varepsilon\right) \leq \varepsilon.$$

ii) Un ensemble E est repoussable par le joueur 2 s'il existe $\delta > 0$ tel que le complémentaire de E^δ soit approchable par le joueur 2.

Un ensemble $E \subset \mathbb{R}^d$ est donc approchable par le joueur 1 si ce dernier possède une stratégie telle que le paiement moyen converge presque sûrement vers E, uniformément par rapport aux stratégies du joueur 2. On appelle stratégie d'approchabilité de E une stratégie σ qui vérifie le point *i)* pour tout $\varepsilon > 0$.

L'utilisation de δ-voisinages dans la définition de la repoussabilité implique qu'un ensemble E ne peut être à la fois approchable par le joueur 1 et repoussable par le joueur 2. Néanmoins, le contraire n'est pas vrai, il existe des ensembles qui ne sont ni approchables ni repoussables comme c'est le cas dans l'exemple suivant, issu de Blackwell [17].

Exemple : Soient le jeu sous forme matricielle A et l'ensemble $E \subset \mathbb{R}^2$ suivants :

$$A = \begin{pmatrix} (0,0) & (0,0) \\ (1,0) & (1,1) \end{pmatrix} \text{ et } E = \{(1/2, y); 0 \leq y \leq 1/4\} \bigcup \{(1, y), 1/4 \leq y \leq 1\}.$$

En jouant durant les N premières étapes la ligne du bas puis durant les N étapes suivantes la ligne du haut (resp. la ligne du bas) si $\overline{\rho}_n^2$ — la seconde coordonnée de $\overline{\rho}_n$ — est plus petite (resp. plus grande) que $1/2$, le joueur 1 s'assure que $\overline{\rho}_{2N}$ est dans E. Il n'est donc pas repoussable. La stratégie du joueur 2 consistant à jouer la colonne de droite (resp. gauche) si $\overline{\rho}_n^1$ est plus petit (resp. plus grand) que $3/4$ empêche que le paiement converge vers E.

Blackwell [17] a donné une condition purement géométrique qui assure qu'un ensemble E est approchable. Pour l'exprimer, on appelle $P^1(y) = \{\rho(x, y); x \in \Delta(I)\}$, l'ensemble des paiements espérés compatibles avec $y \in \Delta(J)$; $P^2(x)$ est défini de manière similaire.

Définition 2.2

Un ensemble fermé $E \subset \mathbb{R}^d$ est un B-set, si pour tout $z \in \mathbb{R}^d$, il existe un projeté $\pi \in \Pi_E(z)$ et $x\,(= x(z)) \in \Delta(I)$ tels que l'hyperplan passant par π et perpendiculaire à $z - \pi$ sépare z de $P^2(x)$, ou formellement :

$$\forall z \in \mathbb{R}^d, \exists \pi \in \Pi_E(z), \exists x \in \Delta(I) : \forall y \in \Delta(J), \quad \langle \rho(x, y) - \pi, z - \pi \rangle \leq 0. \quad (2.1)$$

Blackwell a défini un B-set à partir des points extérieurs à l'ensemble. Il est aussi possible de le faire directement à partir des points appartenant à E. Étant donné un point $p \in E$, $q \in \mathbb{R}^d$ est une normale proximale à E en p (voir par exemple Clarke [32]) s'il existe $\tau > 0$ tel que $p + \tau q$ se projette sur E en p. L'ensemble des normales proximales à E au point $p \in E$ est donc définie par

$$\mathrm{NC}_E(p) = \left\{ q \in \mathbb{R}^d : \exists \tau > 0, p = \Pi_E(p + \tau q) \right\}.$$

Avec cette définition, As Soulaimani, Quincampoix et Sorin [4], théorème 8, ont établi qu'un ensemble E est un B-set si et seulement si tous les points $p \in E$ vérifient la

condition de Blackwell, *i.e.*

$$\forall q \in \mathrm{NC}_E(p), \exists x \in \Delta(I), \forall y \in \Delta(J), \quad \langle \rho(x,y) - p, q \rangle \leq 0. \tag{2.2}$$

Être un B-set est bien une condition suffisante pour être un ensemble approchable avec en plus une vitesse de convergence explicite :

Théorème 2.3 (*Blackwell [17]*)

Si E est un B-set, alors E est approchable par le joueur 1. De plus, la stratégie σ du joueur 1 définie par $\sigma(h^n) = x(\overline{\rho}_n)$ est telle que pour toute stratégie τ du joueur 2 et tout $\eta > 0$:

$$\mathbb{E}_{\sigma,\tau}[d_E^2(\overline{\rho}_n)] \leq \frac{K^2}{n} \quad \text{et} \quad \mathbb{P}_{\sigma,\tau}\left(\sup_{n \geq N} d(\overline{\rho}_n, E) \geq \eta\right) \leq \frac{2K^2}{\eta^2 N}, \tag{2.3}$$

où $K = \min(2M, M + \|E\|)$, $M = \sup_{i,j} \|\rho(i,j)\|$ et $\|E\| = \sup_{z \in E} \|z\|$.

Blackwell [17] a montré que la distance du paiement espéré à l'ensemble est bornée par $2M/\sqrt{n}$. Mertens, Sorin et Zamir [84] (théorème 4.3 page 102) ont montré la convergence presque sûre avec comme constante $K = 2M$. D'ailleurs, même si dans ce cadre précis les deux quantités sont identiques, on peut remplacer $\sup_{i,j} \|\rho(i,j)\|^2$ par $\sup_{x,y} \mathbb{E}_{x,y}[\|\rho(i,j)\|^2]$.

La variante de la démonstration de Blackwell ci-dessous n'a d'intérêt que si les paiements sont grands comparés aux éléments de E ; par exemple, dans le cas $E = \{0\}$ les constantes sont divisées par 4.

Démonstration . Définissons la stratégie du joueur 1 par $\sigma(h^n) = x(\overline{\rho}_n)$. En notant $\delta_n = d(\overline{\rho}_n, E)$ et π_n un élément de $\Pi_E(\overline{\rho}_n)$ donné par la condition de Blackwell (2.1), on obtient :

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

En conditionnant par h^n, en utilisant la condition (2.1), ainsi que la définition de M et $\|E\|$, la dernière inéquation devient :

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

Une récurrence donne $\mathbb{E}_{\sigma,\tau}[\delta_n^2] \leq (M + \|E\|)^2/n$, ce qui implique la convergence en probabilité de $\overline{\rho}_n$ vers C. Pour prouver la convergence presque sûre, il suffit de remarquer que $Z_n = \delta_n^2 + \mathbb{E}_{\sigma,\tau}\left[\sum_{k=n}^{\infty} \|\rho_{k+1} - \pi_k\|^2 / (k+1)^2 \middle| h^n\right]$ est une surmartingale et ensuite conclure de manière identique à Mertens Sorin et Zamir [84]. □

Remarques : La stratégie décrite dans le théorème 2.3 n'utilise pas le fait que le joueur 1 observe les actions du joueur 2 ; il lui suffit en effet d'apprendre à chaque étape le paiement ρ_n. En particulier, la finitude de J ne joue aucun rôle. On pourrait supposer qu'à chaque étape le joueur 2 choisit un vecteur $U = (U^i)_{i\in I}$ dans un compact $\mathcal{U} \subset \left(\mathbb{R}^d\right)^I$ et que le paiement est simplement défini par $\rho(x, U) = x.U = \sum_{i\in I} x^i U^i \in \mathbb{R}^d$. La condition de Blackwell (2.1) pour un ensemble E devient alors :

$$\forall z \in \mathbb{R}^d, \exists p \in \Pi_E(z), \inf_{x\in\Delta(I)} \sup_{U\in\mathcal{U}} \langle x.U - p, z - p\rangle \leq 0.$$

De plus $\rho_n - \mathbb{E}_{\sigma,\tau}[\rho_n|h^{n-1}]$ est une différence de martingale (son espérance conditionnelle à h^{n-1} est nulle) donc en utilisant l'inégalité d'Hoeffding-Azuma [61, 11] pour les sommes de différences de martingales (voir aussi Cesa-Bianchi et Lugosi [29]), on peut montrer qu'il suffit que le joueur 1 observe son paiement espéré $\mathbb{E}_{\sigma,\tau}[\rho_n|h^{n-1}]$ pour approcher un B-set.

Il existe deux caractérisations d'ensembles approchables convexes équivalentes :

Théorème 2.4 (*Blackwell [17]*)

Un ensemble convexe $C \subset \mathbb{R}^d$ est approchable par le joueur 1 si et seulement si tout demi-espace qui le contient est approchable et si et seulement si :

$$P^1(y) \cap C \neq \emptyset, \quad \forall y \in \Delta(J). \tag{2.4}$$

En particulier, un convexe C est soit approchable par le joueur 1 soit repoussable par le joueur 2.

Remarques : En utilisant le théorème de Von Neumann, Blackwell [17] a prouvé qu'un ensemble C vérifiant la condition (2.4) vérifie aussi la condition (2.1) et est donc un B-set. Ce théorème est prouvé ultérieurement dans la section 5.3, sans faire appel à ces outils. En fait, cette caractérisation peut être vue comme l'analogue du théorème du Minmax de Von Neumann : si pour tout $y \in \Delta(J)$, il existe $x \in \Delta(I)$ tel que $\rho(x, y)$ est dans C alors il existe une stratégie du joueur 1 telle que, contre toute stratégie du joueur 2, le paiement moyen converge vers C. En particulier, il y a — au moins — deux façons de prouver qu'un ensemble convexe C est approchable : soit en montrant directement que C est un B-set, ce qui donne une expression d'une stratégie d'approchabilité, soit en montrant que le joueur 2 ne peut repousser C. On sait qu'il existe alors une stratégie d'approchabilité du joueur 1 et calculer cette stratégie revient, étape par étape, à trouver une stratégie optimale dans un jeu auxiliaire à somme nulle.

Il existe également une condition nécessaire et suffisante d'approchabilité pour un ensemble quelconque :

Proposition 2.5 (*Spinat [110]*)

Un ensemble est approchable si et seulement si il contient un B-set.

L'idée principale de la démonstration est la suivante : s'il existe un point p d'un ensemble approchable E ne vérifie pas la condition de Blackwell (2.1) — un tel point est appelé secondaire — alors l'ensemble E privé d'un voisinage de p reste approchable. Grossièrement, en enlevant tous les points secondaires à un ensemble E, on obtient un B-set (la démonstration du théorème 9.20 est similaire est donnée en section 9.2).

Cela dit, vérifier qu'un ensemble donné est un B-set est très coûteux : Mannor et Tsitsiklis [81] ont montré (en le réduisant à 3-SAT) que, même pour que les cas du singleton $\{0\}$ ou d'un orthant, le problème de savoir si un ensemble est repoussable est NP-difficile.

Approchabilité faible

Dans l'exemple donné par Blackwell [17] d'un ensemble qui n'est ni approchable ni repoussable, le joueur 1 a cependant une stratégie (dépendante de N qui doit être assez grand) qui lui assure qu'à l'étape N le paiement moyen $\overline{\rho}_N$ est ε-proche de E. Blackwell a appelé cette propriété *approchabilité faible* et a conjecturé que tout ensemble est soit faiblement approchable, soit faiblement repoussable, ce qui est défini de manière analogue. La démonstration est due à Vieille [113].

Définition 2.6

i) Un ensemble fermé $E \subset \mathbb{R}^d$ est faiblement approchable par le joueur 1 si pour tout $\varepsilon > 0$, il existe un entier $N \in \mathbb{N}$ tel que pour tout $n \geq N$, le joueur 1 ait une stratégie σ_n telle que pour toute stratégie τ du joueur 2 $\mathbb{E}_{\sigma_n,\tau}\left[d(\overline{\rho}_n, E)\right] \leq \varepsilon$.

ii) Un ensemble E est faiblement repoussable par le joueur 2 s'il existe $\delta > 0$ tel que le complémentaire de E^δ soit faiblement approchable par le joueur 2.

Il est important de noter que dans ce cadre les stratégies σ du joueur 1 peuvent dépendre de la longueur du jeu N, ce qui n'était pas le cas pour l'approchabilité classique.

Théorème 2.7 (*Vieille [113]*)

Un ensemble fermé est soit faiblement approchable par le joueur 1, soit faiblement repoussable par le joueur 2.

Vieille [113] a construit un jeu différentiel $\mathcal{D}$ (en temps continu et de durée finie) tel que les répétitions finies de Γ peuvent être vues comme des discrétisations de $\mathcal{D}$. Le résultat se déduit de l'existence d'une valeur du jeu $\mathcal{D}$.

2.2 En dimension infinie

On considère dans cette section une variante de Γ — introduite par Lehrer [69] — où l'espace d'arrivée de la fonction de paiements ρ n'est plus un espace de dimension finie mais $\mathcal{L}_2(\Omega, \mathbb{R})$ où $(\Omega, \mathcal{F}, \lambda)$ est un espace probabilisé. Dans ce contexte, seuls des ensembles convexes seront approchés, ce qui est défini de la façon suivante :

Définition 2.8

Un ensemble convexe fermé C de $\mathcal{L}_2(\Omega, \mathbb{R})$ est approchable par le joueur 1, si ce dernier possède une stratégie σ telle que pour toute stratégie τ du second joueur $\overline{\rho}_n$ converge λ-presque sûrement vers C pour $\mathbb{P}_{\sigma,\tau}$-presque toutes les histoires.

En dimension infinie, on ne requiert pas que la convergence soit uniforme par rapport aux stratégies du joueur 2. Le produit scalaire de $\mathcal{L}_2$, noté $\langle \cdot, \cdot \rangle$, définit une notion de B-set de manière identique au cas de la dimension finie.

Théorème 2.9 (*Lehrer [69]*)

Un B-set convexe C est approchable par le joueur 1.

Ce théorème est une conséquence des deux lemmes suivants, adaptés des résultats de Lehrer [69]. Le premier est l'équivalent du principe géométrique de Blackwell et le second est un résultat de convergence de variables aléatoires.

Lemme 2.10

Soit C un convexe fermé de $\mathcal{L}_2(\Omega, \mathbb{R})$ et supposons que :
i) $(g_n)_{n\in\mathbb{N}}$ est une suite de variables aléatoires bornées λ-ps par $B \in \mathcal{L}_2$;
ii) $\langle \overline{g}_n - \Pi_C(\overline{g}_n), g_{n+1} - \Pi_C(\overline{g}_n) \rangle \leq 0$.
Alors $\overline{g}_n$ converge vers C, λ-ps.

Démonstration . Les arguments de la dimension finie impliquent que, pour tout $n \in \mathbb{N}$, $d^2(\overline{g}_n, C) \leq 4\|B\|^2/n$ et donc $\overline{g}_n$ converge en probabilité vers C. Cependant, si l'on note $f_n = \overline{g}_n - \Pi_C(\overline{g}_n)$ et comme la projection sur un convexe est 1-Lipschitz :

$$\|f_{n+1} - f_n\| \leq \left\|\overline{g}_{n+1} - \overline{g}_n - \left[\Pi_C\left(\overline{g}_{n+1}\right) - \Pi_C\left(\overline{g}_n\right)\right]\right\| \leq 2\left\|\overline{g}_{n+1} - \overline{g}_n\right\| \leq \frac{4\|B\|}{n+1}.$$

Finalement, le résultat est une conséquence directe du lemme suivant. □

Lemme 2.11

Toute suite $f_n \in \mathcal{L}_2(\Omega, \mathbb{R})$ telle que $\|f_n\|^2 = O\left(\frac{1}{n}\right)$ et $\|f_{n+1} - f_n\|^2 = O\left(\frac{1}{n^2}\right)$ converge vers 0, λ-presque sûrement.

Démonstration . Soit $(f_n)_{n\in\mathbb{N}}$ une telle suite et $A > 0$ tels que $\|f_n - f_{n-1}\|^2 \leq An^{-2}$. En notant $\beta_n = n^{5/6}$, la série $\sum_{n=1}^{+\infty} \beta_n \|f_n\|^2/n$ est convergente. Pour tout $n \in \mathbb{N}$, on note $M_n = \left\lceil n^{\frac{6}{5}} \right\rceil$ et on appelle k_n l'entier qui minimise $\|f_k\|$ sur $]M_n, M_{n+1}]$, ainsi :

$$\|f_{k_n}\|^2 \leq \frac{1}{M_{n+1} - M_n} \sum_{k=M_n+1}^{M_{n+1}} \|f_k\|^2 \leq \frac{M_{n+1}}{M_{n+1} - M_n} \sum_{k=M_n+1}^{M_{n+1}} \frac{\|f_k\|^2}{k}.$$

Comme $M_{n+1}(M_{n+1} - M_n)^{-1} \sim 5/6n$ et $\beta_{M_n} \sim n$, alors $\|f_{k_n}\|^2 \leq \sum_{k=M_n+1}^{M_{n+1}} \beta_k \|f_k\|^2/k$ et donc $\sum_{n\in\mathbb{N}} \|f_{k_n}\|^2 < +\infty$. Le lemme de Fatou implique que f_{k_n} converge vers 0 λ-ps. Soit $k \in]M_n, M_{n+1}]$ et $h_k = f_k - f_{k_n}$, alors pour $k > k_n$ (le résultat s'obtient de manière similaire pour $k \leq k_n$) :

$$\|h_k\|^2 \leq \left\| \sum_{j=k_n+1}^{k} (f_j - f_{j-1}) \right\|^2 \leq A \frac{(k - k_n)^2}{M_n^2} \leq A \left(\frac{M_{n+1} - M_n}{M_n} \right)^2$$

et en sommant on obtient :

$$\sum_{n=1}^{\infty} \|h^n\|^2 \leq \sum_{k=1}^{\infty} A \left(\frac{M_{k+1} - M_k}{M_k} \right)^2 (M_{k+1} - M_k) = \sum_{k=1}^{\infty} A \left(\frac{M_{k+1} - M_k}{M_k} \right)^3 M_k < \infty$$

car $\frac{M_{k+1} - M_k}{M_k} = O(k^{-1})$. Donc $(h_k)_{k\in\mathbb{N}}$ converge λ-ps vers 0 et la suite $(f_k)_{k\in\mathbb{N}}$, en tant que somme de deux suites convergentes, converge λ-ps vers 0. □

Remarque : L'hypothèse de convexité de C n'est utilisée que pour avoir une projection Lipschitzienne. Lehrer [69], définition 4, a montré qu'il suffisait en fait que l'ensemble C vérifie la propriété suivante, appelée $\mathcal{L}_2$-boundedness : pour toute suite $(g_n)_{n\in\mathbb{N}}$ dans $\mathcal{L}_2(\Omega, \mathcal{R})$:

1) s'il existe $B_1 \in \mathcal{L}_2$ qui majore tous les g_n alors il existe $B_2 \in \mathcal{L}_2$ qui majore toutes les projections sur C ;
2) s'il existe $B_1 \in \mathcal{L}_2$ telle que pour tout $n \in \mathbb{N}$, $|g_{n+1} - g_n| \leq B/n$, alors il existe $B_2 \in \mathcal{L}_2$ telle que $|\Pi_C(g_{n+1}) - \Pi_C(g_n)| \leq B_2/n$ pour tout n.

Démonstration du théorème 2.9 . On suppose que le joueur 1 utilise la stratégie σ donnée par la définition d'un B-set et on note τ la stratégie du second joueur. Soit $\lambda \otimes \mathbb{P}_{\sigma,\tau}$ la mesure produit sur l'espace produit $\Omega \times \mathcal{H}$, sur lequel on définit les variables aléatoires $\widetilde{\rho}_n$ par $\widetilde{\rho}_n(h, \omega) = \rho_n(i_n, j_n)(\omega)$, où (i_n, j_n) est le couple d'actions joué à la n-ème étape le long de $h \in \mathcal{H}$. Comme ρ est bornée, alors $\widetilde{\rho}_n$ vérifie le point $i)$ du Lemme 2.10 ; l'hypothèse $ii)$ est, quant à elle, vérifiée par définition de σ. Ainsi $\widetilde{\rho}_n$ converge $\lambda \otimes \mathbb{P}_{\sigma,\tau}$-presque sûrement vers C qui est donc approchable par le joueur 1. □

Le théorème 2.9 implique que la caractérisation d'un convexe fermé approchable est donc exactement la même en dimension infinie et en dimension finie. En particulier, un convexe est toujours soit approchable par le joueur 1, soit repoussable par le

joueur 2 et ce dès que les deux joueurs ont un ensemble fini d'actions. Il est possible de relâcher cette dernière hypothèse en suivant Mertens, Sorin et Zamir [84], théorème 4.5 p. 103.

On suppose que les espaces d'actions $(I, \mathcal{I})$ et $(J, \mathcal{J})$ sont deux espaces mesurables et $\rho : I \times J \to L_2(\Omega, \mu, \mathcal{F})$ est la fonction de paiement du joueur 1, étendue comme précédemment à $\Delta(I) \times \Delta(J)$, le produit des espaces de probabilités sur I et J munis de la topologie faible-$\star$. On suppose qu'il existe $B \in L_2(\Omega, \mu, \mathcal{F})$ qui borne $\rho(i, j)$ pour tout $i, j \in I \times J$ et que pour tout $y \in \Delta(J)$, $P^1_f(y)$, l'adhérence de $P^1(y)$, est compacte.

Proposition 2.12

Si pour tout $u \in L_2(\Omega, \mu, \mathcal{F})$, tel que $\sup_{x \in C} \langle x, u \rangle < +\infty$, le jeu à somme nulle où les paiements sont définis par $\langle u, \rho(i, j) \rangle$ a une valeur, alors :

a) *C est approchable par le joueur 1 si et seulement si pour tout $y \in \Delta(J)$, $P^1_f(y) \bigcap C \neq \emptyset$;*

b) *s'il existe y_0 tel que $P^1_f(y_0) \bigcap C = \emptyset$ alors C est repoussable par le joueur 2 ;*

c) *C est approchable par le joueur 1 si et seulement si pour tout $\epsilon > 0$, pour tout $q \in L_2(\Omega)$ et pour tout $y \in \Delta(J)$, il existe $x \in \Delta(I)$ tel que $\langle q - \Pi_C(q), \rho(x, y) - \Pi(q) \rangle \leq \epsilon$.*

La preuve est identique à celle en dimension finie. La condition d'existence de la valeur dans le jeu de paiement $\langle u, \rho(i, j) \rangle$ est par exemple satisfaite dès que I est compact et $\rho(\cdot, j)$ est continue (voir par exemple Sorin [107], Annexe A). De plus, ces hypothèses assurent que $P^1(y)$ est compact.

En l'absence de compacité de $P^1_f(y)$, il se peut que $P^1_f(y_0) \bigcap C$ soit vide pour un certain $y_0 \in \Delta(J)$ et que C soit approchable. La condition d'approchabilité des points a) et c) est alors seulement suffisante.

2.3 Extensions

On introduit quelques versions d'approchabilité dans un contexte continu. Celles-ci fournissent de nouvelles preuves des résultats précédents qui peuvent être généralisées facilement, notamment lorsque la durée des étapes varie. Trois dernières extensions sont considérées, lorsque les paiements ne sont pas bornés, lorsque les coordonnées du paiement peuvent être actives ou non et lorsque l'on suppose que les joueurs ont une rationalité limitée.

Approches continues

Benaïm, Hofbauer et Sorin [15] ont remarqué que la stratégie d'approchabilité d'un B-set E décrite dans le théorème 2.3 vérifiait à l'étape n la relation de récurrence

suivante : conditionnellement à h^n,

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

où $N(z) = \big\{\omega \in \mathbb{R}^d; \|\omega\|_\infty \leq \|\rho\|_\infty \text{ et } \exists p \in \Pi_E(z), \langle z-p, \omega - p\rangle \leq 0\big\}$. La suite des paiements moyens $(\overline{\rho}_n)_{n\in\mathbb{N}}$ est donc une approximation stochastique discrete (A.S.D.) de $\boldsymbol{\rho}$, solution de l'inclusion différentielle associée

$$\dot{\boldsymbol{\rho}} \in N(\boldsymbol{\rho}) - \boldsymbol{\rho}, \quad \boldsymbol{\rho}(0) = \rho_0 \in \mathbb{R}^d.$$

La fonction $g(t) = d\,(\boldsymbol{\rho}(t), E)$ est une fonction de Lyapounov associée à E car sa dérivée vérifie $g'(t) \leq -g(t)$ et donc $g(t) \leq g(0)e^{-t}$. Ainsi $\boldsymbol{\rho}$ converge vers E et en tant qu'A.S.D. $\overline{\rho}_n$ converge presque sûrement vers E. Néanmoins, la vitesse de convergence n'est pas connue et n'est, a priori, pas uniforme (cependant Sorin [109] et Hart et Mas-Colell [58] ont obtenu des bornes explicites pour un cas particulier, présenté en section 4.3).

As Soulaimani, Quincampoix et Sorin [4] ont, quant à eux, associé au jeu Γ un jeu différentiel $\mathcal{D}_a$ où l'espace de contrôles du joueur 1 (resp. joueur 2) est $\mathcal{X} = \Delta(I)$ (resp. $\mathcal{Y} = \Delta(J)$) et la dynamique est donnée par :

$$\frac{d}{dt}\overline{\boldsymbol{\rho}}(t) = \frac{-\overline{\boldsymbol{\rho}}(t) + \rho(x(t), y(t))}{t}, \quad \overline{\boldsymbol{\rho}}(0) = 0;$$

car $\overline{\boldsymbol{\rho}}(t) = \frac{1}{t}\int_{0^t} \rho(x(s), y(s))ds$ est la moyenne temporelle des paiements. Après le changement de variable $t = e^s$ et $\boldsymbol{\rho}(s) = \overline{\boldsymbol{\rho}}(e^s)$, la dynamique devient

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

Cette transformation permet d'obtenir une nouvelle caractérisation d'un B-set, à savoir qu'un ensemble E est un B-set si et seulement si c'est un domaine discriminant du joueur 1 pour la dynamique f, *i.e.*

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

Cette propriété est à mettre en relation avec la condition (2.3) de Blackwell exprimée à partir des normales proximales.

Durées d'étape variables

Dans sa définition originale de l'approchabilité, Blackwell [17] a considéré la suite des moyennes arithmétiques des paiements. On peut également supposer que toutes les étapes n'ont pas la même durée (ou le même poids dans les moyennes successives). On appelle τ_n la durée de l'étape n ; une stratégie du joueur 1 approche alors un

ensemble E si la suite des moyennes pondérées $\sum_{k=1}^{n} \tau_k \rho_k / \left(\sum_{k=1}^{n} \tau_k\right)$ converge presque sûrement vers E.

Il existe deux hypothèses distinctes sur la suite τ_n : soit elle est donnée de manière exogène, soit elle est fonction des actions des joueurs, *i.e.* il existe une fonction τ positive sur $I \times J$ telle que $\tau_n = \tau(i_n, j_n)$.

Dans le premier cas, appelons $\overline{\rho}_{\tau,n}$ la moyenne pondérée des paiements jusqu'à l'étape n. Cette suite satisfait la relation de récurrence

$$\overline{\rho}_{\tau,n+1} - \overline{\rho}_{\tau,n+1} = \frac{\tau_{n+1}}{\sum_{k=1}^{n} \tau_k} \left(\rho_{n+1} - \overline{\rho}_{\tau,n}\right)$$

et donc l'approche continue de Benaïm, Hofbauer et Sorin [15] implique qu'un B-set est approchable dès que la suite des $\tau_n / \left(\sum_{k=1}^{n} \tau_k\right)$ satisfait la condition d'un algorithme de Robbins-Monroe, à savoir :

$$\sum_{n \in \mathbb{N}} \left(\frac{\tau_n}{\sum_{k=1}^{n} \tau_k} \right) = +\infty \quad \text{et} \quad \sum_{n \in \mathbb{N}} \left(\frac{\tau_n}{\sum_{k=1}^{n} \tau_k} \right)^2 < +\infty.$$

Une démonstration semblable à celle des moyennes de Cesaro donne une vitesse de convergence uniforme.

L'approchabilité dans le jeu Γ sous l'hypothèse d'existence d'une fonction τ strictement positive telle que $\tau_n = \tau(i_n, j_n)$, a été étudiée par Mannor et Shimkin [79]. Elle se ramène toutefois à de l'approchabilité classique dans le jeu auxiliaire associé $\widehat{\Gamma}$ où la fonction de paiement est définie pour tout $x \in \Delta(I)$ et $y \in \Delta(J)$ par

$$\rho_\tau(x, y) = \left(\mathbb{E}_{x,y} \left[\tau(i,j) \rho(i,j) \right], \tau(x,y) \right) \in \mathbb{R}^d \times \mathbb{R}.$$

On associe à tout ensemble $E \subset \mathbb{R}^d$ (dans le jeu Γ), l'ensemble $\widehat{E} \subset \mathbb{R}^d \times \mathbb{R}$ (dans $\widehat{\Gamma}$) défini par :

$$\widehat{E} = \left\{ (R, t) \in \mathbb{R}^d \times [\tau_0, \tau_1]; \frac{1}{t} R \in E \right\},$$

où $\tau_0 > 0$ et $\tau_1 > 0$ sont les bornes inférieure et supérieure de τ sur $I \times J$. Il est clair que si C est un ensemble convexe, $\widehat{C}$ est aussi convexe. Cependant, le résultat intéressant est l'équivalence entre les notions d'approchabilité.

Lemme 2.13

E est approchable dans Γ si et seulement si $\widehat{E}$ l'est dans $\widehat{\Gamma}$. Pour tout ensemble convexe C, l'ensemble $\widehat{C}$ est aussi convexe.

Démonstration . Soit (R, t) tel que $d\left((R,t), \widehat{E}\right) := \|(R,t) - (R_e, t_e)\| \leq \varepsilon$. Alors

$$d\left(\frac{R}{t}, E\right) \leq \left\| \frac{R}{t} - \frac{R_e}{t_e} \right\| \leq \frac{1}{t_e} \|R - R_e\| + \|R\| \left| \frac{1}{t} - \frac{1}{t_e} \right| \leq \varepsilon \left(\frac{1}{\tau_0} + \frac{\|\rho\|_\infty}{\tau_0^2} \right).$$

De la même façon si $d(R/t, E) := d(R/t, r_e) \leq \varepsilon$ avec $t \in [\tau_0, \tau_1]$ alors

$$d\left((R,t), \widehat{E}\right) \leq \|(R,t) - (tr_e, t)\| \leq \tau_1 \left\| \frac{R}{t} - r_e \right\| \leq \tau_1 \varepsilon,$$

ce qui prouve la première partie du lemme.

La seconde partie est évidente et provient de la définition de la convexité. □

Une conséquence directe de ce lemme est le théoreme 5.1 de Mannor et Shimkin [79] :

Proposition 2.14

Un ensemble convexe C est approchable si et seulement si pour tout $y \in \Delta(J)$, il existe $x \in \Delta(I)$ tel que $\mathbb{E}_{x,y}\left[\tau(i,j)\rho(i,j)\right] / \tau(x,y) \in C$.

Paiements non bornés et loi forte des grands nombres

Dans la remarque suivant la démonstration du théorème 2.3, on suppose que le joueur 2 choisit à l'étape n un vecteur de paiement U_n dans un compact $\mathcal{U}$ de $\left(\mathbb{R}^d\right)^I$ et que le paiement espéré est égal à $x_n.U_n$ (avec $x_n = \sigma(h^{n-1})$ la loi de i_n). Le fait que tous les U_n appartiennent à $\mathcal{U}$ peut être assoupli en :

$$\sum_{n\in\mathbb{N}} \frac{\|U_n\|^2}{(n+1)^2} < \infty \quad \left(\text{ou } \mathbb{E}_{\sigma,\tau}\left[\sum_{n\in\mathbb{N}} \frac{\|x_n.U_n\|^2}{(n+1)^2}\right] < \infty \right). \tag{2.5}$$

Sous cette hypothèse, la démonstration du théorème 2.3 ne change pas, *i.e.* Z_n est toujours une surmartingale mais

$$\mathbb{E}_{\sigma,\tau}[Z_n] \leq \mathbb{E}_{\sigma,\tau}[\delta_n^2] + \sum_{m=n+1}^{\infty} \frac{\|U_m\|^2}{(m+1)^2} := \frac{4M^2}{n} + O_n$$

où O_n est une suite qui décroit strictement vers 0.

L'inégalité maximale de Doob pour les surmartingales (voir Neveu [85], prop. IV.2.7) devient :

$$\mathbb{P}_{\sigma,\tau}\left(\exists n \geq N, Z_n \geq \varepsilon\right) \leq \frac{\mathbb{E}_{\sigma,\tau}[Z_n]}{\varepsilon} \leq \frac{4M^2}{\varepsilon n} + \frac{O_n}{\varepsilon}.$$

L'approchabilité (avec paiements bornés) d'un ensemble peut être vue comme une extension de la loi des grands nombres (voir Mertens, Sorin et Zamir [84], exercice 4 p. 104) : soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes et bornées (ou de variances bornées). Alors la suite $\overline{X}_n - \mathbb{E}[\overline{X}_n]$ approche $\{0\}$, car pour tout $n \in \mathbb{N}$ la condition de Blackwell 2.1 est vérifiée :

$$\left\langle \left(\overline{X}_n - \mathbb{E}_{\sigma,\tau}[\overline{X}_n]\right) - 0, \mathbb{E}\left[(X_{n+1} - \mathbb{E}[X_{n+1}])\right] - 0 \right\rangle = 0.$$

L'approchabilité avec des paiements non bornés, mais vérifiant une des deux conditions (2.5) peut être vue, quant à elle, comme une extension de la loi forte des grands nombres de Kolmogorov : si X_n est une suite de variables indépendantes vérifiant *le critère de Kolmogorov*, *i.e.* telle que la suite des variances v_n vérifie $\sum_{n\in\mathbb{N}} v_n/n^2$, alors $\overline{X}_n - \mathbb{E}[\overline{X}_n]$ converge presque sûrement vers 0 (voir par exemple Feller [40], chapitre X.7).

Approchabilité avec activations

On suppose dans cette section que seules certaines coordonnées du vecteur de paiement sont actives à chaque étape, bien qu'elles soient presque toutes actives une infinité de fois. Formellement, pour toute histoire finie h^n, on définit une variable aléatoire $\mathcal{X}[h^n] \in \mathcal{L}_2(\Omega, \mathbb{R})$ à valeurs dans $\{0, 1\}$. La coordonnée ω est active après l'histoire h^n si $\mathcal{X}[h^n](\omega) = 1$ et l'on suppose que, le long de chaque histoire infinie, $\sum_{n\in\mathbb{N}} \mathcal{X}[h^n](\omega) = +\infty$, pour λ-presque toutes les coordonnées $\omega \in \Omega$. Le paiement de l'étape n est donné sous ces hypothèses par $\rho_n = \rho(i_n, j_n)\mathcal{X}[h^n] \in \mathcal{L}_2(\Omega, \mathbb{R})$.

On se restreint dans cette section aux pavés de $\mathcal{L}_2(\Omega, \mathbb{R})$, *i.e.* aux ensembles de la forme :

$$C = \{f \in \mathcal{L}_2(\Omega, \mathbb{R}); c_0 \mathbb{1}_{W_0} \leq f\mathbb{1}_{W_0} \text{ et } f\mathbb{1}_{W_1} \leq c_1 \mathbb{1}_{W_1}\}$$

où $c_0, c_1 \in \mathcal{L}_2(\Omega, \mathbb{R})$ et $\mathbb{1}_{W_0}, \mathbb{1}_{W_1}$ sont des indicatrices d'ensembles mesurables de Ω.

Théorème 2.15

Soit C un pavé de $\mathcal{L}_2$ et σ une stratégie du joueur 1 telle que pour toute stratégie τ du joueur 2 et tout $n \in \mathbb{N}$, en notant $x_{n+1} = \sigma(h^n)$ et $y_{n+1} = \tau(h^n)$:

$$\left\langle \frac{\mathcal{X}[h^{n+1}]}{\sum_{m=1}^{n+1} \mathcal{X}[h^m]} \left(\overline{\rho}_n - \Pi_c(\overline{\rho}_n) \right), \rho(x_{n+1}, y_{n+1}) - \Pi_C(\overline{\rho}_n) \right\rangle \leq 0.$$

Alors σ est une stratégie d'approchabilité de C.

Ceci est une conséquence du lemme suivant (voir Lehrer [69] théorème 4).

Lemme 2.16

Soit C un pavé de $\mathcal{L}_2(\Omega, \mathbb{R})$ et supposons que :

i) $(g_n)_{n\in\mathbb{N}}$ est une suite de variables aléatoires bornées λ-ps par $B \in \mathcal{L}_2$;

ii) $(\mathcal{X}_n)_{n\in\mathbb{N}}$ est une suite de variables aléatoires à valeurs dans $\{0, 1\}$ telles que $\sum_{n\in\mathbb{N}} \mathcal{X}_n = +\infty$, λ-presque sûrement ;

iii) $\overline{g}_n = \sum_{k=1}^{n} \mathcal{X}_k g_k / \overline{\mathcal{X}}_n$;

iv) $\langle \mathcal{X}_{n+1} (\overline{g}_n - \Pi_C(\overline{g}_n)), (g_{n+1} - \Pi_C(\overline{g}_n)) / \overline{\mathcal{X}}_{n+1} \rangle \leq 0$.

Alors $\overline{g}_n$ converge vers C, λ-ps.

Démonstration . La démonstration est très proche de celle du Lemme 2.10 (elle repose d'ailleurs aussi sur le Lemme 2.11) et est une application de Lehrer [69], théorème 4. Pour tout $n \in \mathbb{N}$, l'inégalité suivante, où l'on a noté $f_n = \overline{g}_n - \Pi_C(\overline{g}_n)$, est une simple conséquence du point *iv)* :

$$\|f_{n+1}\|^2 \leq \|f_n\|^2 - 2\left\langle \mathcal{X}_{n+1}\frac{f_n}{\overline{\mathcal{X}}_{n+1}}, f_n\right\rangle + \left\|\frac{\mathcal{X}_{n+1}}{\overline{\mathcal{X}}_{n+1}}\left(g_{n+1} - \overline{g}_n\right)\right\|^2.$$

Ainsi comme g_{n+1} et $\overline{g}_n$ sont bornées par B et $\mathcal{X}_n \in \{0,1\}$:

$$2\left\langle \mathcal{X}_{n+1}\frac{f_n}{\overline{\mathcal{X}}_{n+1}}, f_n\right\rangle \leq \|f_n\|^2 - \|f_{n+1}\|^2 + 4\int_\Omega \frac{\mathcal{X}_{n+1}(\omega)}{\overline{\mathcal{X}}^2_{n+1}(\omega)}B^2(\omega)d\lambda(\omega).$$

Or, pour λ-presque tout ω, $\sum_{n\in\mathbb{N}} \mathcal{X}_{n+1}(\omega)/\overline{\mathcal{X}}^2_{n+1}(\omega) = \sum_{n\in\mathbb{N}} 1/(n+1)^2 = \pi^2/6$. Définissons $j(n,\omega) = \inf\left\{m \in \mathbb{N}, \overline{\mathcal{X}}_m(\omega) = n\right\}$ et $\widetilde{f}_n(\omega) = f_{j(n,\omega)}(\omega)$, alors

$$\sum_{n\in\mathbb{N}} \frac{\left\|\widetilde{f}_n\right\|^2}{n+1} = \sum_{n\in\mathbb{N}} \left\langle \mathcal{X}_{n+1}\frac{f_n}{\overline{\mathcal{X}}_{n+1}}, f_n\right\rangle \leq \frac{\|f_0\|^2}{2} + \frac{\pi^2}{3}\|B\|^2.$$

Comme C est un pavé, pour tout $\omega \in \Omega$,

$$\begin{aligned}\left|\widetilde{f}_{n+1}(\omega) - \widetilde{f}_n(\omega)\right| &= \left|\left[\overline{g}_{j_{n+1,\omega}}(\omega) - \overline{g}_{j_{n,\omega}}(\omega)\right] - \left[\Pi_C\left(\overline{g}_{j_{n+1,\omega}}\right)(\omega) - \Pi_C\left(\overline{g}_{j_{n,\omega}}\right)(\omega)\right]\right| \\ &\leq 2\left|\overline{g}_{j_{n+1,\omega}}(\omega) - \overline{g}_{j_{n,\omega}}(\omega)\right| = 2\left|\frac{g_{j_{n+1,\omega}}(\omega) - \overline{g}_{j_{n,\omega}}(\omega)}{n+1}\right| \leq \frac{4B(\omega)}{n+1}\end{aligned}$$

et donc $\|\widetilde{f}_{n+1} - \widetilde{f}_n\|^2 = 16\|B\|^2/(n+1)^2$.

En utilisant alors le Lemme 2.11, avec β_n une suite croissante telle que $\sum \beta_n\|\widetilde{f}\|^2/n$ converge et $M_{n+1} = \left\lceil \frac{\beta_{M_n}}{\beta_{M_n}+1}\right\rceil$, on obtient la convergence λ-presque sure de $\widetilde{f}_n$ et donc celle de f_n. □

On a utilisé le fait que C est un pavé pour borner $\left\|\widetilde{f}_{n+1} - \widetilde{f}_n\right\|$. La condition de $\mathcal{L}_2$-boundedness n'est pas suffisante pour cette démonstration. En effet si l'on définit la fonction $\widetilde{g}_n(\omega) = g_{j(n,\omega)}(\omega)$ alors $\Pi_C\left(\widetilde{g}_n\right)(\omega)$ n'est pas, a priori, égale à $\Pi_C\left(g_{j(n,\omega)}\right)(\omega)$.

Par ailleurs, de la même façon que lorsqu'il n'y a pas d'activation, on peut remplacer l'hypothèse *iv)* par

$$\sum_{n\in\mathbb{N}} \left\langle \frac{\mathcal{X}_{n+1}}{\overline{\mathcal{X}}_{n+1}}\left(\overline{g}_n - \Pi_C(\overline{g}_n)\right), g_{n+1} - \Pi_C(\overline{g}_n)\right\rangle < +\infty.$$

Stratégies à mémoire bornée

La stratégie d'approchabilité de Blackwell ne nécessite pas de connaître, à chaque étape, tous les paiements passés, mais simplement leur moyenne. Néanmoins, afin de pouvoir actualiser cette moyenne il est nécessaire de garder en mémoire le nombre d'étapes jouées, ce qui prend une place de plus en plus grande. On peut alors se demander s'il est possible d'approcher un ensemble E avec des stratégies plus simples, par exemple implémentables par des automates ou ayant une mémoire bornée.

On dit qu'une stratégie σ a une mémoire de taille $M \in \mathbb{N}$, si pour toute histoire finie $h^n \in H_n$, $\sigma(h^n)$ ne dépend que de $(i_{n-M+1}, j_{n-M+1}, \ldots, i_n, j_n)$, les M derniers profils d'actions jouées. Lehrer et Solan [73, 75] ont montré que si un convexe C est approchable par le joueur 1, alors en se restreignant à des stratégies avec une mémoire de taille $M \in \mathbb{N}$ ce dernier peut approcher le $O(1/\sqrt{M})$-voisinage de C. A fortiori, il peut donc faire de même avec des stratégies implémentables par des automates.

L'idée est relativement naturelle, il suffit de jouer la stratégie de Blackwell sur un bloc de M étapes, puis de recommencer. Il est nécessaire de coder le début et la fin d'un bloc, mais cela peut se faire avec $\sqrt{M}$ étapes, par exemple en jouant tout le temps la même action et en s'assurant qu'il n'y ait pas de telle suite dans le même bloc. Le paiement moyen de chaque bloc sera proche de C qui est convexe donc le paiement moyen sur l'ensemble des blocs sera lui-aussi proche de C.

D'un autre côté, Zapechelnyuk [121] a considéré la stratégie à mémoire bornée (naturellement adaptée à partir de celle de Blackwell) définie par $\sigma(h^n) = x\left(\overline{\rho}_n^M\right)$, où $x(\cdot)$ est donnée par la Définition 2.2 et $\overline{\rho}_n^M$ est la moyenne des paiements sur les M dernières étapes. Considérons le jeu où les paiements du joueur 1 (le joueur ligne) sont donnés par la matrice suivante,

	G	D
H	(0,-1)	(0,1)
B	(1,0)	(-1,0)

et où σ est une stratégie d'approchabilité de $C = \mathbb{R}^2_-$. Pour M assez grand, il existe une stratégie du joueur 2 telle que la suite $\left(\overline{\rho}_n^M\right)_{n\in\mathbb{N}}$ entre dans un cycle (de longueur $2M$ ou $2M+2$). Grossièrement, ce dernier est défini par 4 blocs d'une durée égale à $M/2$ (ou $M/2+1$) étapes où chaque bloc est constitué d'une répétition du même couple d'action (sauf éventuellement sur une seule étape). L'ordre des actions jouées sur les blocs est (H,G), (H,D), (B,D) et (B,G).

À la fin des blocs (B,D) et (H,G), $\overline{\rho}_n^M$ est proche, respectivement, de $(-1/2, 1/2)$ ou $(1/2, -1/2)$ donc il est à une distance d'environ $1/2$ de C. La suite $(\overline{\rho}_n^M)_{n\in\mathbb{N}}$ des moyennes sur les M dernières étapes ne converge donc pas vers C.

Cependant, rien n'indique que la moyenne des paiements $\overline{\rho}_n$ ne converge pas, quant à elle, vers C (ce qui est d'ailleurs le cas dans l'exemple de Zapechelnyuk [121]).

3

Tests de Théories et Calibration

On rappelle dans ce chapitre les notions de tests de théories ainsi que certaines conditions sous lesquelles les tests sont ou ne sont pas manipulables. On détaille en particulier les tests de calibration, leurs différentes variantes et les liens avec les notions de fusion.

Sommaire

3.1 Apprentissage non-Bayésien - Test de théories **26**
Manipulabilité des tests 26
Tests non manipulables 27
Plusieurs prédicteurs 28
Tests de calibration 29
3.2 Apprentissage Bayésien - Fusion **32**
Jeu entre l'inspecteur et un prédicteur 34
3.3 Calibration par rapport à une grille et ε-calibration **35**
Presque-calibration 37

Une suite $s^\infty \in S^{\mathbb{N}}$ — appelée *trajectoire* dans l'espace fini d'états S — est générée par la nature suivant la loi μ_0. Cette dernière appartient à $\Delta\left(S^{\mathbb{N}}\right)$ l'ensemble des mesures de probabilité sur $S^{\mathbb{N}}$, muni de la tribu produit. Une théorie sur μ_0 est formulée, *i.e.* un élément μ de $\Delta\left(S^{\mathbb{N}}\right)$ est annoncé par un prédicteur qui peut être de deux types : soit c'est un *expert* qui choisit $\mu = \mu_0$, soit c'est un *aveugle* qui ne sait rien.

Un inspecteur est chargé de découvrir le type du prédicteur, en soumettant sa théorie (dont l'ensemble, muni de la topologie faible-$\star$, est noté $\mathcal{T}$) à un test T, *i.e.* une fonction (Borel) mesurable de $\mathcal{T} \times S^{\mathbb{N}}$ à valeurs dans $\{0, 1\}$. Si $T(\mu, s^\infty) = 0$, le test T rejette la théorie μ en s^∞ et si $T(\mu, s^\infty) = 1$ il accepte la théorie μ en s^∞ — on dit aussi que la théorie réussit le test T en s^∞.

Il est souvent imposé aux tests d'accepter la vérité avec probabilité $1-\varepsilon$, *i.e.* pour toute théorie $\mu \in \mathcal{T}$, $\mathbb{P}_\mu \left\{ s^\infty \in \Delta\left(S^{\mathbb{N}}\right), T(\mu, s^\infty) = 1 \right\} \geq 1-\varepsilon$. Ceci signifie que si un expert annonce μ (qui est en fait la loi de s^∞), alors T accepte cette théorie avec une grande μ-probabilité. Par exemple, le test naïf qui consiste à n'accepter le long de la trajectoire s^∞, uniquement δ_{s^∞}, la masse de Dirac en s^∞, n'accepte pas la vérité. En effet, dès que μ n'est pas une masse de Dirac, $\mathbb{P}_\mu \left(T(\mu, s^\infty) = 1\right) = 0$.

D'après le théorème d'extension de Kolmogorov, toute fonction μ (appelée théorie de comportement et dont l'ensemble est noté $\mathcal{T}_c$) de l'ensemble des histoires finies $\bigcup_{n\in\mathbb{N}} \left(\Delta(S) \times S\right)^n$ à valeur dans $\Delta(S)$ induit une unique théorie (aussi notée μ). Les choix d'une théorie de comportement $\mu \in \mathcal{T}_c$ et d'une trajectoire $s^\infty = (s_n)_{n\in\mathbb{N}}$ définissent une unique histoire infinie $(\mu, s)^\infty = (\mu_1, s_1, \mu_2, s_2, \dots) \in (\Delta(S) \times S)^{\mathbb{N}}$, où $\mu_{n+1} \in \Delta(S)$ est la probabilité de s_{n+1} conditionnellement à l'histoire passée, appelée *prédiction* de l'étape $n+1$.

Un test T n'utilise que les données observées si $T(\mu, s^\infty) = T\left[(\mu, s)^\infty\right] \in \{0, 1\}$ pour toute trajectoire et toute théorie de comportement μ ; lorsque l'on considère cette classe de tests, on entend implicitement que les prédicteurs sont restreints à $\mathcal{T}_c$ afin d'éviter les problèmes d'existence des suites de prédictions. Dans ce cadre, un test est donc une fonction mesurable de $\mathcal{T}_c \times S^{\mathbb{N}}$ (chacun muni de sa tribu produit) à valeurs dans $\{0, 1\}$.

3.1 Apprentissage non-Bayésien - Test de théories

L'interaction entre prédicteur et inspecteur peut se représenter sous forme de jeu. L'inspecteur choisit un test T qu'il annonce au prédicteur puis celui-ci choisit une théorie $\mu \in \mathcal{T}$. La nature génère alors une trajectoire s^∞ selon la loi $\mu_0 \in \mathcal{T}$, qui est nécessairement égale à μ si le prédicteur est un expert. Le paiement du prédicteur est $T(\mu, s^\infty)$.

Manipulabilité des tests

Une question essentielle est de distinguer, parmi tous les tests, ceux qui peuvent être réussis par un prédicteur aveugle et qui ne permettent donc pas de distinguer un expert d'un aveugle. Plus précisément, on dit qu'un test peut être *réussi de manière aveugle avec probabilité* $1-\eta$ s'il existe une probabilité ξ sur $\mathcal{T}$ telle que pour toute histoire $s^\infty \in S^{\mathbb{N}}$, $\xi\left\{\mu \in \mathcal{T}, T(\mu, s^\infty) = 1\right\} \geq 1-\eta$.

1. Tout test à **horizon fini** (*i.e.* il existe $n \in \mathbb{N}$ tel que $T(\mu, s^\infty) = T(\mu, s^n)$, avec s^n le préfixe de taille n de s^∞), n'utilisant **que les données observées** et acceptant la vérité avec probabilité $1-\varepsilon$ peut être réussi de manière aveugle avec probabilité $1-\varepsilon$ (Sandroni [101]).

2. Tout test à **horizon infini**, n'utilisant **que les données observées** et acceptant la vérité avec probabilité $1-\varepsilon$ peut être réussi de manière aveugle avec probabilité à $1-\varepsilon-\delta$, pour tout $\delta > 0$ (Shmaya [105]).
3. Tout test **indépendant du futur** à **rejet en temps fini** (*i.e.* si, le long de toute trajectoire, les prédictions générées par deux théories μ et μ' coincident jusqu'à l'étape n et $T(\mu, s^\infty) = T(\mu, s^n) = 0$ alors $T(\mu', s^n) = 0$) et acceptant la vérité avec probabilité $1-\varepsilon$ peut être réussi de manière aveugle avec probabilité $1-\varepsilon$ (Olszewski et Sandroni [87]).

Ces trois résultats ne donnent pas de théories explicites. Leurs preuves sont basées, pour le point 1 et 3, sur le théorème de min-max de Fan [36] et pour le point 2 sur le théorème de décidabilité de Martin [82].

Tests non manipulables

En réponse aux résultats précédents négatifs, il existe aussi des tests non manipulables, parfois au coût d'hypothèses ou contraintes supplémentaires sur le prédicteur — ou d'en relâcher sur le test. On dit qu'un test ne peut être manipulé, s'il ne peut être réussi de manière aveugle avec une probabilité strictement positive.

On rappelle qu'un sous-ensemble E d'un espace mesurable $(\Omega, \mathcal{F})$ est universellement mesurable si E appartient à la complétion de $\mathcal{F}$ par rapport à toutes les mesures de probabilité, voir Billingsley [16], *i.e.* pour toute probabilité λ, il existe 2 ensembles mesurables B et A tels que $A \subset E \subset B$ et $\lambda(B \backslash A) = 0$; $\lambda(E)$ est alors défini par $\lambda(E) = \lambda(A) = \lambda(B)$). Le test T est universellement mesurable si $T^{-1}(\{1\})$ est universellement mesurable (il n'est donc pas nécessairement mesurable).

1. Sous **l'hypothèse du continu**, il existe un test T qui accepte la vérité avec probabilité 1 et tel que pour tout $\xi \in \Delta(\mathcal{T})$ il existe un ensemble $\mathcal{S} \subset S^{\mathbb{N}}$ non dénombrable tel que pour tout $s^\infty \in \mathcal{S}$, $\xi\{\mu, T(\mu, s^\infty) = 1\} = 0$ (Dekel et Feinberg [34]).
2. Sous **l'axiome du choix**, il existe un test **universellement mesurable** qui accepte la vérité et qui ne peut être manipulé de manière aveugle (Shmaya [105]).
3. Il existe un test T qui accepte la vérité avec probabilité 1 et tel que pour tout $\xi \in \Delta(\mathcal{T})$ l'ensemble de révélation

$$R_\xi = \left\{ s^\infty \in S^{\mathbb{N}}, \xi\{\mu, T(\mu, s^\infty) = 1\} \leq \varepsilon \right\}$$

   est un **ensemble *gras***, *i.e.* le complémentaire d'une union dénombrable de fermé d'intérieur vide — appelé ensemble de catégorie I de Baire (Olsewski et Sandroni [89]).
4. Il existe un test T et, pour tout $n \in \mathbb{N}$, un **sous-ensemble de théories admissibles** $\Gamma_n \subset \mathcal{T}$ tels que pour tout $\varepsilon > 0$, et m suffisamment grand (*i.e.* supérieur à un certain $\overline{m}_\varepsilon$) T accepte la vérité si elle appartient à Γ_m avec probabilité $1-\varepsilon$ et rejette $\mu \neq \mu_0$ avant l'étape m avec probabilité au moins

$1-\varepsilon$. De plus, les ensembles Γ_m sont *grands* au sens suivant : chaque Γ_m contient un ouvert O_m tel que pour tout voisinage V de n'importe quel $\tau \in \mathcal{T}$, il existe $\underline{m}$ et $\underline{\tau}$ tels que $\underline{\tau} \in O_m \cap V$ pour $m \geq \underline{m}$ (Olszewski et Sandroni [90]).

5. Supposons qu'il existe une famille $(\mu_\theta)_{\theta\in\Theta}$ de **théories admissibles paramétrée** par (Θ, λ) où λ est une probabilité sur Θ. Si μ_θ converge (au sens *fusion faible*, voir section 3.2) vers μ_0 et $\lim_{n\to\infty} \sup_{\Omega\in\mathcal{F}_n^\infty} |\mu_\theta(\Omega|h^n) - \mu_\theta(\Omega)| = 0$ pour λ-presque tout θ, où $\mathcal{F}_n^\infty$ est la tribu générée par l'ensemble des suffixes de trajectoire $\left\{(s_t)_{t\geq n}, s^\infty \in S^{\mathbb{N}}\right\}$. Il existe alors un test qui accepte la vérité avec probabilité $1-\varepsilon$ et qui ne peut être ε-manipulé (Al-Najjar, Sandroni, Smorodinsky et Weinstein [1]).

Les résultats 1 et 2 donnent l'existence de tests généraux, mais qui ne peuvent être construits. Le point 3 aborde le problème sous un angle différent, en cherchant non pas à construire un test qui ne peut pas être manipulé avec une grande probabilité, mais sur un ensemble topologiquement *large*. Pour les résultats 4 et 5, c'est en réduisant l'ensemble des théories admissibles du prédicteur que l'on parvient à construire un test qui n'est pas manipulable (mais Olszewski et Sandroni [90] et Najjar, Sandroni, Smorodinsky et Weinstein [1] ont ainsi pu construire ces tests de manière explicite).

Plusieurs prédicteurs

Supposons maintenant que deux prédicteurs émettent une théorie et qu'au plus l'un d'entre eux soit un expert. Un test T est une fonction de $\mathcal{T} \times \mathcal{T} \times S^{\mathbb{N}}$ à valeurs dans $\{0,1\}^2$. Un test *choisit un prédicteur* s'il accepte sa théorie et refuse l'autre.

Sous certaines conditions, il est possible de déterminer qui est l'expert soit avec une grande probabilité, soit sur un ensemble de trajectoires topologiquement gras, au sens de Baire :

1. Supposons qu'il y ait un expert. Alors il existe un test T qui le choisit **en temps infini** avec probabilité 1 ou alors la différence des prédictions tend vers 0, Al-Najjar et Weinstein [2].
2. Dans le même cadre, pour tout $\varepsilon > 0$ il existe un entier K et un test qui choisit **en temps fini** l'expert ou alors les prédictions sont toutes ε-proches sauf sur K périodes, Al-Najjar et Weinstein [2].
3. Il existe un test (appelé *cross-calibration*) qui accepte la théorie d'un prédicteur aveugle en présence d'un expert seulement sur un sous-ensemble de $\Delta(S^{\mathbb{N}})$ **de catégorie I**. S'il n'y a pas d'expert, deux prédicteurs réussissent simultanément le test sur un sous-ensemble de catégorie I, Feinberg et Stewart [39].
4. Tout test **indépendant du futur** et qui **accepte la vérité avec probabilité** $1-\varepsilon$, **peut être réussi** par deux prédicteurs aveugles avec probabilité $1-\varepsilon-\delta$, pour tout $\delta > 0$, Olszewski et Sandroni [88].

La différence notable entre les troisième et quatrième points est, comme remarqué par Olszewski et Sandroni [87], que le test de *cross-calibration* peut rejeter la vérité.

Tests de calibration

Les tests empiriques de calibration (introduits notamment par Dawid [33]) comparent la moyenne empirique des états à la *moyenne des prédictions* d'une théorie $\mu \in \mathcal{T}_c$. La moyenne de $\mu_1, \ldots, \mu_n \in \Delta(S)$ est notée $\overline{\mu}_n \in \Delta(S)$ et elle est définie par

$$\overline{\mu}_n(A) = \frac{\sum_{m=1}^{n} \mu_m(A)}{n} \text{ pour tout ensemble Borel mesurable } A. \tag{3.1}$$

La justification de l'étude de ces tests est due, dans un premier temps, à Dawid [33] :

Considérons un expert qui connaît donc la loi μ_0 d'une trajectoire $(s_n)_{n\in\mathbb{N}}$ à valeur dans $\{0,1\}$. À l'étape n, il prédit $\mu_0(s_{n+1} = 1|s^n)$, la probabilité conditionnelle à l'histoire $s^n = (s_1, \ldots, s_n)$ que l'état suivant soit 1. Soient $p \in [0,1]$ fixé et $\varepsilon > 0$ quelconque. Sur l'ensemble des étapes où la prédiction est ε-proche de p, la moyenne empirique des états est asymptotiquement ε-proche de p, dès que ce nombre d'étapes n'est pas borné (voir Dawid [33], p. 607). En conclusion, un prédicteur qui ne réussit pas un test de ce type ne peut être un expert.

De plus, Oakes [86] (dont l'argument a ensuite été simplifié par Dawid [86]) a montré qu'un prédicteur aveugle ne peut pas réussir ces tests s'il choisit une théorie μ de manière déterministe. En effet, considérons l'histoire s^∞ définie récursivement par $s_{n+1} = 0$ si $\mu_{n+1} = \mu(1|s^n) \geq 1/2$ et $s_{n+1} = 0$ sinon. Alors, s'il y a une infinité de prédictions strictement plus petites que $1/2$ (resp. supérieures ou égales à $1/2$), sur ces dates là, la moyenne empirique des états est exactement égale à 1 (resp. 0) et le test rejette la théorie. Par opposition, Foster et Vohra [43] ont construit un algorithme σ choisissant des prédictions aléatoirement, *i.e.* σ est une probabilité sur $\Delta\left(S^{\mathbb{N}}\right)$, et de manière aveugle telle que la suite de prédictions réussit chacun des tests considérés par Dawid [33], presque sûrement.

Plusieurs critiques peuvent être formulées envers la simplicité des tests de calibration. En effet, la théorie qui prédit à chaque étape une probabilité de $1/2$ le long de la suite alternée de 0 et de 1 réussit ce test (on dit aussi qu'elle est *naïvement*-calibrée). Par contre, celle qui prédit successivement 0.01 et 0.99, et qui donc donne a priori plus d'information sur la suite, ne l'est pas. Il est donc nécessaire — pour pouvoir distinguer un expert d'un aveugle — de faire également un test restreint à l'ensemble des dates paires, et de manière similaire sur les dates impaires, puis les multiples de trois, quatre, etc.

Supposons maintenant que la suite des états est gouvernée par une chaîne de Markov, où la probabilité de changer d'état est de 0.01 (et donc la probabilité de rester dans le même est de 0.99). Prédire $1/2$ à chaque étape est encore calibré, tandis que si l'on regarde uniquement les étapes suivant celles où l'état était 1, alors la prédiction calibrée est 0.99. Il est donc également nécessaire de faire des tests de calibration sur certaines étapes choisies en fonction de la suite des états mais aussi (à cause de l'exemple d'Oakes) en fonction de la suite des prédictions.

On appelle $H^n = (S \times \Delta(S))^n$ l'ensemble des histoires finies (*i.e.* la suite des états et des prédictions) jusqu'à la date n, et $\mathcal{H}$ l'ensemble des histoires muni de la topologie produit. Un test de calibration repose sur un couple de fonctions définies sur $H := \bigcup_{n\in\mathbb{N}} H^n$, appelé *règle d'inspection* :

Définition 3.1

Une règle d'inspection $\mathrm{RI} = (U, C)$ *est un couple de fonctions sur* H *telles que :*

1) $U(h^{n+1})$ *et* $C(h^{n+1})$ *sont deux événements mesurables à la même date finie* $n+m$ *(i.e. ils sont générés par les cylindres de taille* $n+m$*) ;*
2) $C(h^{n+1})$ *est un sous-ensemble de* $U(h^{n+1})$.

Étant données une partie h^∞ et l'histoire finie h^n (le préfixe de h^∞), l'événement $U(h^n)$ est l'univers d'activation du test, *i.e.* ce dernier est actif à l'étape $n+1$ si $h^\infty \in U(h^n)$ et, le cas échéant, l'événement testé est $C(h^n)$.

Définition 3.2

Étant donnée une règle d'inspection $\mathrm{RI} = (U, C)$*, une théorie* μ *est* RI*-calibrée le long de la partie* $h^\infty \in \mathcal{H}$ *si*

$$\lim_{n\to\infty} \frac{\sum_{m=1}^{n} \mathbb{1}\left\{h^\infty \in U(h^n)\right\}\left[\mathbb{1}\left\{h^\infty \in C(h^n)\right\} - \mu(C(h^n)|U(h^n), h^{n-1})\right]}{\sum_{m=1}^{n} \mathbb{1}\left\{h^\infty \in U(h^n)\right\}} = 0$$

lorsque $\sum_{m=1}^{n} \mathbb{1}\left\{h^\infty \in U(h^{n+1})\right\} = \infty$.

Dans cette définition, on utilise les conventions suivantes : $\mu(C|U) = +\infty$ si $\mu(U) = 0$, $0/0 = 1$ et $0.\infty = 0$. Par ailleurs, on remplacera parfois le dénominateur par n, mais cela sera précisé le cas échéant. Lorsque $h^\infty \in C(h^n)$, on dit que $C(h^n)$ est réalisé.

Intuitivement, une théorie est RI-calibrée si la moyenne des prédictions conditionnelles et la moyenne empirique des réalisations sont asymptotiquement proches, dès que le test est actif une infinité de fois.

Une règle d'inspection est dite

i) indépendante des prédictions si $U(h^{n+1})$ et $C(h^{n+1})$ ne dépendent pas de la prédiction à l'étape $n+1$ et
ii) à court terme si l'événement $C(h^n)$ est mesurable à la date $n+1$ pour toute histoire $h^n \in H^n$.

L'ensemble RI_{ic} des règles d'inspection indépendantes des prédictions et à court terme est muni de la tribu générée par les ouverts (les ensembles contenant toutes les règles d'inspection qui coincident sur un nombre fini d'histoires). On définit de manière similaire RI_c l'ensemble des règles d'inspection à court terme, qui peuvent donc dépendre des prédictions.

Le théorème suivant, dont la démonstration est basée sur l'inégalité de Ky Fan [37], donne l'existence d'une théorie qui manipule simultanément de nombreux tests de calibration.

Théorème 3.3 (*Lehrer [68]*)

Soit λ une probabilité sur RI_{ic}. Il existe une théorie μ qui passe λ-presque tous les tests de calibration le long de chaque partie h^∞. Une telle théorie est dite λ-calibrée.

Dans la version de ce théorème prouvée par Lehrer [68] le numérateur à l'intérieur de la somme est multiplié par $\mu(U(h^n)|h^n)$ afin qu'il soit linéaire en fonction des prédictions et des états (voir Lehrer [68], Remarque 4).

Ce résultat est doublement important, car étant donnée une probabilité λ sur RI_{ic}, le prédicteur aveugle peut non seulement choisir une théorie λ-calibrée, mais de plus il peut le faire de manière déterministe (la démonstration, proche de celle de Lehrer, est donnée en section 5.3).

Bien sûr, ceci n'est pas possible dès que U n'est pas indépendant des prédictions, comme l'ont montré Oakes [86] et Dawid [33]. Dans ce cadre, Sandroni, Smorodinsky et Vohra [102] ont néanmoins obtenu un résultat similaire au théorème 3.3 en réduisant les règles d'inspection admissibles et en autorisant le prédicteur à choisir sa théorie de manière aléatoire, *i.e.* en choisissant une fonction ξ définie sur H et à valeurs dans $\Delta(\Delta(S))$. Un tel choix induit pour toute histoire infinie h^∞ une unique probabilité $\xi_\star^{h^\infty}$ sur $\mathcal{T}_c$ (d'après le théorème d'extension de Kolmogorov).

Les règles d'inspection considérées sont à court terme et sans univers d'activation, *i.e.* $U(h^n) \in \{\emptyset, S\}$ pour toute histoire finie. De plus, on suppose qu'il existe une fonction $A : H \mapsto \{0,1\}$ et D, un sous-ensemble de S, tels que la règle d'inspection est active à l'étape $n+1$ si $A(h^n) = 1$ et si $\mu_{n+1} \in D$. Les choix de A et D génèrent $|D|$ règles d'inspection $\{A_d^D, d \in D\}$ où la règle d'inspection A_d^D est définie par $C(h^n) = d$ pour toute histoire finie. Une théorie est dite A^D-calibrée si elle est A_d^D-calibrée pour tout $d \in D$ et on appelle $\mathrm{RI}_{\mathcal{A}^\mathcal{D}}$ l'ensemble de ces règles d'inspection.

Théorème 3.4 (*Sandroni, Smorodinsky et Vohra [102]*)

Soit $\mathbf{A^D}$ un sous-ensemble dénombrable de $\mathrm{RI}_{\mathcal{A}^\mathcal{D}}$. Il existe une théorie aléatoire ξ telle que, le long de chaque partie h^∞, $\xi_\star^{h^\infty}$-presque toutes les théories réussissent tous les tests de calibration de $\mathbf{A^D}$.

Pour résumer, si les règles d'inspection utilisées par l'inspecteur ne dépendent pas des prédictions alors le test est manipulable de façon déterministe, tandis que que si elles en dépendent, le test n'est manipulable que de façon aléatoire. En particulier, il n'existe pas de théorie déterministe telle que le long de toute histoire h^∞ et pour

tout $p \in \Delta(S)$ et $\varepsilon > 0$

$$\frac{1}{n}\sum_{m=1}^{n} \mathbb{1}_{p,\varepsilon}(\mu_m)\,(s_m - \mu_m) \text{ converge vers } 0 \tag{3.2}$$

avec μ_m la prédiction faite à l'étape m (conditionnellement à h^{n-1}) et $\mathbb{1}_{p,\varepsilon}$ l'indicatrice de la boule fermée centrée en p de rayon ε. On rappelle que Foster et Vohra [43] ont donné un algorithme qui indique comment choisir aléatoirement μ_m de sorte que que la propriété (3.2) soit vraie presque sûrement.

3.2 Apprentissage Bayésien - Fusion

Les tests de calibration sont des tests empiriques qui visent à déterminer si une suite de prédictions est faite par un expert ou par un prédicteur aveugle. Ils ne mesurent pas a priori la précision asymptotique d'une théorie μ par rapport à μ_0.

Dans ce cadre — on parle d'apprentissage Bayésien — un prédicteur annonce une théorie μ (on emploie aussi le terme de croyance) et l'inspecteur vérifie comment, conditionnellement à l'histoire h^n, les deux mesures $\mu(\cdot|h^n)$ et $\mu_0(\cdot|h^n)$ se comportent asymptotiquement. Blackwell et Dubins [19] ont défini une notion de convergence entre mesures, appelé *fusion* (merging dans la littérature anglaise). On appelle $\mathcal{B}$ la tribu engendrée sur $\mathcal{H}$ par l'ensemble des histoires finies et $(\mathcal{B}_n)_{n\in\mathbb{N}}$ la filtration de $(\mathcal{H}, \mathcal{B})$ définie par les ensembles mesurables à la date n.

Définition 3.5

Une mesure μ fusionne *avec μ_0 si pour μ_0-presque toute histoire h^∞ :*

$$\lim_{n\to\infty} \sup_{A\in\mathcal{B}} |\mu(A|h^n) - \mu_0(A|h^n)| = 0.$$

Intuitivement, si la croyance d'un prédicteur μ fusionne avec μ_0 alors, après avoir observé une histoire finie suffisamment longue h^n, la mise-à-jour de sa croyance (*i.e.* $\mu(\cdot|h^n)$) est très proche de la loi réelle $\mu_0(\cdot|h^n)$. L'observation de h^n permet bien de prédire μ_0.

Blackwell et Dubins [19] ont montré que si μ_0 est absolument continue par rapport à μ (*i.e.* si pour tout $B \in \mathcal{F}$, $\mu(B) = 0$ implique que $\mu_0(B) = 0$, ce que l'on note $\mu \ll \mu_0$), alors μ fusionne vers μ_0. Réciproquement, Kalai et Lehrer [64] ont montré que si μ fusionne avec μ_0 (le long de toute filtration) alors $\mu_0 \ll \mu$. Absolue continuité et fusion sont donc deux notions équivalentes.

La condition de fusion est, dans certains cas, trop forte pour déterminer s'il y a apprentissage ou non de μ_0, comme l'illustrent les exemples 5 et 6 de Lehrer et Smorodinsky [72]. Considérons le processus stochastique déterminé par une succession d'experiences de Bernouilli de paramètre p. On note $\mu_0(p)$ la loi des suites de $\{0,1\}^{\mathbb{N}}$

générées ainsi. Supposons que le prédicteur sache que p est rationnel mais qu'il n'en connaisse pas la valeur exacte. En énumérant $\mathbb{Q} \cap [0,1] = \{p_n, n \in \mathbb{N}\}$, la théorie μ définie par $\mu = \sum_{n\in\mathbb{N}} \mu_0(p_n)2^{-(n+1)}$ fusionne avec μ_0 puisque $\mu_0 \ll \mu$.

Par contre, supposons que le paramètre p soit choisi uniformément sur $[0,1]$. Après un grand nombre d'étapes, p sera presque connu, au sens où il appartient avec une probabilité arbitrairement grande à un intervalle $[p-\varepsilon, p+\varepsilon]$ arbitrairement petit, mais μ n fusionne pas avec μ_0. En effet l'événement défini par $A = \left\{h^\infty : \lim_{n\to\infty} \frac{1}{n}\sum_{m=1}^{n} s_m = p\right\}$ est tel que $\mu_0(A) = 1$ et $\mu(A) = 0$. Le problème vient du fait que A est un *événement asymptotique*. C'est pourquoi la notion de *fusion faible* a été introduite par Kalai et Lehrer [64].

Définition 3.6

Une mesure μ fusionne *faiblement avec μ_0 si pour μ_0-presque toute histoire h^∞ :*

$$\lim_{n\to\infty} \sup_{A\in\mathcal{B}_{n+1}} |\mu(A|h^n) - \mu_0(A|h^n)| = 0. \tag{3.3}$$

La définition, qui dépend de la filtration choisie, est équivalente si au lieu de prendre le supremum sur les événements mesurables dans dans un *futur immédiat*, *i.e.* à la date $n+1$, mais il est pris sur ceux mesurables dans *un futur proche*, *i.e.* à une date $n+k$ avec k fixé. Kalai et Lehrer [64] ont montré qu'une mesure μ fusionnant faiblement le long de n'importe quelle filtration vers μ_0, fusionne vers μ_0.

Un ensemble $\mathbb{N}_\varepsilon \subset \mathbb{N}$ est *plein* si sa *densité supérieure* notée $\mathrm{UD}(\mathbb{N}_\varepsilon)$ vérifie

$$\mathrm{UD}(\mathbb{N}_\varepsilon) := \limsup_{n\to\infty} \frac{|\mathbb{N}_\varepsilon \cap \{0,\ldots,n\}|}{n} = 1.$$

Définition 3.7

Une mesure μ fusionne *presque-faiblement avec μ_0 si pour tout $\varepsilon > 0$, et pour μ_0-presque toute histoire h^∞, il existe $\mathbb{N}_\varepsilon$ un sous-ensemble plein de $\mathbb{N}$ tel que :*

$$|\mu_0(A|h^n) - \mu(A|h^n)| < \varepsilon, \forall n \in \mathbb{N}_\varepsilon, \forall A \in \mathcal{B}_{n+1}.$$

Là encore, la définition dépend de la filtration choisie. Lehrer et Smorodinsky [71] ont montré que si μ est telle que pour μ_0-presque toutes les histoires, il existe un ensemble plein $\mathcal{N}'$ tel que :

$$\liminf_{n\in\mathbb{N}'} \left(\frac{\mu(h^{n+1})}{\mu_0(h^{n+1})}\right)^{1/n} \geq 1 \tag{3.4}$$

alors μ fusionne presque-faiblement vers μ_0. La condition (3.4) n'est pas nécessaire, cependant il existe une condition nécessaire et suffisante très proche mais qui requiert

encore plus de notations (voir Lehrer et Smorodinsky [72] pour une étude poussée de ces convergences et des exemples illustratifs).

Kalai, Lehrer et Smorodinsky [65] ont relié les notions de fusion aux tests de calibration où les règles d'inspection appartiennent aux ensembles suivants :

1. RI_{iu} est l'ensemble des règles d'inspection indépendantes des prédictions et sans univers d'activation ;
2. $\mathrm{RI}_{iuc} \subset \mathrm{RI}_{iu}$ est l'ensemble des règles d'inspection de RI_{iu} à court terme ;
3. $\mathrm{RI}_{iuca} \subset \mathrm{RI}_{iuc}$ est l'ensemble des règles d'inspection de RI_{iuc} telles que pour presque toute partie h^∞, $\liminf_{n\to\infty} \frac{1}{n}\sum_{m=1}^{n} \mathbb{1}\{h^\infty \in U(h^{n+1})\} > 0$.

Théorème 3.8 (*Kalai, Lehrer et Smorodinsky [65]*)

1. *Une mesure μ* fusionne *avec μ_0 si et seulement si μ est* RI-*calibrée pour toute règle d'inspection* $\mathrm{RI} \in \mathrm{RI}_{iu}$, *$\mu_0$-ps.*
2. *Une mesure μ* fusionne *faiblement avec μ_0 si et seulement μ est* RI-*calibrée pour toute règle d'inspection* $\mathrm{RI} \in \mathrm{RI}_{iuc}$, *$\mu_0$-ps.*
3. *Une mesure μ* fusionne *presque-faiblement avec μ_0 si et seulement μ est* RI-*calibrée pour toute règle d'inspection* $\mathrm{RI} \in \mathrm{RI}_{iuca}$, *$\mu_0$-ps.*

Jeu entre l'inspecteur et un prédicteur

Les deux sections précédentes présentent des résultats à première vue contradictoires. En effet, d'un côté on a affirmé que, étant donnée une distribution de règles d'inspection à court terme, on peut construire une théorie de comportement μ calibrée de manière aveugle, *i.e.* sans information sur μ_0. D'un autre côté, une théorie ne peut réussir tous ces tests qu'à la condition que μ fusionne-faiblement vers μ_0, et donc il est nécessaire que μ et μ_0 soient très proches (car la notion de fusion-faible est plus faible que la fusion et plus forte que la fusion-presque faible).

Une réponse à ce paradoxe peut être trouvée si l'on considère le jeu entre l'inspecteur et le prédicteur (introduit par Lehrer [68], section 4) où l'espace d'actions de l'inspecteur est $\Delta(\mathrm{RI}_{iuc})$ — l'ensemble des probabilités sur les règles d'inspection indépendantes des prédictions sans univers d'activation et à court terme, munie de la tribu produit — et celui du prédicteur est $\mathcal{T}_c$. Étant données une partie h^∞ choisie par la nature, une théorie μ et une règle d'inspection RI, le paiement du prédicteur, noté $V(\mu, \mathrm{RI}, h^\infty)$, est égal à 1 si μ est RI-calibrée le long de h^∞ et 0 sinon.

D'après le théorème 3.3, pour toute stratégie λ de l'inspecteur, il existe une stratégie μ du prédicteur aveugle telle que pour toute histoire infinie h^∞, $V(\mu, \mathrm{RI}, h^\infty) = 1$ soit :

$$\min_{\lambda \in \Delta(\mathrm{RI}_{iuc})} \max_{\mu \in \mathcal{T}_c} \mathbb{E}_{\mu_0,\lambda}\left[V(\mu, \mathrm{RI}, h^\infty)\right] = 1.$$

Notons $\mathcal{M}_f(\mu_0)$ l'ensemble des mesures qui ne fusionnent pas faiblement avec μ_0. Le théorème 3.8 stipule que pour toute stratégie μ du prédicteur qui appartient à

$\mathcal{M}_f(\mu_0)$, il existe un test de calibration RI tel que μ n'est pas RI-calibrée le long de toutes les histoires h^∞. Ainsi :

$$\inf_{\lambda \in \Delta(\mathrm{RI}_{iuc})} \mathbb{E}_{\mu_0,\lambda}\left[V(\mu, \mathrm{RI}, h^\infty)\right] < 1.$$

Le premier élément de réponse au paradoxe serait de dire que dans ce jeu entre prédicteur aveugle et inspecteur, le *supinf* n'est pas égal au *maxmin*. Cependant, ce n'est a priori pas nécessairement vrai, même si l'on suppose que la stratégie μ appartient à $\mathcal{M}_f(\mu_0)$. Il suffit par contre de supposer l'existence d'un $\varepsilon > 0$ tel que l'ensemble des histoires ne vérifiant pas la condition (3.3) — données par la définition de fusion faible — a une μ_0-probabilité plus grande que ε. Sous ces hypothèses, le *supinf* est en effet inférieur à $1 - \varepsilon$ tandis que le *minmax* vaut 1.

Le second élément de réponse est de remarquer que, pour certaines lois μ_0, le choix des tests peut forcer toute théorie les réussissant de manière aveugle à être proche de μ_0. En effet, supposons que $S = \{0, 1\}$, $\mu_0 = \delta_{1,1,1,\ldots}$ et notons RI_1 la règle d'inspection toujours active telle que $C(h^n) = \{1\}$ pour toute histoire finie. Pour qu'une stratégie μ soit RI_1-calibrée, il est nécessaire que pour tout $\varepsilon > 0$, $\mu(1|h^n) < 1 - \varepsilon$ seulement sur un ensemble d'histoires de densité supérieure nulle. En particulier, toute stratégie RI_1-calibrée fusionne presque-faiblement vers μ_0.

3.3 Calibration par rapport à une grille et ε-calibration

Dans la section précédente, une théorie (de comportement) aléatoire est définie comme une application de l'ensemble des histoires finies dans $\Delta(\Delta(S))$. Il s'agit donc d'une stratégie de comportement dans le jeu répété à deux joueurs où l'espace d'actions pures du joueur 1 est $\Delta(S)$ et celui du joueur 2 (aussi appelé Nature) est S.

Étant données une stratégie σ du joueur 1 et une histoire $h^\infty \in (\Delta(S) \times S)^{\mathbb{N}}$, on appelle $\mu_n \in \Delta(S)$ la prédiction faite à l'étape n, dont la loi est $\sigma(h^{n-1}) \in \Delta\left(\Delta(S)\right)$. Pour tout $p \in \Delta(S)$, on note $N_n(p, \varepsilon) = \sum_{m=1}^{n} \mathbb{1}\left\{\|\mu_m - p\| \leq \varepsilon\right\}$ l'ensemble des étapes avant la n-ème, où la prédiction $\mu_m \in \Delta(S)$ est ε-proche de p. La distribution empirique des états sur $N_n(p)$ est $\overline{s}_n(p, \varepsilon) \in \Delta(S)$ et la moyenne des prédictions — au sens de 3.1 — sur $N_n(p)$ est $\overline{\mu}_n(p, \varepsilon) \in \Delta(S)$.

Définition 3.9

Une stratégie σ du joueur 1 est ε-calibrée si pour tout $p \in \Delta(S)$ et toute stratégie τ de la nature :

$$\limsup_{n\to\infty} \frac{N_n(p,\varepsilon)}{n}\left(\|\overline{\mu}_n(p,\varepsilon) - \overline{s}_n(p,\varepsilon)\|^2 - \varepsilon^2\right) \leq 0, \quad \mathbb{P}_{\sigma,\tau}\text{−ps.}$$

Une stratégie est dite naïvement calibrée si et seulement si elle est ε-calibrée pour tout $\varepsilon > 0$.

La norme utilisée est la norme $\mathcal{L}_2$ de $\mathbb{R}^S$ et, bien évidemment, il est possible d'enlever les carrés dans la définition 3.9. Néanmoins, c'est précisément cette formulation qui sera utile ultérieurement. Intuitivement, une stratégie du joueur 1 est ε-calibrée si sur l'ensemble des dates où la prédiction est ε-proche de p (et si cet ensemble a une densité positive non nulle) la moyenne des prédictions est ε-proche de la moyenne empirique des états.

La construction d'une stratégie naïvement calibrée peut ainsi être ramenée à la construction pour tout $\varepsilon > 0$ d'une stratégie ε-calibrée. Il suffit ensuite de concaténer les stratégies, en utilisant par exemple l'argument classique de *doubling-trick* (voir par exemple Sorin [106], proposition 3.2 p. 56). L'existence de ces stratégies a été montrée par Foster et Vohra [43] (et aussi par Fudenberg et Levine [49]) en considérant des stratégies calibrées par rapport à une ε-grille de $\Delta(S)$, définie comme suit.

Définition 3.10

Un sous ensemble $\mathcal{L} = \{x(l), l \in L\}$ fini de $K \subset \mathbb{R}^d$ est une ε-grille de K si pour tout $x \in K$ il existe $l \in L$ tel que $\|x - x(l)\| \leq \varepsilon$.

Une grille est *régulière* s'il existe $\{e_1, \ldots, e_d\}$, d vecteurs indépendants, tels que

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

On suppose dorénavant que le joueur 1 ne peut faire des prédictions que dans une grille donnée $\mathcal{L} = \{\mu(l), l \in L\}$. Une stratégie est donc une application de l'ensemble des histoires dans $\Delta(L)$.

La distribution empirique des états sur $N_n(l) = \sum_{m=1}^{n} \mathbb{1}\{\mu_m = \mu(l)\}$, l'ensemble des étapes avant la n-ème où la prédiction est $\mu(l)$, est notée $\overline{s}_n(l)$.

Définition 3.11

Une stratégie σ du joueur 1 est calibrée par rapport à une grille $\mathcal{L}$, si pour tout $\mu(l), \mu(k) \in \mathcal{L}$ et toute stratégie τ de la nature :

$$\limsup_{n \to \infty} \frac{N_n(l)}{n} \left(\|\mu(l) - \overline{s}_n(l)\|^2 - \|\mu(k) - \overline{s}_n(l)\|^2 \right) \leq 0, \quad \mathbb{P}_{\sigma,\tau}\text{−ps.}$$

En d'autres termes, une stratégie σ est calibrée par rapport à $\mathcal{L}$ si, sur les dates où $\mu(l)$ est prédite, la distribution empirique des états est plus proche de $\mu(l)$ que de n'importe quel autre $\mu(k)$. Il est clair qu'une stratégie calibrée par rapport à une grille sera ε-calibrée pour ε assez petit.

Théorème 3.12 (*Foster et Vohra [43]*)

Pour toute grille $\mathcal{L}$, il existe une stratégie calibrée par rapport à $\mathcal{L}$. Il existe donc une stratégie ε-calibrée pour tout $\varepsilon > 0$ et une stratégie naïvement calibrée.

Leur démonstration repose sur l'exhibition d'un algorithme qui fait tendre le score de Brier vers zéro. En section 5.2, on introduira un raffinement de cette notion que l'on appelle calibration par rapport à un graphe.

Presque-calibration

On rappelle qu'un algorithme déterministe ne peut être utilisé pour construire une stratégie naïvement calibrée, il est nécessaire qu'il soit aléatoire.

Il est par contre possible d'obtenir une propriété très proche de l'ε-calibration, appelée *presque-calibration*. En effet, un algorithme de Kakade et Foster [63] (ainsi que celui de Vovk, Nouretdinov, Takemura et Shafer [118]) assure que pour toute fonction lipschitzienne ω de $\Delta(S)$ dans $[0,1]$:

$$\frac{1}{n}\sum_{m=1}^{n} \omega(\mu_m)\,(s_m - \mu_m) \to_{n\to\infty} 0. \tag{3.5}$$

La propriété (3.5) et voisine de la propriété (3.2) puisqu'il est possible d'approcher $\mathbb{1}_{p,\varepsilon}$ d'aussi près que l'on veut par des fonctions lipschitziennes (c'est d'ailleurs la raison pour laquelle, dans l'équation (3.2), on ne divise pas par $\sum_{m\leq n} \mathbb{1}_{p,\varepsilon}(\mu_m)$ mais par n). Le principal intérêt de ce résultat est qu'il est possible de construire une théorie aléatoire ε-calibrée (*i.e.* la suite dans l'équation (3.2), au lieu de converger vers 0, est bornée asymptotiquement par $\varepsilon > 0$ presque sûrement) à partir d'une théorie déterministe presque calibrée.

En reprenant la construction de Foster et Kakade [63], on considère une triangulation finie $\{K \in \mathcal{K}\}$ de $\Delta(S)$ telle que chaque simplexe de la triangulation est de diamètre plus petit que ε. Pour tout $p \in \Delta(S)$, on note $K(p) \in \mathcal{V}$ le simplexe qui contient p (s'il appartient à plusieurs simplexes, on en choisit un arbitrairement) et $V(p)$ l'ensemble des sommets de $K(p)$. Avec ces notations, tout point p de $\Delta(S)$ s'écrit de manière unique $p = \sum_{v\in V(p)} \omega_v(p)$ où $\omega_v(\cdot)$, définie sur $K(p)$, est à valeur dans $[0,1]$ et est étendue aux simplexes ne contenant pas p par $\omega_v = 0$. Ainsi pour tout sommet v de la triangulation, la fonction ω_v est lipschitzienne, vaut 1 en v et s'annule sur tous les simplexes qui ne contiennent pas v ; en particulier, on peut la voir comme une approximation de $\mathbb{1}_{v,\varepsilon}$.

Ces fonctions ω_v permettent de construire $\widetilde{\mu}$ une théorie aléatoire calibrée à partir de la théorie déterministe μ donnée par l'algorithme déterministe presque-calibré, de la façon suivante. Au lieu de prédire de manière déterministe μ_n, le prédicteur va choisir aléatoirement $\widetilde{\mu}_n = v$, un des sommets du simplexe contenant μ_n, avec probabilité $\omega_v(\mu_n)$. Ainsi, $\widetilde{\mu}_n$ est ε-proche de μ_n et son espérance est exactement égale à μ_n. Cette méthode est appelée *arrondissement aléatoire* car, par exemple, au

lieu de prédire 0.529 il suffit de prédire 0.5 avec probabilité 0.65 et 0.6 avec probabilité 0.34.

En conclusion, il n'est pas possible de construire une stratégie ε-calibrée en faisant des prédictions déterministes. Par contre, il suffit de faire des perturbations aléatoires — arbitrairement petites — pour pouvoir y arriver.

4

Non-Regret

On rappelle les notions de regret (ou consistance) externe ainsi que les différents raffinements, proposés notamment par Foster et Vohra [42], Fudenberg et Levine [51], et ainsi que Lehrer [70], etc. Les liens avec les concepts d'équilibres de jeux sont étudiés et ce chapitre se clôt avec des exemples non usuels d'utilisation des notions de regret.

Sommaire

4.1 Non-regret externe **40**
Prédictions avec conseils d'experts 41
Regret externe et jeux 42
4.2 Non-regret interne **43**
Équilibres corrélés 45
4.3 Extensions **45**
Du regret externe au regret *swap* 45
Notions plus fines de regret 48
Approches continues 49
Autres utilisations du regret 51

Hannan [56] a introduit la notion de regret externe dans les jeux finis à deux joueurs afin de fournir un critère d'évaluation de stratégies dans un cadre non-Bayésien. Formellement, un joueur n'a pas de regret externe (on dit aussi que sa stratégie est consistante extérieurement) si, asymptotiquement, il n'aurait pas pu gagner strictement plus s'il avait connu — avant le commencement du jeu — la distribution empirique des actions de son adversaire. Cette notion a été raffinée par la consistence interne une première fois par Foster et Vohra [42] (ainsi que Fudenberg et Levine [51]) : un joueur n'a pas de regret interne si, pour chacune de ses actions, il n'a pas de regret externe sur l'ensemble des dates où il l'a jouée. Lehrer [70] et Fudenberg et Levine [51] ont défini un second raffinement du regret en contraignant

le joueur 1 à ne pas avoir de regret externe sur des sous-ensembles d'étapes dépendant plus finement des actions jouées.

La notion de regret fut aussi largement utilisée dans d'autres domaines des jeux répétés, notamment par Foster et Young [46] (voir aussi Germano et Lugosi [52]) afin de construire des procédures qui convergent vers des équilibres de Nash et par Halpern et Pass [55] afin de définir un nouveau concept d'étude des jeux finis.

4.1 Non-regret externe

Considérons un jeu répété à deux joueurs Γ_e où, à l'étape $n \in \mathbb{N}$, le joueur 1 (resp. le joueur 2) choisit l'action $i_n \in I$ (resp. $j_n \in J$) où I et J sont finis. Ces choix génèrent un paiement $\rho_n = \rho(i_n, j_n) \in \mathbb{R}$ où ρ est une fonction (à valeur réelle) de $I \times J$ dans $\mathbb{R}$, étendue sur $\Delta(I) \times \Delta(J)$ par $\rho(x, y) = \mathbb{E}_{x,y}[\rho(i, j)]$. Aucune hypothèse n'est faite sur les paiements ni les objectifs du joueur 2.

De manière usuelle, on appelle $H_n = (I \times J)^n$ l'ensemble des histoires finies de taille n et une stratégie σ du joueur 1 est une fonction de l'ensemble des histoires finies $H = \bigcup_{n\in\mathbb{N}} H_n$ dans $\Delta(I)$, l'ensemble des probabilités sur I. Un couple de stratégies (σ, τ), avec τ définie de manière similaire, induit une probabilité $\mathbb{P}_{\sigma,\tau}$ sur l'ensemble des parties $\mathcal{H} = (I \times J)^\infty$ munie de la tribu produit.

Les choix de i_n et j_n définissent également $r_n \in \mathbb{R}^I$, le vecteur de regret externe instantané de l'étape n, donné par :

$$r_n = r(i_n, j_n) := (\rho(1, j_n) - \rho(i_n, j_n), \ldots, \rho(I, j_n) - \rho(i_n, j_n)) \in \mathbb{R}^I.$$

Intuitivement, le regret r_n représente la différence entre ce que le joueur 1 aurait pu obtenir en choisissant une autre action et ce qu'il a effectivement obtenu. Hannan demande à une stratégie que chaque composante de la moyenne des regrets soit asymptotiquement négative. Le cas échéant, le joueur ne pourra ainsi pas se dire "*si j'avais su* [la moyenne empirique des actions de l'autre joueur] *j'aurais tout le temps joué l'action i*". En effet, par linéarité de ρ, la moyenne des n premiers regrets instantanés, appelée regret externe à l'étape n et notée $\overline{r}_n$, vérifie :

$$\overline{r}_n = \big(\rho(1, \overline{j}_n) - \overline{\rho}_n, \ldots, \rho(I, \overline{j}_n) - \overline{\rho}_n\big) \in \mathbb{R}^I.$$

Définition 4.1

Une stratégie σ du joueur 1 est consistante extérieurement (ou n'a pas de regret externe) si pour toute stratégie τ du joueur 2 :

$$\limsup_{n\to\infty} \overline{r}_n \leq 0, \quad \mathbb{P}_{\sigma,\tau}\text{−ps} \tag{4.1}$$

où l'inégalité doit être comprise composante par composante.

Une autre formulation évidemment équivalente à (4.1) est que

$$\limsup_{n\to\infty} \max_{i\in I} \rho(i, \bar{\jmath}_n) - \overline{\rho}_n \leq 0 \quad \mathbb{P}_{\sigma,\tau}\text{−ps.}$$

Théorème 4.2 (*Hannan [56]*)

Il existe une stratégie σ consistante extérieurement telle que pour toute stratégie τ du joueur 2 :

$$\mathbb{E}_{\sigma,\tau}\left[\left\|(\overline{r}_n)^+\right\|\right] = O\left(\frac{1}{\sqrt{n}}\right).$$

Dans cette définition, U^+ dénote la partie positive du vecteur $U \in \mathbb{R}^I$: sa i-ème composante est $(U^+)^i = \max(0, U^i)$.

Prédictions avec conseils d'experts

Une interprétation — mais qui est aussi une généralisation — des résultats précédents concerne les jeux de prédictions avec conseils d'experts (étudiés en détail par Cesa-Bianchi et Lugosi [29]). À chaque étape $n \in \mathbb{N}$, un agent doit prendre une décision ω_n dans un ensemble topologique compact convexe Ω et il est conseillé par I experts, *i.e.* l'expert i lui propose de choisir l'action ω_n^i. Une fois son choix fait, la nature révèle l'état du monde s_n, ce qui entraîne une perte $L_n = L(\omega_n, s_n)$.

Après n étapes, l'agent a subi une perte moyenne de $\overline{L}_n = \frac{1}{n}\sum_{m\leq n} L(\omega_m, s_m)$ tandis que le meilleur expert a, quant à lui, subi la plus petite perte moyenne égale à $\bar{L}_n^\star = \frac{1}{n}\min_{i\in I}\sum_{m\leq n} L(\omega_m^i, s_m)$. Un critère d'évaluation d'une stratégie sera donc de comparer ces deux pertes moyennes. Le résultat de Hannan — correspondant au cas où Ω est un I-simplexe et $L(\cdot, s)$ est linéaire sur Ω — peut être généralisé :

Théorème 4.3 (*Auer, Cesa-Bianchi et Gentile [7]*)

Si L est convexe et à valeurs dans $[0,1]$, il existe un algorithme tel que :

$$\overline{L}_n - \bar{L}_n^\star \leq 2\sqrt{\frac{1}{2n}\ln(I)} + \frac{1}{n}\sqrt{\frac{\ln(I)}{8}}.$$

Par exemple, l'algorithme (défini en section 4.3) d'Auer, Cesa-Bianchi, Freund et Schapire [6] appelé *exponential weight algorithm* avec un paramètre qui décroît avec le temps vérifie ce résultat.

Pour une certaine classe de fonctions de perte régulières (mais qui ne contient pas les fonctions bilinéaires), appelées *mixables*, Vovk [116] a montré que la différence entre la perte moyenne de l'agent et celle du meilleur expert décroît encore plus rapidement en $a\frac{\ln(I)}{n}$, avec une caractérisation explicite de la constante $a \in \mathbb{R}$.

Regret externe et jeux

L'existence de stratégies sans regret externe peut être utilisée pour démontrer deux résultats classiques en théorie des jeux : la non-vacuité de l'ensemble de Hannan [56] d'un jeu fini et le théorème de min-max de Von Neumann [115] (en généralisant les résultats de Blum et Mansour [22] pour le cas des fonctions linéaires et de Cesa-Bianchi et Lugosi [29] pour le cas des fonctions concaves-convexes).

Soit G le jeu à L joueurs où l'ensemble fini d'actions du joueur l est noté I_l et sa fonction de paiement est $\rho_l : \pi_{l\in L} I_l \to \mathbb{R}$. L'ensemble de Hannan du joueur 1 est le sous-ensemble de $\Delta(\Pi_{l\in L} I_l)$ défini par :

$$\begin{aligned} H_1 &= \left\{ z \in \Delta(\Pi_{l\in L} I_l); \rho_1(i, z^{-1}) \leq \rho(z), \forall i \in I_1 \right\} \\ &= \left\{ z \in \Delta(\Pi_{l\in L} I_l); \left(\rho(1, z^{-1}) - \rho(z), \ldots, \rho(I_1, z^{-1}) - \rho(z)\right) \in \mathbb{R}_-^{I_1} \right\} \end{aligned}$$

où $\rho(z) = \mathbb{E}_z[\rho(i_1, \ldots, i_L)]$ et $\rho(i, z^{-1}) = \mathbb{E}_{z^{-1}}[\rho(i, i_2, \ldots, i_L)]$ avec z^{-1} la première marginale de z. Une distribution z sur l'ensemble des profils d'actions est donc dans l'ensemble H^1 si le joueur 1 — en supposant le comportement de l'ensemble des autres joueurs fixé — n'a pas intérêt à dévier et à toujours jouer une même action.

Par définition de H_1 et par linéarité du paiement, si la stratégie du joueur 1 est consistante extérieurement, alors la distribution empirique des profils d'actions converge vers H_1. Cette propriété est qualifiée d'**unilatérale** car elle ne suppose rien sur les stratégies des autres joueurs, qui peuvent être consistantes extérieurement ou non. Si l'on définit $H_2, \ldots, H_L$ de manière similaire et que l'on suppose que tous les joueurs utilisent une stratégie consistante extérieurement de manière unilatérale (*i.e.* on ne fait aucune hypothèse sur la procédure jointe ; on peut aussi dire que chaque joueur choisit l'algorithme consistant qu'il veut) alors la distribution empirique des profils d'actions converge vers l'ensemble de Hannan $\overline{H} = \cap_{l\in L} H_l$ qui est donc non-vide.

Dans les jeux à somme nulle, $z \in \Delta(I \times J)$ appartient à l'ensemble de Hannan $\overline{H}$ si et seulement si :

$$\min_{z^2 \in \Delta(J)} \max_{i\in I} \rho(i, z^2) \leq \max_{i\in I} \rho(i, z^{-1}) \leq \rho(z) \leq \min_{j\in J} \rho(z^{-2}, j) \leq \max_{z^1 \in \Delta(I)} \min_{j\in J} \rho(z^1, j).$$

Ainsi $\overline{H}$ est égal à l'ensemble des équilibres de Nash et tout jeu fini à somme nulle admet une valeur, égale à $\rho(z)$.

Pour le cas particulier des jeux de potentiel à deux joueurs, Hart et Mas-Colell [59] ont construit des stratégies particulières sans regret externe (voir section 5.1) telles que le produit des distributions empiriques des actions des joueurs converge vers l'ensemble des équilibres de Nash, et plus précisément vers un sous-ensemble des équilibres dont les paiements sont égaux.

Cependant, alors que la convergence vers l'ensemble des équilibres de Nash dans les jeux à somme nulle est une propriété unilatérale, il s'agit dans ce cas d'une propriété

globale : le résultat n'est vrai que pour la procédure mise en place, qui indique simultanément à chacun des joueurs comment jouer.

Par ailleurs, ce résultat n'est a priori pas valable pour n'importe quelle stratégie sans regret externe et Hart et Mas-Colell [60] ont aussi montré qu'il ne s'étend pas à tous les jeux (même ceux qui ne possèdent qu'un unique équilibre de Nash).

Revenons aux jeux à somme nulle. Le théorème ci-dessous, dû à Fan [36], est une généralisation du théorème du minmax de Von Neumann [115].

On rappelle qu'une fonction continue sur $X \times Y$ est dite *concave-like convexe-like* si pour tous $x^1, x^2 \in X$ (resp. $y^1, y^2 \in Y$) et $\alpha \in [0,1]$, il existe $x^0 \in X$ (resp. $y^0 \in Y$) tel que $\rho(x^0, \cdot) \geq \alpha\rho(x^1, \cdot) + (1-\alpha)\rho(x^2, \cdot)$ (resp. $\rho(\cdot, y^0) \leq \alpha\rho(\cdot, y^1) + (1-\alpha)\rho(\cdot, y^2)$)

Théorème 4.4

Soient X et Y deux ensembles compacts et ρ une fonction concave-like convexe-like sur $X \times Y$. Alors le jeu sur $X \times Y$ a une valeur v.

La valeur est définie de manière usuelle par :

$$v = \max_{x \in X} \min_{y \in Y} \rho(x,y) = \min_{y \in Y} \max_{x \in X} \rho(x,y).$$

Démonstration . Soient $\varepsilon > 0$ et $\delta = \omega(\varepsilon)$ où $\omega(\cdot)$ est le module de continuité de la fonction ρ qui est continue sur le compact X. On considère le jeu auxiliaire où les espaces d'actions pures des deux joueurs sont respectivement $\mathcal{X}$ et $\mathcal{Y}$, deux δ-grilles de X et Y et l'on suppose que les deux joueurs ont des stratégies consistantes extérieurement de sorte que :

$$\max_{x \in \mathcal{X}} \rho(x, \bar{y}_n^0) \leq \max_{x \in \mathcal{X}} \rho(x, \overline{y}_n) \leq \overline{\rho}_n \leq \min_{y \in \mathcal{Y}} \rho(\overline{x}_n, y) \leq \min_{y \in \mathcal{Y}} \rho(\bar{x}_n^0, y)$$

où $\bar{y}_n^0$ et $\bar{x}_n^0$ sont donnés par la définition de concave-like, convexe-like. Cette série d'inégalité, avec la définition de $\mathcal{X}$ et ω impliquent que pour tout $\varepsilon > 0$,

$$\min_{y \in Y} \max_{x \in X} \rho(x,y) - \varepsilon \leq \max_{x \in X} \min_{y \in Y} \rho(x,y) + \varepsilon.$$

Comme le $\max_{y \in Y} \min_{x \in X} \rho(x,y)$ est toujours plus petit que le $\min_{x \in X} \max_{y \in Y} \rho(x,y)$, ce jeu a une valeur. □

4.2 Non-regret interne

Le regret interne est un raffinement du regret externe introduit par Foster et Vohra [42] (ainsi que Fudenberg et Levine [51], définition 8.3, en tant que cas particulier de la consistance universelle conditionnelle). À l'étape $n \in \mathbb{N}$, les choix de i_n et j_n génèrent en plus du paiement $\rho_n \in \mathbb{R}$, et du regret externe $r_n = r(i_n, j_n) \in \mathbb{R}^I$,

la matrice carrée (de taille $I \times I$) du regret interne instantané $R_n := R(i_n, j_n)$ dont la (i,k)-ème coordonnée est :

$$R_n^{i,k} = \begin{cases} \rho(k, j_n) - \rho(i_n, j_n) & \text{si } i = i_n \\ 0 & \text{sinon.} \end{cases}$$

Autrement dit, R_n est une matrice dont toutes les lignes sont nulles sauf la i_n-ème qui est r_n.

Définition 4.5

Une stratégie σ du joueur 1 est consistante intérieurement si pour toute stratégie τ du joueur 2

$$\limsup_{n\to\infty} \overline{R}_n \leq 0, \mathbb{P}_{\sigma,\tau}\text{−ps.} \tag{4.2}$$

Là encore, l'inégalité doit être comprise composante par composante. Cette définition équivaut au fait que $\overline{R}_n$ converge vers $\mathbb{R}_-^c$ (avec $c = |I \times I|$), ou à ce que pour tout $i \in I$:

$$\limsup_{n\to\infty} \frac{|N_n(i)|}{n} \left(\overline{\rho}_n(i) - \max_{k\in K} \rho(k, \overline{\jmath}_n(i)) \right) \leq 0,$$

ce qui veut dire que le joueur 1 n'a pas de regret externe sur $N_n(i) = \{m \leq n, i_m = i\}$, l'ensemble des dates où il a joué l'action i, dès que la densité supérieure de ce dernier est strictement positive. On parlera de *ε-consistance interne* si la condition (4.2) n'est vérifiée qu'à ε près, *i.e.* si

$$\limsup_{n\to\infty} \overline{R}_n \leq \varepsilon, \mathbb{P}_{\sigma,\tau}\text{−ps.}$$

Le regret interne est un concept plus fin que le regret externe car quelles que soient les stratégies des deux joueurs $\left\|(\overline{r}_n)^+\right\|_\infty \leq I \left\|\left(\overline{R}_n\right)^+\right\|_\infty$ pour tout $n \in \mathbb{N}$.

Théorème 4.6 (*Foster et Vohra [42]*)

Il existe une stratégie consistante intérieurement telle que que pour toute stratégie τ du joueur 2 :

$$\mathbb{E}_{\sigma,\tau}\left[\left\|\left(\overline{R}_n\right)^+\right\|\right] = O\left(\frac{1}{\sqrt{n}}\right).$$

La démonstration de ce résultat (il s'agit d'une adaptation par Sorin [108] de celle de Foster et Vohra [42]) est donnée en section 5.1. Il est également possible de définir une stratégie consistante comme une stratégie satisfaisant :

$$\limsup_{n\to\infty} \overline{\rho}_n(i) - \max_{k\in K} \rho(k, \overline{\jmath}_n(i)) \leq 0, \text{pour tous les } i \text{ tels que } |N_n(i)| \to_{n\to\infty} \infty.$$

Cette formulation ne sera pas utilisée pour les raisons suivantes : il n'existe pas — à ma connaissance — de constructions explicites de ces stratégies, les vitesses de convergence ne sont pas connues et surtout à l'étape n le joueur 1 ne peut pas encore évaluer avec certitude son regret car, a priori, il ne sait pas encore si l'action i sera utilisée une infinité de fois.

Équilibres corrélés

Dans un jeu répété à L joueurs avec espaces d'actions finis, on a vu dans la section 4.1 que si chaque joueur utilise une stratégie extérieurement consistante, alors la distribution empirique des profils d'actions converge vers l'ensemble de Hannan. Un résultat similaire existe dans le cas où les joueurs n'ont pas de regret interne. On rappelle la définition d'un équilibre corrélé, notion introduite par Aumann [8] :

Une distribution $z \in \Delta(\Pi_{l \in L} I_l)$ est un équilibre corrélé si pour tout joueur $l \in L$ et toute action $i \in I_l$:

$$\rho_l(k, z^{-l}(i)) - \rho_l(i, z^{-l}(i)) \leq 0, \quad \forall k \in I$$

où $z^{-l}(i) = z(\cdot | i_l = i)$ est la probabilité sur $\Pi_{l' \neq l} I_{l'}$ induite par z sachant que $i_l = i$.

Si les joueurs utilisent des stratégies consistantes intérieurement, la distribution empirique des actions converge vers l'ensemble des équilibres corrélés. Il s'agit là encore d'une propriété unilatérale et la convergence a lieu vers l'ensemble des équilibres corrélés et non pas vers un équilibre corrélé particulier.

4.3 Extensions

Du regret externe au regret *swap*

Stoltz et Lugosi [111] ont construit un algorithme consistant intérieurement à partir d'un algorithme consistant extérieurement (et en utilisant ensuite un argument de point fixe).

Soit $\Gamma^{\rightarrow}$ le jeu dans lequel l'espace d'actions du joueur 1 est $\left\{\xi^{i \to k}, i, k \in I\right\}$ et celui du joueur 2 est $\mathcal{U}$, où $\mathcal{U}$ est un ensemble compact de $\mathbb{R}^I$. On suppose qu'il existe une suite exogène $p_n \in \Delta(I)$ telle que les choix à l'étape n de $\xi_n^{i \to k}$ et de $U_n \in \Delta(I) \times \mathcal{U}$ induisent le paiement $V_n^{i \to k} \in \mathbb{R}$ défini par :

$$V_n^{i \to k} = \begin{cases} \sum_{j \neq \{i,k\}} p_n(j) U_n^j + (p_n(i) + p_n(k))\, U_n^k := \langle p_n^{i \to k}, U_n \rangle & \text{si } i \neq k \\ \sum_{i \in I} p_n(i) U_n^i = \langle p_n, U_n \rangle & \text{sinon.} \end{cases}$$

L'intuition est la suivante : l'action $\xi^{i \to k}$ représente dans $\Gamma^{\rightarrow}$ le fait de jouer l'action k à la place de l'action i dans le jeu Γ. En effet, si dans ce dernier à l'étape n les choix des joueurs sont $p_n \in \Delta(I)$ et $U_n \in (U)$, le paiement espéré est $\langle p_n, U_n \rangle$; par contre, si à chaque fois qu'il devrait jouer l'action i le joueur 1 joue l'action k, son paiement espéré est $\langle p_n^{i \to k}, U_n \rangle$.

On appelle $\overline{r}_n$ le regret externe moyen (jusqu'à l'étape n) de la stratégie Θ du joueur 1 dans $\Gamma^{\rightarrow}$ et on note $\Theta_n^{i\rightarrow k} = \Theta[\xi^{i\rightarrow k}|h^{n-1}]$ le poids donné par Θ à l'action $\xi^{i\rightarrow k}$ à l'étape n. Alors, pour tout $i_0, k_0 \in I$,

$$\mathbb{E}_\Theta\left[\bar{r}_n^{i_0,k_0}\right] = \frac{\sum_{m=1}^n \langle p_m^{i_0\rightarrow k_0}, U_m\rangle}{n} - \frac{\sum_{m=1}^n \langle \sum_{i,k\in I} \Theta_m^{i\rightarrow k} p_m^{i\rightarrow k}, U_m\rangle}{n}.$$

Dans le jeu Γ, on suppose que la stratégie σ est définie par $\sigma(h^{n-1}) = p_n$ et on appelle $\overline{R}_n$ le regret interne moyen (jusqu'à l'étape n) de σ. Alors, pour tout $i_0, k_0 \in I$

$$\mathbb{E}_\sigma\left[\bar{R}_n^{i_0,k_0}\right] = \sum_{m=1}^n p_m^{i_0}\left(U_m^{k_0} - U_m^{i_0}\right) = \frac{\sum_{m=1}^n \langle p_m^{i_0\rightarrow k_0}, U_m\rangle}{n} - \frac{\sum_{m=1}^n \langle p_m, U_m\rangle}{n}.$$

Ainsi, le choix de $p_n = \sum_{i,k\in I} \Theta_n^{i\rightarrow k} p_n^{i\rightarrow k}$ implique que $\mathbb{E}_\Theta[\overline{r}_n] = \mathbb{E}_\sigma\left[\overline{R}_n\right]$; une stratégie Θ consistante extérieurement dans $\Gamma^{\rightarrow}$ induit donc une stratégie σ consistante intérieurement dans Γ. L'existence d'un tel point fixe p_n est assurée par le théorème de Brouwer.

On obtient la convergence presque sûre du regret interne vers 0 à partir de la convergence en espérance en remarquant que $\overline{R}_n - \mathbb{E}_\sigma\left[\overline{R}_n\right]$ est une moyenne de différences bornées de martingales. Il suffit ensuite d'appliquer le lemme suivant (voir Hall et Heyde [54], théorème 2.7 p. 41 et exemple 1 p. 19).

Lemme 4.7

Soit $\{X_n\}_{n\in\mathbb{N}}$ une suite de différences bornées de martingales adaptée à la filtration $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, i.e. il existe $B > 0$ tel que, pour tout $n \in \mathbb{N}$, $|X_n| \leq B$ presque sûrement et $\mathbb{E}\left[X_n\middle|\mathcal{F}_{n-1}\right] = 0$.

Alors $\frac{1}{n}\sum_{m=1} X_m$ converge presque sûrement vers 0.

Blum et Mansour [22] ont eux combiné un nombre fini d'algorithmes consistants extérieurement afin de construire un algorithme n'ayant pas de regret *swap*, qui est une notion plus fine que le regret interne :

Definition 4.8

La stratégie σ du joueur 1 n'a pas de regret ϕ-swap (avec ϕ une application de I dans I dite d'échange), si pour toute stratégie τ du joueur 2 :

$$\limsup_{n\rightarrow\infty} \overline{R}_n(\phi) := \frac{\sum_{m=1}^n \rho\left(\phi(i_m), j_m\right) - \rho(i_m, j_m)}{n} \leq 0, \mathbb{P}_{\sigma,\tau}\text{−ps.}$$

Si la stratégie σ n'a pas de regret ϕ-swap pour toute les fonctions d'un ensemble Φ, alors on dit que cette stratégie n'a pas de regret Φ-swap ; si Φ est l'ensemble de toutes les fonctions d'échanges, alors elle n'a pas de regret swap.

Cette définition est plus fine que les notions de regrets externe et interne car :

1. une stratégie est consistante extérieurement si elle est Φ_e-*swap* consistante, avec $\Phi_e = \{\phi_k, \forall k \in I : \phi_k(i) = k, \forall i \in I\}$;
2. une stratégie est consistante intérieurement si elle est Φ_i-*swap* consistante, avec $\Phi_i = \{\phi_{i,k}, \forall i, k \in I : \phi_{i,k}(i) = k \text{ et } \phi_{i,k}(l) = l, \text{ si } l \neq i\}$;

De la même façon que le regret interne est plus fin que le regret externe, si l'on note $\mathbf{\Phi} = \{\phi : I \to I\}$ alors

$$\sup_{i,k \in I} \overline{R}_n^{ik} \leq \sup_{\phi \in \mathbf{\Phi}} \overline{R}_n(\phi) \leq I \sup_{i,k \in I} \overline{R}_n^{i,k}.$$

Revenons à l'algorithme Θ de Blum et Mansour [22]. Il est basé sur l'exécution parallèle de I sous-algorithmes $\{\theta^i\}_{i \in I}$ consistants extérieurement (et sur un argument de point fixe). À l'étape $n+1$, étant données les probabilités $q_{n+1}^i \in \Delta(I)$ spécifiées par ces sous-algorithmes θ^i, l'algorithme général Θ spécifie une probabilité p_{n+1} définie comme un point fixe par

$$p_{n+1} := (p_{n+1}(1), \ldots, p_{n+1}(I)) = \sum_{i \in I} p_{n+1}(i) q_{n+1}^i.$$

Le choix j_{n+1} du joueur 2 génère le vecteur de paiements $U_{n+1} = (\rho(i, j_n))_{i \in I} \in \mathbb{R}^I$; cependant, le sous-algorithme θ^i est appliqué, quant à lui, à la suite de vecteurs $p_n(i) U_{n+1}$. Par définition de la consistance extérieure de θ^i, pour tout $k \in I$:

$$\frac{\sum_{m=1}^n p_m(i) U_m^k}{n} - \frac{\langle q_m^i, p_m(i) U_m \rangle}{n} \leq o(1). \tag{4.3}$$

La somme, sur l'ensemble des sous-algorithmes des seconds termes est égale à

$$\frac{\sum_{m=1}^n \langle \sum_{i \in I} q_m^i p_m(i), U_m \rangle}{n} = \frac{\sum_{m=1}^n \langle p_m, U_m \rangle}{n},$$

qui est la moyenne des gains espérés de l'algorithme général Θ. Soit $\phi : I \to I$ une fonction d'échange quelconque, alors la somme sur I des équations (4.3), prises en $k = \phi(i)$, donne :

$$\mathbb{E}\left[\overline{R}_n(\phi)\right] = \frac{\sum_{m=1}^n \sum_{i \in I} p_m(i) U_m^{\phi(i)}}{n} - \frac{\sum_{m=1}^n \langle p_m, U_m \rangle}{n} \leq o(1).$$

L'algorithme général Θ n' donc pas de regret swap espéré. La convergence presque sûre s'obtient encore grâce au lemme 4.7.

L'introduction du regret *swap* permet aussi de généraliser la définition du regret interne lorsque l'ensemble d'actions I n'est pas fini. En effet dans ce cadre il est possible de construire des stratégies trivialement consistantes intérieurement (au sens

4.5) en n'utilisant qu'une seule fois chaque action ; chacune des suites de fréquences $(N_n(i)/n)_{n\in\mathbb{N}}$ convergeant évidemment vers 0.

Dans le cas où I est un ensemble compact convexe normé et ρ concave, Stoltz et Lugosi [112] ont obtenu l'existence de stratégies déterministes Φ_c-*swap* consistantes avec Φ_c l'ensemble des fonctions continues de I dans I. On montre en section 5.1 une généralisation de ce résultat :

Théorème 4.9

Si I est un ensemble métrique convexe et compact et $\rho(\cdot, y)$ est continue pour tout $y \in \Delta(J)$ et uniformément bornée, alors il existe une stratégie Φ_c-swap consistante.

Notions plus fines de regret

Fudenberg et Levine [51] puis Lehrer [70] ont également généralisé la notion de regret interne, dans deux directions. La première consiste à définir des ensembles plus fins sur lesquels le regret est calculé et la seconde revient à calculer ce dernier pas nécessairement par rapport à une stratégie constante. Formellement, une fonction d'activation A définie sur $H \times I$ et à valeur dans $\{0, 1\}$ indique si après l'histoire h^n et étant donné le choix i_{n+1} du joueur 1, l'étape $n+1$ est active (*i.e.* si $A(h^n, i_{n+1}) = 1$ alors cette étape va compter dans le calcul du regret). Une fonction de remplacement ϕ est une application de $H \times I$ dans I qui, quant à elle, indique après l'histoire h^n quelle action jouer à la place de i_{n+1}.

On dit qu'une stratégie σ n'a pas de (A, ϕ)-regret si pour toute stratégie τ du joueur 2 :

$$\limsup_{n\to\infty} \frac{\sum_{m=1}^{n} A(h_m, i_{m+1}) \left[\rho\big(\phi(h_m, i_{m+1})\big) - \rho_n\right]}{\sum_{m=1}^{n} A(h_m, i_{m+1})} \leq 0, \quad \mathbb{P}_{\sigma,\tau}\text{−ps} \tag{4.4}$$

dès que $\sum_{m=0}^{n} A(h_m, i_{m+1}) \to \infty$. Étant donnée une probabilité λ sur l'ensemble des couples fonctions d'activations/fonctions de remplacement (muni de la tribu produit), Lehrer [70] a montré l'existence d'une stratégie sans (A, ϕ)-regret, pour λ-presque tous les couples. En particulier, pour un ensemble dénombrable de couples, il existe une stratégie qui n'a pas de regret vis-à-vis de tous les couples.

Il s'agit bien d'une généralisation des notions précédentes en considérant les ensembles suivants :

Regret externe : $\mathcal{E}$ est l'ensemble des couples $(\mathbb{1}, i^\star)$ où $\mathbb{1}$ est la fonction toujours active et $i^\star$ la fonction de remplacement toujours égale à i ;

Regret interne : $\mathcal{I}$ est l'ensemble de cardinal I^2 des couples $(\mathbb{1}_i, k^\star)$ avec $\mathbb{1}_i(h_n, i_{n+1}) = 1$ lorsque $i_{n+1} = i$;

Regret swap : $\mathcal{S}$ est l'ensemble de cardinal I^I des couples $(\mathbb{1}, \phi^\star)$ tels qu'il existe une fonction ϕ de I dans I avec $\phi^\star(h_n, \cdot) = \phi(\cdot)$;

Consistence Universelle Conditionnelle : étant donnée $B_1, \ldots, B_K$ une partition finie de $\mathbb{N}$, $\mathcal{CUD}$ est l'ensemble de cardinal KI des couples $(\mathbb{1}_{B_k}, i^\star)$ où $\mathbb{1}_{B_k}(h_n, i_{n+1}) = 1$ si l'étape $n+1$ appartient à B_k.

Deux raisons expliquent le choix de ne pas multiplier par la fréquence de la fonction d'activation (*i.e.* de ne pas diviser par n au lieu de $\sum_{m=1}^{n} A(h_m, i_{m+1})$) dans cette définition. Même si l'on divisait par n, les stratégies vérifiant la propriété (4.4) ne sont pas construites explicitement (voir la section 5.1). Par ailleurs, les ensembles $\mathcal{E}$, $\mathcal{I}$ et $\mathcal{S}$ sont finis, donc multiplier par la fréquence des fonctions d'activation n'est pas très important. Par contre, dans le cas général, il serait possible de construire des stratégies triviales telles que toutes les fréquences soit nulles.

Approches continues

Les résultats des deux prochaines extensions, attribués à leurs auteurs originaux, ont été largement étendus et réinterprétés (notamment en terme d'A.S.D., voir Benaim, Hofbauer et Sorin [15]) par Sorin [108].

Hart et Mas-Colell [58] ont prouvé l'existence de stratégies consistantes en se basant sur l'étude d'un potentiel P de $D := \mathbb{R}^I_-$. Une fonction P est un potentiel si elle est de classe $\mathcal{C}^1$, positive et de gradient positif (composante par composante), nulle exactement sur D et telle que $\langle \nabla P(x), x \rangle > 0$ pour tout $x \notin D$. La stratégie consiste à jouer à l'étape n selon la loi x_n qui est proportionnelle à $\nabla P(\overline{r}_n)$ (et de façon arbitraire si $\overline{r}_n$ est dans D).

En remarquant que $\langle x, \mathbb{E}_x[R_{n+1}] \rangle = 0$ quelle que soit j_{n+1} (voir section 5.1), on en déduit que la suite $\overline{r}_n$ est une approximation discrète stochastique (une A.S.D.) de l'inclusion différentielle associée $\dot{r} \in N(r) - r$ où l'on a défini

$$N(r) = \left\{ \omega \in \mathbb{R}^I ; \langle \nabla P(r), \omega \rangle = 0 \text{ et } \|\omega\| \leq \|\rho\|_\infty \right\}.$$

En conséquence,

$$\frac{d}{dt} P(r(t)) \in \langle P(r(t)), N(r(t)) \rangle - \langle P(r(t)), r(t) \rangle = -\langle P(r(t)), r(t) \rangle < 0.$$

La fonction $P(r(\cdot))$ est donc de Lyapounov, et $r(\cdot)$, ainsi que $\overline{r}_n$ (en tant qu'A.S.D.), convergent vers D. En conséquence, la stratégie de Hart et Mas-Colell est bien consistante extérieurement. Si l'on choisit comme potentiel la fonction $P(x) = \|x^+\|_2^2$, on obtient l'*algorithme à poids polynomiaux* (voir Cesa-Bianchi et Lugosi [29])

Cesa-Bianchi et Lugosi [28] ont utilisé cette technique afin de construire une stratégie consistante intérieurement : à l'étape n, il suffit de prendre pour x_n n'importe quelle mesure invariante de $\left(\overline{R}_n\right)^+$ (comme proposé par Foster et Vohra [44]). On rappelle que μ est une mesure invariante d'une matrice A, de taille $I \times I$ et à coefficients positifs A^{ik}, si pour tout $i \in I$, $\mu(i) \sum_{k \in I} A^{ik} = \sum_{k \in I} \mu(k) A^{ki}$. On peut alors vérifier (voir section 5.1) que $\langle A, \mathbb{E}_\mu[R_{n+1}] \rangle = 0$ pour n'importe quel choix de j_{n+1} et donc la suite $\overline{R}_n$ est une A.S.D. de l'inclusion différentielle $\dot{R} \in N(R) - R$ et est donc

consistante intérieurement. Ces deux résultats sont bien sûr à mettre en relation avec l'approchabilité en temps continu (voir section 2.3) et ils fournissent d'ailleurs une intuition des résultats du chapitre suivant.

Pour les mêmes raisons que dans le chapitre 2, ces résultats s'étendent immédiatement au cas où l'étape n a une durée $\tau(n)$, sous la condition usuelle que la suite $(\tau_n/\sum_{m=1}^{n}\tau_m)_{n\in\mathbb{N}}$ soit dans $l^2(\mathbb{N})$ mais pas dans $l^1(\mathbb{N})$.

Fudenberg et Levine [50], Proposition 4.5, ont étudié un processus appelé *smooth fictitious play* qui n'a pas de regret externe. On définit ρ^ε, une ε-perturbation de la fonction de paiement ρ par :

$$\rho^\varepsilon(x,y) = \rho(x,y) + \varepsilon\psi(x), \forall y \in \Delta(J).$$

Faisons le changement de variable $U = [\rho(i,y)]_{i\in I}$ qui est donc un vecteur de l'ensemble compact $\mathcal{U} = [-\|\rho\|_\infty, \|\rho\|_\infty]^I \subset \mathbb{R}^I$. La fonction ρ^ε se réécrit en

$$\rho^\varepsilon(x,U) = \langle x, U\rangle + \varepsilon\psi(x)$$

où $\psi : \Delta(I) \to \mathbb{R}$ est choisie de telle sorte que :

i) ψ soit une fonction $\mathcal{C}^1$ de norme infinie inférieure à 1 ;

ii) la fonction $\mathrm{BR}^\varepsilon : \mathcal{U} \to \Delta(I)$ définie par $\mathrm{BR}^\varepsilon(U) = \mathrm{argmax}_{x\in X}\langle x, U\rangle + \varepsilon\psi(x)$ soit univoque et continue ;

iii) pour tout $U \in \mathcal{U}$, $D_1\rho^\varepsilon(\cdot, U) = 0$ en $\mathrm{BR}^\varepsilon(U)$.

La stratégie associée à cette perturbation est définie par $\sigma^\varepsilon(h^n) = \mathrm{BR}^\varepsilon(\overline{U}_n)$.

Notons $W^\varepsilon(U) = \sup_{x\in X}\rho^\varepsilon(x,U) = \langle \mathrm{BR}^\varepsilon(U), U\rangle + \varepsilon\psi\left(\mathrm{BR}^\varepsilon(U)\right)$. Une stratégie σ est en particulier ε-consistante extérieurement si $\limsup W^\varepsilon(\overline{U}_n) - \overline{\rho}_n \leq 0$.

La dynamique continue associée à $(\overline{U}_n, \overline{\rho}_n)$ s'écrit :

$$(\dot{U}, \dot{\omega}) \in \{(V, \langle \mathrm{BR}^\varepsilon(U), V\rangle)\,; V \in \mathcal{U}\} - (U, \omega).$$

Si l'on note $q(t) = W^\varepsilon(U(t)) - \omega(t)$, en différentiant q, on obtient $\dot{q}(t) + q(t) \leq \varepsilon$ et donc $q(t) \leq \varepsilon + Me^{-t}$ pour une certaine constante M.

En conséquence, l'ensemble

$$\{(U,\omega) \in \mathbb{R}^I \times \mathbb{R}; \langle \mathrm{BR}^\varepsilon(U), U\rangle + \varepsilon\,\mathrm{BR}^\varepsilon(U) - \omega \leq \varepsilon\}$$

est un attracteur global (voir par exemple Benaïm, Hofbauer et Sorin [15]).

On a ainsi montré l'existence, pour tout $\eta > 0$, de stratégies η-consistantes ; plus précisément, il existe $\underline{\varepsilon}$ tel que σ^ε soit η-consistante pour tout $\varepsilon < \underline{\varepsilon}$.

Le *smooth fictitious play* est une généralisation de deux classes d'algorithmes, appelés *exponential weight algorithm* et *follow the perturbed leader* (voir Cesa-Bianchi et Lugosi [29] sections 4.2 et 4.3). En effet, pour la première classe, il suffit de choisir l'entropie comme pénalisation, *i.e.* $\psi(x) = -\sum_{k\in I} x(k)\ln(x(k))$, pour avoir

$$\mathrm{BR}^\varepsilon(U)^i = \frac{\exp(U^i/\varepsilon)}{\sum_{k\in I}\exp(U^k/\varepsilon)}$$

ce qui caractérise bien l'*exponential weight algorithm*.

Le lien avec l'algorithme *Follow the Perturbed Leader* (qui suit le processus appelé *Stochastic Fictitious Play* de Fudenberg et Kreps [48]) est plus complexe. Cet algorithme ne choisit pas une pénalisation $\varepsilon\psi$ déterministe, mais perturbe chaque composante de $\overline{U}_n$ par une variable aléatoire ε_n^i, où la densité jointe $f : \mathbb{R}^n \to \mathbb{R}$ du vecteur $(\varepsilon_n^i)_{i\in I}$, est indépendante de $\overline{U}_n$ et de l'étape. Le choix de i_{n+1} est celui de la composante qui maximise $\left(\overline{U}_n\right)^i + \varepsilon_n^i$. En particulier, l'action i est choisie à l'étape $n+1$ avec probabilité $C_i(\overline{U}_n)$, où C_i est définie par :

$$C^i(U) = \mathbb{P}\left(\operatorname{argmax}_{k\in I} U^k + \varepsilon^k = i \middle| h^n\right).$$

Ainsi l'algorithme *Follow the Perturbed Leader* génère un processus discret stochastique, qui est une A.S.D. de l'inclusion différentielle :

$$(\dot{U}, \dot{\omega}) \in \{(V, \langle C(U), V\rangle)\,; V \in \mathcal{U}\} - (U, \omega).$$

Il s'agit bien d'un cas particulier du *Smooth Fictitious Play* car si f est strictement positive et si C est de classe $\mathcal{C}^1$, alors Hofbauer et Sandholm [62] ont montré qu'il existe alors une perturbation déterministe $\varepsilon\psi$ telle que $C(U) = \mathrm{BR}^\varepsilon(U)$.

L'inconvénient majeur de cette approche est qu'elle ne permet pas de connaître les vitesses uniformes de convergence du regret (ni même si la convergence est uniforme). Cela dit, en bornant précisément les écarts entre la version discrète et la version continue de l'*exponential weight algorithm*, Sorin [109] a montré que les bornes obtenues en temps continu s'appliquent à la version discrète. De plus cette analyse est valable à la fois pour la consistance externe, et pour la consistence interne.

Autres utilisations du regret

Congestion de réseaux : Un réseau est un graphe $G = (V, E)$ où V est un ensemble de noeuds avec une origine o et une destination d et E un ensemble d'arêtes orientées. L'ensemble des chemins (sans boucles) partant de o et allant à d est noté $\mathcal{P}$. Il y a un continuum de joueurs représenté par le segment $[0, 1]$ et, chaque jour, chacun d'entre eux choisit de parcourir un chemin de $\mathcal{P}$. On appelle $f_n(P)$ la proportion des joueurs qui prennent le chemin P à la date $n \in \mathbb{N}$.

Toute fonction $f : \mathcal{P} \to \mathbb{R}^+$, normalisée par $\sum_{p\in\mathcal{P}} f(P) = 1$ est appelé flux et elle induit une congestion $f(e) = \sum_{P:e\in P} f(P)$ sur l'arête e. Celle-ci a un coût de congestion positif, continu et croissant $l_e : \mathbb{R} \to \mathbb{R}$ fonction de sa congestion. Le coût total d'un chemin P est donc $l_P(f) = \sum_{e\in P} l_e(f(e)))$.

À l'étape n, on note $L_P(f_n)$ la congestion de l'arête P généré par le flux f_n. De manière classique, on dit qu'un agent n'a pas de regret externe si :

$$\limsup_{n\to\infty} \frac{\sum_{m\le n} L_{P_m}(f_m)}{n} \le \min_{P\in\mathcal{P}} \frac{\sum_{m\le n} L_P(f_m)}{n}.$$

Grâce à la structure particulière de ces jeux, Blum, Even-Dar et Ligett [21] ont montré que si les coûts de congestion sont Lipschitz, et que tous les joueurs utilisent une stratégie sans regret externe alors pour tout $\varepsilon > 0$ il existe un entier N_ε tel que le flux moyen $\overline{f}_n$ est un ε-équilibre de Nash pour $n \geq N_\varepsilon$.

On rappelle que Wardrop [119] a caractérisé les équilibres de Nash : un flux f est un équilibre de Nash si et seulement si $f(P) > 0$ implique que $L_P(f) \leq \min_{P' \in \mathcal{P}} l_{P'}(f)$ et f est un ε-équilibre de Nash si $\sum_{P \in \mathcal{P}} f(P) L_P(f) \leq \min_{P \in \mathcal{P}} l_P(f) + \varepsilon$.

Recherche d'équilibres de Nash : Foster et Young [46, 45] (voir aussi Germano et Lugosi [52]) ont construit des stratégies σ_ε non couplées (indépendantes des paiements des autres joueurs), basées sur un estimateur de type regret et qui convergent vers un équilibre de Nash. Lorsque l'on parle de regret dans cette section et la suivante, il ne s'agit pas d'une propriété asymptotique des stratégies, mais de l'évaluation de la distance entre un profil d'actions et un équilibre de Nash.

Formellement, considérons un jeu à L joueurs où l'espace d'actions du joueur $l \in L$ est noté I_l et sa fonction de paiement est $\rho^l : \Pi_{l \in L} I_l \to \mathbb{R}$. Le regret subi $r^l(x)$ par le joueur $l \in L$ sur le profil d'actions $x = (x_1, \ldots, x_L) \in \Pi_{l \in L} \Delta(I_l)$ et le regret maximal $R^l(x_l)$ subi de l'action $x_l \in \Delta(I)$ sont respectivement :

$$r^l(x) = \sup_{i \in I_l} \rho^l(i, x^{-l}) - \rho^l(x) \text{ et } R^l(x_l) = \sup_{x^{-l} \in \Pi_{k \neq l} \Delta(I_k)} r^l(x_l, x^{-l})$$

avec, de manière usuelle, $x^{-l} = (x_1, \ldots, x_{l-1}, x_{l+1}, \ldots, x_L)$. L'éloignement, au sens du regret, entre un profil d'actions $x \in \Pi_{l \in L} \Delta(I_l)$ et l'ensemble des équilibres de Nash est simplement $\max_{l \in L} r^l(x)$, le plus grand regret subi par les joueurs.

Pour tout $\varepsilon > 0$, la stratégie σ_ε est construite de la façon suivante : le joueur joue par blocs de N_ε étapes et sur un bloc il joue toujours selon la même probabilité p, qui appartient à une ε-grille finie et fixée de $\Delta(I)$.

À la fin d'un bloc, un joueur calcule son regret externe moyen sur ce bloc ; si celui-ci est petit (*i.e.* plus petit que $\tau_\varepsilon > 0$) alors le joueur garde la même stratégie, qui est a priori *peu éloigné* d'un équilibre de Nash. Si le regret est grand (*i.e.* plus grand que $\tau_\varepsilon > 0$) alors il choisit une nouvelle stratégie uniformément sur la ε-grille pour le bloc suivant.

Ainsi, s'il y a suffisamment de blocs, il y a une grande probabilité de trouver un profil de stratégies proche d'un ε-équilibre de Nash (en effet si le profil est loin d'un ε-équilibre, alors un des joueurs va dévier). Et les stratégies sont construites de sorte que la probabilité de sortir d'un voisinage d'un équilibre est très faible. En faisant tendre ε-suffisamment lentement vers 0, on assure que les profils convergent bien vers l'ensemble des équilibres de Nash.

L'avantage de cette construction est que l'on peut facilement la modifier si l'on suppose qu'un joueur n'observe que ses paiements et non le vecteur de regret — qu'il doit alors estimer. Son inconvénient majeur est qu'elle est essentiellement basée sur

une exploration exhaustive de l'ensemble des profils d'actions, ce qui ne représente pas un véritable apprentissage.

Définition d'équilibres : Halpern et Pass [55] (ainsi que Renou et Schlag [96]) ont proposé d'utiliser la notion de regret comme concept d'équilibre. Cela permet en effet de rationaliser certains comportements observés expérimentalement, en considérant une élimination itérative des stratégies faisant subir le regret maximal le plus important.

Leur exemple le plus frappant est le dilemme des voyageurs imaginé par Basu [13] : deux clients d'une compagnie aériennes ont un bagage (supposé identique) qui a été abîmé. Ils doivent annoncer une évaluation (notée m_1 et m_2 comprise entre 2 et 100 euros) de ceux-ci afin d'être indemnisés. Si $m_1 = m_2$ alors les deux clients reçoivent m_1 ; sinon celui qui a demandé le plus petit montant $\underline{m}$ reçoit $\underline{m} + 2$ tandis que celui qui a demandé le plus gros montant reçoit $\underline{m} - 2$. Après élimination des stratégies strictement dominées, il ne reste plus que les choix $m_1 = m_2 = 2$; ce couple forme donc l'unique équilibre de Nash du jeu. Cependant, des études expérimentales (les personnes ayant par exemple répondu à celle de Becker, Carter et Naeve [14] étaient des membres de la Game Theory Society) montrent que les stratégies rapportant le plus consistent à demander une forte évaluation (97 euros pour l'expérience citée).

Pour appliquer l'élimination des stratégies proposée par Halpern et Pass, il suffit de calculer, pour tout $m_1 \in \{2, \ldots, 100\}$, $R^1(m_1)$ et de supposer que le joueur 1 n'utilise aucune des stratégies qui le maximise. On rappelle que

$$R^1(m_1) = \sup_{m_2 \in \{2,\ldots,100\}} r^1(m_1, m_2) = \sup_{m \in \{2,\ldots,100\}} \rho^1(m_1, m) - \rho^1(m_1, m_2).$$

De simples calculs montrent que $R^1(m_1) = 3$ si $m_1 \in \{96, \ldots, 100\}$ et $R^1(m_1) \geq 4$ sinon.

En effet, $r(m_1, m_2) = (m_2 + 1) - (m_2 - 2) = 3$ pour tout $m_2 \in \{2, \ldots, m_1 - 1\}$, $r(m_1, m_1) = m_1 - (m_1 + 1) = 1$ et $r(m_1, m_2) = (m_2 + 1) - (m_1 + 2) \leq 100 - 96 - 1 = 3$. Si $m_1 \leq 95$, alors $r^1(m_1, 100) = 101 - (m + 2) = 99 - m \geq 4$. On remarque d'ailleurs que la stratégie de l'unique équilibre de Nash donne le regret maximal $R^2(2) = 97$ le plus grand.

Ainsi l'ensemble des stratégies $\{2, \ldots, 95\}$ est éliminé. En répétant cette procédure (où l'on suppose donc que les deux joueurs ne choisissent leurs stratégie que dans $\{96, \ldots, 100\}$) il ne reste que la stratégie 97.

Cependant, cette élimination a un inconvénient majeur : elle ne prend pas en compte les paiements des autres joueurs et on peut ainsi aboutir à des résultats absurdes ; par exemple, considérons le jeu suivant :

	L	R
T	(0,100)	(0,0)
B	(99,100)	(-100,0)

Le regret généré par T (resp. B) est de 99 (resp. 100) donc, d'après l'analyse précédente le joueur 1 devrait jouer T et le joueur 2 devrait, quant à lui, jouer L. Une élimination des stratégies strictement dominées (en deux étapes) implique que la case jouée sera (B, L) qui maximise le paiement des deux joueurs.

5

Liens entre Approchabilité, Calibration et non-Regret

Construire une stratégie d'approchabilité d'un convexe peut se ramener à la construction d'une stratégie calibrée dans un premier jeu auxiliaire. À son tour, cela se ramène à la construction d'une stratégie consistante dans un second jeu auxiliaire, à nouveau ramenée à la construction d'une stratégie d'approchabilité d'un orthant dans un troisième jeu auxiliaire. Ainsi on peut dire qu'en quelque sorte, les trois notions d'approchabilité, de calibration et de non-regret sont équivalentes.
La dernière sous-section est une traduction de la section 2 de l'article Calibration and Internal no-Regret with Partial Monitoring *dont un résumé étendu est publié dans les* Proceedings of the 20th conference on Algorithmic Learning Theory

Sommaire

5.1 Approchabilité et non-regret **56**
Espace d'actions du joueur 1 fini 56
Espaces d'actions infinies et regret généralisé 59
5.2 Non-regret et calibration **61**
Calibration par rapport à un graphe 62
5.3 Calibration et approchabilité **65**
De l'approchabilité à la calibration 65
De la calibration à l'approchabilité 69

L'APPROCHABILITÉ a été introduite par Blackwell [17] dans les jeux répétés à paiements vectoriels, le non-regret par Hannan [56] dans les jeux répétés à paiements réels et la calibration par Dawid [33] dans les jeux répétés de prédictions. Il n'y a donc, à première vue, pas de liens évidents entre ces trois notions. Cela dit, une conséquence des résultats d'approchabilité est l'existence de stratégies consistantes extérieurement et intérieurement, en dimension finie ou non (voir le chapitre 4). Cette

dernière implique l'existence de stratégies ε-calibrées (voir le chapitre 3), ce qui entraîne l'existence de stratégies d'approchabilité d'ensembles convexes (voir le chapitre 2). Il est également possible de montrer directement l'existence de stratégies calibrées à partir de la caractérisation de l'approchabilité d'ensembles convexes. C'est pour ces raisons que l'on parle d'équivalence entre ces trois notions.

5.1 Approchabilité et non-regret

Espace d'actions du joueur 1 fini

Blackwell [18] fut le premier à prouver que l'existence de stratégies consistantes extérieurement était une conséquence directe de son théorème d'approchabilité. En effet, pour tout $y \in \Delta(J)$, notons $\widetilde{\rho}(y) = \max_{x\in\Delta(I)} \rho(x,y) \in \mathbb{R}$ le paiement maximal du joueur 1 sachant que le joueur 2 joue y et $C \subset \mathbb{R}^J \times \mathbb{R}$ l'ensemble convexe défini par

$$C = \left\{(y,z) \in \mathbb{R}^J \times \mathbb{R}; y \in \Delta(J), z \geq \widetilde{\rho}(y)\right\}.$$

Un couple (y,z) appartient à C si y est une action mixte du joueur 2 — vue comme un vecteur de $\mathbb{R}^J$ — et z un paiement plus grand que $\widetilde{\rho}(y)$. En conséquence, si le couple $(\overline{\jmath}_n, \overline{\rho}_n)$ appartient à C alors $\overline{\rho}_n \geq \max_{i\in I} \rho(i, \overline{\jmath}_n)$ et le joueur 1 n'a pas de regret externe.

Proposition 5.1 (*Blackwell [18]*)

Il existe une stratégie σ consistante extérieurement telle que pour toute stratégie τ du joueur 2 et tout $\eta > 0$:

$$\mathbb{E}_{\sigma,\tau}\left[\left\|(\overline{r}_n)^+\right\|\right] = O\left(\frac{1}{\sqrt{n}}\right) \text{ et } \mathbb{P}_{\sigma,\tau}\left(\sup_{n\geq N}\left\|(\overline{r}_n)^+\right\| \geq \eta\right) \leq O\left(\frac{1}{\eta^2 N}\right).$$

La démonstration, reprise par Luce et Raiffa [77], annexe A.8.6 et Mertens, Sorin et Zamir [84], exercice 7 p. 107, est élémentaire. Il suffit de remarquer que dans le jeu où le paiement de l'étape n est $(j_n, \rho_n) \in \mathbb{R}^J \times \mathbb{R}$ — où j_n est vu comme le j_n-ème vecteur de la base canonique de $\mathbb{R}^J$ — le convexe C n'est pas repoussable par le joueur 2. En effet pour toute action $y \in \Delta(J)$ du joueur 2, il existe une action $i \in I$ du joueur 1 tel que $(y, \rho(i,y)) \in C$. En conséquence C est approchable par le joueur 1.

Ce résultat est vrai car l'application $y \mapsto \widetilde{\rho}(y)$ est continue et donc il suffit bien que $(\overline{\jmath}_n, \overline{\rho}_n)$ converge vers C pour que la stratégie du joueur 1 soit consistante extérieure. Par ailleurs, il est plus fort que la simple existence de stratégies consistantes, car la convergence est uniforme par rapport aux stratégies du second joueur, ce qui n'est pas a priori requis par la définition 4.1 de la consistance interne.

La stratégie de Blackwell ne s'exprime pas grâce à une formule explicite contrairement à la stratégie suivante, remarquée par Hart et Mas-Colell [57]. On rappelle qu'une stratégie est consistante extérieurement si la partie positive de $\overline{r}_n$ converge vers 0, donc si $\overline{r}_n$ converge vers l'orthant négatif de $\mathbb{R}^I$. Les résultats suivants sont démontrés en exhibant des algorithmes explicites construisants des stratégies consistantes extérieurement et intérieurement.

Proposition 5.2 (*Hart et Mas-Colell [57]*)

La stratégie σ où à l'étape n le joueur 1 joue proportionnellement à $(\overline{r}_n)^+$ (et arbitrairement si $\overline{r}_n$ est dans l'orthant négatif) est une stratégie consistante extérieurement. Pour toute stratégie τ du joueur 2 et tout $\eta > 0$:

$$\mathbb{E}_{\sigma,\tau}\left[\left\|(\overline{r}_n)^+\right\|\right] = \frac{4\sqrt{I}\|\rho\|_\infty}{\sqrt{n}} \text{ et } \mathbb{P}_{\sigma,\tau}\left(\sup_{n\geq N}\left\|(\overline{r}_n)^+\right\| \geq \eta\right) \leq \frac{32I\|\rho\|_\infty^2}{\eta^2 N}.$$

Démonstration . On considère la stratégie définie par $\sigma(h^n) = (\overline{r}_n)^+ / \left\|(\overline{r}_n)^+\right\|_1$ si $\overline{r}_n$ n'est pas dans l'orthant négatif et $\sigma(h^n)$ choisie arbitrairement sinon. Cette stratégie vérifie la condition de Blackwell (2.1), grâce au principe géométrique suivant.

Lemme 5.3

Quel que soit le choix de $x \in \Delta(I)$,

$$\langle x, \mathbb{E}_x\left[r(i,j)\right]\rangle = 0, \quad \forall j \in J \tag{5.1}$$

Démonstration du Lemme 5.3 .

Par linéarité de ρ, la k-ème composante de $\mathbb{E}_x\left[r(i,j)\right]$ est $\rho(k,j) - \rho(x,j)$ et donc le produit scalaire vaut $\sum_{k\in I} x^k\left[\rho(k,j) - \rho(x,j)\right] = \rho(x,j) - \rho(x,j) = 0$. □

Comme $x_{n+1} = \sigma(h^n)$ est proportionnel à $(\overline{r}_n)^+ = \overline{r}_n - (\overline{r}_n)^-$, le lemme 5.3 (avec le fait que $\langle X - X^-, X^-\rangle = \langle X^+, X^-\rangle = 0$) implique que :

$$\langle \overline{r}_n - (\overline{r}_n)^-, \mathbb{E}_{\sigma,\tau}\left[r_{n+1}|h^n\right] - (\overline{r}_n)^-\rangle = 0.$$

En conséquence, l'orthant négatif de $\mathbb{R}^I$ est un B-set et la stratégie décrite une stratégie d'approchabilité. Elle est donc consistante extérieurement. □

Ce résultat correspond à l'approche continue du regret définie en section 2.3 pour le potentiel $P(x) = \|x^+\|^2$. Les avantages de cette construction sont triples : elle est explicite, elle se généralise facilement à la consistance interne et ne dépend pas de l'espace d'action du joueur 2 (ni sur la description, ni sur les vitesses de convergence).

En particulier, la démonstration est strictement identique dans le *cas compact*, *i.e.* lorsque le joueur 2 choisit à l'étape n un vecteur de paiement $U_n \in [-B, B]^I$

— avec B une constante positive — tel que le paiement du joueur 1 est la i_n-ème composante de ce vecteur et le regret est donné par $r_n := r(i_n, U_n) = \left(U_n^k - U_n^{i_n}\right)_{k\in I}$. Similairement, on note le regret interne de l'étape n par $R_n = R(i_n, U_n) \in \mathbb{R}^{I\times I}$, où $R(i, U)$ est la matrice donc la i-ème ligne est $r(i, U)$ et toutes les autres sont nulles.

Proposition 5.4 (*Hart et Mas-Colell [57]*)

La stratégie σ où à l'étape n le joueur 1 joue une mesure invariante de $\left(\overline{R}_n\right)^+$ (et arbitrairement si $\overline{R}_n$ est dans l'orthant négatif) est une stratégie consistante intérieurement. Pour toute stratégie τ du joueur 2 et tout $\eta > 0$:

$$\mathbb{E}_{\sigma,\tau}\left[\left\|\left(\overline{R}_n\right)^+\right\|_2\right] = \frac{4B\sqrt{I}}{\sqrt{n}} \text{ et } \mathbb{P}_{\sigma,\tau}\left(\sup_{n\geq N}\left\|\left(\overline{R}_n\right)^+\right\|_2 \geq \eta\right) \leq \frac{32IB^2}{\eta^2 N}.$$

Démonstration . Considérons la stratégie σ telle que $x_{n+1} := \sigma(h^n) \in \Delta(I)$ est une mesure invariante de $\left(\overline{R}_n\right)^+$. On rappelle que x est une mesure invariante d'une matrice $A = (A^{ik})_{i,k\in I}$ à coefficients positifs si pour tout $i \in I$

$$\sum_{k\in I} x^k A^{ki} = x^i \sum_{k\in K} A^{ik}$$

et que son existence est une conséquence du théorème de Perron-Frobenius (voir par exemple Seneta [103]).

Cette stratégie vérifie la condition de Blackwell (2.1), grâce à un principe géométrique équivalent à (5.1) :

Lemme 5.5

Pour toute matrice A à coefficient positifs et toute mesure invariante x de A,

$$\langle A, \mathbb{E}_x\left[R(i,U)\right]\rangle = 0, \quad \forall\, U \in \mathbb{R}^I \tag{5.2}$$

Démonstration du Lemme 5.5 . La (i,k)-ème composante de $\mathbb{E}_x\left[R(i,U)\right]$ est $x^i\left(U^k - U^i\right)$ donc le produit scalaire vaut $\sum_{i,k} A^{ik}x^i\left(U^k - U^i\right)$. Le coefficient de U^i dans cette somme est

$$\sum_{k\in I} x^k A^{ki} - x^i \sum_{k\in K} A^{ik} = 0$$

car x est une mesure invariante de A. □

Comme $x_{n+1} = \sigma(h^n)$ est une mesure invariante de $\left(\overline{R}_n\right)^+ = \overline{R}_n - \left(\overline{R}_n\right)^-$ alors le lemme 5.5 implique que :

$$\langle \overline{R}_n - \left(\overline{R}_n\right)^-, \mathbb{E}_{\sigma,\tau}\left[R_{n+1}|h^n\right] - \left(\overline{R}_n\right)^-\rangle = 0.$$

En conséquence, l'orthant négatif est un B-set et la stratégie décrite une stratégie d'approchabilité ; elle est donc consistante intérieurement. □

Cette stratégie issue de Sorin [108] était aussi celle décrite par Foster et Vohra [42] bien que leur démonstration ne soit pas une application directe du théorème d'approchabilité (mais basée sur le score de Brier).

Espaces d'actions infinies et regret généralisé

Le théorème 5.6 suivant est l'extension du théorème 4.9 de la section 4.2 au cas *compact*. On suppose que l'espace d'action du joueur 1 est un ensemble métrique convexe compact I et qu'à l'étape n, le joueur 2 choisit une fonction U_n parmi un ensemble $\mathcal{U}$ de fonctions équicontinues sur I et bornées par $B \in L_2$. Les choix de U_n et de $i_n \in I$ génèrent le paiement $U_n(i_n)$. On parle de cas compact car le théorème d'Arzela-Ascoli implique que $\mathcal{U}$ est relativement compact.

Théorème 5.6

Dans le cas compact, il existe une stratégie Φ_c-swap, où Φ_c l'ensemble des fonctions continues de I dans I.

On rappelle que Stoltz et Lugosi [112] ont démontré une version finie de ce théorème.

Démonstration . Soit $\widehat{\Gamma}$ le jeu auxiliaire où l'espace d'action du joueur 1 est I et celui du joueur 2 est $\mathcal{U}$. Les choix de $i \in I$ et $U \in \mathcal{U}$ génèrent un paiement $\widehat{U}(i) \in L_2(\Phi_c, \lambda)$ — où λ est une probabilité quelconque sur $(\Phi_c, \|\cdot\|_\infty)$ muni de la tribu borélienne — défini par :

$$\widehat{U}(i)[\phi] = U[\phi(i)] - U[i], \ \forall \phi \in \Phi_c.$$

Le pavé $L_2^-(\Phi_c, \lambda) := \{U \in L_2, U \leq 0\}$ n'est pas repoussable par le joueur 2 ; en effet, pour tout $U \in \mathcal{U}$, il existe $i \in I$ — n'importe quel minimisateur global de U — tel que $\widehat{U}(i)$ appartienne à $L_2^-(\Phi_c, \lambda)$. Comme c'est un ensemble convexe, la proposition 2.12 implique qu'il est approchable par le joueur 1 qui n'a donc pas de regret ϕ-swap pour λ-presque toutes les fonctions $\phi \in \Phi_c$.

Or l'ensemble des fonctions continues sur I à valeurs dans I est séparable (voir Rudin [99] ou Stoltz et Lugosi [112]), il existe donc $\{\phi_k, k \in \mathbb{N}\}$ un sous-ensemble dénombrable dense de Φ_c et on définit la probabilité $\lambda = \sum_{k \in \mathbb{N}} 2^{-k} \delta_{\phi_k}$.

Comme $\mathcal{U}$ est une famille équicontinue, pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que :

$$\forall i, i' \in I, \forall U \in \mathcal{U}, d(i, i') \leq \delta \Longrightarrow |U(i) - U(i')| \leq \varepsilon.$$

Soit $\phi \in \Phi_c$, il existe ϕ_k tel que $\|\phi - \phi_k\|_\infty \leq \delta$ et donc

$$\frac{\sum_{m=1}^n U_m\left[\phi(i_m)\right] - U_m[i]}{n} \leq \frac{\sum_{m=1}^n U_m\left[\phi_k(i_m)\right] - U_m[i]}{n} + \varepsilon.$$

Comme σ n'a pas de regret ϕ_k-*swap*, elle a un regret ϕ-*swap* plus petit que ε, pour tout $\varepsilon > 0$. La stratégie σ n'a donc pas de regret Φ_c-swap. □

Jouer à chaque étape proportionnellement à la partie positive du regret externe est toujours une stratégie consistante extérieurement. Par contre, pour la consistance interne il n'est a priori pas possible de jouer une mesure invariante de la partie positive du regret car son existence n'est pas garantie. Cependant, pour tout $q \in L_2(\Phi, \lambda)$ et $\varepsilon > 0$, il existe $x \in \Delta(I)$ tel que pour tout $U \in \mathcal{U}$, $\left\langle q - q^-, \widehat{U}(x) - q^- \right\rangle \leq \varepsilon$, où on a noté q^- la projection de q sur $L_2^-(\Phi, \lambda)$ et $q^+ = q - q^-$.

On rappelle que le module de continuité ω_ϕ d'une fonction ϕ uniformément continue sur I (à valeurs dans un espace métrisé par d_1) est la plus grande (point par point) fonction réelle croissante vérifiant pour tout $\varepsilon > 0$,

$$\forall i, k \in I, d(i,k) \leq \omega_\phi(\varepsilon) \Longrightarrow d_1(\phi(i), \phi(k)) \leq \varepsilon.$$

Par hypothèse d'équicontinuité, il existe $\omega_{\mathcal{U}}$, une fonction croissante sur $\mathbb{R}_+$ strictement positive, telle que pour tout $U \in \mathcal{U}$, $\omega_{\mathcal{U}}(\cdot) \leq \omega_U(\cdot)$. Pour tout $\eta > 0$, on définit Φ_η, un sous-ensemble fermé de Φ_c, par

$$\Phi_\eta = \{\phi \in \Phi_c; \omega_\phi [\omega_{\mathcal{U}}(\varepsilon)] \geq \eta\}.$$

Comme Φ_η est une suite croissante de fermés dont l'union est Φ_c, il existe η tel que $\int_{\Phi_c \setminus \Phi_\eta} q^+[\phi] d\lambda \leq \varepsilon$, ce qui implique que pour tout U et tout $i \in I$:

$$\left| \left(\int_{\Phi_\eta} q^+[\phi] U[\phi(i)] - U[i] \int_{\Phi_\eta} q^+(\phi) d\lambda \right) - \left\langle q - q^-, \widehat{U}(i) - q^- \right\rangle \right| \leq 2B\varepsilon.$$

Soit $\mathcal{I}$ une η-grille finie de I et pour tout $\phi \in \Phi_c$, notons $\widehat{\phi} : I \to \mathcal{I}$ une fonction qui minimise (point par point) $d(\widehat{\phi}(i), \phi(i))$. Par conséquent, pour tout $i \in I$, par définition de Φ_η :

$$\left| \int_{\Phi_\eta} q^+[\phi] U[\widehat{\phi}(i)] d\lambda - \int_{\Phi_\eta} q^+[\phi] U[\phi(i)] d\lambda \right| \leq \varepsilon \int_\Phi q^+[\phi] d\lambda.$$

Pour tous les couples $(i,k) \in \mathcal{I}^2$, on définit l'ensemble $\Phi_\eta^{i,k} = \{\phi \in \Phi_\eta : \widehat{\phi}(i) = k\}$ de telle sorte que :

$$\int_{\Phi_\eta} q^+[\phi] U[\widehat{\phi}(i)] d\lambda = \sum_{k \in \mathcal{I}} \int_{\Phi_\eta^{i,k}} q^+[\phi] d\lambda U[k].$$

On note aussi $\theta^{i,k} := \int_{\Phi_\eta^{i,k}} q^+(\phi) d\lambda$ et on appelle x_q une mesure invariante de la matrice θ. Alors, $\sum_{i \in \mathcal{I}} x_q(i) \sum_{k \in \mathcal{I}} \theta^{i,k} (U(k) - U(i)) = 0$ et donc :

$$\begin{aligned} \left\langle q - q^-, \widehat{U}(x_q) - q^- \right\rangle &\leq \sum_{i \in \mathcal{I}} x(i) \sum_{k \in \mathcal{I}} \theta^{i,k} (U(k) - U(i)) + \varepsilon \left(2B + \int_\Phi q^+[\phi] d\lambda \right) \\ &\leq \varepsilon \left(2B + \int_\Phi q^+[\phi] d\lambda \right). \end{aligned}$$

Cette démonstration, basée sur l'approchabilité en dimension infinie, peut facilement être adaptée pour montrer l'existence de stratégie sans regret avec activations (comme définie en section 2.3, voir aussi Lehrer [70]).

5.2 Non-regret et calibration

Pour construire une stratégie calibrée par rapport à une grille, Foster et Vohra [43] ont utilisé une stratégie consistante intérieurement dans un jeu auxiliaire. Sorin [108] a simplifié leur démonstration grâce au lemme 5.8 ci-dessous. Tout d'abord, rappelons rapidement le modèle général du jeu de prédictions défini en chapitre 3.

À chaque étape $n \in \mathbb{N}$, le joueur 2 choisit s_n parmi un ensemble fini d'états S. Le joueur 1 fait une prédiction sur s_n en choisissant une probabilité μ_n dans une grille finie $\mathcal{L} = \{\mu(l), l \in L\}$. Explicitement, une stratégie du joueur 1 est une application de l'ensemble des histoires finies $H := \bigcup_{n\in\mathbb{N}} (L \times S)^n$ dans $\Delta(L)$, l'ensemble des probabilités sur L. Les stratégies du joueur 2 sont définies de manière duale.

Une stratégie σ du joueur 1 est calibrée (par rapport à $\mathcal{L}$) si pour toute stratégie τ du joueur 2 et pour tout $l, k \in L$, la moyenne empirique des états lorsque le joueur 1 a prédit $\mu(l)$ est plus proche de $\mu(l)$ que de n'importe quel autre $\mu(k)$. Formellement, si pour tout $l, k \in L$

$$\limsup_{n\to\infty} \frac{N_n(l)}{n}\left(\|\overline{s}_n(l) - \mu(l)\|^2 - \|\overline{s}_n(l) - \mu(k)\|^2 \right) \leq 0, \mathbb{P}_{\sigma,\tau}\text{−ps}$$

où $N_n(l) = \{m \leq n, l_m = l\}$ est l'ensemble des dates jusqu'à la n-ème où le joueur 1 a prédit $\mu(l)$ et $\overline{s}_n(l) = \sum_{m\in N_n(l)} s_m/N_n(l)$ est la moyenne empirique des états sur $N_n(l)$. L'ensemble $\Delta(S)$ est vu comme un sous-ensemble de $\mathbb{R}^S$ muni de la norme 2.

Théorème 5.7

Il existe une stratégie σ calibrée par rapport à $\mathcal{L}$ telle que pour toute stratégie τ du joueur 2 et tout $\eta > 0$:

$$\mathbb{E}_{\sigma,\tau}\left[\sup_{l,k\in L} \frac{N_n(l)}{n}\left(\|\overline{s}_n(l) - \mu(l)\|^2 - \|\overline{s}_n(l) - \mu(k)\|^2 \right)\right] \leq \frac{6\sqrt{L}}{\sqrt{n}} \quad \text{et}$$

$$\mathbb{P}_{\sigma,\tau}\left(\sup_{n\geq N}\sup_{l,k\in L} \frac{N_n(l)}{n}\left(\|\overline{s}_n(l) - \mu(l)\|^2 - \|\overline{s}_n(l) - \mu(k)\|^2 \right) \geq \eta\right) \leq \frac{72L}{\eta^2 N}.$$

La démonstration fait appel à la construction de stratégies consistantes et utilise le simple lemme suivant :

Lemme 5.8

Pour toute suite $(s_m)_{m\in\mathbb{N}}$ et tout $l, k \in L$

$$\sum_{m\in N_n(l)} \frac{\|s_m - \mu(l)\|^2}{N_n(l)} - \frac{\|s_m - \mu(k)\|^2}{N_n(l)} = \|\overline{s}_n(l) - \mu(l)\|^2 - \|\overline{s}_n(l) - \mu(k)\|^2.$$

Démonstration . Il suffit de développer les sommes des deux côtés. □

Pour que le lemme 5.8 soit vrai, il est nécessaire que la norme soit euclidienne (comme remarqué par Cesa-Bianchi et Lugosi [29] p. 89 et exercice 4.14).

Démonstration du théorème 5.7 . Notons Γ_r le jeu répété auxiliaire où les espaces d'actions sont L et S. Les choix à l'étape n de s_n et de $l_n \in L$ génèrent le regret interne défini par $R_n = R(l_n, U_n)$ où $U_n = -\|s_n - \mu(l)\|^2 \in \mathbb{R}^L$. Considérons σ une stratégie consistante intérieurement du joueur 1, alors pour toute stratégie τ du joueur 2,

$$\limsup_{n\to\infty} \sup_{l,k} \frac{|N_n(l)|}{n} \left(\sum_{m\in N_n(l)} \frac{\|s_m - \mu(l)\|^2}{|N_n(l)|} - \frac{\|s_m - \mu(k)\|^2}{|N_n(l)|} \right) \leq 0.$$

Le lemme 5.8 implique que σ est calibrée par rapport à $\mathcal{L}$. Les vitesses de convergence sont données par celles de l'algorithme consistant intérieurement. □

Remarque : La stratégie σ construite est telle que pour tout $l \in L$, $\overline{s}_n(l)$ est asymptotiquement plus proche de $\mu(l)$ que de n'importe quel autre $\mu(k)$, dès que $|N_n(l)|/n$ n'est pas trop petit.

Le fait que s_n appartienne à un ensemble fini S et que les $\{\mu(l), l \in L\}$ soient des probabilités sur S ne joue aucune rôle : on montre de la même façon que pour tout ensemble fini $\{a(l) \in \mathbb{R}^d, l \in L\}$, le joueur 1 a une stratégie σ telle que pour toute suite bornée $(a_m)_{m\in\mathbb{N}}$ de $\mathbb{R}^d$ et tout l et k :

$$\limsup_{n\to\infty} \frac{|N_n(l)|}{n} \left(\|\overline{a}_n(l) - a(l)\|^2 - \|\overline{a}_n(l) - a(k)\|^2 \right) \leq 0.$$

Réciproquement, Foster et Vohra [42] ont construit une stratégie consistante intérieurement à partir d'une stratégie calibrée.

Calibration par rapport à un graphe

La notion de calibration peut être affaiblie en définissant la calibration par rapport à un graphe $\mathcal{G} = (\mathcal{L}, \mathcal{E})$ où $\mathcal{L} = \{\mu(l), l \in L\}$ est un sous-ensemble fini de $\Delta(S)$ et $\mathcal{E}$ est un ensemble d'arêtes orientées, *i.e.* de couples (l, k) avec $(l, k) \in L^2$. Notons $V(l) = \{k \in L, (l, k) \in \mathcal{E}\}$ l'ensemble des voisins de $\mu(l) \in \mathcal{L}$.

Definition 5.9

Une stratégie σ du joueur 1 est calibrée par rapport à $\mathcal{G} = (\mathcal{L}, \mathcal{E})$ si pour toute stratégie τ du joueur 2 et pour tout $l \in L$

$$\limsup_{n\to\infty} \frac{N_n(l)}{n} \left(\|\overline{s}_n(l) - \mu(l)\|^2 - \|\overline{s}_n(l) - \mu(k)\|^2 \right) \leq 0, \forall k \in V(l), \mathbb{P}_{\sigma,\tau}\text{–ps.}$$

L'existence d'une telle stratégie est évidemment impliquée par l'existence d'une stratégie calibrée. De plus,

Lemme 5.10

Il existe une stratégie du joueur 1 calibrée par rapport au graphe $\mathcal{G} = (\mathcal{L}, \mathcal{E})$. Les vitesses de convergence du théorème 5.7 restent valables en remplaçant respectivement $6\sqrt{L}$ par $12\sqrt{\sup_{l\in L} |V(l)|\delta_{\mathcal{G}}}$ et $72L$ par $288 \sup_{l\in L} |V(l)|\delta_{\mathcal{G}}^2$ où on a noté $\delta_{\mathcal{G}} = \sup_{(l,k)\in\mathcal{E}} \|\mu(l) - \mu(k)\|$ est le diamètre du graphe.

Démonstration . Il suffit de considérer le jeu auxiliaire d'espace d'actions L et S où, à l'étape n, le regret induit par le graphe $\mathcal{G}$ est une matrice R'_n dont la (l,k)-ème coordonnée est :

$$(R'_n)^{lk} = \begin{cases} \|s_n - \mu(l)\|^2 - \|s_n - \mu(k)\|^2 & \text{si } l_n = l \text{ et } k \in V(l) \\ 0 & \text{sinon.} \end{cases}$$

Notons σ la stratégie du joueur 1 consistant à jouer à l'étape $n+1$ selon $x_{n+1} = \sigma(h^n)$ une mesure invariante de $\left(\overline{R}'_n\right)^+$. Le lemme 5.5 implique que

$$\left\langle \left(\overline{R}'_n\right)^+, \mathbb{E}_\sigma\left[R'_{n+1}|s_{n+1}\right]\right\rangle = \left\langle \left(\overline{R}'_n\right)^+, \mathbb{E}_\sigma\left[R_{n+1}|s_{n+1}\right]\right\rangle = 0$$

où R_{n+1} est le regret interne usuel de l'étape $n+1$. Ainsi σ approche l'orthant négatif de $\mathbb{R}^{L\times L}$ et elle est calibrée par rapport à $\mathcal{G}$.

Les bornes s'obtiennent en remarquant que $\|R'_{n+1}\| \leq 6 \sup_{l\in L} |V(l)|\delta_{\mathcal{G}}$. □

Ce lemme simple a pour conséquence le théorème 5.11 ci-dessous (plus précisément, il implique le lemme 5.12 suivant).

Théorème 5.11 (*Mannor et Stoltz [80]*)

Il existe une stratégie σ naïvement calibrée et telle que pour toute stratégie τ du joueur 2 et tout ensemble Borel-mesurable $B \subset \Delta(S)$:

$$\lim_{n\to\infty} \frac{n^{\frac{1}{S+1}}}{\sqrt{\ln(n)}} \left\| \frac{1}{n} \sum_{m=1}^{n} \mathbb{1}\left\{\mu_m \in B\right\} (\mu_m - s_m) \right\| \leq \Gamma(S), \quad \mathbb{P}_{\sigma,\tau}\text{−ps}$$

où $\Gamma(S)$ est une constante dépendant de S.

La démonstration de Mannor et Stoltz (celle du théorème 5.14 en est une modification) repose sur un argument de doubling-trick (voir Sorin [106], proposition 3.2 p. 56) ajouté au lemme 5.12 qui suit.

L'algorithme que l'on propose repose sur la résolution de systèmes linéaires creux (et non pas de programmes linéaires).

Lemme 5.12

Pour tout $\varepsilon > 0$, il existe σ une stratégie ε-calibrée telle que pour tout $p \in \Delta(S)$, toute stratégie τ du second joueur et tout $n \in \mathbb{N}$:

$$\mathbb{E}_{\sigma,\tau}\left[\sum_{p\in\Delta(S)} \frac{|N_n(p,\varepsilon)|}{n}\Big(\|\overline{\mu}_n(p,\varepsilon) - \overline{s}_n(p,\varepsilon)\| - \varepsilon\Big)\right] \leq \left(\frac{2}{\varepsilon}\right)^{\frac{S-1}{2}} \frac{4\gamma(S)}{\sqrt{n}},$$

où l'on a reprit les notations de la définition 3.11, avec $\gamma(S)$ une constante dépendante de S, de l'ordre de $1/S!$.

Démonstration . Soit $\varepsilon > 0$ fixé, la stratégie σ considérée est simplement la stratégie calibrée par rapport à une grille bien choisie de $\Delta(S)$ donnée par le Lemme 5.10.

On rappelle que l'ensemble des probabilités sur S peut être défini comme un sous-ensemble de $\mathbb{R}^{S-1}$ de la façon suivante :

$$\Delta(S) = \left\{x = (x_1, \ldots, s_{S-1}) \in \mathbb{R}^{S-1} : x_1, \ldots, x_{S-1} \geq 0 \text{ et } \sum x_s \leq 1\right\}.$$

On appelle $\mathbf{e}_s$ le vecteur de $\mathbb{R}^{S-1}$ dont toutes les coordonnées sont nulles sauf la s-ème qui vaut 1 et $\mathcal{L}_\varepsilon$ la grille régulière définie par

$$\mathcal{L}_\varepsilon = \left\{\mu = \sum_{k=1}^{S-1} \frac{2\varepsilon n_k}{\sqrt{S-1}}\mathbf{e}_s \in \Delta(S); n_k \in \mathbb{N}\right\} := \left\{\mu(l) = \sum_{k=1}^{S-1} \frac{2\varepsilon n_k(l)}{\sqrt{S-1}}\mathbf{e}_s, l \in L_\varepsilon\right\}. \tag{5.3}$$

On suppose qu'une arête relie les deux sommets $\mu(l)$ et $\mu(l')$ si $|n_k(l) = n_k(l')| = 0$ pour tout $k \in \{1, \ldots, S-1\}$ sauf pour un unique k_0 tel que $|n_{k_0}(l) - n_{k_0}(l')| = 1$.

Le diamètre de ce graphe est $\frac{2\varepsilon}{\sqrt{S-1}}$ et $|V(l)| \leq 2^{S-1}$ pour tout $l \in L$. Le lemme 5.10 implique qu'il existe une stratégie σ du joueur 1 telle que pour toute stratégie τ du joueur 2 et tout $n \in \mathbb{E}$,

$$\mathbb{E}_{\sigma,\tau}\left[\left\|\left(\overline{R}'_n\right)^+\right\|_2\right] \leq \frac{8.2^{\frac{S+1}{2}}\varepsilon}{\sqrt{S-1}} \quad \text{avec, on rappelle,}$$

$$\left(\bar{R}'_n\right)^+ = \frac{|N_n(l)|}{n}\Big(\|\overline{s}_n(l) - \mu(l)\|^2 - \|\overline{s}_n(l) - \mu(k)\|^2\Big)^+_{k\in V(l), l\in L}.$$

La géométrie de la grille $\mathcal{L}_\varepsilon$ implique que

$$\begin{aligned}\|\overline{s}_n(l) - \mu(l)\|_2 &\leq \varepsilon + \sqrt{\sum_{k\in V(l)} \max\left(\left\langle \overline{s}_n(l) - \frac{\mu(l)+\mu(k)}{2}, \frac{\mu(k)-\mu(l)}{\|\mu(k)-\mu(l)\|}\right\rangle, 0\right)^2} \\ &= \varepsilon + \sqrt{\sum_{k\in V(l)} \frac{\max\left(\|\overline{s}_n(l) - \mu(l)\|^2 - \|\overline{s}_n(l) - \mu(k)\|^2, 0\right)^2}{4\|\mu(k)-\mu(l)\|^2}}\end{aligned}$$

Comme $\|\mu(k) - \mu(l)\| = 2\varepsilon/\sqrt{S-1}$ si $k \in V(l)$, en multipliant l'inégalité par $|N_n(l)|/n$, puis en sommant sur $l \in L_\varepsilon$ et en prenant la racine carrée, on obtient

$$\|\widetilde{R}_n\|_2 := \sqrt{\sum_{l \in L_\varepsilon} \left[\frac{|N_n(l)|}{n}\Big(\|\overline{s}_n(l) - \mu(l)\|_2 - \varepsilon\Big)\right]^2} \leq \frac{\sqrt{S-1}}{4\varepsilon}\left\|\left(\overline{R}'_n\right)^+\right\|.$$

Comme $\|\widetilde{R}_n\|_1 \leq \sqrt{L_\varepsilon}\|\widetilde{R}_n\|_2$ car $\widetilde{R}_n \in \mathbb{R}^{L_\varepsilon}$, le résultat découle du fait qu'il existe une constante $\gamma(S)$, de l'ordre de $1/S!$, telle que $L_\varepsilon \leq \gamma(S)\varepsilon^{-(S-1)}$. □

5.3 Calibration et approchabilité

De l'approchabilité à la calibration

Il existe de nombreuses démonstrations de l'existence (voire leur construction) de stratégies calibrées utilisant le théorème d'approchabilité de Blackwell, à la fois avec des règles d'inspection (comme Lehrer [68] ou Sandroni, Smorodinsky et Vohra [102]) ou pour la calibration naïve (celle de Foster [41] ou encore celle de Mannor et Stoltz [80]). On va dans cette section démontrer le théorème 3.3 et une version différente du théorème 5.11.

Règles d'inspection indépendantes des prédictions

On rappelle qu'étant donnée une règle d'inspection $\mathrm{RI} = (U, C)$ à court terme et indépendante des prédictions, une stratégie σ est RI-calibrée le long de la partie $h^\infty \in \mathcal{H}$ si

$$f_n(h^\infty, \sigma, \mathrm{RI}) := \frac{\sum_{m=1}^n \mathbb{1}\{s_{n+1} \in U(h^n)\}\left[\mathbb{1}\{s_{n+1} \in C(h^n)\} - \sigma(C(h^n)|U(h^n), h^n)\right]}{\sum_{m=1}^n \mathbb{1}\{s_{n+1} \in U(h^n)\}}$$

converge vers 0 dès que $\sum_{m=1}^\infty \mathbb{1}\{s_{n+1} \in U(h^n)\} = \infty$.

Le théorème 3.3 dit que pour toute probabilité λ sur $\mathcal{RI}_{ic}$ (l'ensemble des règles d'inspection à court terme et indépendante des prédictions), il existe une stratégie pure σ (*i.e.* à chaque étape, le joueur choisit $\mu_n \in \Delta(S)$ de façon déterministe) qui passe, le long de chaque partie h^∞, λ-presque tous les tests de calibration.

La démonstration de ce résultat est très proche de celle de Lehrer [68], seul change le Lemme 5.13.

Démonstration du théorème 3.3 . La démonstration repose essentiellement sur le théorème d'approchabilité en dimension infinie (avec activations). Soit λ une probabilité sur $\mathcal{RI}_{ic}$, on va construire, par induction, une stratégie σ qui approche $\{0\} \subset L_2(\mathcal{RI}_{ic}, \lambda)$. Soient donc h^∞ une histoire fixée et $n \in \mathbb{N}$.

On considère le jeu auxiliaire Γ_n^c à somme nulle où l'espace d'actions du joueur 2 est $\Delta(S)$ et celui du joueur 1 est $\Delta(S)_0$, l'intérieur de $\Delta(S)$, vu comme un sous-ensemble de $\mathbb{R}^{S-1}$. On restreint le joueur 2 à cet espace d'action afin de s'assurer que $\sigma(U(h^n)|h^n)$ soit toujours non nul et donc $\sigma(C(h^n)|U(h^n), h^n)$ ait toujours un sens.

Les choix de $\mu \in \Delta(S)_0$ et $s' \in S$ génèrent un paiement défini par :

$$g_n(\mu, s) = \int_{\mathcal{RI}_{ic}} \frac{f_n(h^\infty, \sigma, \mathrm{RI})}{1 + \sum_{k=1}^n \mathbb{1}_{\{s_{k+1} \in U(h^k)\}}} \mathbb{1}_{s' \in U(h^n)} \left[\mathbb{1}_{s' \in C(h^n)} - \mu\left(C(h^n)|U(h^n)\right)\right] d\lambda;$$

La fonction g_n est étendue — en sa seconde variable — par linéarité sur $\Delta(S)$. Comme pour toute règle d'inspection $\mathrm{RI} \in \mathcal{RI}_{ic}$

$$\frac{f_n(h^\infty, \sigma, \mathrm{RI})}{1 + \sum_{k=1}^n \mathbb{1}_{\{s_{k+1} \in U(h^k)\}}} = \frac{f_n(h^\infty, \sigma, \mathrm{RI})}{\sum_{k=1}^{n+1} \mathbb{1}_{\{s_{k+1} \in U(h^k)\}}}$$

et d'après le théorème 2.12, il suffit pour que σ approche le pavé $\{0\}$ que pour tout $\varepsilon > 0$, il existe $\mu \in \Delta(S)_0$ telle que $g_n(\mu, \mu') \leq \varepsilon$ pour tout $\mu' \in \Delta(S)$.

Ceci est une conséquence du lemme 5.13 suivant, qui est une application de l'inégalité de Ky Fan [37]. □

Lemme 5.13

Soit G un jeu à deux joueurs à somme nulle d'espace d'actions $X \subset \mathbb{R}^d$ convexe compact pour les deux joueurs, de fonction de gain g telle que $g(x, x) = 0$ pour tout x dans X_0, l'intérieur de X.

Si pour tout $x \in X$, $g(\cdot, x)$ est affine et $g(x, \cdot)$ est continue et uniformément bornée par B sur X_0 alors pour tout $\varepsilon > 0$, il existe $\underline{x} \in X_0$ tel que $g(x, \underline{x}) \leq \varepsilon$ pour tout $x \in X$.

Démonstration . Rappelons le lemme connu sous le nom d'inégalité de Ky Fan [37]. Soit K un ensemble compact convexe d'un ensemble euclidien, $g : K \times K \to \mathbb{R}$ telle que pour tout $y \in K$, $g(\cdot, y)$ soit concave sur K et pour tout $x \in K$, $g(x, \cdot)$ soit continue sur K. Si de plus $g(x, x) = 0$ pour tout $x \in K$, alors il existe $\underline{x} \in K$ telle que $g(x, \underline{x}) \leq 0$ pour tout $x \in K$.

Soit G, X, g donnés par les hypothèses du lemme. Pour tout $\eta > 0$, on appelle X_η l'ensemble convexe compact défini par $X_\eta = \{x \in X, d(x, \partial X) \geq \eta\}$ où ∂X est la frontière de X. Alors g vérifie les hypothèses de l'inégalité de Ky Fan sur X_η. Il existe donc $\underline{x}_\eta$ tel que pour tout $x \in X_\eta$, $g(x, \underline{x}_\eta) \leq 0$. Comme g est bornée par B sur X_η :

$$\forall y \in X, g(x, \underline{x}_\eta) \leq \frac{B(d-1)}{1 - (d-1)\eta}\eta.$$

Ainsi, en posant $\eta = \varepsilon / \left[(d-1)(B+\varepsilon)\right]$ on obtient le résultat. □

CALIBRATION NAÏVE

Le théorème suivant est à mettre en relation avec le théorème 5.11. Le résultat de ce dernier est plus fort, mais les convergences sont plus lentes.

Théorème 5.14

Il existe une base dénombrable de voisinages $\mathcal{V} = \{V_k; k \in \mathbb{N}\}$ de $\Delta(S)$ et une stratégie σ du joueur 1 telle que pour tout $V \in \mathcal{V}$ et toute stratégie τ du joueur 2 :

$$\left\| \frac{1}{n} \sum_{m=1}^{n} \mathbb{1}\left\{\mu_m \in V\right\} (\mu_m - s_m) \right\| \leq \gamma''(S) \frac{\ln(n)^{\frac{3}{2}}}{\sqrt{n}}, \quad \mathbb{P}_{\sigma,\tau}\text{−ps}$$

où $\gamma''(S)$ est une constante dépendante de S.

La démonstration est une modification de celle de Mannor et Stoltz [80]. Elle est basée sur la construction de stratégies calibrées par rapport à des grilles de plus en plus fines et choisies avec soin (qui vont définir la base adéquate de voisinages), ainsi que d'un argument classique de doubling trick.

La base de voisinages est un ensemble dénombrable de boules pour la norme infinie tel que :

1) l'ensemble des centres des boules est dense dans $\Delta(S)$;
2) pour chacun de ces centres v, il existe dans la base de voisinages une boule centrée en v de rayon arbitrairement petit.

Démonstration . Comme dans la démonstration du lemme 5.12, on représente $\Delta(S)$ par

$$\Delta(S) = \left\{ x = (x_1, \ldots, s_{S-1}) \in \mathbb{R}^{S-1} : x_1, \ldots, x_{S-1} \geq 0 \text{ et } \sum x_s \leq 1 \right\}.$$

Soit m_k une suite strictement croissante d'entiers impairs tels que pour tout $k \in \mathbb{N}$ m_{k+1} soit un multiple de m_k. Pour tout $k \in \mathbb{N}$, on note $\mathcal{L}_k$ la grille régulière de $\Delta(S)$ définie par :

$$\mathcal{L}_k = \left\{ \mu = \sum_{s=1}^{S-1} \frac{n_s}{m_k} \mathbf{e}_s \in \Delta(S); n_k \in \mathbb{N} \right\} := \{\mu(l,k), l \in L_k\} \subset \Delta(S)$$

où $\mathbf{e}_s$ est le s-ème vecteur canonique de $\mathbb{R}^{S-1}$. Les grilles sont de plus en plus fines, au sens où $\mathcal{L}_k \subset \mathcal{L}_{k+1}$ pour tout $k \in \mathbb{N}$.

La cellule de Voronoï $V(l,k)$ associée à $\mu(l,k)$ est l'ensemble des points de $\Delta(S)$ qui sont plus proches, pour la norme 2, de $\mu(l,k)$ que de n'importe quel autre $\mu(l',k)$. Formellement, par définition de $\mathcal{L}_k$, pour tout $l \in L_k$:

$$V(l,k) = \left\{ \omega \in \Delta(S), \|\omega - \mu(l,k)\|_\infty \leq \frac{1}{2.m_k} \right\}.$$

L'ensemble de ces cellules est notée $\mathcal{V}_k$ et leur union (sur tous les $k \in \mathbb{N}$) forme la base de voisinage $\mathcal{V}$. Par construction, toute cellule de $\mathcal{V}_k$ est une union de cellules de $\mathcal{V}_{k+1}$ et le diamètre de $V(l,k)$ est $\frac{\sqrt{S-1}}{m_k}$.

Pour tout $\mu \in \Delta(S)$ et $s \in S$, on définit par récurrence $G^K(\mu, s)$ un vecteur de $\mathbb{R}^{S(L_1+\cdots+L_K)}$ par :

$$G^1(\mu, s) = \left[\mathbb{1}\{\mu \in V(l,1)\}(\mu - \mathbf{e}_S)\right]_{l \in L_1} \in \mathbb{R}^{SL_1} \text{ et}$$

$$G^K(\mu, s) = \left[G^{K-1}(\mu, s), \left(\mathbb{1}\{\mu \in V(l,k)\}(\mu - \mathbf{e}_S)\right)_{l \in L_K}\right] \in \mathbb{R}^{S(L_1+\cdots+L_K)}.$$

On introduit le jeu Γ_K où l'espace d'actions du joueur 1 est $\mathcal{L}_K$, celui du joueur 2 est S et le paiement est G^K. Vu que les grilles sont régulières et de plus en plus fines et que chaque m_k est impair, pour tout $k \in \mathbb{N}$ un élément $\mu(l,k) \in \Delta(S)$ ne peut appartenir qu'à une seule cellule de $\mathcal{V}_{k'}$ pour tout $k' \in \mathbb{N}$. Ainsi, pour tout $\mu(l,k) \in \mathcal{L}_K$ et $s \in S$:

$$\left\|G^K(\mu(l,k), s)\right\|_2 \leq 2\sqrt{K}.$$

Dans ce jeu, si on note $B(0,\eta)_\infty$ la boule fermé pour la norme infinie centrée en 0 et de rayon η, l'ensemble convexe C^K défini par

$$C^K := \Pi_{l \in L_k, k \leq K} B\left(0, \frac{\sqrt{S-1}}{2.m_K}\right)_\infty = B\left(0, \frac{\sqrt{S-1}}{2.m_K}\right)_\infty^{L_1+\cdots+L_K} \subset \mathbb{R}^{(S-1)(L_1+\cdots+L_K)}$$

n'est pas repoussable par le joueur 2. En effet, pour tout $\mu \in \Delta(S)$, il existe $\mu(l,K)$ tel que $\mu \in V(l,K)$, donc $\|\mu - \mu(l,K)\| \leq \frac{\sqrt{S-1}}{2m_K}$. Ainsi $\mathbb{E}_\mu\left[G^K(\mu(l,K), s)\right]$ appartient bien à C^K et ce dernier est approchable par le joueur 1.

En conséquent, il existe une stratégie σ^K telle que pour toute stratégie τ du joueur 2 :

$$\mathbb{E}_{\sigma^K,\tau}\left[d\left(\overline{G}_n^K, C^K\right)\right] \leq 4\sqrt{\frac{K}{n}}, \text{ avec } G_n^K \text{ le paiement de l'étape } n.$$

Considérons la stratégie σ construite en jouant par bloc, en suivant la concaténation usuelle de stratégies appelée doubling trick. Précisément, le joueur 1 joue sur le K-ème bloc selon σ^K. La longueur de ce bloc est N^K et on appelle $\tau^K = N^1 + \cdots + N^k$ — ces longueurs sont explicitées ultérieurement. Soit $V = V(l,k) \in \mathcal{V}$ un des voisinages fixé de la base et $\underline{k}$ le plus petit entier tel que $\mu(l,k) \in \mathcal{L}_{\underline{k}}$. Par construction des grilles, sur tous les blocs avant le $\underline{k}$-ème, aucune des différentes prédictions μ_n n'appartient à V.

Ainsi, pour tout $n \in \mathbb{N}$ et en notant M l'entier tel que $\tau^M < n \leq \tau^{M+1}$:

$$\mathbb{E}_{\sigma,\tau}\left\|\frac{1}{n}\sum_{t=1}^{n}\mathbb{1}\left\{\mu_t \in V\right\}(\mu_t - s_t)\right\| \leq \sum_{k=1}^{M}\frac{N^k}{n}\mathbb{E}_{\sigma^k,\tau}\left\|\frac{1}{N^k}\sum_{t=\tau^k+1}^{\tau^{k+1}}\mathbb{1}\left\{\mu_t \in V\right\}(\mu_t - s_t)\right\|$$
$$+\frac{1}{n}\mathbb{E}_{\sigma,\tau}\left\|\sum_{t=\tau^M+1}^{n}\mathbb{1}\left\{\mu_t \in V\right\}(\mu_t - s_t)\right\|$$

Par définition de C^k :

$$\begin{aligned}\mathbb{E}_{\sigma^k,\tau}\left\|\frac{1}{N^k}\sum_{t=\tau^k+1}^{\tau^{k+1}}\mathbb{1}\left\{\mu_t\in V\right\}(\mu_t-s_t)\right\| \leq & \mathbb{E}_{\sigma^k,\tau}\left[d\left(\overline{G}^k_{N^k},C^k\right)\right]+\frac{\sqrt{S-1}}{m_k}\\ \leq & 4\sqrt{\frac{k}{N^k}}+\frac{\sqrt{S-1}}{m_k}\end{aligned}$$

Donc en sommant cette dernière inégalité sur k, on obtient :

$$\begin{aligned}\mathbb{E}_{\sigma,\tau}\left\|\frac{1}{n}\sum_{t=1}^{n}\mathbb{1}\left\{\mu_t\in V\right\}(\mu_t-s_t)\right\| \leq & \sum_{k=1}^{M}\frac{N^k}{n}4\sqrt{\frac{k}{N^k}}+\frac{n-\tau^M}{n}4\sqrt{\frac{M+1}{n-\tau^M}}\\ & +\sum_{k=1}^{M}\frac{N_k}{n}\frac{\sqrt{S-1}}{m_k}+\frac{n-\tau^M}{n}\frac{\sqrt{S-1}}{m_{M+1}}\\ \leq & \frac{4}{\sqrt{n}}\sum_{k=1}^{M+1}\sqrt{k}+\frac{\sqrt{S-1}}{n}\sum_{k=1}^{M+1}\frac{N^k}{m_k}.\end{aligned}$$

En particulier, si $N^k = m_k = 3^k$ alors $M+1 \leq \ln(n)/\ln(3)$ donc il existe une constante $\gamma'(S)$ telle que pour tout $V \in \mathcal{V}$:

$$\mathbb{E}_{\sigma}\left\|\frac{1}{n}\sum_{t=1}^{n}\mathbb{1}\left\{\mu_t\in V\right\}(\mu_t-s_t)\right\| \leq \gamma'(S)\frac{\ln(n)^{\frac{3}{2}}}{\sqrt{n}}.$$

On obtient la convergence presque sûre de la même façon que Mannor et Stoltz [80] en remarquant que $\overline{G}^K_n - \mathbb{E}_\sigma\left[\overline{G}^K_n\right]$ est une moyenne de différences de martingales et en utilisant l'inégalité d'Hoeffding-Azuma en dimension quelconque (prouvée par Chen et White [31]). □

De la calibration à l'approchabilité

On a montré dans les sections précédentes que l'on peut obtenir une stratégie calibrée à partir de la construction d'une stratégie d'approchabilité d'un ensemble convexe (un orthant ou un pavé suivant les preuves) dans un jeu auxiliaire.

Réciproquement, on va montrer que l'approchabilité d'un B-set convexe peut se ramener à l'existence d'une stratégie calibrée. L'avantage de cette nouvelle preuve du théorème 2.4, contrairement aux précédentes, est qu'elle s'étend naturellement au cas d'observations partielles. L'idée de la démonstration suivante est assez proche de celle de Foster et Vohra [42].

Démonstration alternative du théorème 2.4 . Supposons que la condition (2.4) est satisfaite et reformulée en

$$\forall y \in \Delta(J), \exists x(=x_y) \in \Delta(I), \rho(x_y, y) \in C. \tag{5.4}$$

Si le joueur 1 connaissait par avance y_n alors il n'aurait qu'à jouer à l'étape n selon x_{y_n} pour que le paiement espéré $\mathbb{E}_{\sigma,\tau}[\rho_n]$ soit dans C. Comme cet ensemble est convexe, le paiement moyen espéré serait aussi dans C. Bien sûr, le joueur 1 ne connaît pas y_n, mais en utilisant la calibration il peut le prédire *assez précisément*.

Puisque ρ est multilinéaire et donc continue sur $\Delta(I) \times \Delta(J)$, pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que :

$$\forall y, y' \in \Delta(J), \|y - y'\|_2 \leq 2\delta \Rightarrow \rho(x_y, y') \in C^\varepsilon.$$

On introduit le jeu auxiliaire Γ_a où le joueur 2 choisit une action (ou un état) $j \in J$ et le joueur 1 le prédit en utilisant $\{y(l), l \in L\}$, une δ-grille finie de $\Delta(J)$. Considérons une stratégie σ calibrée du joueur 1 par rapport à cette grille. Par définition, $\overline{\jmath}_n(l)$, la distribution empirique des actions du joueur 2 sur $N_n(l)$, est asymptotiquement δ-proche de $y(l)$.

Il ne reste qu'à définir la stratégie du joueur 1 dans le jeu initial Γ. Si dans le jeu Γ_a, à l'étape n, la stratégie σ indique de jouer $l_n = l$ alors le joueur 1 joue dans Γ selon $x_{y(l)} = x(l) \in \Delta(I)$ — donné par (5.4). Puisque les choix des actions des joueurs sont indépendants, $\overline{\rho}_n(l)$ va être proche de $\rho\left(x(l), \overline{\jmath}_n(l)\right)$, donc proche de $\rho(x(l), y(l))$ (car σ est calibrée) et finalement proche de C^ε, dès que la densité supérieure de $N_n(l)$ n'est pas nulle.

En effet, par construction de σ, pour tout $\eta > 0$ il existe $N^1 \in \mathbb{N}$ tel que, pour toute stratégie τ du joueur 2 :

$$\mathbb{P}_{\sigma,\tau}\left(\forall l \in L, \forall n \geq N^1, \frac{|N_n(l)|}{n}\left(\|\overline{\jmath}_n(l) - y(l)\|_2^2 - \delta^2\right) \leq \eta\right) \geq 1 - \eta.$$

Ceci implique qu'avec probabilité au moins $1 - \eta$, pour tout $l \in L$ et $n \geq N^1$, soit $\|\overline{\jmath}_n(l) - y(l)\| \leq 2\delta$ ou $N_n(l)/n \leq \eta/3\delta^2$, donc :

$$\mathbb{P}_{\sigma,\tau}\left(\forall l \in L, \forall n \geq N^1, \frac{|N_n(l)|}{n} d\left(\rho(x(l), \overline{\jmath}_n(l)), C\right) \leq \varepsilon \frac{|N_n(l)|}{n} + \frac{\eta}{3\delta^2}\right) \geq 1 - \eta. \quad (5.5)$$

L'inégalité d'Hoeffding-Azuma [11, 61] pour les sommes de différences bornées de martingales implique que pour tout $\eta > 0$, $n \in \mathbb{N}$, σ et τ :

$$\mathbb{P}_{\sigma,\tau}\left(|\overline{\rho}_n(l) - \rho(x(l), \overline{\jmath}_n(l))| \leq \eta \big| |N_n(l)|\right) \leq 2\exp\left(-\frac{|N_n(l)|\eta^2}{2}\right),$$

donc :

$$\mathbb{P}_{\sigma,\tau}\left(\frac{|N_n(l)|}{n}|\overline{\rho}_n(l) - \rho(x(l), \overline{\jmath}_n(l))| \leq \eta\right) \leq 2\exp\left(-\frac{n\eta^2}{2}\right)$$

et en sommant sur les $n \geq N$ et $l \in L$, on obtient

$$\mathbb{P}_{\sigma,\tau}\left(\exists n \geq N, \exists l \in L, \frac{|N_n(l)|}{n}|\overline{\rho}_n(l) - \rho(x(l), \overline{\jmath}_n(l))| \leq \eta\right) \leq \frac{4L}{\eta^2}\exp\left(-\frac{N\eta^2}{2}\right). \quad (5.6)$$

Ainsi pour tout $\eta > 0$, il existe $N^2 \in \mathbb{N}$ tel que pour tout $n \geq N^2$:

$$\mathbb{P}_{\sigma,\tau}\left(\forall m \geq n, \forall l \in L, \frac{|N_n(l)|}{n} |\overline{\rho}_n(l) - \rho(x(l), \overline{\jmath}_n(l))| \leq \eta\right) \geq 1 - \eta.$$

Puisque C est un ensemble convexe, $d(\cdot, C)$ est convexe et donc avec probabilité au moins $1 - 2\eta$, pour tout $n \geq \max(N^1, N^2)$:

$$\begin{aligned} d(\overline{\rho}_n, C) &= d\left(\sum_{l \in L} \frac{|N_n(l)|}{n} \overline{\rho}_n(l), C\right) \leq \sum_{l \in L} \frac{|N_n(l)|}{n} d(\overline{\rho}_n(l), C) \\ &\leq \sum_{l \in L} \frac{|N_n(l)|}{n} d(\rho(x(l), \overline{\jmath}_n(l)), C) + \sum_{l \in L} \frac{|N_n(l)|}{n} |\overline{\rho}_n(l) - \rho(x(l), \overline{\jmath}_n(l))| \\ &\leq \varepsilon + L\eta\left(\frac{1}{3\delta^2} + 1\right). \end{aligned}$$

D'où l'ensemble C est bien approchable par le joueur 1.

Si au contraire, il existe y tel que $P^2(y) \cap C = \emptyset$, alors le joueur 2 peut approcher $P^2(y)$ en jouant à chaque étape selon y. L'ensemble C n'est donc pas approchable par le joueur 1. □

Remarque : Pour déduire que $\overline{\rho}_n$ appartient à C^ε du fait que $\overline{\rho}_n(l)$ appartient à C^ε pour tout $l \in L$, il est nécessaire que C (ou $d(\cdot, C)$) soit convexe. Donc cette démonstration n'est pas valable si C n'est pas convexe.

REMARQUES SUR L'ALGORITHME

a) Blackwell a prouvé le théorème 2.4 en utilisant le théorème de minmax de Von Neumann, ce dernier permettant de montrer qu'un ensemble convexe C qui vérifie la condition (5.4) est un B-set. En effet, soit z un point hors de C. Puisque C est convexe, pour tout $c \in C$, $\langle c - \Pi_C(z), z - \Pi_C(z)\rangle \leq 0$, où $\Pi_C(z)$ est la projection de z sur C. Donc comme pour tout $y \in \Delta(J)$, il existe $x \in \Delta(I)$ tel que $\rho(x_y, y) \in C$,

$$\forall y \in \Delta(J), \exists x \in \Delta(J), \langle \mathbf{E}_{x,y}[\rho(i,j)] - \Pi_C(z), z - \Pi_C(z)\rangle \leq 0$$

et si on définit $G(x,y) = \langle \mathbf{E}_{x,y}[\rho(i,j)] - \Pi_C(z), z - \Pi_C(z)\rangle$ alors G est linéaire en ses deux variables de sorte que

$$\min_{x \in \Delta(I)} \max_{y \in \Delta(J)} G(x,y) = \max_{y \in \Delta(J)} \min_{x \in \Delta(J)} G(x,y) \leq 0,$$

ce qui implique que C est un B-set.

La stratégie σ définie par $\sigma(h^n) = x_n$ où x_n minimise $\max_{y \in \Delta(J)} G(x,y)$ est une stratégie d'approchabilité appelé *implicite* car elle ne peut être exprimée

facilement. En effet, elle requiert de trouver, étape par étape, une action optimale dans un jeu à somme nulle, ou de manière équivalente de résoudre un programme linéaire, ce qui est assez coûteux. Il existe des algorithmes polynomiaux (voir Khachiyan [53]) cependant leurs vitesses de convergence sont plus grandes que celle de l'élimination gaussienne et leurs constantes peuvent être trop grandes pour n'importe quelle application pratique. Il est toutefois possible de trouver une solution ε-optimale en répétant un nombre polynomial de fois un *exponential weight algorithm* (voir Cesa-Bianchi et Lugosi [29], section 7.2).

Pour $\varepsilon > 0$ fixé, la stratégie que l'on a décrite et qui approche C^ε nécessite de calculer à chaque étape une mesure invariante d'une matrice à coefficients positifs. Ceci se ramène évidemment à la résolution d'un système d'équations linéaires qui est assuré d'avoir une solution. En particulier, une élimination gaussienne (comme l'ont proposé Foster et Vohra [42]) trouve une solution en un nombre polynomial (en $|L|$) d'étapes.

Supposons que les paiements soient bornés par 1. L'objectif de la stratégie que l'on a construite est que le paiement moyen converge vers C^ε. On peut ainsi choisir pour $\{y(l), l \in L\}$ n'importe quelle $\varepsilon/2$-grille de $\Delta(J)$, et donc $|L|$ est borné par $(\varepsilon/2)^{J-1}$. Il n'est pas non-plus nécessaire de déterminer exactement $x(l)$ et on peut les choisir dans n'importe quelle $\varepsilon/2$-grille de $\Delta(I)$.

En conclusion, l'algorithme *implicite* de Blackwell construit une stratégie qui approche (exactement) un convexe C en résolvant étape par étape un programme linéaire sans phase d'initialisation. Pour tout $\varepsilon > 0$, notre algorithme *explicite* construit une stratégie qui approche C^ε en résolvant, étape par étape, un système d'équations linéaires avec une phase d'initialisation, trouver les couples $(x(l), y(l)$, ce qui nécessite au plus $(\varepsilon/2)^{I+J}$ étapes.

b) Le théorème 2.4 de Blackwell dit que si pour chaque action $y \in \Delta(J)$ du joueur 2, le joueur 1 a une action $x \in \Delta(I)$ telle que $\rho(x, y) \in C$ alors ce dernier peut approcher C. En d'autres termes, si dans le jeu (d'espaces d'actions $\Delta(I)$ et $\Delta(J)$) en un coup où le joueur 2 joue en premier et le joueur 1 en second, ce dernier a une stratégie telle que le paiement est dans le convexe C, alors il a aussi une stratégie d'approchabilité de C dans le jeu répété où le joueur 2 joue en second et lui en premier.
L'utilisation de la calibration permet de transformer cet argument implicite en un argument explicite : en exécutant une stratégie calibrée (dans un jeu auxiliaire où J joue le rôle de l'ensemble d'états), le joueur 1 peut forcer la distribution empirique moyenne des actions du joueur 2 sur $N_n(l)$ à être proche de $y(l)$. Ainsi il n'a qu'à (et il ne pourrait faire mieux) jouer selon $x_{y(l)}$ sur ces étapes.

c) La construction d'une stratégie d'approchabilité de C^ε se ramène à la construction d'une stratégie calibrée dans un premier jeu auxiliaire, donc à la construction d'une stratégie consistante intérieurement dans un second jeu auxiliaire,

donc à l'approchabilité d'un orthant dans un troisième jeu auxiliaire. Ainsi, l'approchabilité d'un ensemble convexe arbitraire se ramène à l'approchabilité d'un orthant. En particulier, ceci implique que $\mathbf{E}_{\sigma,\tau}\left[d\left(\overline{\rho}_n, C\right) - \varepsilon\right] \leq O\left(n^{-1/2}\right)$. La constante peut dépendre crucialement de ε.

d) La réduction de l'approchabilité d'un ensemble convexe $C \subset \mathbb{R}^d$ dans un jeu Γ à l'approchabilité d'un orthant dans un jeu auxiliaire Γ' peut aussi être faite de la façon suivante. Pour tout $\varepsilon > 0$, il existe un ensemble fini de demi-espaces $\{H(l), l \in L\}$ tels que $C \subset \cap_{l \in L} H(l) \subset C^{\varepsilon}$. Pour tout $l \in L$, on définit $c(l) \in \mathbb{R}^d$ et $b(l) \in \mathbb{R}$ par

$$H(l) = \left\{\omega \in \mathbb{R}^d, \langle \omega, c(l) \rangle \leq b(l)\right\}$$

et le jeu auxiliaire Γ' avec les paiements donnés par

$$\widehat{\rho}(i,j) = \left(\langle \rho(i,j), c(l) \rangle - b(l)\right)_{l \in L} \in \mathbb{R}^L.$$

Une stratégie qui approche l'orthant négatif dans Γ' va approcher, dans le jeu Γ, l'ensemble $\bigcap_{l \in L} H(l)$ et donc C^{ε}. Cependant, une telle stratégie ne sera peut être pas basée sur la consistance interne et pourra ne pas être explicite.

Deuxième partie

Jeux à Observations Partielles

6

Internal Consistency with Partial Monitoring

Ce chapitre introduit les jeux avec observations partielles. On rappelle les définitions de regrets externe et interne et on construit des stratégies consistantes. Il est issu de l'article Calibration and Internal no-Regret with Partial Monitoring*, dont un résumé étendu est publié dans les* Proceedings of the 20-th conference on Algorithmic Learning Theory.

Sommaire

6.1 Full monitoring case : from approachability to calibration **79**
From approachability to internal no-regret 84
From internal regret to calibration 85
From calibration to approachability 86
6.2 Internal regret in the partial monitoring framework **90**
External regret 91
Internal regret 91
On the strategy space 96
6.3 Back on payoff space **96**
The worst case fulfills Assumption 6.15 97
Compact case 98
Regret in terms of actual payoffs 99
External and internal consistency 101

CALIBRATION, approachability and regret are three notions widely used both in game theory and machine learning. There are, at first glance, no obvious links between them. Indeed, calibration has been introduced by Dawid [33] for repeated games of predictions : at each stage, Nature chooses an outcome s in a finite set S and Predictor forecasts it by announcing, stage by stage, a probability over S. A strategy is calibrated if the empirical distribution of outcomes on the set of stages where Predictor made a specific forecast is close to it. Foster and Vohra [43] proved the

existence of such strategies. Approachability has been introduced by Blackwell [17] in two-person repeated games, where at each stage the payoff is a vector in $\mathbb{R}^d$: a player can approach a given set $E \subset \mathbb{R}^d$, if he can insure that, after some stage and with a great probability, the average payoff will always remain close to E. This is possible, see Blackwell [17], as soon as E satisfies some geometrical condition (it is then called a B-set) and this gives a full characterization for the special case of convex sets. No-regret has been introduced by Hannan [56] for two-person repeated games with payoffs in $\mathbb{R}$: a player has no external regret if his average payoff could not have been asymptotically better by knowing in advance the empirical distribution of moves of the other players. The existence of such strategies was also proved by Hannan [56].

Blackwell [18] (see also Luce and Raifa [77], A.8.6 and Mertens, Sorin and Zamir [84], Exercice 7 p. 107) was the first to notice that the existence of externally consistent strategies (strategies that have no external regret) can be proved using his approachability theorem. As shown by Hart and Mas-Colell [57], the use of Blackwell's theorem actually gives not only the existence of externally consistent strategies but also a construction of strategies that fulfill a stronger property, called internal consistency : a player has asymptotically no internal regret, if for each of his actions, he has no external regret on the set of stages where he played it (as long as this set has a positive density). This more precise definition of regret has been introduced by Foster and Vohra [42] (see also Fudenberg and Levine [51]).

Foster and Vohra [43] (see also Sorin [108] for a shorter proof) constructed calibrated strategy by computing, in an auxiliary game, a strategy with no internal regret. These results are recalled in section 6.1 and we also refer to Cesa-Bianchi and Lugosi [29] for more complete survey on sequential prediction and regret.

We provide in section 6.1 a kind of converse result : an explicit ε-approachability strategy for a convex B-set is constructed through the use of a calibrated strategy, in some auxiliary game. This last statement proves that the construction of an approachability strategy of a convex set can be deduced from the construction of a calibrated strategy, which is deduced from the construction of an internally consistent strategy, itself deduced from the construction of an approachability strategy. So calibration, regret and approachability are, in some sense, *equivalent*.

In section 6.2, we consider repeated games with partial monitoring, *i.e.* where players do not observe the action of their opponents, but receive random signals. The idea behind the proof that, in the full monitoring case, approachability follows from calibration can be extended to this new framework to construct consistent strategies in the following sense. A player has asymptotically no external regret if his average payoff could not have been better by knowing in advance the empirical distribution of signals (see Rustichini [100]). The existence of strategies with no external regret was proved by Rustichini [100] while Lugosi, Mannor and Stoltz [78] constructed explicitly such strategies. The notion of internal regret was introduced by Lehrer and Solan [74] and they proved the existence of consistent strategies (in the special case where the signal received do not depend on the player's action). Our main result is the

construction of such strategies even when the signal depends on the action played. We show in section 6.3 that our algorithm also works when the opponent is not restricted to a finite number of actions, discuss our assumption on the regularity of the payoff function (see Assumption 6.15) and extend our framework to more general cases.

6.1 Full monitoring case : from approachability to calibration

We recall the main results about calibration of Foster and Vohra [43], approachability of Blackwell [17] and regret of Hart and Mas-Colell [57]. We will prove some of these results in details, since they give the main ideas about the construction of strategies in the partial monitoring framework, given in section 6.2.

CALIBRATION

We consider a two-person repeated game where, at stage $n \in \mathbb{N}$, Nature (Player 2) chooses an outcome s_n in a finite set S and Predictor (Player 1) forecasts it by choosing μ_n in $\Delta(S)$, the set of probabilities over S. We assume furthermore that μ_n belongs to a finite set $\mathcal{M} = \{\mu(l), l \in L\}$. The prediction at stage n is then the choice of an element $l_n \in L$, called the *type of that stage*. The choices of l_n and s_n depend on the past observations $h_{n-1} = (l_1, s_1, \ldots, l_{n-1}, s_{n-1})$ and may be random. Explicitly, the set of finite histories is denoted by $H = \bigcup_{n\in\mathbb{N}} (L \times S)^n$, with $(L \times S)^0 = \emptyset$ and a behavioral strategy σ of Player 1 is a mapping from H to $\Delta(L)$. Given a finite history $h_n \in (L \times S)^n$, $\sigma(h_n)$ is the law of l_{n+1}. A strategy τ of Nature is defined similarly as a mapping from H to $\Delta(S)$. A couple of strategies (σ, τ) generates a probability, denoted by $\mathbb{P}_{\sigma,\tau}$, over $\mathcal{H} = (L \times S)^{\mathbb{N}}$, the set of plays endowed with the cylinder σ-field.

We will use the following notations. For any families $\mathbf{a} = \{a_m \in \mathbb{R}^d\}_{m\in\mathbb{N}}$ and $\mathbf{l} = \{l_m \in L\}_{m\in\mathbb{N}}$ and any integer $n \in \mathbb{N}$, $N_n(l) = \{1 \leq m \leq n, l_m = l\}$ is the set of stages of type l (before the n-th), $\overline{a}_n(l) = \frac{1}{N_n(l)} \sum_{m\in N_n(l)} a_m$ is the average of $\mathbf{a}$ on this set and $\overline{a}_n = \frac{1}{n} \sum_{m=1}^{n} a_m$ is the average of $\mathbf{a}$ over the n first stages.

Definition 6.1 (*Dawid [33]*)

A strategy σ of Player 1 is calibrated with respect to $\mathcal{M}$ if for every $l \in L$ and every strategy τ of Player 2 :

$$\limsup_{n\to+\infty} \frac{|N_n(l)|}{n} \left(\|\overline{s}_n(l) - \mu(l)\|_2^2 - \|\overline{s}_n(l) - \mu(k)\|_2^2 \right) \leq 0, \quad \forall k \in L, \mathbb{P}_{\sigma,\tau}\text{-as} \ , \quad (6.1)$$

where $\Delta(S)$ is seen as a subset of $\mathbb{R}^{|S|}$.

In words, a strategy of Player 1 is calibrated with respect to $\mathcal{M}$ if $\overline{s}_n(l)$, the empirical distribution of outcomes when $\mu(l)$ was predicted, is asymptotically closer to $\mu(l)$ than to any other $\mu(k)$ (or conversely, that $\mu(l)$ is the closest possible prediction to $\overline{s}_n(l)$), as long as $|N_n(l)|/n$, the frequency of l, does not go to 0. Foster and Vohra [43] proved the existence of such strategies with an algorithm based on the Expected Brier Score.

An alternative (and more general) way of defining calibration is the following. Player 1 is not restricted to make prediction in a finite set $\mathcal{M}$ and, at each stage, he can choose any probability in $\Delta(S)$. Consider any finite partition $\mathcal{P} = \{P(k), k \in K\}$ of $\Delta(S)$ with a diameter small enough (we recall that the diameter of a partition is defined as $\max_{k\in K} \max_{x,y\in P(k)} \|x - y\|$). A strategy is ε-calibrated if the empirical distribution of outcomes (denoted by $\overline{s}_n(k)$) when the prediction is in $P(k)$ is asymptotically ε-close to $P(k)$ (as long as the frequency of $k \in K$ does not go to zero). Formally :

Definition 6.2

A strategy σ of Player 1 is ε-calibrated if there exists $\overline{\eta} > 0$ such that for every finite partition $\mathcal{P} = \{P(k), k \in K\}$ of $\Delta(S)$ with diameter smaller than $\overline{\eta}$ and every strategy τ of Player 2 :

$$\limsup_{n\to+\infty} \frac{|N_n(k)|}{n}\left(d^2\big(\overline{s}_n(k), P(k)\big) - \varepsilon^2\right) \leq 0, \quad \forall k \in k, \mathbb{P}_{\sigma,\tau}\text{-as} \ , \tag{6.2}$$

where for every set $E \subset \mathbb{R}^d$ and $z \in \mathbb{R}^d$, $d(z, E) = \inf_{e\in E} \|z - e\|_2$.

The following Lemma 6.3 states a calibrated strategy with respect to a grid (as in Definition 6.1) is ε-calibrated (as in Definition 6.2), therefore we will only use the first formulation.

Lemma 6.3

For every $\varepsilon > 0$, there exists a finite set $\mathcal{M} = \{\mu(l), l \in L\}$ such that any calibrated strategy with respect to $\mathcal{M}$ is ε-calibrated.

Proof. Let $\mathcal{M} = \{\mu(l), l \in L\}$ be a finite ε-grid of $\Delta(S)$: for every probability $\mu \in \Delta(S)$, there exists $\mu(l) \in \mathcal{M}$ such that $\|\mu - \mu(l)\| \leq \varepsilon$. In particular, for every $l \in L$ and $n \in \mathbb{N}$, there exists $l' \in L$ such that $\|\overline{s}_n(l) - \mu(l')\| \leq \varepsilon$. Equation (6.1) implies then that

$$\limsup_{n\to\infty} \frac{|N_n(l)|}{n}\left(d^2(\overline{s}_n(l), \mu(l)) - \varepsilon^2\right) \leq 0, \mathbb{P}_{\sigma,\tau}\text{-as} \ .$$

Let $2\overline{\eta}$ be the smallest distance between any two different $\mu(l)$ and $\mu(l')$. In any finite partition $\mathcal{P} = \{P(k), k \in K\}$ of $\Delta(S)$ of diameter smaller $\overline{\eta}$, $\mu(l)$ belongs to at most one $P(k)$. Hence σ is obviously ε-calibrated. $\square$

Remark : Lemma 6.3 implies that one can construct an ε-calibrated strategy as soon as he can construct a calibrated strategy with respect to a finite ε-grid of $\Delta(S)$. The size of this grid is in the order of $\varepsilon^{-|S|}$ (exponential in ε) and it is not known yet if there exists an efficient algorithm (polynomial in ε) to compute ε-calibration. The results holds with condition (6.2) replaced by

$$\limsup_{n\to+\infty} \frac{|N_n(k)|}{n}\Big(d\big(\overline{s}_n(k), P(k)\big) - \varepsilon\Big) \leq 0, \quad \forall k \in k, \mathbb{P}_{\sigma,\tau}\text{-as}$$

however Lemma 6.3 is trivially true with the square terms $d^2(\overline{s}_n(k), P(k))$ and ε^2.

APPROACHABILITY

We will prove in the following subsection that calibration follows from no-regret and that no-regret follows from approachability (proofs originally due to, respectively, Foster and Vohra [43] and Hart and Mas-Colell [57]). We present here the notion of approachability introduced by Blackwell [17].

Consider a two-person game repeated in discrete time with vector payoffs, where at stage $n \in \mathbb{N}$, Player 1 (resp. Player 2) chooses the action $i_n \in I$ (resp. $j_n \in J$), where both I and J are finite. The corresponding vector payoff is $\rho_n = \rho(i_n, j_n)$ where ρ is a mapping from $I \times J$ into $\mathbb{R}^d$. As usual, a behavioral strategy σ (resp. τ) of Player 1 (resp. Player 2) is a mapping from the set of finite histories $H = \bigcup_{n\in\mathbb{N}} (I \times J)^n$ to $\Delta(I)$ (resp. $\Delta(J)$).

For a closed set $E \subset \mathbb{R}^d$ and $\delta \geq 0$, we denote by $E^\delta = \{z \in \mathbb{R}^d, d(z, E) \leq \delta\}$ the δ-neighborhood of E and by $\Pi_E(z) = \{e \in E, d(z, E) = \|z - e\|\}$ the set of closest points to z in E.

Definition 6.4

i) A closed set $E \subset \mathbb{R}^d$ is approachable by Player 1 if for every $\varepsilon > 0$, there exist a strategy σ of Player 1 and $N \in \mathbb{N}$, such that for every strategy τ of Player 2 and every $n \geq N$:

$$\mathbf{E}_{\sigma,\tau}\left[d(\overline{\rho}_n, E)\right] \leq \varepsilon \quad \textit{and} \quad \mathbb{P}\left(\sup_{n\geq N} d(\overline{\rho}_n, E) \geq \varepsilon\right) \leq \varepsilon \ .$$

Such a strategy σ, independent of ε, is called an approachability strategy of E.

ii) A set E is excludable by Player 2, if there exists $\delta > 0$ such that the complement of E^δ is approachable by Player 2.

In words, a set $E \subset \mathbb{R}^d$ is approachable by Player 1, if he has a strategy such that the average payoff converges almost surely to E, uniformly with respect to the strategies of Player 2.

Blackwell [17] noticed that a closed set E that fulfills a purely geometrical condition (see Definition 6.5) is approachable by Player 1. Before stating it, let us denote by $P^1(x) = \{\rho(x,y), y \in \Delta(J)\}$, the set of expected payoffs compatible with $x \in \Delta(I)$ and we define similarly $P^2(y)$.

Definition 6.5

A closed subset E of $\mathbb{R}^d$ is a B-set, if for every $z \in \mathbb{R}^d$, there exist $p \in \Pi_E(z)$ and $x\,(= x(z,p)) \in \Delta(I)$ such that the hyperplane through p and perpendicular to $z - p$ separates z from $P^1(x)$, or formally :

$$\forall z \in \mathbb{R}^d, \exists p \in \Pi_E(z), \exists x \in \Delta(I), \langle \rho(x,y) - p, z - p \rangle \leq 0, \quad \forall y \in \Delta(J) \ . \tag{6.3}$$

Informally, from any point z outside E there is a closest point p and a probability $x \in \Delta(I)$ such that, no matter the choice of Player 2, the expected payoff and z are on different sides of the hyperplane through p and perpendicular to $z - p$. To be precise, this definition (and the following theorem) does not require that J is finite : one can assume that Player 2 chooses an outcome vector $U \in [-1,1]^{|I|}$ so that the expected payoff is $\rho(x,U) = \langle x, U \rangle$.

Theorem 6.6 (*Blackwell [17]*)

If E is a B-set, then E is approachable by Player 1. Moreover, the strategy σ of Player 1 defined by $\sigma(h_n) = x(\overline{\rho}_n)$ is such that, for every strategy τ of Player 2 :

$$\mathbf{E}_{\sigma,\tau}[d_E^2(\overline{\rho}_n)] \leq \frac{4B}{n} \quad \text{and} \quad \mathbb{P}_{\sigma,\tau}\left(\sup_{n \geq N} d(\overline{\rho}_n, E) \geq \eta\right) \leq \frac{8B}{\eta^2 N} \ , \tag{6.4}$$

with $B = \sup_{i,j} \|\rho(i,j)\|^2$.

In the case of a convex set C, a complete characterization is available :

Corollary 6.7 (*Blackwell [17]*)

A closed convex set $C \subset \mathbb{R}^d$ is approachable by Player 1 if and only if :

$$P^2(y) \cap C \neq \emptyset, \quad \forall y \in \Delta(J) \ . \tag{6.5}$$

In particular, a closed convex set C is either approachable by Player 1, or excludable by Player 2.

Remark : Corollary 6.7 implies that there are (at least) two different ways to prove that a convex set is approachable. The first one, called direct proof, consists in proving that C is a B-set while the second one, called undirect proof, consists

in proving that C is not excludable by Player 2, which reduces to find, for every $y \in \Delta(J)$, some $x \in \Delta(I)$ such that $\rho(x, y) \in C$.

Consider a two-person repeated game in discrete time where, at stage $n \in \mathbb{N}$, Player 1 chooses $i_n \in I$ as above and Player 2 chooses a vector $U_n \in [-1, 1]^c$ (with $c = |I|$). The associated payoff is $U_n^{i_n}$, the i_n-th coordinate of U_n. The internal regret of the stage is the matrix $R_n = R(i_n, U_n)$, where R is the mapping from $I \times [-1, 1]^c$ to $\mathbb{R}^{c^2}$ defined by :

$$R(i, U)^{(i',j)} = \begin{cases} 0 & \text{if} \quad i' \neq i \\ U^j - U^i & \text{otherwise.} \end{cases}$$

With this definition, the average internal regret $\overline{R}_n$ is defined by :

$$\overline{R}_n = \left[\frac{\sum_{m \in N_n(i)} (U_m^j - U_m^i)}{n} \right]_{i,j \in I} = \left[\frac{|N_n(i)|}{n} \left(\overline{U}_n(i)^j - \overline{U}_n(i)^i \right)_{j \in I} \right]_{i \in I} .$$

Definition 6.8 (*Foster and Vohra [42]*)

A strategy σ of Player 1 is internally consistent if for any strategy τ of Player 2 :

$$\limsup_{n \to \infty} \overline{R}_n \leq 0, \quad \mathbb{P}_{\sigma,\tau}\text{-as} \ .$$

In words, a strategy is internally consistent if for every $i \in I$ (with a positive frequency), Player 1 could not have increased his payoff if he had known, before the beginning of the game, the empirical distribution of Player 2's actions on $N_n(i)$. Stated differently, when Player 1 played action i, it was his best (stationary) strategy. The existence of such strategies have been first proved by Foster and Vohra [42] and Fudenberg and Levine [51].

Theorem 6.9

There exist internally consistent strategies.

Hart and Mas-Colell [57] noted that an internally consistent strategy can be obtained by constructing a strategy that approaches the negative orthant $\Omega = \mathbb{R}_-^{c^2}$ in the auxiliary game where the vector payoff at stage n is R_n. Such a strategy, derived from approachability theory, is stronger than just internally consistent since the regret converges to the negative orthant uniformly with respect to Player 2's strategy (which was not required in Definition 6.8).

The proof of the fact that Ω is a B-set relies on the two followings lemmas : Lemma 6.10 gives a geometrical property of Ω and Lemma 6.11 gives a property of the function R.

From approachability to internal no-regret

Lemma 6.10

Let $\Pi_\Omega(\cdot)$ be the projection onto Ω. Then, for every $A \in \mathbb{R}^{c^2}$:

$$\langle \Pi_\Omega(A), A - \Pi_\Omega(A) \rangle = 0 \ . \tag{6.6}$$

Proof. Note that since $\Omega = \mathbb{R}_-^{c^2}$ then $A^+ = A - \Pi_\Omega(A)$ where $A_{ij}^+ = \max(A_{ij}, 0)$ and similarly $A^- = \Pi_\Omega(A)$. The result is just a rewriting of $\langle A^-, A^+ \rangle = 0$. □

For every $(c \times c)$-matrix $A = (a_{ij})_{i,j \in I}$ with non-negative coefficients, $\lambda \in \Delta(I)$ is an invariant probability of A if for every $i \in I$:

$$\sum_{j \in I} \lambda(j) a_{ji} = \lambda(i) \sum_{j \in I} a_{ij} \ .$$

The existence of an invariant probability follows from the similar result for Markov chains, implied by Perron-Frobenius Theorem (see *e.g.* Bapat and Raghavan [12], Theorem 1.4.4 p. 17 and Theorem 1.7.3 p. 35).

Lemma 6.11

Let $A = (a_{ij})_{i,j \in I}$ be a non-negative matrix. Then for every λ, invariant probability of A, and every $U \in \mathbb{R}^c$:

$$\langle A, \mathbf{E}_\lambda [R(\cdot, U)] \rangle = 0 \ . \tag{6.7}$$

Proof. The (i,j)-th coordinate of $\mathbf{E}_\lambda [R(\cdot, U)]$ is $\lambda(i)(U^j - U^i)$, therefore :

$$\langle A, \mathbf{E}_\lambda [R(\cdot, U)] \rangle = \sum_{i,j \in I} a_{ij} \lambda(i) \left(U^j - U^i \right)$$

and the coefficient of each U^i is $\sum_{j \in I} a_{ij} \lambda(i) - \sum_{j \in I} a_{ji} \lambda(j) = 0$, because λ is an invariant measure of A. Therefore $\langle A, \mathbf{E}_\lambda [R(\cdot, U)] \rangle = 0$. □

Proof of Theorem 6.9. Summing equations (6.6) (with $A = \overline{R}_n$) and (6.7) (with $A = \left(\overline{R}_n\right)^+$) gives :

$$\left\langle \mathbf{E}_{\lambda_n} [R(\cdot, U)] - \Pi_\Omega(\overline{R}_n), \overline{R}_n - \Pi_\Omega(\overline{R}_n) \right\rangle = 0 \ ,$$

for every λ_n invariant probability of $\overline{R}_n^+$ and every $U \in [-1, 1]^I$.

Define the strategy σ of Player 1 by $\sigma(h_n) = \lambda_n$. The expected payoff at stage $n+1$ (given h_n and $U_{n+1} = U$) is $\mathbf{E}_{\lambda_n} [R(\cdot, U)]$, so Ω is a B-set and is approachable by Player 1. □

Remark : The construction of the strategy is based on approachability properties therefore the convergence is uniform with respect to the strategies of Player 2. Theorem 6.6 implies that for every $\eta > 0$, and for every strategy τ of Player 2 :

$$\mathbb{P}_{\sigma,\tau}\left(\exists n \geq N, \exists i, j \in i, \frac{|N_n(i)|}{n}\left(\overline{U}_n(i)^j - \overline{U}_n(i)^i\right) > \eta\right) = O\left(\frac{1}{\eta^2 N}\right)$$

$$\text{and} \quad \mathbf{E}_{\sigma,\tau}\left[\sup_{i\in I}\frac{|N_n(l)|}{n}\left(\overline{U}_n(i)^j - \overline{U}_n(i)^i\right)^+\right] = O\left(\frac{1}{\sqrt{n}}\right) \ .$$

Although they are not required by definition 6.8, those bounds will be useful to prove that calibration implies approachability.

From internal regret to calibration

The construction of calibrated strategies can be reduced to the construction of internally consistent strategies. The proof of Sorin [108] simplifies the one originally due to Foster and Vohra [42] by using the following lemma :

Lemma 6.12

Let $(a_m)_{m\in\mathbb{N}}$ be a sequence in $\mathbb{R}^d$ and α, β two points in $\mathbb{R}^d$. Then for every $n \in \mathbb{N}^*$:

$$\frac{\sum_{m=1}^n \|a_m - \beta\|_2^2 - \|a_m - \alpha\|_2^2}{n} = \|\overline{a}_n - \beta\|_2^2 - \|\overline{a}_n - \alpha\|_2^2 \ , \tag{6.8}$$

with $\|\cdot\|_2$ the Euclidian norm of $\mathbb{R}^d$.

Proof. Develop the sums in equation (6.8) to get the result. □

Now, we can prove the following :

Theorem 6.13 (*Foster and Vohra [42]*)

For every finite grid $\mathcal{M}$ of $\Delta(S)$, there exist calibrated strategies of Player 1 with respect to $\mathcal{M}$. In particular, for every $\varepsilon > 0$ there exist ε-calibrated strategies.

Proof. We start with the framework described in 6.1. Consider the auxiliary two-person game with vector payoff defined as follows. At stage $n \in \mathbb{N}$, Player 1 (resp. Player 2) chooses the action $l_n \in L$ (resp. $s_n \in S$) which generates the vector payoff $R_n = R(l_n, U_n) \in \mathbb{R}^d$, where R is as in 6.1, with :

$$U_n = \left(-\|s_n - \mu(l)\|_2^2\right)_{l\in L} \in \mathbb{R}^c \ .$$

By definition of R and using Lemma 6.12, for every $n \in \mathbb{N}^*$:

$$\begin{aligned}\overline{R}_n^{lk} &= \frac{|N_n(l)|}{n}\left(\frac{\sum_{m\in N_n(l)}\|s_m-\mu(l)\|_2^2-\|s_m-\mu(k)\|_2^2}{|N_n(l)|}\right)\\ &= \frac{|N_n(l)|}{n}\left(\|\overline{s}_n(l)-\mu(l)\|_2^2-\|\overline{s}_n(l)-\mu(k)\|_2^2\right).\end{aligned}$$

Let σ be an internally consistent strategy in this auxiliary game, then for every $l \in L$ and $k \in L$:

$$\limsup_{n\to\infty}\frac{|N_n(l)|}{n}\left(\|\overline{s}_n(l)-\mu(l)\|_2^2-\|\overline{s}_n(k)-\mu(k)\|_2^2\right)\leq 0,\quad \mathbb{P}_{\sigma,\tau}\text{-as}\ .$$

Therefore σ is calibrated, with respect to $\mathcal{M}$; if it is an ε-grid of $\Delta(S)$, then σ is ε-calibrated. □

Remark : We have proved that σ is such that, for every $l \in L$, $\overline{s}_n(l)$ is closer to $\mu(l)$ than to any other $\mu(k)$, as soon as $|N_n(l)|/n$ is not too small.

The facts that s_n belongs to a finite set S and $\{\mu(l)\}$ are probabilities over S are irrelevant : one can show that for any finite set $\{a(l) \in \mathbb{R}^d, l \in L\}$, Player 1 has a strategy σ such that for any bounded sequence $(a_m)_{m\in\mathbb{N}}$ in $\mathbb{R}^d$ and for every l and k :

$$\limsup_{n\to\infty}\frac{|N_n(l)|}{n}\left(\|\overline{a}_n(l)-a(l)\|^2-\|\overline{a}_n(l)-a(k)\|^2\right)\leq 0\ .$$

From calibration to approachability

The proof of Theorem 6.13 shows that the construction of a calibrated strategy can be obtained through an approachability strategy of an orthant in an auxiliary game.

Conversely, we will show that the approachability of a convex B-set can be reduced to the existence of a calibrated strategy in an auxiliary game, and so give a new proof of Corollary 6.7 (and mainly construct explicit strategies).

Alternative proof of Corollary 6.7 : The idea of the proof is very natural : assume that condition (6.5) is satisfied and rephrased as :

$$\forall y \in \Delta(J), \exists x(=x_y) \in \Delta(I), \rho(x_y, y) \in C\ . \tag{6.9}$$

If Player 1 knew in advance y_n then he would just have to play accordingly to x_{y_n} at stage n so that the expected payoff $\mathbf{E}_{\sigma,\tau}[\rho_n]$ would be in C. Since C is convex, the average payoff would also be in C. Obviously Player 1 does not know y_n but, using calibration, he can make *good* predictions about it.

Since ρ is multilinear and therefore continuous on $\Delta(I) \times \Delta(J)$, for every $\varepsilon > 0$, there exists $\delta > 0$ such that :

$$\forall y, y' \in \Delta(J), \|y - y'\|_2 \leq 2\delta \Rightarrow \rho(x_y, y') \in C^\varepsilon \ .$$

We introduce the auxiliary game Γ where Player 2 chooses an action (or outcome) $j \in J$ and Player 1 forecasts it by using $\{y(l), l \in L\}$, a finite δ-grid of $\Delta(J)$. Let σ be a calibrated strategy for Player 1, so that $\bar{\jmath}_n(l)$, the empirical distribution of actions of Player 2 on $N_n(l)$, is asymptotically δ-close to $y(l)$.

Define the strategy of Player 1 in the initial game by performing σ and if $l_n = l$ by playing accordingly to $x_{y(l)} = x(l) \in \Delta(I)$, as depicted in (6.9). Since the choices of actions of the two players are independent, $\overline{\rho}_n(l)$ will be close to $\rho\left(x(l), \bar{\jmath}_n(l)\right)$, hence close to $\rho(x(l), y(l))$ (because σ is calibrated) and finally close to C^ε, as soon as $|N_n(l)|$ is not too small.

Indeed, by construction of σ, for every $\eta > 0$ there exists $N_1 \in \mathbb{N}$ such that, for every strategy τ of Player 2 :

$$\mathbb{P}_{\sigma,\tau}\left(\forall l \in L, \forall n \geq N_1, \frac{|N_n(l)|}{n}\left(\|\bar{\jmath}_n(l) - y(l)\|_2^2 - \delta^2\right) \leq \eta\right) \geq 1 - \eta \ .$$

This implies that with probability greater than $1 - \eta$, for every $l \in L$ and $n \geq N_1$, either $\|\bar{\jmath}_n(l) - y(l)\| \leq 2\delta$ or $N_n(l)/n \leq \eta/3\delta^2$, so :

$$\mathbb{P}_{\sigma,\tau}\left(\forall l \in L, \forall n \geq N_1, \frac{|N_n(l)|}{n} d\left(\rho(x(l), \bar{\jmath}_n(l)), C\right) \leq \varepsilon \frac{|N_n(l)|}{n} + \frac{\eta}{3\delta^2}\right) \geq 1 - \eta \ . \tag{6.10}$$

Hoeffding-Azuma [11, 61] inequality for sums of bounded martingale differences implies that for any $\eta > 0$, $n \in \mathbb{N}$, σ and τ :

$$\mathbb{P}_{\sigma,\tau}\left(|\overline{\rho}_n(l) - \rho(x(l), \bar{\jmath}_n(l))| \leq \eta \big| |N_n(l)|\right) \leq 2\exp\left(-\frac{|N_n(l)|\eta^2}{2}\right) \ ,$$

therefore :

$$\mathbb{P}_{\sigma,\tau}\left(\frac{|N_n(l)|}{n}|\overline{\rho}_n(l) - \rho(x(l), \bar{\jmath}_n(l))| \leq \eta\right) \leq 2\exp\left(-\frac{n\eta^2}{2}\right)$$

and summing over $n \in \{N, \ldots, \}$ and $l \in L$ gives

$$\mathbb{P}_{\sigma,\tau}\left(\exists n \geq N, \exists l \in L, \frac{|N_n(l)|}{n}|\overline{\rho}_n(l) - \rho(x(l), \bar{\jmath}_n(l))| \leq \eta\right) \leq \frac{4L}{\eta^2}\exp\left(-\frac{N\eta^2}{2}\right) \ . \tag{6.11}$$

So for every $\eta > 0$, there exists $N_2 \in \mathbb{N}$ such that for every $n \geq N_2$:

$$\mathbb{P}_{\sigma,\tau}\left(\forall m \geq n, \forall l \in L, \frac{|N_n(l)|}{n}|\overline{\rho}_n(l) - \rho(x(l), \bar{\jmath}_n(l))| \leq \eta\right) \geq 1 - \eta \ .$$

Since C is a convex set, $d(\cdot, C)$ is convex and with probability at least $1-2\eta$, for every $n \geq \max(N_1, N_2)$:

$$\begin{aligned} d(\overline{\rho}_n, C) &= d\left(\sum_{l\in L} \frac{|N_n(l)|}{n}\overline{\rho}_n(l), C\right) \leq \sum_{l\in L} \frac{|N_n(l)|}{n} d(\overline{\rho}_n(l), C) \\ &\leq \sum_{l\in L} \frac{|N_n(l)|}{n} d(\rho(x(l), \overline{\jmath}_n(l)), C) + \sum_{l\in L} \frac{|N_n(l)|}{n} |\overline{\rho}_n(l) - \rho(x(l), \overline{\jmath}_n(l))| \\ &\leq \varepsilon + L\eta\left(\frac{1}{3\delta^2}+1\right). \end{aligned}$$

And C is approachable by Player 1.

On the other hand, if there exists y such that $P^2(y) \cap C = \emptyset$, then Player 2 can approach $P^2(y)$, by playing at every stage accordingly to y. Therefore C is not approachable by Player 1. □

Remark :: To deduce that $\overline{\rho}_n$ is in C^ε from the fact that $\overline{\rho}_n(l)$ is in C^ε for every $l \in L$, it is necessary that C (or $d(\cdot, C)$) is convex. So this proof does not work if C is not convex.

REMARKS ON THE ALGORITHM

a) Blackwell proved Corollary 6.7 using Von Neumann's minmax theorem, the latter allowing to show that a convex set C that fulfills condition (6.9) is a B-set. Indeed, let z be a point outside C. Recall that for every $y \in \Delta(J)$ there exists $x_y \in \Delta(I)$ such that $\rho(x_y, y) \in C$. Since C is convex, if we denote by $\Pi_C(z)$ the projection of z onto it, then for every $c \in C$ $\langle c - \Pi_C(z), z - \Pi_C(z)\rangle \leq 0$, . Therefore,

$$\forall y \in \Delta(J), \exists x \in \Delta(J), \langle \mathbf{E}_{x,y}[\rho(i,j)] - \Pi_C(z), z - \Pi_C(z)\rangle \leq 0$$

and if we define $g(x,y) = \langle \mathbf{E}_{x,y}[\rho(i,j)] - \Pi_C(z), z - \Pi_C(z)\rangle$ then g is linear in both of its variable so

$$\min_{x\in\Delta(I)} \max_{y\in\Delta(J)} g(x,y) = \max_{y\in\Delta(J)} \min_{x\in\Delta(J)} g(x,y) \leq 0,$$

which implies that C is a B-set.
The strategy σ defined by $\sigma(h_n) = x_n$ where x_n minimizes $\max_{y\in\Delta(J)} G(x,y)$ is an approachability strategy, said to be *implicit* since there are no easy way to construct it. Indeed computing σ would require to find, stage by stage, an optimal action in a zero-sum game or equivalently to solve a Linear Program. There exist polynomial algorithms (see Khachiyan [53]) however their rates of convergence are bigger than the one of Gaussian elimination and their constants can be too huge for any practical use. However, it is possible to find ε-optimal

solution by repeating an polynomial number of time the *exponential weight algorithm* (see Cesa-Bianchi and Lugosi [29], Section 7.2).
For a fixed $\varepsilon > 0$, the strategy (that approaches C^ε) we described computes at each stage an invariant measure of a matrix with non-negative coefficients. This obviously reduces to solve a system of linear equations which is guaranteed to have a solution. And this is solved polynomially (in $|L|$) by, for example and as proposed by Foster and Vohra [42], a Gaussian elimination. If payoffs are bounded by 1, then one can take for $\{y(l), l \in L\}$ any arbitrarily $\varepsilon/2$-grid of $\Delta(J)$, so $|L|$ is bounded by $(2/\varepsilon)^{|J|}$. Moreover, the strategy aims to approach C^ε, so it is not compulsory to determine exactly $x(l)$, one can choose them in any $\varepsilon/2$-grid of $\Delta(I)$.
In conclusion, Blackwell's *implicit* algorithm constructs a strategy that approaches (exactly) a convex C by solving, stage by stage, a Linear Program without any initialization phase. For every $\varepsilon > 0$, our *explicit* algorithm constructs a strategy that approaches C^ε by solving, stage by stage, a system of linear equations with an initialization phase (the matchings between $x(l)$ and $y(l)$) requiring at most $(2/\varepsilon)^{I+J}$ steps.

b) Blackwell's Theorem states that if for every move $y \in \Delta(J)$ of Player 2, Player 1 has an action $x \in \Delta(I)$ such that $\rho(x, y) \in C$ then C is approachable by Player 1. In other words, if in the one-stage (expected) game where Player 2 plays first and Player 1 plays second, Player 1 has a strategy such that the payoff is in a convex C, then he also has a strategy, in the repeated (expected) game where Player 2 plays second and Player 1 plays first, such that the average payoff converges to C .
The use of calibration transforms this implicit statement into an explicit one : while performing a calibrated strategy (in an auxiliary game where J plays the role of the set of outcomes), Player 1 can enforce the property that, for every $l \in L$, the average move of Player 2 is almost $y(l)$ on $N_n(l)$. So he just has to play $x_{y(l)}$ on these stage and he could not do better.

c) We stress out the fact that the construction of an approachability strategy of C^ε reduces to the construction of a calibrated strategy in an auxiliary game, hence to the construction of an internally-consistent strategy in a second auxiliary game, therefore to the construction of an approachability strategy of a negative orthant in a third auxiliary game. In conclusion, the approachability of an arbitrary convex set reduces to the approachability of an orthant. Along with equations (6.10) and (6.11), this implies that $\mathbf{E}_{\sigma,\tau}\left[d\left(\overline{\rho}_n, C\right) - \varepsilon\right] \leq O\left(n^{-1/2}\right)$. However, as said before, the constant depends on $\varepsilon^{|J|}$.

d) The reduction of the approachability of a convex set $C \subset \mathbb{R}^d$ in a game Γ to the approachability of an orthant in an auxiliary game Γ' can also be done via the following scheme : for every $\varepsilon > 0$, find a finite set of half-spaces $\{H(l), l \in L\}$ such that $C \subset \cap_{l \in L} H(l) \subset C^\varepsilon$. For every $l \in L$, define $c(l) \in \mathbb{R}^d$ and $b(l) \in \mathbb{R}$

such that :

$$H(l) = \left\{\omega \in \mathbb{R}^d, \langle \omega, c(l)\rangle \leq b(l)\right\}$$

and the auxiliary game Γ' with payoffs defined by

$$\widehat{\rho}(i,j) = (\langle \rho(i,j), c(l)\rangle - b(l))_{l\in L} \in \mathbb{R}^L.$$

Obviously, a strategy that approaches the negative orthant in Γ' will approach, in the game Γ, the set $\bigcap H(l)$ and therefore C^ε. However, such a strategy might not be based on regret and might not be explicit.

6.2 Internal regret in the partial monitoring framework

Consider a two person game repeated in discrete time. At stage $n \in \mathbb{N}$, Player 1 (resp. Player 2) chooses $i_n \in I$ (resp. $j_n \in J$), which generates the payoff $\rho_n = \rho(i_n, j_n)$ where ρ is a mapping from $I \times J$ to $\mathbb{R}$. Player 1 does not observe this payoff, he receives a signal $s_n \in S$ whose law is $s(i_n, j_n)$ where s is a mapping from $I \times J$ to $\Delta(S)$. The three sets I, J and S are finite and the two functions ρ and s are extended to $\Delta(I) \times \Delta(J)$ by $\rho(x,y) = \mathbb{E}_{x,y}[\rho(i,j)] \in \mathbb{R}$ and $s(x,y) = \mathbb{E}_{x,y}[s(i,j)] \in \Delta(S)$.

We define the mapping $\mathbf{s}$ from $\Delta(J)$ to $\Delta(S)^I$ by $\mathbf{s}(y) = (s(i,y))_{i\in I}$ and we call such a vector of probability a flag. Player 1 cannot distinguish between two different probabilities y and y' in $\Delta(J)$ that induces the same flag $\mu \in \Delta(S)^I$, *i.e.* such that $\mu = \mathbf{s}(y) = \mathbf{s}(y')$. Thus we say that $\mu = \mathbf{s}(y)$, although unobserved, is the *relevant or maximal* information available to Player 1 about the choice of Player 2. We stress out that a flag μ is not observed since given $x \in \Delta(I)$ and $y \in \Delta(J)$, Player 1 has just an information about μ^i which is only one component of μ (the i-th one, where i is the realization of x). Moreover, this component is the law of a random variable whose realization (*i.e.* the signal $s \in S$) is the only observation of Player 1.

Example. Label efficient prediction : Consider the following game (Example 6.4 in Cesa-Bianchi and Lugosi [29]). Nature chooses an outcome G or B and Player 1 can either observe the actual outcome (action o) or choose to not observe it and to pick a label g or b. If he chooses the right label, his payoff is 1 and otherwise 0. Payoffs and laws of signals received by Player 1 can be resumed by the following matrices (where a, b and c are three different probabilities over a finite set S).

Payoffs :

	G	B
o	0	0
g	0	1
b	1	0

and Signals :

	G	B
o	a	b
g	c	c
b	c	c

Action G, whose best response is g, generates the flag (a,c,c) and action B, whose best response is b, generates the flag (b,c,c). In order to distinguish between those two actions, Player 1 needs to know the entire flag and therefore to know $s(o,y)$ although action o is never a best response (but is said to be *purely informative*).

As usual, a behavioral strategy σ of Player 1 (resp. τ of Player 2) is a function from the set of finite histories for Player 1, $H^1 = \bigcup_{n\in\mathbb{N}} (I \times S)^n$, to $\Delta(I)$ (resp. from $H^2 = \bigcup_{n\in\mathbb{N}} (I \times S \times J)^n$ to $\Delta(J)$). A couple (σ, τ) generates a probability $\mathbb{P}_{\sigma,\tau}$ over $\mathcal{H} = (I \times S \times J)^{\mathbb{N}}$.

External regret

Rustichini [100] defined external consistency in the partial monitoring framework as follows : a strategy σ of Player 1 has no external regret if $\mathbb{P}_{\sigma,\tau}$-as :

$$\limsup_{n\to+\infty} \max_{x\in\Delta(I)} \min_{\begin{cases} y \in \Delta(J), \\ \mathbf{s}(y) = \mathbf{s}(\bar{\jmath}_n) \end{cases}} \rho(x, y) - \overline{\rho}_n \leq 0.$$

where $\mathbf{s}(\bar{\jmath}_n) \in \Delta(S)^I$ is the average flag. In words, the average payoff of Player 1 could not have been uniformly better if he had known the average distribution of flags before the beginning of the game.

Given a flag $\mu \in \Delta(S)^I$, the function $\min_{y\in\mathbf{s}^{-1}(\mu)} \rho(\cdot, y)$ may not be linear. So the best response of Player 1 might not be a pure action in I, but a mixed action $x \in \Delta(I)$ and any pure action in the support of x may be a bad response. This explains why, in Rustichini's definition, the maximum is taken over $\Delta(I)$ and not just over I as in the usual definition of external regret.

Example. Matching Penny in the dark : Player 1 chooses either Tail or Heads and flips a coin. Simultaneously, Nature chooses on which face the coin will land. If Player 1 guessed correctly his payoff equals 1, otherwise -1. We assume that Player 1 does not observe the coin.

Payoffs and signals are resumed in the following matrices :

Payoffs :

	T	H
T	1	-1
H	-1	1

and Signals :

	T	H
T	c	c
H	c	c

Every choice of Nature generates the same flag (c, c). For every $x \in \Delta(\{H, T\})$, $\min_{y\in\Delta(J)} \rho(x, y)$ is non-positive and equals zero only if $x = (1/2, 1/2)$. Therefore the only best response of Player 1 is $(1/2, 1/2)$, while both T or H give the worst payoff of -1.

Internal regret

We consider here a generalization of the previous's framework : at stage $n \in \mathbb{N}$, Player 2 chooses a flag $\mu_n \in \Delta(S)^I$ while Player 1 chooses an action i_n and receives a signal s_n whose law is the i_n-th coordinate of μ_n. Given a flag μ and $x \in \Delta(I)$, Player 1 evaluates the payoff through an evaluation function G from $\Delta(I) \times \Delta(S)^I$ to $\mathbb{R}$, which is not necessarily linear.

Recall that with full monitoring, a strategy has no internal regret if each action $i \in I$ is the best response to the average empirical observation on the set of stages where i was actually played. With partial monitoring, best responses are elements of $\Delta(I)$ and not elements of I, so if we want to define internal regret in this framework, we have to distinguish the stage not as a function of the action actually played (*i.e.* $i_n \in I$) but as a function of its law (*i.e.* $x_n \in \Delta(I)$). We assume that the strategy of Player 1 can be described by a finite family $\{x(l) \in \Delta(I), l \in L\}$ such that, at stage $n \in \mathbb{N}$, Player 1 chooses a type l_n and, given this choice, i_n is drawn accordingly to $x(l_n)$. We assume that L is finite since otherwise Player 1 have trivial strategies that guarantee that the frequency of every l converges to zero. Note that since the choices of l_n can be random, any behavioral strategy can be described in such a way.

Definition 6.14 (*Lehrer-Solan [74]*)

For every $n \in \mathbb{N}$ and every $l \in L$, the average internal regret of type l at stage n is

$$\mathcal{R}_n(l) = \sup_{x \in \Delta(I)} \left[G(x, \overline{\mu}_n(l)) - G(\overline{\imath}_n(l), \overline{\mu}_n(l))\right].$$

A strategy σ of Player 1 is (L, ε)-internally consistent if for every strategy τ of Player 2 :

$$\limsup_{n \to +\infty} \frac{|N_n(l)|}{n} \Big(\mathcal{R}_n(l) - \varepsilon\Big) \leq 0, \quad \forall l \in L, \quad \mathbb{P}_{\sigma,\tau}\text{-as.}$$

Remark : Note that this definition, unlike in the full monitoring case, is not intrinsic. It depends on the choice (which can be assumed to be made by Player 1) of $\{x(l), l \in L\}$, and is based uniquely on the potential observations (*i.e.* the sequences of flags $(\mu_n)_{n \in \mathbb{N}}$) of Player 1.

In order to construct (L, ε)-internally consistent strategies, some regularity over G is required :

Assumption 6.15

For every $\varepsilon > 0$, there exist $\big\{\mu(l) \in \Delta(S)^I, x(l) \in \Delta(I), l \in L\big\}$ two finite families and $\eta, \delta > 0$ such that :

1. $\Delta(S)^I \subset \bigcup_{l \in L} B(\mu(l), \delta)$;
2. *For every $l \in L$, if $\|x - x(l)\| \leq 2\eta$ and $\|\mu - \mu(l)\| \leq 2\delta$, then $x \in BR_\varepsilon(\mu)$,*

where $BR_\varepsilon(\mu) = \big\{x \in \Delta(I) : G(x, \mu) \geq \sup_{z \in \Delta(I)} G(z, \mu) - \varepsilon\big\}$ is the set of ε-best response to $\mu \in \Delta(S)^I$ and $B(\mu, \delta) = \big\{\mu' \in \Delta(S)^I, \|\mu' - \mu\| \leq \delta\big\}$.

In words, Assumption 6.15 implies that G is regular with respect to μ and with respect to x : given ε, the set of flags can be covered by a finite number of balls

centered in $\{\mu(l), l \in L\}$, such that $x(l)$ is an ε-best response to any μ in this ball. And if x is close enough to $x(l)$, then x is also an ε-best response to any μ close to $\mu(l)$. Without loss of generality, we can assume that $x(l)$ is different from $x(l')$ for any $l \neq l'$.

Theorem 6.16

Under Assumption 6.15, there exist (L, ε)-internally consistent strategies.

Some parts of the proof are quite technical, however the insight is very simple, so we give firstly the main ideas. Assume for the moment that Player 1 fully observes the flag at each stage. If, in the one stage game, Player 2 plays first and his choice generates a flag $\mu \in \Delta(S)^I$, then Player 1 has an action $x \in \Delta(I)$ such that x belongs to $BR_\epsilon(\mu)$. Using a minmax argument (like Blackwell did for the proof of Theorem 6.7, recall Remark 6.1 *b*) one could prove that Player 1 has an (L, ε)-internally consistent strategy (as did Lehrer and Solan [74], with some extra assumptions on $\mathbf{s}$).

The idea is to use calibration to transform this implicit proof into a constructive one, as in the alternative proof of Corollary 6.7. Fix $\varepsilon > 0$ and consider the game where Player 1 predicts the sequence $(\mu_n)_{n\in\mathbb{N}}$ using the δ-grid $\{\mu(l), l \in L\}$ given by Assumption 6.15. A calibrated strategy of Player 1 chooses a sequences $(l_n)_{n\in\mathbb{N}}$ in such a way that $\overline{\mu}_n(l)$ is asymptotically δ-close to $\mu(l)$. Hence Player 1 just has to play accordingly to $x(l) \in BR_\varepsilon(\mu(l))$ on these stages.

Indeed, since the choices of action are independent, $\overline{\imath}_n(l)$ will be asymptotically η-close to $x(l)$ and the regularity of G will imply then that $\overline{\imath}_n(l) \in BR_\varepsilon(\overline{\mu}_n(l))$ and so the strategy will be (L, ε)-internally consistent.

The only issue is that in the current framework the signal depends on the action of Player 1 since the law of s_n is the i_n component of μ_n, which is not observed. Signals (that belong to S) and predictions (that belong to $\Delta(S)^I$) are in two different spaces, so the existence of calibrated strategies is not straightforward. However, it is well known that, up to a slight perturbation of $x(l)$, the information available to Player 1 after a long time is close to $\overline{\mu}_n(l)$ (as in the multi-armed bandit problem, some calibration and no-regret frameworks, see *e.g.* Cesa-Bianchi and Lugosi [29] chapter 6 for a survey on these techniques).

For every $x \in \Delta(I)$, define $x_\eta \in \Delta(I)$, the η-perturbation of x by $x_\eta = (1-\eta)x + \eta\mathbf{u}$ with $\mathbf{u}$ the uniform probability over I and for every n define $\widehat{s}_n$ by :

$$\widehat{s}_n = \left(\frac{\mathbb{1}\{s_n = s\}\mathbb{1}\{i_n = i\}}{x_\eta(l_n)[i_n]} \right) \in \mathbb{R}^{SI},$$

with $x_\eta(l_n)[i_n] \geq \eta > 0$ the weight put by $x_\eta(l_n)$ on i_n. We denote by $\widetilde{s}_n(l)$, instead of $\overline{\widehat{s}}_n(l)$, their average on $N_n(l)$.

Lemma 6.17

For every $\theta > 0$, there exists $N \in \mathbb{N}$ such that, for every $l \in L$:

$$\mathbb{P}_{\sigma,\tau}\left(\forall m \geq n, \|\widetilde{s}_n(l) - \overline{\mu}_n(l)\| \leq \theta | \, N_n(l) \geq N\right) \geq 1 - \theta.$$

Proof. Since for every $n \in \mathbb{N}$, the choices of i_n and μ_n are independent :

$$\begin{aligned}
\mathbb{E}_{\sigma,\tau}\left[\widehat{s}_n | \, h_{n-1}, l_n, \mu_n\right] &= \sum_{i\in I}\sum_{s\in S} \mu_n^i[s] x_\eta(l_n)[i] \left(0, \ldots, \frac{s}{x_\eta(l_n)[i]}, \ldots, 0\right) \\
&= \sum_{i\in I}\sum_{s\in S} \mu_n^i[s] \left(0, \ldots, s, \ldots, 0\right) = \sum_{i\in I} \left(0, \ldots, \mu_n^i, \ldots, 0\right) \\
&= \left(\mu_n^1, \ldots, \mu_n^I\right) = \mu_n,
\end{aligned}$$

where μ_n is seen as an element of $\mathbb{R}^{SI}$. Therefore $\widetilde{s}_n(l)$ is an unbiased estimator of $\overline{\mu}_n(l)$ and Hoeffding-Azuma's inequality (actually its multidimensionnal version by Chen and White [31] together with the fact that $\sup_{n\in\mathbb{N}} \|\widehat{s}_n\| \leq \eta^{-1} < \infty$) implies that for every $\theta > 0$ there exists $N \in \mathbb{N}$ such that, for every $l \in L$:

$$\mathbb{P}_{\sigma,\tau}\left(\forall m \geq n, \|\widetilde{s}_n(l) - \overline{\mu}_n(l)\| \leq \theta | \, |N_n(l)| \geq N\right) \geq 1 - \theta.$$

□

Assume now that Player 1 uses a calibrated strategy to predict the sequences of $\widehat{s}_n$ (this game is in full monitoring), then he knows that asymptotically $\widetilde{s}_n(l)$ is closer to $\mu(l)$ than to any $\mu(k)$ (as soon as the frequency of l is big enough), therefore it is δ-close to $\mu(l)$. Lemma 6.17 implies that $\overline{\mu}_n(l)$ is asymptotically close to $\widetilde{s}_n(l)$ and therefore 2δ-close to $\mu(l)$.

Proof of Theorem 6.16 Let the families $\{x(l) \in \Delta(I), \mu(l) \in \Delta(S)^I, l \in L\}$ and $\eta, \delta > 0$ be given by Assumption 6.15 for a fixed $\varepsilon > 0$.

Let Γ' be the auxiliary repeated game where, at stage n, Player 1 (resp. Player 2) chooses $l_n \in L$ (resp. $\mu_n \in \Delta(S)^I$). Given these choices, i_n (resp. s_n) is drawn accordingly to $x_\eta(l_n)$ (resp. $\mu_n^{i_n}$). By Lemma 6.17, for every $\theta > 0$, there exists $N_1 \in \mathbb{N}$ such that for every $l \in L$:

$$\mathbb{P}_{\sigma,\tau}\left(\forall m \geq n, \|\widetilde{s}_n(l) - \overline{\mu}_n(l)\| \leq \theta | \, |N_n(l)| \geq N_1\right) \geq 1 - \theta. \tag{6.12}$$

Let σ be a calibrated strategy associated to $(\widetilde{s}_n)_{n\in\mathbb{N}}$ in Γ'. For every $\theta > 0$, there exists $N_2 \in \mathbb{N}$ such that with $\mathbb{P}_{\sigma,\tau}$-probability greater than $1 - \theta$:

$$\forall n \geq N_2, \forall l, k \in L, \frac{|N_n(l)|}{n}\left(\|\widetilde{s}_n(l) - \mu(l)\|^2 - \|\widetilde{s}_n(l) - \mu(k)\|^2\right) \leq \theta. \tag{6.13}$$

Since $\{\mu(k), k \in L\}$ is a δ-grid of $\Delta(S)^I$, for every $n \in \mathbb{N}$ and $l \in L$, there exists $k \in L$ such that $\|\widetilde{s}_n(l) - \mu(k)\| \leq \delta$. Therefore, combining equation (6.12) and (6.13), for every $\theta > 0$ there exists $N_3 \in \mathbb{N}$ such that :

$$\mathbb{P}_{\sigma,\tau}\left(\forall n \geq N_3, \forall l \in L, \frac{|N_n(l)|}{n}\left(\|\overline{\mu}_n(l) - \mu(l)\|^2 - \delta^2\right) \leq \theta,\right) \geq 1 - \theta. \tag{6.14}$$

For every stage of type $l \in L$, i_n is drawn accordingly to $x_\eta(l)$ and by definition $\|x_\eta(l) - x(l)\| \leq \eta$. Therefore Hoeffding-Azuma's inequality implies that, for every $\theta > 0$ there exists $N_4 \in \mathbb{N}$ such that :

$$\mathbb{P}_{\sigma,\tau}\left(\forall n \geq N_4, \forall l \in L, \frac{|N_n(l)|}{n}\left(\|\bar{\imath}_n(l) - x(l)\| - \eta\right) \leq \theta,\right) \geq 1 - \theta. \tag{6.15}$$

Combining equation (6.14), (6.15) and using Assumption 6.15, for every $\theta > 0$, there exists $N \in \mathbb{N}$ such that for every strategy τ of Player 2 :

$$\mathbb{P}_{\sigma,\tau}\left(\forall n \geq N, \forall l \in L, \frac{|N_n(l)|}{n}\left(\mathcal{R}_n(l) - \varepsilon\right) \leq \theta,\right) \geq 1 - \theta, \tag{6.16}$$

and σ is (L, ε)-internally consistent. □

Remark : Lugosi, Mannor and Stoltz [78] provided an algorithm that constructs, by block of size $m \in \mathbb{N}$, a strategy that has no external regret. We can describe it as follows. Play at every stage of the k-th block B_k according to the same probability $x_k \in \Delta(I)$. Then compute (using Lemma 6.17) an estimator of the average flag on this block and denote it by $\widetilde{\mu}_k$. Knowing this flag, compute the average regret accumulated on this specific block and aggregate it to the previous regret in order to estimate the average regret from the beginning of the game. Decide next what action is going to be played on the following block according to a classic exponential weight algorithm. With a fine tuning of $m \in \mathbb{N}$ (and $\eta > 0$), the external regret of this strategy converges to zero at the rate $O\left(n^{-1/5}\right)$ (the optimal rate is known to be $n^{-1/3}$).

Instead of trying to compute (or at least approximate) the sequence of payoffs from the sequence of signals, our algorithm consider an abstract auxiliary game defined on the signal space (*i.e.* on the relevant information, the observations). We define payoffs in this abstract game in order to transform it into a game with full monitoring : the action set of Player 2 are flags, that are (almost) observed by Player 1.

The strategy constructed is based on ε-calibration and Hoeffding-Azuma's inequality, therefore one can show that :

$$\mathbb{E}_{\sigma,\tau}\left[\sup_{l \in L} \frac{|N_n(l)|}{n}\left(\mathcal{R}_n(l) - \varepsilon\right)\right] \leq O\left(\frac{1}{\sqrt{n}}\right).$$

So given $\varepsilon > 0$, one can construct a strategy such that the internal regret converges quickly to ε, but maybe very slowly to zero (because the constants depend, once again, drastically on ε^J).

Remark : Since $\widetilde{s}_n$ converges to $\overline{\mu}_n$, the regret can be defined in terms of observed empirical flags instead of unobserved average flag. For the same reason, $x(l)$ can be used to define regret.

On the strategy space

One might object that behavioral strategies of Players 1 are defined as mappings from the set of past histories $H^1 = \bigcup_{n\in\mathbb{N}} (I \times S)^n$ into $\Delta(I)$ while in Definition 6.14 (and Theorem 6.16) strategies considered are defined as mappings from $\bigcup_{n\in\mathbb{N}} (I \times S \times L)^n$ into $\Delta(L)$, with the specification that given $l_n \in L$, the law of i_n is $x(l_n)$ — for a fixed family $\{x(l), l \in L\}$. Hence, they can be defined as mappings from $\bigcup_{n\in\mathbb{N}} (X \times I \times S)^n$ into $\Delta(X)$ (where $X = \Delta(I)$ and $\Delta(X)$ is embedded with the star-weak topology) and thus are behavioral strategies in the game where Player 1's action set is X and he receives at each stage a signal in $I \times S$.

Therefore, they are equivalent to (i.e., following Mertens Sorin and Zamir [84], Theorem 1.8 p. 55, generate the same probability on the set of plays as) mixed strategies, which are mixtures of pure strategies, *i.e.* mappings from $\bigcup_{n\in\mathbb{N}} (X \times I \times S)^n$ into X. These latter are equivalent to applications from $\bigcup_{n\in\mathbb{N}} (I \times S)^n$ into X. Indeed, consider for example $\sigma : \bigcup_{n\in\mathbb{N}} (X \times T)^n \to X$ and define $\widetilde{\sigma} : \bigcup_{n\in\mathbb{N}} T^n \to X$ recursively by $\widetilde{\sigma}(\emptyset) = \sigma(\emptyset)$ and

$$\widetilde{\sigma}(t_1, \ldots, t_n) = \sigma(\widetilde{\sigma}(\emptyset), t_0, \ldots, \widetilde{\sigma}(t_0, \ldots, t_{n-1}), t_n).$$

Finally, they are, in the game where Player 1's action set is I and he receives at each stage a signal in S, mixtures of behavioral strategies — also called general strategies — so are equivalent to behavioral strategies.

In conclusion, given a strategy defined as in Definition 6.14, there exists a behaviorial strategy that generates the same probability on the set of plays (for every strategy τ of Player 2).

In these general strategies, Player 1 uses two types of signals : the signals generated by *the game*, *i.e.* the sequence $(i_n, s_n)_{n\in\mathbb{N}}$ and some private signals generated by *his own strategy*, *i.e.* the sequences of l_n. We can compute internal regret in Theorem 6.16 not only because the choices of μ_n and l_n are independent given the past, but mainly because the choices of μ_n and i_n are independent, even when l_n is known.

6.3 Back on payoff space

In the section we give simple condition on G that ensures it fulfills Assumption 6.15. We also extend the framework to the so-called *compact case*. Finally, we prove that an internally consistent strategy (in a sense to be specified) is also externally consistent.

The worst case fulfills Assumption 6.15

Proposition 6.18

Let $G : \Delta(I) \times \Delta(S)^I$ be such that for every $\mu \in \Delta(S)^I$, $G(\cdot, \mu)$ is continuous and the family of function $\{G(x, \cdot), x \in \Delta(I)\}$ is equicontinuous.

Then G fulfills Assumption 6.15.

Proof. Since $\{G(x, \cdot), x \in \Delta(I)\}$ is equicontinuous and $\Delta(S)^I$ compact, for every $\varepsilon > 0$, there exists $\delta > 0$ such that :

$$\forall x \in \Delta(I), \forall \mu, \mu' \in \Delta(S)^I, \|\mu - \mu'\| \leq 2\delta \Rightarrow |G(x, \mu) - G(x, \mu')| \leq \frac{\varepsilon}{2}.$$

Let $\{\mu(l), l \in L\}$ be a finite δ-grid of $\Delta(S)^I$ and for every $l \in L$, $x(l) \in BR(\mu(l))$ so that $G(x(l), \mu(l)) = \max_{z \in \Delta(I)} G(z, \mu(l))$. Since $G(x(l), \cdot)$ is continuous, there exists $\eta(l) > 0$ such that :

$$\|x - x(l)\| \leq \eta(l) \Rightarrow |G(x, \mu(l)) - G(x(l), \mu(l))| \leq \varepsilon/2.$$

Define $\eta = \min_{l \in L} \eta(l)$ and let $x \in \Delta(I)$, $\mu \in \Delta(S)^I$ and $l \in L$ such that $\|x - x(l)\| \leq \eta$ and $\|\mu - \mu(l)\| \leq \delta$, then :

$$G(x, \mu) \geq G(x, \mu(l)) - \frac{\varepsilon}{2} \geq G(x(l), \mu(l)) - \varepsilon = \max_{z \in \Delta(I)} G(z, \mu(l)) - \varepsilon,$$

and $x \in BR_\varepsilon(\mu)$. □

This proposition implies that the evaluation function used by Rustichini fulfills Assumption 6.15 (see also Lugosi, Mannor and Stoltz [78], Lemma 3.1 and Proposition A.1). Before proving that, we introduce $\mathcal{S}$, the range of $\mathbf{s}$, which is a closed convex subset of $\Delta(S)^I$, and $\Pi_{\mathcal{S}}(\cdot)$ the projection onto it.

Corollary 6.19

Define $W : \Delta(I) \times \Delta(S)^I \to \mathbb{R}$ by :

$$W(x, \mu) = \begin{cases} \inf_{y \in \mathbf{s}^{-1}(\mu)} \rho(x, y) & \text{if} \quad \mu \in \mathcal{S} \\ W(x, \Pi_{\mathcal{S}}(\mu)) & \text{otherwise.} \end{cases}$$

Then W fulfills Assumption 6.15.

Proof. The function $\mathbf{s}$ can be extended linearly to $\mathbb{R}^J$ by $\mathbf{s}(y) = \sum_{j \in J} y(j)\mathbf{s}(j)$ where $y = (y(j))_{j \in J}$. Therefore (Aubin and Frankowska [5], Theorem 2.2.1, p. 57) the multivalued application $\mathbf{s}^{-1} : \mathcal{S} \rightrightarrows \Delta(J)^I$ is λ-Lipschitz, and since $\Pi_{\mathcal{S}}$ is 1-Lipschitz (because $\mathcal{S}$ is convex), $W(x, \cdot)$ is also λ-Lipschitz, for every $x \in \Delta(I)$. Therefore, $\{G(x, \cdot), x \in \Delta(I)\}$ is equicontinuous. For every $\mu \in \Delta(S)^I$, $W(\cdot, \mu)$ is r-Lipschitz (where $r = \|\rho\|$, see *e.g.* Lugosi, Mannor and Stoltz [78]), therefore continuous. Hence, by Proposition 6.18, W fulfills Assumption 6.15. □

Compact case

Assumption 6.15 does not require that Player 1 faces only one opponent, nor that his opponents have only a finite set of actions. As long as G is regular then Player 1 has a (L, ε)-internally consistent strategy, for every $\varepsilon > 0$. We consider in this section a particular framework, referred as the *compact case* (as mentioned in section 6.1).

Player 1's action set is still denoted by I, but we now assume that the action set of Player 2 is $[-1,1]^I$. The payoff mapping ρ from $\Delta(I) \times [-1,1]^I$ to $\mathbb{R}$ is simply defined by $\rho(x, U) = \langle x, U \rangle$. Let $\mathbf{s}$ be a multivalued application from $[-1,1]^I$ to $\Delta(S)^I$. Given the choices of i and U, Player 1 does not observe U but receives a signal $s \in S$, whose law is the i-th component of μ which belongs to $\mathbf{s}(U)$. If $\mathbf{s}(U)$ is not a singleton then we can assume either that μ is chosen by Nature (a third player) or by Player 2.

A multivalued application $\mathbf{s}$ is closed-convex if $\lambda\mathbf{s}(x)+(1-\lambda)\mathbf{s}(z) \subset \mathbf{s}(\lambda x+(1-\lambda)z)$ and its graph is closed and its inverse is defined by $\mathbf{s}^{-1}(\mu) = \{U \in [-1,1]^I, \mu \in \mathbf{s}(U)\}$. It is clear that if $\mathbf{s}$ is closed-convex then $\mathbf{s}^{-1}$ is also closed-convex.

Proposition 6.20

Define the worst case mapping as in Corollary 6.19. If $\mathbf{s}$ *is closed-convex and its range is a polytope (the convex hull of a finite number of points), then* W *fulfills Assumption 6.15.*

Proof. We follow Aubin et Frankowska [5] : let μ_0 be in $\mathcal{S}$ the range of $\mathbf{s}$, U_0 be in $\mathbf{s}^{-1}(\mu_0)$ and g be the mapping defined by :

$$\begin{array}{rcl} g : \mathcal{S} & \mapsto & \mathbb{R} \\ \mu & \to & g(\mu) = \inf\limits_{U \in \mathbf{s}^{-1}(\mu)} \|U - U_0\| = d\left(U_0, \mathbf{s}^{-1}(\mu)\right). \end{array}$$

Since $\mathbf{s}$ is convex, so is $\mathbf{s}^{-1}$ (in the multivalued sense) and g (in the univalued sense). The sections $\{\mu | g(\mu) \leq \lambda\}$ are closed (see Aubin and Frankowska [5], Lemma 2.2.3 p.$\tilde{5}$9) so g is lower semi-continuous. Since the domain of g is a polytope, g is also upper semi-continuous (see Rockafellar [97], Theorem 10.2 p. 84). Therefore g is continuous over $\mathcal{S}$ and there exists $\delta(U_0)$ such that if $\|\mu - \mu_0\| \leq \delta(U_0)$ then $d\left(U_0, \mathbf{s}^{-1}(\mu)\right) \leq \varepsilon$.

Since $\mathbf{s}^{-1}(\mu_0)$ is compact, for every $\varepsilon > 0$, there exists a finite set $\mathcal{U}$ such that $\mathbf{s}^{-1}(\mu_0) \subset \bigcup_{U \in \mathcal{U}} B(U, \varepsilon)$. Define $\delta(\mu_0) = \inf_{U \in \mathcal{U}} \delta(U_0)$, then for every μ in $\Delta(S)^I$, $\|\mu - \mu_0\| \leq \delta(\mu_0)$ implies that $\mathbf{s}^{-1}(\mu_0) \subset \mathbf{s}^{-1}(\mu) + 2\varepsilon B$ (with B the unit ball). The graph of $\mathbf{s}^{-1}$ is compact so for every $\varepsilon > 0$ there exists $0 < \delta'(\mu_0) < \delta(\mu_0)$ such that if $\|\mu - \mu_0\| \leq \delta'(\mu_0)$ then $\mathbf{s}^{-1}(\mu) \subset \mathbf{s}^{-1}(\mu_0) + 2\varepsilon B$.

There exists a finite set M such that the compact set $\mathcal{S}$ is included in the union of open balls $\bigcup_{\mu \in M} B(\mu, \delta'(\mu)/3)$. If we denote by $\delta = \inf_{\mu \in M} \delta'(\mu)/3$ then for every μ and μ' in $\mathcal{S}$, if $\|\mu - \mu'\| \leq \delta$, there exists $\mu_1 \in M$ such that μ and μ' belongs to $B(\mu_1, \delta'(\mu_1))$ hence $\mathbf{s}^{-1}(\mu) \subset \mathbf{s}^{-1}(\mu_1) + 2\varepsilon B \subset \mathbf{s}^{-1}(\mu') + 4\varepsilon B$.

Let μ and μ' in $\Delta(S)^I$ such that $\|\mu - \mu'\| \leq \delta$. Then since $\mathcal{S}$ is a convex set $\|\Pi_{\mathcal{S}}(\mu) - \Pi_{\mathcal{S}}(\mu')\| \leq \delta$ and for every $x \in \Delta(I)$

$$W(x,\mu) = \inf_{U \in \mathbf{s}^{-1}(\Pi_{\mathcal{S}}(\mu))} \langle x, U\rangle \geq \inf_{U \in \mathbf{s}^{-1}(\Pi_{\mathcal{S}}(\mu'))} \langle x, U\rangle - 4\varepsilon = W(x,\mu') - 4\varepsilon.$$

Let x and x' in $\Delta(I)$ such that $\|x - x'\| \leq \varepsilon$ then for all $\mu \in \Delta(S)^I$

$$W(x,\mu) = \inf_{U \in \mathbf{s}^{-1}(\Pi_{\mathcal{S}}(\mu))} \langle x, U\rangle \geq \inf_{U \in \mathbf{s}^{-1}(\Pi_{\mathcal{S}}(\mu))} \langle x', U\rangle - \varepsilon = W(x',\mu) - \varepsilon.$$

Hence if $x(l)$ is a ε-best response to $\mu(l)$, $\|x - x(l)\| \leq \varepsilon$ and $\|\mu - \mu(l)\| \leq \delta$ then

$$\begin{aligned} W(x,\mu) \geq W(x(l),\mu) - \varepsilon \geq W(x(l),\mu(l)) - 5\varepsilon &\geq \sup_{z \in \Delta(I)} W(z,\mu(l)) - 6\varepsilon \\ &\geq \sup_{z \in \Delta(I)} W(z,\mu) - 10\varepsilon, \end{aligned}$$

so x is a 10ε-best response to μ. □

Remark on the assumptions over $\mathbf{s}$: $\mathbf{s}$ is assumed to be multivalued since in the finite case, there might be two different probabilities y and y' in $\Delta(J)$ that generate the same outcome vector $\rho(y) = (\rho(i,y))_{i \in I} = \rho(y')$ but two different flags $\mathbf{s}(y)$ and $\mathbf{s}(y')$.

It is also convex : if Player 2 can generate a flag μ by playing $y \in \Delta(J)$ and a flag μ' by playing y', then a convex combination of y and y' should generate the same convex combination of flags. This assumption is specifically needed with repeated game : for example, Player 2 can play y on odd stages and y' on even stages. Player 1 must know that the average empirical flag can be generated by $1/2y + 1/2y'$.

The fact that the range of $\mathbf{s}$ is a polytope (or at least that it is locally simplicial, see Rockafellar [97] p. 84 for formal definitions and examples) is needed for the proof that W is continuous. It is obviously true in the finite dimension case that the graph is a polytope.

Regret in terms of actual payoffs

As Rustichini [100], we can define regret in term of unobserved average payoff.

Definition 6.21

A strategy σ of Player 1 is (L,ε)-internally consistent with respect to the actual payoffs if for every $l \in L$:

$$\limsup_{n \to +\infty} \frac{|N_n(l)|}{n} \left(\sup_{x \in \Delta(I)} [W(x, \overline{\mu}_n(l)) - \overline{\rho}_n(l)] - \varepsilon \right) \leq 0, \quad \mathbb{P}_{\sigma,\tau}\text{-as.}$$

Proposition 6.22

For every $\varepsilon > 0$, there exist (L, ε)-internally consistent strategies with respect to the actual payoffs.

Proof. Consider the strategy σ given by Theorem 6.16 with the worst case mapping. By definition of W and using the independence of the choices of $x(l)$ and μ_n, one can easily show that asymptotically $W(x(l), \overline{\mu}_n(l)) \leq \overline{\rho}_n(l)$. Therefore the strategy σ is also (L, ε)-consistent with respect to the actual payoffs. □

Now we can define 0-internally consistent strategies (see Lehrer and Solan [74] definition 10) :

Definition 6.23

A strategy σ of Player 1 is 0-internally consistent if for every $\varepsilon > 0$, there exists $\delta > 0$ such that for every finite partition $\{P(l), l \in L\}$ of $\Delta(I)$ with diameter smaller than δ and every $l \in L$:

$$\limsup_{n \to +\infty} \frac{|N_n(l)|}{n} \left(\sup_{x \in \Delta(I)} [W(x, \overline{\mu}_n(l)) - \overline{\rho}_n(l)] - \varepsilon \right) \leq 0, \quad \mathbb{P}_{\sigma,\tau}\text{-as},$$

where $N_n(l) = \{m \leq n, x_n \in P(l)\}$ with x_n the law (that might be chosen at random by Player 1) of i_n given the past history and $\overline{\mu}_n(l)$ (resp. $\overline{\imath}_n(l)$) is the average flag (resp. action of Player 1) on $N_n(l)$.

The average flag $\overline{\mu}_n$ belongs to $\Delta(S)$ and is defined by :

$$\overline{\mu}_n[s] = \frac{\sum_{m=1}^{n} \mu_m[s]}{n} \text{ for any } s \in S.$$

Proposition 6.24

There exist 0-internally consistent strategies with respect to the actual payoffs.

Proof. The proof relies uniquely on a classic doubling trick (see *e.g.* Sorin [106], Proposition 3.2 p. 56) recalled below.

Denote by σ_k the strategy given by Proposition 6.22 for $\varepsilon_k = 2^{-(k+3)}$. Consider the strategy σ of player defined by block : on the first block of length N_1, Player 1 plays accordingly to σ_1, then on the second block of length N_2 accordingly to σ_2, and so on. Formally, for n such that $\sum_{k=1}^{p-1} N_k \leq n \leq \sum_{k=1}^{p} N_k$, $\sigma(h_n) = \sigma_p(h_n^p)$ where $h_n^p = (i_m, l_m, s_m)_{m \in \{\sum_{k=1}^{p-1} N_k, \ldots, n\}}$ is the partial history on the last block. Remark 6.2 implies that for every $p \in \mathbb{N}$ there exists $M_p \in \mathbb{N}$ such that

$$\mathbb{E}_{\sigma,\tau} \left[\sup_{l \in L} \frac{|N_n(l)|}{n} \Big(\mathcal{R}_n(l) \Big) \right] \leq \frac{1}{2^{p+1}}.$$

Let $(N_k)_{k\in\mathbb{N}}$ be a sequence such that $\sum_{p=1}^{k-1} N_p = o(N_k)$ and $M_{k+1} = o(N_k)$ (where $u_n = o(v_n)$ means that $v_n > 0$ and $\lim_{n\to\infty} \frac{u_n}{v_n} = 0$). With this definition, the m-th block is way longer than all the previous blocks, and longer than the time required by σ_{k+1} to be ε_{k+1}-consistent (in expectation). So the (maybe high) regret accumulated during the first M_n stages of the n-th block is negligible compared to the small regret accumulated before (during the first $(n-1)$-blocks). After these M_n stages, the regret (on the n-th block) is smaller than ε_n and at the end of this block, the cumulative regret is very close to ε. □

Remark : The use of a doubling trick prevents us to easily find a bound on the rate of convergence of the regret. The proof of Proposition 6.24 requires that the sum of the regret on two different block is smaller than the average regret. This is why we restrict this definition to internally consistent strategies with respect to the actual payoffs. One may compare Definition 6.23 of 0-consistency to the Definition 6.2 of ε-calibrated strategies.

External and internal consistency

With full monitoring, by linearity of the payoff function, a strategy that is internally consistent is also externally consistent. This properties holds in partial monitoring, when we consider regret in terms of actual payoffs :

Proposition 6.25

For every $\varepsilon > 0$ and $\{x(l), l \in L\}$ of $\Delta(I)$, every (L,ε)-internally consistent strategy with respect to the actual payoffs is ε-externally consistent with respect to the actual payoffs, i.e. $\mathbb{P}_{\sigma,\tau}$-ps :

$$\limsup_{n\to+\infty} \max_{x\in\Delta(I)} W(x,\overline{\mu}_n) - \overline{\rho}_n \leq \varepsilon.$$

Proof. Let $\varepsilon > 0$, $L \subset \Delta(I)$ and σ be an (L,ε)-internally consistent strategy with respect to the actual payoffs. Since $\mathbf{s}^{-1}(\cdot)$ is convex then, for every $x \in \Delta(I)$, the mapping $\mu \mapsto W(x,\mu)$ is convex and so is the mapping $\mu \mapsto \max_{x\in\Delta(I)} W(x,\mu)$. Hence

$$\begin{aligned}\limsup_{n\to+\infty} \max_{x\in\Delta(I)} W(x,\overline{\mu}_n) - \overline{\rho}_n &\leq \sum_{l\in L} \limsup_{n\to+\infty} \frac{|N_n(l)|}{n}\left(\max_{x\in\Delta(I)} W(x,\overline{\mu}_n(l)) - \overline{\rho}_n(l)\right)\\ &\leq \limsup_{n\to+\infty} \sum_{l\in L} \frac{|N_n(l)|}{n}\varepsilon \leq \varepsilon\end{aligned}$$

and so σ is ε-externally consistent. □

Proposition 6.25 holds for the compact case under the assumption that σ is closed-convex. Note that the proof relies on the fact that W is convex and the actual payoffs

are linear. It is clear that this result does not extend to any evaluation function. Indeed, consider the optimistic function defined by (for $\mu \in \mathcal{S}$) :

$$O(x, \mu) = \sup_{y \in \mathbf{s}^{-1}(\mu)} \rho(x, y),$$

then the more information about $\bar{\jmath}_n$ that Player 1 gets, the less he evaluates his payoff. So an internally consistent strategy (*i.e.* a strategy that is consistent with a more precise knowledge on the moves of Player 2) might not be externally consistent.

Concluding remarks

In the full monitoring framework, many improvements have been made in the past years about calibration and regret (see for instance [70, 102, 117]). Here, we aimed to clarify the links between the original notions of approachability, internal regret and calibration in order to extend applications (in particular, to get rid of the finiteness of J), to define the internal regret with signals as calibration over an appropriate space and to give a proof derived from no-internal regret in full monitoring, itself derived from the approachability of an orthant in this space.

Acknowledgments : I deeply thank my advisor Sylvain Sorin for his great help and numerous comments. I also acknowledge very helpful remarks from Gilles Stoltz.

An extended abstract of this paper appeared in the *Proceedings of the 20th International Conference on Algorithmic Learning Theory*, Springer, 2009.

7

Calibration-Based Optimal Algorithms

On modifie dans ce chapitre l'algorithme naïf du chapitre précédent, qui construit des stratégies consistantes intérieurement, afin de créer, lorsque l'évaluation est du type "worst-case" un algorithme dont la vitesse de convergence en $O\left(n^{-1/3}\right)$ est optimale. Il requiert à chaque étape de résoudre un système d'équations linéaires de taille uniformément bornée. On fournit également quelques améliorations de cet algorithme pour diminuer les constantes de convergence. Ce chapitre est issu de l'article No-Regret with Partial Monitoring : Calibration-Based Optimal Algorithms.

Sommaire

7.1 Full monitoring **105**
Model and definitions 105
A naïve algorithm, based on calibration 107
Calibration and Laguerre diagram 111
Optimal algorithm with full monitoring 114
7.2 Partial monitoring **116**
Definitions 116
A naïve algorithm 119
Optimal algorithms 119
7.3 Concluding remarks **124**
Second algorithm : calibration and polytopial complex. 124
Extension to compact case 125
Strengthening of the constants 126
7.4 Proofs of technical results **128**
Proof of proposition 7.18 128
Proof of Lemma 7.11 131

Hannan [56] introduced the notion of regret in repeated games with full monitoring : a player has no external regret if, asymptotically, his average payoff could

not have been greater by knowing before the beginning of the game the empirical distribution of moves of the other player. The existence of strategies with this property, originally proved by Hannan [56], has been rediscovered by Blackwell [18] using his approachability theorem. A generalization of this result and a more precise notion of regret are due to Foster and Vohra [42] (see also Fudenberg and Levine [51]) : there exist strategies such that a player has no internal regret, *i.e.* for each of his action, he has no external regret on the set of stages where he actually played it. Hart and Mas-Colell [57] also used Blackwell's approachability theorem to construct explicit strategies such that the internal (and therefore the external) regret at stage n is bounded by $O\left(n^{-1/2}\right)$.

Some of those results have been extended to games with partial monitoring, where players receive random signals whose laws might depend on the unobserved chosen actions. In this framework, Rustichini [100] defined — and proved the existence of — strategies with no external regret. In words, a player has no regret if his average payoff could not have been greater if he had known the empirical distribution of signals before the beginning of the game. Under some strong assumptions on the signalling structure (Cesa-Bianchi, Lugosi and Stoltz [30] assumed that a player can deduce from the signals observed his payoff and Lugosi, Mannor and Stoltz [78] considered deterministic feedback), there exist strategies that bound the expected regret in $O\left(n^{-1/3}\right)$, which is the optimal bound (see Cesa-Bianchi and Lugosi [29], Theorem 6.7). When no such assumption is made, Lugosi, Mannor and Stoltz [78] provided an algorithm (based on the exponential weight algorithm) that bounds regret in $O\left(n^{-1/5}\right)$.

In this framework, internal regret was defined by Lehrer and Solan [74] ; stages are no longer distinguished as a function of the action played by Player 1 (as in the full monitoring case) but as a function of the law of player 1's action. Indeed, the evaluation of the payoff is not linear with respect to the distribution of signals. So given any of those, a best response might consist only in a mixed action (*i.e.* a probability over the set of actions). Lehrer and Solan [74] also proved the existence of strategies with no internal regret, if the laws of the signals do not depend on Player 1's action. Perchet [92] provided an explicit algorithm based on calibration, introduced by Dawid [33], that does not require this last assumption. Roughly speaking, Player 1 ε-discretizes arbitrarily the space of distributions of signals and each point of the discretization is called a possible prediction. Then, stage after stage, player 1 predicts what will be the law of the next signal and plays a *best response* — in a sense to be defined — to it. If the sequence of predictions is calibrated then the average signals, on the set of stages where he made a specific prediction, will be close to this prediction. The continuity of payoff and signaling functions implies that such a strategy can bound the regret in $\varepsilon+O\left(n^{-1/2}\right)$ where the constants depend drastically on ε. Definitions and proofs of this algorithm are recalled in section 2.2.

We provided in section 7.2 and section 7.3 two algorithms that bound the expected

internal regret in $O\left(n^{-1/3}\right)$. The first one does not use an arbitrary discretization but constructs carefully a specific one and then computes, stage by stage, the solution of a system of linear equations. The second one does not use any discretization, but requires at each stage to solve a linear program.

Section 1 is devoted to the full monitoring case. We recall definitions of calibration, no-regret and give a naïve algorithm (based on the idea of Foster and Vohra [42]) to construct strategies with internal regret asymptotically smaller than ε. We show how to modify this algorithm to bound the regret in $O\left(n^{-1/2}\right)$. Section 2 is concerned with the partial monitoring case. We also recall a naïve algorithm and its modification in order to reach the optimal bound of $O\left(n^{-1/3}\right)$. Some extensions (the second algorithm, the so-called *compact case* and some variants to strengthen the constants for the rates of convergence) are presented in section 3. Technical proofs are presented in the section 7.4.

7.1 Full monitoring

Model and definitions

Consider a two-person game repeated in discrete time, where at stage $n \in \mathbb{N}$, Player 1 (resp. Nature) chooses an action $i_n \in I$ (resp. $j_n \in J$) where both I and J are finite. This generates a payoff $\rho_n = \rho(i_n, j_n)$, where ρ is a mapping from $I \times J$ to $\mathbb{R}$, and a regret $r_n \in \mathbb{R}^{|I|}$ defined by :

$$r_n = \big(\rho(i, j_n) - \rho(i_n, j_n)\big)_{i \in I} \in \mathbb{R}^{|I|}.$$

The choices of i_n and j_n depend on the past observations (also called finite history) $h_{n-1} = (i_1, j_1, \ldots, i_{n-1}, j_{n-1})$ and may be random. Explicitly, the set of finite histories is denoted by $H = \bigcup_{n \in \mathbb{N}} (I \times J)^n$, with $(I \times J)^0 = \emptyset$ and a strategy σ of Player 1 is a mapping from H to $\Delta(I)$, the set of probability distributions over I. Given the history $h_n \in (I \times J)^n$, $\sigma(h_n)$ is the law of i_{n+1}. A strategy τ of Nature is defined similarly. A couple of strategies (σ, τ) generates a probability, denoted by $\mathbb{P}_{\sigma,\tau}$, over $\mathcal{H} = (I \times J)^{\mathbb{N}}$, the set of plays endowed with the cylinder σ-field.

We extend the payoff mapping ρ to $\Delta(I) \times \Delta(J)$ by $\rho(x, y) = \mathbb{E}_{x,y}[\rho(i, j)]$ and for any sequence $a = (a_m)_{m \in \mathbb{N}}$ and any $n \in \mathbb{N}_*$, we denote by $\overline{a}_n = \frac{1}{n} \sum_{m=1}^{n} a_m$ the average of a up to stage n.

Definition 7.1 (*Hannan [56]*)

Given $\varepsilon \geq 0$, a strategy σ of Player 1 is ε-externally consistent if for every strategy τ of Nature :

$$\limsup_{n \to \infty} \overline{r}_n^i \leq \varepsilon, \quad \forall i \in I, \quad \mathbb{P}_{\sigma,\tau}\text{−as}.$$

In words, a strategy σ is ε-externally consistent if Player 1 could not have had a greater payoff if he had known, before the beginning of the game, the empirical distribution of actions of Nature.

Foster and Vohra [42] (see also Fudenberg and Levine [51]) defined a more precise notion of regret. The internal regret of the stage n, denoted by $R_n \in \mathbb{R}^{|I\times I|}$, is also generated by the choices of i_n and j_n and its (i,k)-th component is defined by :

$$R_n^{ik} = \begin{cases} \rho(k, j_n) - \rho(i, j_n) & \text{if } i = i_n \\ 0 & \text{otherwise.} \end{cases},$$

stated differently, every row of the matrix R_n are null except the i_n-th which is r_n.

Definition 7.2 (*Foster and Vohra [42]*)

Given $\varepsilon \geq 0$, a strategy σ of Player 1 is ε-internally consistent if for every strategy τ of Nature :

$$\limsup_{n\to\infty} \overline{R}_n^{ik} \leq \varepsilon, \quad \forall i, k \in I, \quad \mathbb{P}_{\sigma,\tau}\text{−as.}$$

The following notations will allow us to provide some other equivalent formulations of internal consistency. Denote by $N_n(i)$ the set of stages before the n-th where Player 1 played the action i and $\bar{\jmath}_n(i) \in \Delta(J)$ the empirical distribution of actions of Nature on this set. Formally,

$$N_n(i) = \{m \leq n, i_m = i\} \quad \text{and} \quad \bar{\jmath}_n(i) = \frac{\sum_{m\in N_n(i)} j_m}{|N_n(i)|}. \tag{7.1}$$

Definition 7.2 is equivalent to the fact that for every $i, k \in I$

$$\limsup_{n\to\infty} \frac{|N_n(i)|}{n}\Big(\rho\big(k, \bar{\jmath}_n(i)\big) - \rho\big(i, \bar{\jmath}_n(i)\big) - \varepsilon\Big) \leq 0, \quad \mathbb{P}_{\sigma,\tau}\text{−as.}$$

If we define, for every $\varepsilon \geq 0$, the ε-best response correspondence by :

$$\mathrm{BR}_\varepsilon(y) = \left\{x \in \Delta(I), \rho(x, y) \geq \max_{z\in\in\Delta(I)} \rho(z, y) - \varepsilon\right\},$$

then a strategy of Player 1 is ε-internally consistent if any of his action i is either an ε-best response to the empirical distribution of Nature's actions on $N_n(i)$ or the frequency of i is very small. We will simply denote BR_0 by BR and call it the best response correspondence.

From now on, given two sequences $\left\{l_m \in L, \omega_m \in \mathbb{R}^d\right\}_{m\in\mathbb{N}}$ where L is a finite set, we will define the subset of integers $N_n(l)$ and the average $\overline{\omega}_n(l)$ as in equation (7.1).

Proposition 7.3 (*Foster and Vohra [42], Hart and Mas-Colell [57]*)

For every $\varepsilon \geq 0$, there exist ε-internally consistent strategies.

Foster and Vohra [42] and Hart and Mas-Colell [57] proved the existence of 0-internally consistent strategies using different algorithms (based respectively on the Expected Brier Score and Blackwell's approachability theorem). In some sense, we merge these two proofs in order to provide a new one — given in the following section — that can be extended quite easily to the partial monitoring framework (unlike the previous algorithms).

A naïve algorithm, based on calibration

The algorithm that constructs an ε-internally consistent strategy is based on this simple fact : if Player 1 can, stage by stage, foresee the law of Nature's next action, say y, then he just has to play any best response to y at the following stage. The continuity of ρ implies that Player 1 does not need to forecast precisely y_n, but up to some $\delta > 0$.

Let $\{y(l), l \in L\}$ be a δ-grid of $\Delta(J)$ (*i.e.* a finite set such that for every $y \in \Delta(J)$ there exists $l \in L$ such that $\|y - y(l)\| \leq \delta$) and $i(l)$ be a best response to $y(l)$, for every $l \in L$. Then if δ is small enough :

$$\|y - y(l)\| \leq 2\delta \Rightarrow i(l) \in \mathrm{BR}_{2\varepsilon}(y)$$

On can compute a *good sequence of forecasts* by computing a calibrated strategy (introduced by Dawid [33]) in an auxiliary game. Definition and existence of calibrated strategies are recalled in the subsection 7.1.

CALIBRATION

Consider a two-person repeated game where, at stage n, Nature chooses the state of the world $s_n \in S$, where S is a finite set, and Player 1 (the predictor) forecasts it by choosing $\mu(l_n)$ where $\{\mu(l), l \in L\}$ is a finite δ-grid of $\Delta(S)$. As usual, a behavioral strategy σ of Player 1 (resp. τ of Nature) is a mapping from the set of finite histories $H = \bigcup_{n\in\mathbb{N}} (L \times S)^n$ to $\Delta(L)$ (resp. $\Delta(S)$) and we also denote by $\mathbb{P}_{\sigma,\tau}$ the probability generated by the couple (σ, τ) over $\mathcal{H}$ the set of plays endowed with the cylinder topology.

Definition 7.4 (*Dawid [33]*)

A strategy σ of Player 1 is calibrated (with respect to $\{\mu(l), l \in L\}$) if for every strategy τ of Nature, $\mathbb{P}_{\sigma,\tau}$-as :

$$\limsup_{n\to\infty} \frac{|N_n(l)|}{n} \left(\|\overline{s}_n(l) - \mu(l)\|^2 - \|\overline{s}_n(l) - \mu(k)\|^2 \right) \leq 0, \quad \forall k, l \in L,$$

where $\|\cdot\|$ is the Euclidian norm of $\mathbb{R}^{|S|}$.

In words, a strategy is calibrated if for every $l \in L$, either the frequency of l is small, or the empirical distribution of states, on the set of stages where $\mu(l)$ was predicted, is closer to $\mu(l)$ than to any other $\mu(k)$. Given a finite grid of $\Delta(S)$, the existence of calibrated strategies has been proved by Foster and Vohra [43] and Hart [43] using respectively the Expected Brier Score or a minmax theorem. We gave here the construction of Sorin [108] (related but simpler than the one of Foster and Vohra).

Proposition 7.5 (*Foster and Vohra [43], Hart and Mas-Colell [43]*)

For any finite grid $\mathcal{L}$ of $\Delta(S)$, there exist calibrated strategies with respect to $\mathcal{L}$ such that :

$$\mathbb{E}_{\sigma,\tau}\left[\max_{l,k\in L}\frac{|N_n(l)|}{n}\left(\|\overline{s}_n(l)-\mu(l)\|^2-\|\overline{s}_n(l)-\mu(k)\|^2\right)\right]\leq O\left(\frac{1}{\sqrt{n}}\right).$$

Proof. Consider the auxiliary game Γ_c where, at stage $n \in \mathbb{N}$, Player 1 (resp. Nature) chooses $l_n \in L$ (resp. $s_n \in S$) and the vector payoff is the matrix $U_n \in \mathbb{R}^{|L\times L|}$ where

$$U_n^{lk}=\begin{cases}\|s_n-\mu(l)\|^2-\|s_n-\mu(k)\|^2 & \text{if } l=l_n\\ 0 & \text{otherwise.}\end{cases}$$

A strategy σ is calibrated with respect to L if $\overline{U}_n$ converges to the negative orthant in Γ_v. Indeed for every $l,k \in L$, the (l,k)-th component of $\overline{U}_n$ is

$$\begin{aligned}\overline{U}_n^{lk} &= \frac{|N_n(l)|}{n}\frac{\sum_{m\in N_n(l)}\|s_m-\mu(l)\|^2-\|s_m-\mu(k)\|^2}{|N_n(l)|}\\ &= \frac{|N_n(l)|}{n}\left(\|\overline{s}_n(l)-\mu(l)\|^2-\|\overline{s}_n(l)-\mu(k)\|^2\right).\end{aligned}$$

Denote by $\overline{U}_n^+ := \left\{\max\left(0,\overline{U}_n^{lk}\right)\right\}_{l,k\in L} := \overline{U}_n-\left(\overline{U}_n\right)^-$ the positive part of $\overline{U}_n$ and by $\lambda_n \in \Delta(L)$ any invariant measure of $\overline{U}_n^+$. We recall that λ is an invariant measure of a nonnegative matrix U if $\sum_{k\in L}\lambda(k)U^{kl}=\lambda(l)\sum_{k\in L}U^{lk}$, for every $l \in L$ (its existence is a consequence of Perron-Frobenius Theorem, see *e.g.* Seneta [103]).

Define the strategy σ of Player 1 inductively as follows. Choose arbitrarily $\sigma(\emptyset)$, the law of Player 1's first action and at stage $n+1$, play accordingly to any invariant measure of $\overline{U}_n$. We claim that this strategy is an approachability strategy of the negative orthant of $\mathbb{R}^{|L\times L|}$ because it satisfies Blackwell's [17] sufficient condition :

$$\forall n\in\mathbb{N},\langle\left(\overline{U}_n\right)-\left(\overline{U}_n\right)^-,\mathbb{E}_{\lambda_n}\left[U_{n+1}|s_{n+1}\right]-\left(\overline{U}_n\right)^-\rangle\leq 0.$$

Indeed, for every possible $s_{n+1} \in S$:

$$\langle\left(\overline{U}_n\right)^+,\mathbb{E}_{\lambda_n}\left[U_{n+1}|s_{n+1}\right]\rangle=0=\langle\left(\overline{U}_n\right)^+,\left(\overline{U}_n\right)^-\rangle, \tag{7.2}$$

where the second equality follows from the definition of positive and negative parts. Consider the first equality. The (l,k)-th component of the matrix $\mathbb{E}_{\lambda_n}[U_{n+1}|s_{n+1}]$ is $\lambda_n(l)\left(\|s_{n+1}-\mu(l)\|^2-\|s_{n+1}-\mu(k)\|^2\right)$, therefore the coefficient of $\|s_{n+1}-\mu(l)\|^2$ is equal to $\sum_{k\in L}\lambda_n(k)\left(\overline{U}_n^+\right)^{kl}-\lambda_n(l)\sum_{k\in L}\left(\overline{U}_n^+\right)^{lk}=0$ since λ_n is an invariant measure of $\left(\overline{U}_n\right)^+$.

Blackwell's [17] result also implies that $\mathbb{E}_{\sigma,\tau}\left[\|\overline{U}_n^+\|\right]\leq 2M_n n^{-1/2}$ for any strategy τ of Nature where $M_n^2=\sup_{m\leq n}\mathbb{E}_{\sigma,\tau}\left[\|U_m\|^2\right]=4|L|$. □

Remark : The strategy σ constructed in Theorem 7.5 depends only on the sequence of averages $(\overline{U}_n)_{n\in\mathbb{N}}$. To compute σ, Player 1 has to find, stage by stage, an invariant measure of a matrix and this can be done, using Gaussian elimination, in $O\left(L^3\right)$ operations.
Precisely, Gaussian elimination requires $L(L+1)/2$ divisions, $(2L^3+3L^2-5L)/6$ multiplications and $(2L^3+3L^2-5L)/6$ substractions, see *e.g.* Farebrother [38], section 1.5 *arithmetical cost*.

Corollary 7.6

For any finite grid $\mathcal{L}$ of $\Delta(S)$, there exists σ a calibrated strategy with respect $\mathcal{L}$ such that for every strategy τ of Nature with $\mathbb{P}_{\sigma,\tau}$ probability at least $1-\delta$:

$$\max_{l,k\in L}\frac{|N_n(l)|}{n}\left(\|\overline{s}_n(l)-\mu(l)\|^2-\|\overline{s}_n(l)-\mu(k)\|^2\right)\leq\frac{2M_n}{\sqrt{n}}+\Theta_n,$$

where $K_n=\sup_{m\leq n}\sup_{l,k\in L}\left|U_n^{lk}-\mathbb{E}_{\sigma,\tau}\left[U_n^{lk}\right]\right|\leq 4$

$$\Theta_n=\min\left\{\frac{v_n}{\sqrt{n}}\sqrt{2\ln\left(\frac{2|L|^2}{\delta}\right)}+\frac{2}{3}\frac{K_n}{n}\ln\left(\frac{2|L|^2}{\delta}\right),\frac{K_n}{\sqrt{n}}\sqrt{2\ln\left(\frac{2|L|^2}{\delta}\right)}\right\}$$

$$M_n=\sup_{m\leq n}\sqrt{\mathbb{E}_{\sigma,\tau}\left[\|U_m\|^2\right]}\leq 3\sqrt{|L|}$$

$$v_n^2=\sup_{m\leq n}\sup_{l,k\in L}\mathbb{E}_{\sigma,\tau}\left[\left|U_n^{lk}-\mathbb{E}_{\sigma,\tau}\left[U_n^{lk}\right]\right|^2\right]\leq 4.$$

Proof. Proposition 7.5 implies that $\mathbb{E}_{\sigma,\tau}\left[\overline{U}_n\right]\leq 2M_s n^{-1/2}$. Hoeffding-Azuma's inequality [61, 11] (see Lemma 7.24 below) implies that with probability at least $1-\delta$:

$$\overline{U}_n^{lk}-\mathbb{E}_{\sigma,\tau}\left[\overline{U}_n^{lk}\right]\leq\frac{K_n}{\sqrt{n}}\sqrt{2\ln\left(\frac{2}{\delta}\right)}$$

and Freedman's inequality (an analogue of Bernstein's inequality for martingale see [47], Proposition 2.1 or Cesa-Bianchi and Lugosi [29], Lemma A.8) implies that

with probability at least $1-\delta$:

$$\overline{U}_n^{lk} - \mathbb{E}_{\sigma,\tau}\left[\overline{U}_n^{lk}\right] \leq \frac{v_n}{\sqrt{n}}\sqrt{2\ln\left(\frac{2}{\delta}\right)} + \frac{2}{3}\frac{K_n}{n}\ln\left(\frac{2}{\delta}\right).$$

The result is a consequence of these two inequalities and of Theorem . □

The definition of Θ_n as a minimum (and the use of Freedman's inequality) will be useful when we will refer to this corollary in the subsequent sections. Obviously, in the current framework, $\Theta_n = \frac{4}{\sqrt{n}}\sqrt{2\ln\left(\frac{2|L|^2}{\delta}\right)}$.

BACK TO THE NAÏVE ALGORITHM

Let us now go back to the construction of ε-consistent strategies in Γ. Consider the auxiliary game Γ_c^1, where Nature chooses at stage n his action $j_n \in J$ and Player 1 forecasts it using a δ-grid $\mathcal{Y} = \{y(l), l \in L\}$ of $\Delta(J)$. Compute σ, a calibrated strategy with respect to $\mathcal{Y}$ in the abstract game Γ_c^1, and whenever Player 1 should play the action l in Γ_c^1, he plays $i(l) \in \mathrm{BR}(y(l))$ in the game Γ. We claim that this defines a strategy σ_ε in Γ which is 2ε-internally consistent.

Proposition 7.7 (*Foster and Vohra [42]*)

For every $\varepsilon > 0$, the strategy σ_ε is 2ε-internally consistent.

Proof. By definition of a calibrated strategy, $\mathbb{P}_{\sigma,\tau}$-as, for every $\eta > 0$, there exists $N \in \mathbb{N}$ such that for every $l, k \in L$ and for every $n \geq N$:

$$\frac{|N_n(l)|}{n}\left(\|\bar{\jmath}_n(l) - y(l)\|^2 - \|\bar{\jmath}_n(l) - y(k)\|^2\right) \leq \eta.$$

Since $\{y(k), k \in L\}$ is a δ-grid of $\Delta(J)$, for every $l \in L$ and every $n \in \mathbb{N}$, there exists $k \in L$ such that $\|\bar{\jmath}_n(l) - y(k)\|^2 \leq \delta^2$, hence $\|\bar{\jmath}_n(l) - y(l)\|^2 \leq \delta^2 + \eta\frac{n}{|N_n(l)|}$. So, by definition of $\{y(l), l \in L\}$:

$$\frac{|N_n(l)|}{n} \geq \frac{\eta}{\delta^2} \Rightarrow \rho(i, \bar{\jmath}_n(l)) - \rho(i(l), \bar{\jmath}_n(l)) \leq 2\varepsilon, \quad \forall i \in I, \forall l \in L, \forall n \geq N.$$

The (i,k)-th component of $\overline{R}_n$ satisfies

$$\begin{aligned}
\frac{|N_n(i)|}{n}\left(\overline{R}_n^{ik} - 2\varepsilon\right) &= \frac{1}{n}\sum_{m \in N_n(i)}\left(\rho(k, j_m) - \rho(i, j_m) - 2\varepsilon\right) \\
&= \frac{1}{n}\sum_{l:i(l)=i}\sum_{m \in N_n(l)}\left(\rho(k, j_m) - \rho(i, j_m) - 2\varepsilon\right) \\
&= \sum_{l:i(l)=i}\frac{|N_n(l)|}{n}\left(\rho(k, \bar{\jmath}_n(l)) - \rho(i(l), \bar{\jmath}_n(l)) - 2\varepsilon\right).
\end{aligned}$$

Recall that either $|N_n(l)|/n \geq \eta/\delta^2$ and $\rho(k, \bar{\jmath}_n(i)) - \rho(i(l), \bar{\jmath}_n(l)) - 2\varepsilon \leq 0$, or $|N_n(l)|/n < \eta/\delta^2$. Since ρ is bounded (by $M_\rho > 0$), then :

$$\frac{|N_n(i)|}{n}\left(\overline{R}_n^{ik} - 2\varepsilon\right) \leq \eta\frac{M_\rho|L|}{\delta^2}, \quad \forall i \in I, \forall k \in I, \forall n \geq N,$$

which implies that σ is 2ε-internally consistent. □

Remark : This algorithm only achieves ε-consistency and Proposition 7.5 implies that

$$\mathbb{E}_{\sigma,\tau}\left[\max_{i,k\in I}\left(\overline{R}_n^{ik} - \varepsilon\right)^+\right] \leq O\left(\frac{1}{\sqrt{n}}\right).$$

It is therefore weaker than the algorithms proposed by Foster and Vohra [42] or Hart and Mas-Colell [57]. Moreover, the constant depends on ε, therefore it is not possible to obtain 0-internally consistency with the same rate with a classic doubling trick argument (*i.e.* play a 2^{-k}-internally consistent strategy on N_k stages, then switch to a 2^{k+1}-internally consistent strategy, and so on, see *e.g.* Sorin [106], Proposition 3.2 p. 56) to . However it can be easily extended to obtain 0-consistency, both with full and partial monitoring.

Calibration and Laguerre diagram

Given a finite subset $\{\mu(l), l \in L\}$ (the Voronoï sites) of $\mathbb{R}^d$, we define the l-th Voronoï cell $V(l)$ (or the cell associated to $\mu(l)$) as the set of points closer to $\mu(l)$ than to any other $\mu(k)$:

$$V(l) = \left\{X \in \mathbb{R}^d, \|X - \mu(l)\|^2 \leq \|X - \mu(k)\|^2, \quad \forall k \in L\right\},$$

where $\|\cdot\|$ is the Euclidian norm of $\mathbb{R}^d$. Each $V(l)$ is a polyhedron (as the intersection of a finite number of half-spaces) and $\{V(l), l \in L\}$ is a covering of $\mathbb{R}^d$. A calibrated strategy with respect to $\{\mu(l), l \in L\}$ has the property that for every $l \in L$, either the frequency of l (*i.e.* $|N_n(l)|/n$) goes to zero, or $\bar{s}_n(l)$, the empirical distribution of states on $N_n(l)$, converges to $V(l)$.

The naïve algorithm uses the Voronoï diagram associated to an arbitrary grid of $\Delta(J)$ and assigns to every small cell a mixed action which is an ε-best reply to every point of this cell. This is possible since ρ is continuous. A calibrated strategy ensures that $\bar{\jmath}_n(l)$ converges to $V(l)$ (or the frequency of l is small) and so playing $i(l)$ on $N_n(l)$ was indeed a ε-best response to $\bar{\jmath}_n(l)$. With this approach, we cannot construct immediately 0-internally consistent strategy. Indeed, this would require that for every $l \in L$ there exists a 0-best response $i(l)$ to any y in $V(l)$. However, there is no reason for every element of $V(l)$ to share a common best response because $\{\mu(l), l \in L\}$ is arbitrarily.

On the other hand, consider the game Matching Penny. We recall that both players have two action Heads and Tails, so $\Delta(J) = \Delta(I) = [0, 1]$ (seen as the probability of

choosing T). The payoff is 1 if both players choose the same action and -1 otherwise. Action H (resp. T) is a best response for Player 1 to any y in $[0, 1/2]$ (resp. in $[1/2, 1]$). These two segments are exactly the cells of the Voronoï diagram associated to $\{y(1) = 1/4, y(2) = 3/4\}$, therefore, performing a calibrated strategy with respect to $\{y(1), y(2)\}$ and playing H (resp. T) on the stages of type 1 (resp. 2) is a strategy of Player 1 that is 0-internally consistent.

This idea can be generalized to any game. Indeed, by Lemma 7.9 below, $\Delta(J)$ can be decomposed into polytopial best-response areas (a polytope is the convex hull of a finite number of points, its vertices). Given such a polytopial decomposition, one can find a finer Voronoï diagram (*i.e.* any best-response area is an union of Voronoï cells) and finally use a calibrated strategy to ensure convergence with respect to this diagram.

However, the proof of the existence of this diagram is quite complicated and the number of cells can be quite huge. Thus, we will consider instead a generalization of Voronoï diagrams, called Laguerre diagrams. Given a subset of Laguerre sites $\{\mu(l), l \in L\}$ and weights $\{\omega(l) \in \mathbb{R}, l \in L\}$, the l-th Laguerre cell $P(l)$ is defined by :

$$P(l) = \left\{X \in \mathbb{R}^d, \|X - \mu(l)\|^2 - \omega(l) \leq \|X - \mu(k)\|^2 - \omega(k), \quad \forall k \in L\right\},$$

where $\|\cdot\|$ is the Euclidian norm of $\mathbb{R}^d$. Each $P(l)$ is a polyhedron (an intersection of a finite number of halfspaces) and $\mathcal{P} = \{P(l), l \in L\}$ is a covering of $\mathbb{R}^d$.

Definition 7.8

$\{K^i, i \in I\}$ is a polytopial complex of a polytope K with non-empty interior if, for every i, j in the finite set I, K^i is a polytope with non-empty interior and the polytope $K^i \cap K^j$ has empty interior.

This definition extends naturally to a polytope K with empty interior, if we consider the affine subspace generated by K.

Lemma 7.9

There exists $I' \subset I$ such that $\{B^i, i \in I'\}$ is a polytopial complex of $\Delta(J)$, where B^i is the i-th best response area defined by

$$B^i = \{y \in \Delta(J), i \in \mathrm{BR}(y)\}.$$

Proof. For any $y \in \Delta(J)$, $\rho(\cdot, y)$ is linear on $\Delta(I)$ and attains its maximum on I, so $\bigcup_{i \in I} B^i = \Delta(J)$. Without loss of generality, we can assume that each B^i is non-empty, otherwise we drop the index i. For every $i, k \in I$, $\rho(i, \cdot) - \rho(k, \cdot)$ is linear on

$\Delta(J)$ therefore B^i, defined by

$$\begin{aligned} B^i &= \{y \in \Delta(J), \rho(i,y) \geq \rho(k,y), \forall k \in K\} \\ &= \bigcap_{k \in I} \{y \in \mathbb{R}^{|J|}, \rho(i,y) - \rho(k,y) \geq 0\} \cap \Delta(J), \end{aligned}$$

is the intersection of a finite number of half-spaces and the polytope $\Delta(J)$, thus is also a polytope. Moreover if B_0^{ik}, the interior of $B^i \cap B^k$, is non-empty then $\rho(i,\cdot)$ equals $\rho(k,\cdot)$ on the subspace generated by B_0^{ik} and therefore on $\Delta(J)$, hence $B^i = B^k$. Denote by I' any subset of I such that for every $i \in I$, there exists exactly one $i' \in I'$ such that $B^i = B^{i'} \neq \emptyset$, then $\{B^i, i \in I'\}$ is a polytopial complex of $\Delta(J)$. □

Proposition 7.10

Let $\mathcal{K} = \{K^i, i \in I\}$ be a polytopial complex of a polytope $K \subset \mathbb{R}^d$. Then there exists $\{\mu(l) \in \mathbb{R}^d, \omega(l) \in \mathbb{R}, l \in L\}$, a finite set of Laguerre sites and weights, such that the Laguerre diagram $\mathcal{P} = \{P(l), l \in L\}$ refines $\mathcal{K}$, i.e. every K^i is a finite union of cells $P(l)$.

Proof. Let $\mathcal{K} = \{K^i, i \in I\}$ be a polytopial complex of $K \subset \mathbb{R}^d$. Each K^i is a polytope, thus defined by a finite number of hyperplanes. Denote by $\mathcal{H} = \{H_t, t \in T\}$ the set of all defining hyperplanes and $\widehat{\mathcal{K}} = \{\widehat{K}^l, l \in L\}$ the decomposition of $\mathbb{R}^d$ induces by $\mathcal{H}$ (usually called arrangement of hyperplanes). It is clear that $\widehat{\mathcal{K}}$ refines $\mathcal{K}$. Theorem 3 and Corollary 1 of Aurenhamer [10] imply that $\widehat{\mathcal{K}}$ is the Laguerre diagram associated to some $\{\mu(l), \omega(l), l \in L\}$ whose exact computation requires the following notation :

i) for every $t \in T$, let $c_t \in \mathbb{R}^d$ and $b_t \in \mathbb{R}$ (which can, without loss of generality, be assumed to be non zero) such that

$$H_t = \left\{X \in \mathbb{R}^d; \langle X, c_t \rangle = b_t\right\}.$$

ii) For every $l \in L$ and $t \in T$, $\sigma_t(l) = 1$ if the origin of $\mathbb{R}^d$ and $\widehat{K}^l$ are in the same halfspace defined by H_t and $\sigma_t(l) = -1$ otherwise.

iii) For every $l \in L$, we define :

$$\mu(l) = \frac{\sum_{t \in T} \sigma_t(l) c_t}{|T|} \quad \text{and} \quad \omega(l) = \|\mu(l)\|^2 + 2\frac{\sum_{t \in T} \sigma_t(l) b_t}{|T|}. \tag{7.3}$$

□

Note that one can add the same constant to every weight $\omega(l_0)$

Buck [26] proved that the number of cells defined by $|T|$ hyperplanes in $\mathbb{R}^d$ is bounded by $\sum_{k=0}^{d} \binom{|T|}{k} := \phi(|T|, d)$. Moreover, $|T|$ is clearly smaller than $|I|(|I|-1)/2$ (in the case where each K^i has a non-empty intersection with every other polytope), so $|L| \leq \phi\left(\frac{|I|^2}{2}, d\right)$. Note that $\phi(n,d) \leq n^d$, if $d \geq n$ then $\phi(n,d) = 2^n$ and for a fixed d, $\phi(n,d) = O\left(n^d/d!\right)$.

Lemma 7.11

Let $\mathcal{P} = \{P(l), l \in L\}$ be a Laguerre diagram associated to the set of sites and weights $\{\mu(l) \in \mathbb{R}^d, \omega(l) \in \mathbb{R}, l \in L\}$. Then, there exists a positive constant $M_P > 0$ such that for every $X \in \Delta(S)^I$ if

$$\|X - \mu(l)\|^2 - \omega(l) \leq \|X - \mu(k)\|^2 - \omega(k) + \theta, \quad \forall l, k \in L \tag{7.4}$$

then $d(X, P(l))$ is smaller than $M_P\theta$.

The proof can be found in section 7.4; the constant M_P depends on the Laguerre diagram, and more precisely on the scalar products $\langle c_t, c_{t'} \rangle$, for every $t, t' \in T$.

Optimal algorithm with full monitoring

We reformulate Proposition 7.5 and Corollary 7.6 in terms of Laguerre diagram.

Theorem 7.12

For any set of sites and weights $\{\mu(l) \in \mathbb{R}^S, \omega(l) \in \mathbb{R}, l \in L\}$ there exists a strategy σ of Player 1 such that for every strategy τ of Nature :

$$\mathbb{E}_{\sigma,\tau}\left[\left\|\left(\overline{U}_{\omega,n}\right)^+\right\|\right] \leq O\left(\frac{1}{\sqrt{n}}\right) \text{ where } U_{\omega,n} \text{ is defined by :}$$

$$(U_{\omega,n})^{lk} = \begin{cases} \left[\|s_n - \mu(l)\|^2 - \omega(l)\right] - \left[\|s_n - \mu(k)\|^2 - \omega(k)\right] & \text{if } l = l_n \\ 0 & \text{otherwise} \end{cases}$$

The proof is identical to the one of Proposition 7.5.

Corollary 7.13

For any set $\{\mu(l) \in \mathbb{R}^S, \omega(l) \in \mathbb{R}, l \in L\}$, there exists a strategy σ of Player 1 such that for every strategy τ of Nature, with $\mathbb{P}_{\sigma,\tau}$ probability at least $1 - \delta$:

$$\max_{l,k \in L} \frac{|N_n(l)|}{n}\left(\left[\|\overline{s}_n(l) - \mu(l)\|^2 - \omega(l)\right] - \left[\|\overline{s}_n(l) - \mu(k)\|^2 - \omega(k)\right]\right) \leq \frac{2M_n}{\sqrt{n}} + \Theta_n$$

where $M_n = \sup_{m \leq n} \sqrt{\mathbb{E}_{\sigma,\tau}\left[\|U_{\omega,m}\|^2\right]} \leq 4\sqrt{L}\|(b,c)\|_\infty$;

$$\Theta_n = \min\left\{\frac{v_n}{\sqrt{n}}\sqrt{2\ln\left(\frac{2L^2}{\delta}\right)} + \frac{2}{3}\frac{K_n}{n}\ln\left(\frac{2L^2}{\delta}\right), \frac{K_n}{\sqrt{n}}\sqrt{2\ln\left(\frac{2L^2}{\delta}\right)}\right\};$$

$$v_n^2 = \sup_{m \leq n}\sup_{l,k \in L} \mathbb{E}_{\sigma,\tau}\left[\left|U_{\omega,m}^{lk} - \mathbb{E}_{\sigma,\tau}\left[U_{\omega,m}^{lk}\right]\right|^2\right] \leq 16\|c\|_\infty^2;$$

$$K_n = \sup_{m \leq n}\sup_{l,k \in L}\left|U_{\omega,m}^{lk} - \mathbb{E}_{\sigma,\tau}\left[U_{\omega,m}^{lk}\right]\right| \leq 8\|c\|_\infty,$$

$$\|c\|_\infty = \sup_{t \in T}|b_t| \text{ and } \|(b,c)\|_\infty = \sup_{t \in T}\|c_t\| + \sup_{t \in T}|b_t|.$$

Such a strategy is called calibrated with respect to the set $\{\mu(l), \omega(l), l \in L\}$.

Proof. Hoeffding-Azuma's inequality [61, 11] implies that with probability at least $1-\delta$:

$$\overline{U}^{lk}_{\omega,n} - \mathbb{E}_{\sigma,\tau}\left[\overline{U}^{lk}_{\omega,n}\right] \leq \frac{K_n}{\sqrt{n}}\sqrt{2\ln\left(\frac{2}{\delta}\right)}$$

and Freedman's inequality [47] implies that with probability at least $1-\delta$:

$$\overline{U}^{lk}_{\omega,n} - \mathbb{E}_{\sigma,\tau}\left[\overline{U}^{lk}_{\omega,n}\right] \leq \frac{v_n}{\sqrt{n}}\sqrt{2\ln\left(\frac{2}{\delta}\right)} + \frac{2}{3}\frac{K_n}{n}\ln\left(\frac{2}{\delta}\right),$$

The result is a consequence of these two inequalities and of Theorem 7.12. □

The definition of Θ_n as a minimum (and the use of Freedman's inequality) will be useful when we will refer to this corollary in the subsequent sections.

Theorem 7.14

There exists an internally consistent strategy σ of Player 1 such that for every strategy τ of Nature and every $n \in \mathbb{N}$, with $\mathbb{P}_{\sigma,\tau}$ probability greater than $1-\delta$:

$$\left\|\left(\overline{R}_n\right)^+\right\|_\infty \leq O\left(\sqrt{\frac{\ln\left(\frac{1}{\delta}\right)}{n}}\right). \tag{7.5}$$

Proof. Lemma 7.9 and Proposition 7.10 imply the existence of a Laguerre Diagram $\{Y(l), l \in L\}$ associated to the finite set $\{y(l) \in \mathbb{R}^J, \omega(l) \in \mathbb{R}, l \in L\}$ that refines $\{B^i, i \in I\}$. Hence, for every $l \in L$, there exists $i(l)$ such that $Y(l) \subset B^{i(l)}$. If we denote by $\widetilde{\jmath}_n(l)$ the projection of $\bar{\jmath}_n(l)$ onto $Y(l)$ then :

$$\begin{aligned}
\left(\overline{R}_n\right)^{ik} &= \sum_{l:i(l)=i} \frac{N_n(l)}{n}\Big(\rho\big(k, \bar{\jmath}_n(l)\big) - \rho\big(i(l), \bar{\jmath}_n(l)\big)\Big) \\
&\leq \sum_{l:i(l)=i} \frac{N_n(l)}{n}\Big(\Big[\rho\big(k, \bar{\jmath}_n(l)\big) - \rho\big(k, \widetilde{\jmath}_n(l)\big)\Big] + \Big[\rho\big(i(l), \widetilde{\jmath}_n(l)\big) - \rho\big(i(l), \bar{\jmath}_n(l)\big)\Big]\Big) \\
&\leq \sum_{l:i(l)=i} \frac{N_n(l)}{n}\Big(2M_\rho \left\|\widetilde{\jmath}_n(l) - \bar{\jmath}_n(l)\right\|\Big) \\
&\leq (2M_\rho M_P L) \max_{l,k\in L} \frac{N_n(l)}{n}\Big(\left[\|\bar{\jmath}_n(l) - y(l)\|^2 - \omega(l)\right] - \left[\|\bar{\jmath}_n(l) - y(k)\|^2 - \omega(k)\right]\Big)
\end{aligned}$$

where the third inequality is due to the fact that $i(l) \in \mathrm{BR}(\widetilde{\jmath}_n(l))$ and ρ is M_ρ-Lipschitz and the fourth inequality is a consequence of Lemma 7.11. Consider the strategy σ where Player 1 plays the action $i(l)$ on the set of stages of type l, then

Corollary 7.13 yields that for every strategy τ of Nature, with $\mathbb{P}_{\sigma,\tau}$ probability at least $1-\delta$:

$$\max_{l,k} \frac{N_n(l)}{n}\Big(\big[\|\bar{\jmath}_n(l)-y(l)\|^2-\omega(l)\big] - \big[\|\bar{\jmath}_n(l)-y(k)\|^2-\omega(k)\big] \Big) \leq \frac{8\sqrt{|L|}\|(b,c)\|_\infty}{\sqrt{n}} + \frac{8\|c\|_\infty}{\sqrt{n}}\sqrt{2\ln\left(\frac{2|L|^2}{\delta}\right)},$$

therefore with $\Omega_0 = 16M_\rho M_P|L|^{3/2}\|(b,c)\|_\infty$ and $\Omega_1 = 16M_\rho M_P|L|^{1/2}\|c\|_\infty$ one has that for every strategy of Nature and with probability at least $1-\delta$:

$$\max_{i,k} \frac{N_n(i)}{n}\Big(\rho\big(k,\bar{\jmath}_n(i)\big)-\rho\big(i,\bar{\jmath}_n(i)\big)\Big) \leq \frac{\Omega_0}{\sqrt{n}} + \frac{\Omega_1}{\sqrt{n}}\sqrt{2\ln\left(\frac{2|L|^2}{\delta}\right)}.$$

□

Remark : Theorem 7.14 is already well-known. The construction of this internally consistent strategy (adapted from Foster and Vohra [42]) relies on Theorem 7.12, which is implied by the existence of internally consistent strategies... However, to construct a calibrated strategy, our algorithm defines payoffs on the set of actions of Nature and then computes a strategy that has no internal regret with respect to these new payoffs. Therefore this strategy does not require that Player 1 observes his payoffs and this is precisely why our algorithm can be generalized to the partial monitoring framework.

The polytopial decomposition of $\Delta(J)$ induced by $\{b(t), c(t), t \in T\}$ is exactly the same as the one induced by $\{\gamma b(t), \gamma c(t), t \in T\}$ for $\gamma > 0$. Thus, by choosing γ small enough, $\|(b,c)\|_\infty$ — and therefore the constants in Corollary 7.13 — can be arbitrarily small (*i.e.* multiplied by any $\gamma > 0$).

However, these two Laguerre diagrams are associated to the sets of sites and weights $\mathcal{L}(1)$ and $\mathcal{L}(\gamma)$, where $\mathcal{L}(\gamma) = \{\gamma\mu(l), \gamma\omega(l) + \gamma^2\|\mu(l)\|^2 - \gamma\|\mu(l)\|, l \in L\}$. One can easily see that the constant M_P defined in Lemma 7.11 should be divided by γ if one uses $\mathcal{L}(\gamma)$ instead of $\mathcal{L}(1)$. So the constant in the proof of Theorem 7.14 does not, in fact, depend on γ. From now on, we will assume that $\|(b,c)\|_\infty$ and $\|c\|_\infty$ are smaller than 1.

7.2 Partial monitoring

Definitions

In the partial monitoring framework, Player 1 does not observe Nature's actions, he receives at stage n a random signal $s_n \in S$, whose law is $s(i_n, j_n)$ where s is a mapping from $I \times J$ to $\Delta(S)$, known from Player 1. We introduce the mapping $\mathbf{s}$

from $\Delta(J)$ to $\Delta(S)^I$ defined by $\mathbf{s}(y) = \left(\mathbb{E}_y\left[s(i,j)\right]\right)_{i\in I} \in \Delta(S)^I$ and any element of $\Delta(S)^I$ is called a flag (and is a vector of probability distributions).

Given a flag μ in the range $\mathcal{S}$ of $\mathbf{s}$, Player 1 cannot distinguish between any different mixed actions y and y' in $\Delta(J)$ that generate μ, *i.e.* such that $\mathbf{s}(y) = \mathbf{s}(y') = \mu$. Thus $\mathbf{s}$ is the maximal informative mapping about Player 2's action.

The worst payoff compatible with x and $\mu \in \mathcal{S}$ is defined by :

$$W(x,\mu) = \inf_{y\in\mathbf{s}^{-1}(\mu)} \rho(x,y), \tag{7.6}$$

and W is extended to $\Delta(S)^I$ by $W(x,\mu) = W\left(x, \Pi_{\mathcal{S}}(\mu)\right)$.

As in the full monitoring case, we define, for every $\varepsilon \geq 0$, the ε-best response multivalued mapping $\mathrm{BR}_\varepsilon : \Delta(S)^I \rightrightarrows \Delta(I)$ by :

$$\mathrm{BR}_\varepsilon(\mu) = \left\{ x \in \Delta(I), W(x,\mu) \geq \sup_{z\in\Delta(I)} W(z,\mu) - \varepsilon \right\}.$$

Given a flag $\mu \in \Delta(S)^I$, the function $W(\cdot,\mu)$ may not be linear so the best response of Player 1 might not contain any element of I.

Example : Label efficient prediction (Example 6.8 in Cesa-Bianchi and Lugosi [29]) :

Consider the following game. Nature chooses an outcome G or B and Player 1 can either observe the actual outcome (action o) or choose to not observe it and pick a label g or b. If he chooses the right label, his payoff is 1 and 0 otherwise. Payoffs and laws of signals received by Player 1 can be resumed by the following matrices (where a, b and c are three different probabilities over a finite set S).

Payoffs :

	G	B
o	0	0
g	0	1
b	1	0

and Signals :

	G	B
o	a	b
g	c	c
b	c	c

Action G, whose best response is g, generates the flag (a,c,c) and action B, whose best response is b, generates the flag (b,c,c). In order to distinguish between those two actions, Player 1 needs to observe the flag and therefore to know $\mathbf{s}(o,y)$ although action o is never a best response (but is purely informative).

Example : Matching Penny in the dark :

Consider the Matching Penny game where Player 1 does not observe the coin (but receives a signal of law c). Every choice of Nature generates the same flag (c,c). For every $x \in \Delta(\{H,T\})$, the worst compatible payoff $W(x,(c,c)) = \min_{y\in\Delta(J)} \rho(x,y)$ is non-positive and equals zero only for $x = (1/2,1/2)$. Therefore the only best response of Player 1 is to play $(1/2,1/2)$, while playing T or H gives the worst payoff of -1.

The definition of external consistency extends naturally to this framework and is quite similar to the one in the full monitoring case. In words, a strategy of Player 1 is externally consistent if he could not have improved his payoff by knowing, before the beginning of the game, the average flag :

Definition 7.15 (*Rustichini [100]*)

A strategy σ of Player 1 is externally consistent if for every strategy τ of Nature :

$$\limsup_{n\to+\infty} \max_{z\in\Delta(I)} W(z,\overline{\mu}_n) - \overline{\rho}_n \leq 0, \quad \mathbb{P}_{\sigma,\tau}\text{-as.}$$

The main issue is the definition of internally consistency. In the full monitoring case, Player 1 has asymptotically no internal regret if, for every $i \in I$, the action i is a best-response to the empirical distribution of Nature's actions, on the set of stages where i was actually played. In the partial monitoring framework, Player 1's action should be a best response to the empirical distribution of signals. Hence we will (following Lehrer and Solan [74]) distinguish the stages not as a function of the action actually played, but as a function of its law.

We assume that the strategy of Player 1 can be generated by a finite family $\{x(l) \in \Delta(I), l \in L\}$ such that, at stage $n \in \mathbb{N}$, Player 1 chooses a type l_n and, given that type, the law of its action i_n is $x(l_n)$. Roughly speaking, a strategy will be ε-internally consistent (with respect to the set L) if, for every $l \in L$, either $x(l)$ is an ε-best response to $\overline{\mu}_n(l)$ (the average flag on the set of stages where the type was l) or the frequency of the type l converges to zero.

The finiteness of L is required to get rid of trivial strategies that use only once each type $l \in L$. The choice of $\{x(l), l \in L\}$ and the description of the strategies are justified more precisely in section 7.2.

Definition 7.16 (*Lehrer and Solan [74]*)

For every $n \in \mathbb{N}$ and every $l \in L$, the average internal regret of type l at stage n is

$$\mathcal{R}_n(l) = \sup_{x\in\Delta(I)} \left[W(x,\overline{\mu}_n(l)) - \overline{\rho}_n(l)\right].$$

A strategy σ of Player 1 is (L,ε)-internally consistent if for every strategy τ of Nature :

$$\limsup_{n\to+\infty} \frac{|N_n(l)|}{n}\Big(\mathcal{R}_n(l) - \varepsilon\Big) \leq 0, \quad \forall l \in L, \quad \mathbb{P}_{\sigma,\tau}\text{-as.}$$

In words, a strategy is (L,ε)-internally consistent if, for every $l \in L$, Player 1 could not have had an uniform better payoff (of at least ε) if he had known, before the

beginning of the game, the average flag on $N_n(l)$ (as soon as the frequency $|N_n(l)|/n$ is not too small).

A naïve algorithm

Theorem 7.17 (***Lehrer and Solan [74]***)

For every $\varepsilon > 0$, there exist (L, ε)-internally consistent strategies.

Lehrer and Solan [74] proved the existence of such strategies and Perchet [92] gave a naïve algorithm. The main idea behind the latter is very similar to the full monitoring case. For simplicity, we assume in the following sketch of the proof, that Player 1 fully observes the sequence of flags $\mathbf{s}(j_n)$.

Recall that W is continuous (see Lugosi, Mannor and Stoltz [78], Proposition A.1), so for every $\varepsilon > 0$ there exist two finite families $\mathcal{M} = \{\mu(l) \in \Delta(S)^I, l \in L\}$, a δ-grid of $\Delta(S)^I$, and $X = \{x(l) \in \Delta(I), l \in L\}$ such that if μ is δ-close to $\mu(l)$ and x is δ-close to $x(l)$ then x belongs to $\mathrm{BR}_\varepsilon(\mu)$. A calibrated algorithm ensures that :

i) $\overline{\mu}_n(l)$ is asymptotically δ-close to $\mu(l)$ (because it is closer to $\mu(l)$ than to every other $\mu(k)$) ;

ii) $\bar{\imath}_n(l)$ converges to $x(l)$ (as soon as $|N_n(l)|$ is big enough), because on $N_n(l)$ the action choices of Player 1 are independent and identically distributed accordingly to $x(l)$) ;

iii) $\overline{\rho}_n(l)$ converges to $\rho(x(l), \bar{\jmath}_n(l))$ which is greater than $W(x(l), \overline{\mu}_n(l))$ (because $\bar{\jmath}_n(l)$ generates the flag $\overline{\mu}_n(l)$) ;

Therefore, $W(x(l), \overline{\mu}_n(l))$ is close to $W(x(l), \mu(l))$ which is greater than $W(z, \mu(l))$ for any $z \in \Delta(I)$ and as a consequence $\overline{\rho}_n(l)$ is asymptotically greater (up to some $\varepsilon > 0$) than $\sup_z W(z, \overline{\mu}_n(l))$, as long as $N_n(l)$ is big enough.

Once again, this approach cannot be used directly to construct $(L, 0)$-internally consistent strategy, but we will prove that one can define wisely $\{\mu(l), \omega(l), l \in L\}$ and $\{x(l), l \in L\}$ (see Proposition 7.18 and Proposition 7.10) so that $x(l) \in \Delta(I)$ is a 0-best response to any flag μ in $P(l)$, the Laguerre cell associated to $(\mu(l), \omega(l))$. The strategy associated with this choice will be $(L, 0)$-internally consistent.

Optimal algorithms

As in the full monitoring framework (see Lemma 7.9), we define for every $x \in \Delta(I)$ the x-best response area $B(x)$ as the set of flags to which x is a best response :

$$B(x) = \left\{\mu \in \Delta(S)^I,\ x \in \mathrm{BR}(\mu)\right\}.$$

Since W is continuous, $\{B(x), x \in \Delta(I)\}$ is a covering of $\Delta(S)^I$. However, one of its subsets is a finite polytopial complex :

Proposition 7.18

There exist a polytopial complex $\{A(l), l \in L\}$ of $\Delta(S)^I$ and a finite family $\{x(l) \in \Delta(I), l \in L\}$ such that $A(l) \subset B(x(l))$ for every $l \in L$.

The rather technical proof can be found in Section 7.4.

Proposition 7.18 states that there exists a finite set $\mathcal{X} \subset \Delta(I)$ that contains a best response to any flag μ. In particular, if Player 1 observes the flag μ_n before choosing his action x_n then, at every stage, x_n would be in $\mathcal{X}$. So in the description of the strategies of Player 1, the finite set $\{x(l), l \in L\} = \mathcal{X}$ is in fact intrinsic *i.e.* determined by the description of the payoff and signal functions.

OUTCOME DEPENDENT SIGNALS

In this section, we assume that the law of the signal received by Player 1 is independent of its action; formally, for every $i, i' \in I$ and every $y \in \Delta(J)$, the probability $s(i, y)$ and $s(i', y)$ are equal. Therefore, $\mathcal{S}$ (the set of realizable flags) can be seen as a polytopial subset of $\Delta(S)$. Proposition 7.18 holds in this framework, therefore there exists a finite family $\{x(l), l \in L\}$ such that for any flag $\mu \in \mathcal{S}$, for some $l \in L$ $x(l)$ is a best-reply to μ (and for a fixed $l \in L$, the set of such μ is a polytope).

Theorem 7.19

There exists a $(L, 0)$-internally consistent strategy σ such that for every strategy τ of Nature, with $\mathbb{P}_{\sigma,\tau}$-probability at least $1 - \delta$:

$$\sup_{l \in L} \frac{N_n(l)}{n} \mathcal{R}_n(l) \leq O\left(\sqrt{\frac{\ln\left(\frac{1}{\delta}\right)}{n}}\right). \tag{7.7}$$

Proof. Proposition 7.10 and Proposition 7.18 imply that there exist two finite families $\{x(l), l \in L\}$ and $\{\mu(l), \omega(l), l \in L\}$ such that $x(l)$ is a best response to any μ in $P(l)$, the Laguerre cell associated to $\mu(l)$ and $\omega(l)$. Assume, for the moment, that for any two different l and k in L, the probabilities $x(l)$ and $x(k)$ are different.

The strategy σ is defined as follows. Compute a strategy $\widehat{\sigma}$ calibrated with respect to the grid $\{\mu(l), \omega(l), l \in L\}$. When Player 1 should play $l \in L$ accordingly to $\widehat{\sigma}$, he plays accordingly to $x(l)$ in the original game. Corollary 7.13 (with the assumption that $\|c\|_\infty$ and $\|(b, c)\|_\infty$ are smaller than 1) implies that with $\mathbb{P}_{\sigma,\tau}$ probability at

least $1-\delta_1$:

$$\max_{l\in L}\frac{|N_n(l)|}{n}\Big(\left[\|\overline{s}_n(l)-\mu(l)\|^2-\omega(l)\right]-\left[\|\overline{s}_n(l)-\mu(k)\|^2-\omega(k)\right]\Big)\leq \frac{8\sqrt{L}}{\sqrt{n}}+\frac{8}{\sqrt{n}}\sqrt{2\ln\left(\frac{2|L|^2}{\delta_1}\right)},$$

therefore combined with Lemma 7.11, this yields that :

$$\max_{l\in L}\frac{|N_n(l)|}{n}\|\overline{s}_n(l)-\widetilde{\mu}_n(l)\|\leq\frac{8M_P\sqrt{L}}{\sqrt{n}}+\frac{8M_P}{\sqrt{n}}\sqrt{2\ln\left(\frac{2|L|^2}{\delta_1}\right)}, \tag{7.8}$$

where $\widetilde{\mu}_n(l)$ is the projection of $\overline{s}_n(l)$ onto $P(l)$.

Hoeffding-Azuma's inequality [61, 11] implies that with $\mathbb{P}_{\sigma,\tau}$ probability at least $1-\delta_2$:

$$\max_{l\in L}\frac{N_n(l)}{n}\|\overline{s}_n(l)-\overline{\mu}_n(l)\|\leq\sqrt{\frac{2\ln\left(\frac{2|S||L|}{\delta_2}\right)}{n}} \tag{7.9}$$

and with probability at least $1-\delta_3$:

$$\max_{l\in L}\frac{N_n(l)}{n}|\overline{\rho}_n(l)-\rho(x(l),\overline{\jmath}_n(l))|\leq M_\rho\sqrt{\frac{2\ln\left(\frac{2|L|}{\delta_3}\right)}{n}}. \tag{7.10}$$

W is M_W-Lipschitz in μ (see Lugosi, Mannor and Stoltz [78]) and $\mathbf{s}\left(\overline{\jmath}_n(l)\right)=\overline{\mu}_n(l)$ therefore :

$$\overline{\rho}_n(l)\geq W\left(x(l),\widetilde{\mu}_n(l)\right)-|\overline{\rho}_n(l)-\rho(x(l),\overline{\jmath}_n(l))|-M_W\|\overline{\mu}_n(l)-\widetilde{\mu}_n(l)\| \tag{7.11}$$

and

$$\begin{aligned}\max_{x\in\Delta(I)}W\left(x,\overline{\mu}_n(l)\right)\leq&\max_{x\in\Delta(I)}W\left(x,\widetilde{\mu}_n(l)\right)+M_W\left(\|\overline{s}_n(l)-\overline{\mu}_n(l)\|+\|\overline{s}_n(l)-\widetilde{\mu}_n(l)\|\right)\\ \leq&\quad W\left(x(l),\widetilde{\mu}_n(l)\right)+M_W\big(\|\overline{s}_n(l)-\overline{\mu}_n(l)\|+\|\overline{s}_n(l)-\widetilde{\mu}_n(l)\|\big)\end{aligned} \tag{7.12}$$

since $x(l)$ is a best response to $\widetilde{\mu}_n(l)$. Equations (7.11) and (7.12) yield

$$\mathcal{R}_n(l)\leq 2M_W\|\overline{s}_n(l)-\overline{\mu}_n(l)\|+2M_W\|\overline{s}_n(l)-\widetilde{\mu}_n(l)\|+|\overline{\rho}_n(l)-\rho(x(l),\overline{\jmath}_n(l))|. \tag{7.13}$$

Combining equations (7.8), (7.9), (7.10) and (7.13) gives that, with probability at least $1-\delta$ and if we define $\Omega_0=16M_PM_W\sqrt{L}$, $\Omega_1=(2M_W+16M_WM_P+M_\rho)$ and $\Omega_2=L\left(L+S+1\right)$,

$$\sup_{l\in L}\frac{N_n(l)}{n}\mathcal{R}_n(l)\leq\frac{\Omega_0}{\sqrt{n}}+\Omega_1\sqrt{2\ln\left(\frac{2\Omega_2}{\delta}\right)} \tag{7.14}$$

If there exist l and k such that $x(l) = x(k)$, then although Player 1 made two different predictions ($\mu(l)$ and $\mu(k)$), he played accordingly to the same probability $x(l) = x(k)$. Define $N_n(l,k)$ as the set of stages where Player 1 predicts either $\mu(l)$ or $\mu(k)$ up to stage n, $\overline{\mu}_n(l,k)$ as the average flag on this set, $\overline{\rho}_n(l,k)$ as the average payoff and $\mathcal{R}_n(l,k)$ as the regret. Since $W(x,\cdot)$ is convex for every $x \in \Delta(I)$, then $\max_{x\in\Delta(I)} W(x,\cdot)$ is also convex so

$$\frac{|N_n(l,k)|}{n} W(x,\overline{\mu}_n(l,k)) \leq \frac{|N_n(l)|}{n} \max_{x\in\Delta(I)} W(x,\overline{\mu}_n(l)) + \frac{|N_n(k)|}{n} \max_{x\in\Delta(I)} W(x,\overline{\mu}_n(k))$$

$$\text{and} \quad -\frac{|N_n(l,k)|}{n}\overline{\rho}_n(l,k) = -\frac{|N_n(l)|}{n}\overline{\rho}_n(l) - \frac{|N_n(k)|}{n}\overline{\rho}_n(k)$$

so we still have

$$\frac{|N_n(l,k)|}{n}\mathcal{R}_n(l,k) \leq O\left(\sqrt{\frac{\ln\left(\frac{1}{\delta}\right)}{n}}\right).$$

Hence the previous bound holds. □

Remark : Lugosi, Mannor and Stoltz [78] have constructed an externally consistent strategy, *i.e.* such that, asymptotically, for any strategy τ of Nature :

$$\overline{\rho}_n \geq \max_{z\in\Delta(I)} W\left(z,\overline{\mu}_n\right), \quad \mathbb{P}_{\sigma,\tau}\text{−as.}$$

The final argument in the proof of Theorem 7.19 also implies that a $(L,0)$-internally consistent strategy is also externally consistent, hence we can compare bounds between our algorithm.

If the signals are deterministic, Lugosi, Mannor and Stoltz [78] showed that the expected regret is smaller than $O\left(n^{-1/2}\right)$. However this bound became, with random signals, $O\left(n^{-1/4}\right)$. Thus our algorithm, along with computing no internal regret, has a better bound.

ACTION-OUTCOME DEPENDANT SIGNALS

In this section, we drop the assumption that the laws on the signals do not depend of Player 1's actions.

Theorem 7.20

There exists an $(L,0)$-internally consistent strategy σ such that, for every strategy τ of Nature, with $\mathbb{P}_{\sigma,\tau}$ probability at least $1-\delta$:

$$\max_{l\in L} \frac{N_n(l)}{n}\mathcal{R}_n(l) \leq O\left(\frac{1}{n^{1/3}}\sqrt{\ln\left(\frac{1}{\delta}\right)} + \frac{1}{n^{2/3}}\ln\left(\frac{1}{\delta}\right)\right). \tag{7.15}$$

Proof. The proof is essentially the same as for Theorem 7.19, so we can assume that $x(l) \neq x(k)$ for any two different l and k in L. The main difference is due to the assumption that the signals depend on Player 1's action which implies that he just observes one component of μ_n.

Following Auer, Cesa-Bianchi, Freund and Schapire [6], we define for every $l \in L$ and $n \in \mathbb{N}$, the $\widehat{\gamma}_n$-perturbation of $x(l)$ by $\widehat{x}(l,n) = (1-\widehat{\gamma}_n)x(l) + \widehat{\gamma}_n u$ where u is the uniform probability over I and $(\widehat{\gamma}_n)_{n\in\mathbb{N}}$ is a non-negative non-increasing sequence. For every $n \in \mathbb{N}$, let $c_n = \mathbb{1}_{(i,s)=(i_n,s_n)}/\widehat{x}(l_n,n)[i_n] \in \mathbb{R}^{|I||S|}$ where $\widehat{x}(l_n,n)[i_n] \geq \gamma_n = \widehat{\gamma}_n/I$ is the weight put by $\widehat{x}(l_n,n)$ on i_n. With this notation, c_n is an unbiased estimator of μ_n since $\mathbb{E}_{\sigma,\tau}[c_n] = \mu_n$, where μ_n is seen as an element of $\mathbb{R}^{|S||I|}$.

Player 1 will compute $\widetilde{\sigma}$ a calibrated strategy (in an auxiliary game Γ_c) with respect to $\{\mu(l), \omega(l), l \in L\}$ but where the signal at stage n is $c_n \in \mathbb{R}^{|I||S|}$. When he should play l in Γ_c accordingly to $\widetilde{\sigma}$, then he plays accordingly to $\widehat{x}(l)$ in the original game. Conditional variances are bounded as

$$\mathrm{Var}_n\left(\frac{\mathbb{1}_{i_n=i,s_n=s}}{\widehat{x}(l_n,n)(i_n)}\right) \leq \frac{1}{\gamma_n},$$

therefore Freedman's inequality implies that — with $\mathbb{P}_{\sigma,\tau}$ probability at least $(1-\delta_1)$ — for every $l \in L$:

$$\frac{|N_n(l)|}{n}\|\overline{c}_n(l) - \widetilde{\mu}_n(l)\| \leq \frac{2M_0M_P}{\sqrt{\gamma_n n}} + \frac{2v_0M_P}{\sqrt{\gamma_n n}}\sqrt{2\ln\left(\frac{2|L|^2}{\delta_1}\right)} + \frac{2}{3}\frac{K_0M_P}{\gamma_n n}\ln\left(\frac{2|L|^2}{\delta_1}\right)$$

because σ is calibrated and with $M_0 = \sqrt{\gamma_n}M_n \leq 4\sqrt{|L|}$, $K_0 = \gamma_n K_n \leq 8$ and $v_0 = \sqrt{\gamma_n}v_n \leq 4$.

The same argument implies that with probability at least $1-\delta_2$, for every $l \in L$

$$\frac{N_n(l)}{n}\|\overline{c}_n(l) - \overline{\mu}_n(l)\| \leq \sqrt{IS}\left(\sqrt{2\frac{1}{n\gamma_n}\ln\left(\frac{2|L||I||S|}{\delta_2}\right)} + \frac{2}{3n\gamma_n}\ln\left(\frac{2|L||I||S|}{\delta_2}\right)\right)$$

and Hoeffding-Azuma's inequality [61, 11] implies that with probability at least $1-\delta_3$:

$$\max_{l\in L}\frac{N_n(l)}{n}|\overline{\rho}_n(l) - \rho(x(l),\overline{\jmath}_n(l))| \leq M_\rho\sqrt{\frac{2}{n}\ln\left(\frac{2|L|}{\delta_3}\right)} + 2M_\rho\frac{\sum_{m\in N_n(l)}\gamma_m}{n}.$$

Hence, by taking $\gamma_n = n^{-1/3}$, one has $\sum_{m\in N_n(l)}\gamma_m \leq 3/2n^{2/3}$ and for every $l \in L$, with probability at least $1-\delta$:

$$\frac{N_n(l)}{n}\mathcal{R}_n(l) \leq \frac{\Omega_1}{n^{1/3}} + \frac{\Omega_2}{n^{1/3}}\sqrt{2\ln\left(\frac{2\Omega_5}{\delta}\right)} + \frac{\Omega_3}{n^{1/2}}\sqrt{2\ln\left(\frac{2\Omega_5}{\delta}\right)} + \frac{2}{3}\frac{\Omega_4}{n^{2/3}}\ln\left(\frac{2\Omega_5}{\delta}\right)$$

where the constants are $\Omega_1 = 16M_P M_W\sqrt{L} + 3M_W M_\rho$, $\Omega_2 = 2M_W\left(8M_P + \sqrt{IS}\right)$, $\Omega_3 = M_\rho$, $\Omega_4 = 2M_W(8M_P + \sqrt{IS})$ and $\Omega_5 = L\,(L + 2 + 2IS)$.

Those constants can be much smaller if we use concentration inequalities in Hilbert spaces (see section 7.3). □

Remark : In the label efficient prediction game defined in Example 7.2, for every strategy σ of Player there exists a sequence of outcomes such that the forecaster expected regret is greater than $n^{-1/3}/7$ (see Theorem 5.1 in Cesa-Bianchi, Lugosi and Stoltz [30]). Therefore the rate of $n^{-1/3}$ of our algorithm is optimal either for internal and external regret.

7.3 Concluding remarks

Second algorithm : calibration and polytopial complex.

The algorithms we described are quite easy to run stage by stage since Player 1 only needs to compute some invariant measures of non-negative matrices. However, this requires to construct the Laguerre diagram $\mathcal{P} = \{P(l), l \in L\}$ given the set $\{b_t, c_t, t \in T\}$. And we have shown that L which is a factor both in the complexity of the algorithms and in their rate of convergence can be in the order of $T^{|S||I|}$. This can be much bigger than L_0, the number of cells in the original polytopial decomposition of $\Delta(S)^I$,.

This section is devoted to a modification of the algorithm that does not require to compute a Laguerre diagram but which is more difficult, stage by stage, to implement. The only difference between the two algorithms is in the definition of calibration.

Let $\{K(l), l \in L_0\}$ be a finite polytopial complex of $\Delta(S)$ defined by two finite families $\{c_t \in \Delta(S), b_t \in \mathbb{R}, t \in T\}$ such that :

$$K(l) = \{\mu \in \Delta(S), \langle \mu, c_t\rangle \leq b_t, \forall t \in T(l) \subset T\} \quad \forall l \in L_0.$$

Let us define $(c_{t,l}, b_{t,l}) = (c_t, b_t)$ if $t \in T(l)$ and $(c_{t,l}, b_{t,l}) = (0, 0)$ otherwise. Then $K(l) = \{\mu \in \Delta(S), \langle \mu, c_{t,l}\rangle \leq b_{t,l}, \forall t \in T\}$.

Definition 7.21

A strategy σ of Player 1 is calibrated with respect to the complex $\{K(l), l \in L_0\}$ if for every strategy τ of Nature, $\mathbb{P}_{\sigma,\tau}$-as :

$$\limsup_{n\to\infty} \frac{|N_n(l)|}{n}\left(\langle \overline{s}_n(l), c_{t,l}\rangle - b_{t,l}\right) \leq 0, \quad \forall t \in T, \forall l \in L_0.$$

Theorem 7.22

There exist calibrated strategies with respect to any finite polytopial complex $\{K(l), l \in L_0\}$.

Proof. Consider the following auxiliary two-person game Γ'_c, where at stage $n \in \mathbb{N}$ Player 1 (resp. Nature) chooses $l_n \in L_0$ (resp. $\mu_n \in S$) which generates the vector payoff $U_n \in \mathbb{R}^d$ (with $d = |T||L_0|$) defined by :

$$U_n^{lk} = \begin{cases} \langle \mathbb{1}_{(s_n,i_n)=(s,i)}, c_{t,l} \rangle - b_{t,l} & \text{if } l = l_n \\ 0 & \text{otherwise.} \end{cases}$$

Any strategy that approaches the negative orthant Ω_- in Γ'_c is calibrated with respect to the complex $\{K(l), l \in L_0\}$.

Blackwell's characterization of approachable convex sets (see Blackwell [17], Theorem 3) implies that Player 1 can approach the convex set Ω_- if (and only if) for every mixed action of Nature in $\Delta(S)$, he has an action $x \in \Delta(L_0)$ such that the expected payoff is in Ω_-. It is clear that given $\mu_n \in \Delta(S)$, playing $l(\mu_n) \in L_0$, where $l(\mu_n)$ is the index of the polytope that contains μ, ensures that $\mathbb{E}_{\mu_n, l(\mu_n)}[U_n]$ is in Ω_-. Therefore there exist calibrated strategies with respect to any polytopial complex. □

This modification of the definition of calibration does not change the other part of the algorithm nor the remaining of the proof (and, once again, Player 1 has to use γ_n-perturbations of his actions, in order to calibrate the sequence of unobserved flags). The constants in the rates of convergence are smaller for two reasons : $|L_0|$ is smaller than L and, in Γ'_c, $\sqrt{\mathbb{E}[U_n^2]}$ is bounded by $\sqrt{\frac{T_0}{\gamma_n}}$ where $T_0 = \sup_{l \in L_0} |T(l)|$ is the maximum number of hyperplanes defining a polytope of the complex.

Extension to compact case

Assume that instead of choosing j_n at stage $n \in \mathbb{N}$ (which generates the flag $\mu_n = \mathbf{s}(j_n)$ and an outcome vector $(\rho(i, j_n))_{i \in I}$), Nature chooses an outcome vector $O_n \in [-1, 1]^I$ and a flag μ_n which belongs to $\mathbf{s}(O_n)$ where $\mathbf{s} : [-1, 1]^I \rightrightarrows \Delta(S)^I$. As before, Player 1's payoff is $O_n^{i_n}$ (the i_n-th coordinate of O_n) and he receives a signal s_n whose law is $\mu_n^{i_n}$. Strategies of Player 1 and Nature and consistency are defined as before.

Theorem 7.23

Assume that the graph of $\mathbf{s}$ *is a polytope, then there exists an* $(L, 0)$*-internally consistent strategy* σ *such that, for every strategy* τ *of Nature, with* $\mathbb{P}_{\sigma,\tau}$ *probability at least* $1 - \delta$ *:*

$$\max_{l \in L} \frac{N_n(l)}{n} \mathcal{R}_n(l) \leq O\left(\frac{1}{n^{1/3}} \sqrt{\ln\left(\frac{1}{\delta}\right)} + \frac{1}{n^{2/3}} \ln\left(\frac{1}{\delta}\right) \right). \tag{7.16}$$

The proof of this result is identical to the one of Theorem 7.20.

Note that the assumption that the graph of $\mathbf{s}$ is a polyhedron is fulfilled in the finite dimension case. The fact that $\mathbf{s}$ is multivalued is a consequence of the fact that in finite dimension there might exist two different mixed actions y_1, y_1 in $\Delta(J)$ that generate the same outcome vectore (*i.e.* $\rho(\cdot, y_1) = \rho(\cdot, y_2)$) but different flags (*i.e.* $\mathbf{s}(y_1) \neq \mathbf{s}(y_2)$).

Strengthening of the constants

We propose two different ideas to strengthen the constants of our algorithm. First, we can use, instead of one concentration inequality per component of the sequence of vectors $\overline{R}_{\omega,n}$, only one concentration inequality for every component. Second, we can implement sparser vector payoffs (and therefore we decrease its norm) by looking at a slight different definition of calibration.

CONCENTRATION INEQUALITIES IN HILBERT SPACES

The rates of convergence of our algorithms rely mainly on three properties : Blackwell's approachability theorem, Hoeffding-Azuma's and Freedman's inequalities. These tools allowed us to study the convergence of a sequence of vectors $\overline{U}_n^+$ towards 0. Approachability is well defined for sequences of vectors, however it is not the case for the two concentration inequalities that hold only for real valued martingales. In our proof, we use the fact that if a process $\left\{U_n \in \mathbb{R}^d\right\}_{n\in\mathbb{N}}$ is a martingale then, componentwise, each process $\left\{U_n^k \in \mathbb{R}\right\}_{n\in\mathbb{N}}$ is a (real valued) martingale. However this approach does not use the fact that U_n is sparse. Thus, the use of concentration inequalities in Hilbert space can sharpen the constant.

Indeed, recall Hoeffding-Azuma's inequality :

Lemma 7.24 (***Hoeffding-Azuma's inequality [61, 11]***)

Let U_n be a sequence of martingale differences (i.e. $\mathbb{E}_{\sigma,\tau}\left[U_{n+1}|h_n\right] = 0$) bounded by K (i.e. $|U_n| < K$ almost-surely for every $n \in \mathbb{N}$).

Then for every $n \in \mathbb{N}$ and every $\varepsilon > 0$:

$$\mathbb{P}_{\sigma,\tau}\left(\left|\overline{U}_n\right| \geq \varepsilon\right) \leq 2\exp\left(\frac{-n\varepsilon^2}{2K^2}\right),$$

which can be expressed as

$$\mathbb{P}_{\sigma,\tau}\left(\left|\overline{U}_n\right| \leq K\sqrt{2\ln\left(\frac{2}{\delta}\right)}\right) \leq 1-\delta. \tag{7.17}$$

Chen and White [31] proved an equivalent property for vector martingale in $\mathbb{R}^d$.

Lemma 7.25 (*Chen and White [31]*)

Let U_n be a sequence of martingale differences in $\mathbb{R}^d$ bounded almost-surely by $K > 0$. Then for every $n \in \mathbb{N}$ and for every $\varepsilon > 0$:

$$\mathbb{P}_{\sigma,\tau}\left(\left\|\overline{U}_n\right\| \geq \varepsilon\right) \leq 2\max\left\{1, \sqrt{\frac{n\varepsilon^2}{2K^2}}\right\}\exp\left(\frac{-n\varepsilon^2}{2K^2}\right).$$

In the algorithm described in the partial monitoring framework (in the outcome dependent case), by only using Hoeffding-Azuma's inequality, we deduce the following inequality

$$\mathbb{P}_{\sigma,\tau}\left(\max_{l,k}\frac{N_n(l)}{n}\left|\overline{U}_n^{l,k}\right| \geq \varepsilon\right) \leq 2L^2\exp\left(\frac{-n\varepsilon^2}{8}\right).$$

However, Chen and White's result, along with the fact that $\|U_n\| \leq O\left(2\sqrt{|L|}\right)$, implies that :

$$\mathbb{P}_{\sigma,\tau}\left(\max_{l,k}\frac{N_n(l)}{n}\left|\overline{U}_n^{l,k}\right| \geq \varepsilon\right) \leq O\left(2\max\left\{1, \sqrt{\frac{n\varepsilon^2}{8|L|}}\right\}\exp\left(\frac{-n\varepsilon^2}{8|L|}\right)\right),$$

so the dependency in $|L|$ can be quite reduced.

There also exist variants of Bernstein's inequality (see *e.g.* Yurinskii [120]) in Hilbert spaces that can be used in order to get more precise bounds.

CALIBRATION WITH RESPECT OF NEIGHBORHOODS

Definition 7.26

Given a finite set $\mathcal{M} = \{(\mu(l), \omega(l)), l \in L\} \subset \mathbb{R}^d \times \mathbb{R}$, we say that $\mu(k)$ is a neighbor of $\mu(l)$ if $k \neq l$ and the dimension of the intersection between the two Laguerre cells $P(l)$ and $P(k)$ is equal to $d-1$.

We defined a calibrated strategy with respect to $\mathcal{M}$, as a strategy σ such that $\overline{s}_n(l)$ is asymptotically closer to $\mu(l)$ than to any other $\mu(k)$ (or $|N_n(l)|/n$ goes to zero). In fact, $\overline{s}_n(l)$ needs only to be closer to $\mu(l)$ than to any of its neighbors. So one can construct calibrated strategies by modifying the algorithm given in Proposition 7.5 ; the payoff at stage n is now denoted by U'_n and is defined by :

$$(U'_n)^{lk} = \begin{cases} \|s_n - \mu(l)\|^2 - \|s_n - \mu(k)\|^2 & \text{if } l = l_n \text{ and } k \text{ is a neighbor of } l \\ 0 & \text{otherwise} \end{cases}$$

It is clear that the strategy that consists in playing an invariant measure of $\overline{U}'_n$ is calibrated. The squared maximal second order moment $M_n^2 = \sup_{m\leq n}\mathbb{E}_{\sigma,\tau}\left[\|U_m\|^2\right]$ equals $4\mathcal{N}$, where $\mathcal{N}$ is the maximal number of neighbors, instead of $4|L|$.

If we consider ε-calibration, the merit of this modification is even more clear. In order to construct ε-calibrated strategies, we usually take any ε-discretization L of $\Delta(S)$ so $|L| = O\left(\varepsilon^{-(|S|-1)}\right)$. However, it is easy to show that there exists a discretization such that $\mathcal{N} = 2^{-d}$ which is independent of ε.

7.4 Proofs of technical results

This section is devoted to the proofs of previously mentioned results, *i.e.* Proposition 7.18 and Lemma 7.11.

Proof of proposition 7.18

Definition 7.27

Let K^1 be a polytope. A correspondence $B : K^1 \rightrightarrows K^2$ is polytopial constant, if there exists $\{K^1(l), l \in L\}$ a finite polytopial complex of K^1, such that $B(\cdot)$ is constant on the interior of $K^1(l)$, for every $l \in L$.

This definition implies that for every $l \in L$, there exists $x(l)$ that belongs to every $B(\mu)$ for all $\mu \in K^1(l)$. Let us now restate Proposition 7.18 :

Proposition 7.28

BR *is polytopial constant.*

This theorem is well-known and quite useful in the full monitoring case (see for example the Lemke-Howson algorithm [76]). In the *compact case*, Proposition 7.18 becomes :

Proposition 7.29

If $\mathbf{s}$ *has a polytopial graph, then* BR *is polytopial constant.*

The proofs of both propositions rely on polytopial parameterized max-min program defined in the next subsection.

CONSTANT SOLUTION OF A POLYTOPIAL PARAMETERIZED MAX-MIN PROGRAM

A Polytopial Parameterized Max-Min Program (PPMP) is defined as follows : let $\mathcal{X}$ and $\mathcal{Y}$ be two Euclidian spaces of respective dimension d_1 and d_2. Consider the program (P_μ), that depends on a parameter $\mu \in \mathcal{M}$ (a polytope in $\mathbb{R}^{d_3}$), defined by

$$(P_\mu) : \quad \begin{array}{ccc} \max & \min & xAy, \\ x \in \mathcal{X} & y \in \mathcal{Y} & \\ s.t.\ Dx \leq d & s.t.\ E_\mu y \leq e_\mu & \end{array}$$

where A is a $d_1 \times d_2$ matrix, $\{E_\mu, e_\mu, \mu \in \mathcal{M}\}$ is a family of matrices and vectors (we do not specify the sizes the matrices, as long as each inequality makes sense) and D, d are also a fixed matrix and vector such that the admissible set $\mathbf{D} = \{x \in X, Dx \leq d\}$ is a polytope. The solution set of (P_μ) is denoted by $B(\mu) \subset \mathcal{X}$.

Theorem 7.30

Assume that the correspondence S defined by :

$$S : \begin{array}{ccc} \mathcal{M} & \rightrightarrows & \mathcal{Y} \\ \mu & \mapsto & S_\mu = \{y \in \mathcal{Y}, E_\mu y \leq e_\mu\} \end{array}$$

has a polytopial graph $\mathbf{S}$. Then $B : \mathcal{M} \rightrightarrows \mathcal{X}$ is polytopial constant.

Proof. The proof consists in two parts ; first, we prove that there exists a finite subset X of $\mathcal{X}$ such that $B(\mu) \cap X \neq \emptyset$ for every $\mu \in \mathcal{M}$. Second, we show that the family of $B^{-1}(x) := \{\mu \in \mathcal{M}, \text{st } x \in B(\mu)\}$ is a polytopial complex of $\mathcal{M}$. Figure 7.1 illustrates the main parts of the proof for a simple example. It is well known that :

i) a linear program is minimized on a vertex of the polytopial feasible set ;

ii) given $x \in \mathcal{X}$ and $\mu \in K^1$, if y minimizes xAy on S_μ then

$$-xA \in \mathrm{NC}_{S_\mu}(y),$$

where $\mathrm{NC}_C(y)$ is the normal cone to the convex $C \subset \mathbb{R}^d$ at $y \in C$ defined by :

$$N_C(y) = \left\{p \in \mathbb{R}^d; \langle p, z - y\rangle, \forall z \in C\right\};$$

iii) Given a polytope P defined by $P = \{x \in \mathbb{R}^d, \langle c_t, x\rangle \leq b_t, t \in T\}$ (where T is a finite set) and any point x_0 in P, we denote by $T(x_0)$ the set of active constraints at x_0, *i.e.* $T(x_0) = \{t \in T, \langle c_t, x_0\rangle = b_t\}$. The normal cone to P at x_0 is (see *e.g.* Rockafellar and Wets [98], Theorem 6.46) :

$$\mathrm{NC}_P(x_0) = \left\{\sum_{t \in T} \xi_t c_t, \text{ where } \xi_t \geq 0, \forall t \in T(x_0) \text{ and } \xi_t = 0, \forall t \notin T(x_0)\right\}. \tag{7.18}$$

This implies that $\mathrm{NC}_P(x_0)$ is a polyhedral cone and $\{\mathrm{NC}_P(v), v \text{ vertex of } P\}$ is a polyhedral complex of $\mathbb{R}^d$ (*i.e.* a finite family of polyhedra that cover $\mathbb{R}^d$ and such that each pair has an intersection with empty interior) called the normal fan of P, see Ziegler [122], example 7.3 page 193.

iv) Let P be a polytope in $\mathbb{R}^k \times \mathbb{R}^n$. For every $x \in \mathbb{R}^k$, let us define the polytope P_x by $P_x = \{y \in \mathbb{R}^n, (x, y) \in P\}$. Equation (7.18) implies that $\mathrm{NC}_{P_x}(y) = \Pi_{\mathbb{R}^n}\left(\mathrm{NC}_P\left((x, y)\right)\right)$, where $\Pi_{\mathbb{R}^n}$ is the projection from $\mathbb{R}^k \times \mathbb{R}^n$ into $\mathbb{R}^n$.

For every $\mu \in \mathcal{M}$, $(\mu, S_\mu) := \{\mu\} \times S_\mu$ is a polytope whose extreme points (their set is denoted by $\mathcal{E}(\mu, S_\mu)$) belong to faces of $\mathbf{S}$. Let $\mathcal{F} = \{F(l), l \in L\}$ denotes the

set of faces of $\mathbf{S}$. For every subset $L' \subset L$, we call $\mathcal{S}(L')$ the subset of $\mathcal{M}$ defined by :

$$S(L') = \left\{ \mu \in \mathcal{M}; \mathcal{E}\left(\mu, S_\mu\right) \cap F_0(l) \text{ is a singleton }, \forall l \in L \right\} \tag{7.19}$$

where for any face $l \in L$ of $\mathbf{S}$, $F_0(l) \subset F(l)$ is the set of points in $F(l)$ that does not belong to any other face of lower dimension — it is, in fact, equal to the relative interior of $F(l)$. Equation (7.19) defines for every $l \in L$ an affine function :

$$y_l(\cdot) : S(L') \mapsto \mathcal{Y} \quad \text{by} \quad (\mu, y_l(\mu)) := \mathcal{E}\left(\mu, S_\mu\right) \cap F_0(l).$$

Let $L' \subset L$ such that $\mathcal{S}(L')$ is not empty. We know that (see point iv)) :

$$\mathrm{NC}_{S_\mu}\left[y_l(\mu)\right] = \Pi_{\mathcal{Y}}\left(\mathrm{NC}_{\mathbf{S}}\left[(\mu, y_l(\mu))\right] \right) := N_l,$$

since $\mathrm{NC}_{\mathbf{S}}\left[(\mu, y_l(\mu))\right]$ is independent of $\mu \in \mathcal{S}(L')$ (because of equation (7.18)). Therefore $\mathcal{N}(L') = \{N_l, l \in L'\}$, the normal fan of S_μ, is constant on $\mathcal{S}(L')$. Moreover $\mathcal{N}(L')$ refines the normal fan of S_μ, for any μ in the closure of $\mathcal{S}(L')$, denoted by $\overline{\mathcal{S}}(L')$. This is also a direct consequence of equation (7.18) and it allows us to extend $y_l(\cdot)$ by continuity to $\overline{\mathcal{S}}(L')$, for every $l \in L'$.

Let $\mathcal{L}$ be the set defined by :

$$\mathcal{L} = \left\{ L'; L \subset L, \overline{\mathcal{S}}(L') \text{ has a non empty interior.} \right\}$$

Then $\mathcal{K}_1 = \left\{ \overline{\mathcal{S}}(L'), L' \in \mathcal{L} \right\}$ is a polytopial complex of $\mathcal{M}$ and it has the following main property :

for every $L' \in \mathcal{L}$ and for every $\mu \in \overline{\mathcal{S}}(L')$, xAy is minimized on S_μ at $y_l(\mu)$ if and only if $-xA$ belongs to N_l.

Let $L' \in \mathcal{L}$ and $l \in L'$ be fixed. The following equality holds for every $\mu \in \overline{\mathcal{S}}(L')$,

$$(P_{L',l,\mu}) : \quad \max_{x \in \mathbf{D} \text{ st } -xA \in N_l} \min_{y \in S_\mu} xAy = \max_{x \in \mathbf{D} \text{ st } -xA \in N_l} xAy_l(\mu). \tag{7.20}$$

Therefore $(P_{L',l,\mu})$ is a linear program, maximized on a vertex of the following polytope $\{x \in \mathbf{D}, -xA \in N_l\}$. Denote its set of vertices by $\{x_l^k, k \in V(L', l)\}$, so that we can rewrite the linear program :

$$(P_{L',l,\mu}) : \quad \max_{x \in \mathbf{D} \text{ st } -xA \in N_l} \min_{y \in S_\mu} xAy = \max_{k \in V(L',l)} x_l^k A y_l(\mu)$$

Since $\mu \mapsto y_l(\mu)$ is an affine mapping from $\overline{\mathcal{S}}(L')$ into $F(l)$, for every $k \in V(L', l)$, x_l^k is a solution of $(P_{L',l,\mu})$ on a polytopial subset of K_1, denoted by $K_{1,l,k}$. This defines, for every $l \in L$, a polytopial complex $\mathcal{K}_{1,l} = \{K_{1,l,k}, k \in V(L', l)\}$ of K_1.

Define $\mathcal{K}_{1,2}(L')$ the polytopial complex generated by the intersection of $\mathcal{K}_{1,l}$ by :

$$\mathcal{K}_{1,2}(L') = \left\{ K_{1,2} \subset K_1; \forall l \in L', \exists k(l) \in V(L', l), K_{1,2} = \bigcap_{l \in L'} K_{1,l,k(l)} \right\}.$$

Let $K_{1,2}$ be a fixed polytope of the complex $\mathcal{K}_{1,2}(L')$. By definition, there exists a finite family $\left\{ x_l^{k(l)}, y_l(\cdot);\ l \in L',\ k(l) \in V(L', l) \right\}$ such that :

$$\max_{x \in \mathbf{D}} \min_{y \in S_\mu} xAy = \max_{l \in L'} x_l^{k(l)} A y_l(\mu), \qquad \text{if } \mu \in K_{1,2}. \tag{7.21}$$

Since each $y_l(\cdot)$ is affine, the maximum in equation (7.21) is attained at a specific $x_l^{k(l)}$ on $K_{1,2,l}$, a polytopial subset of $K_{1,2}$.

Finally, B is constant on $\mathcal{K}_0 = \{K_{1,2,l}, l \in L', K_{1,2} \in \mathcal{K}_{1,2}(L'), L' \in \mathcal{L}\}$, a polytopial complex of K^1. □

We can now prove simultaneously Propositions 7.28 and 7.29 :

Proof of Propositions 7.28 and 7.29 Since $\mathbf{s}$ is linear, its graph, denoted by $\mathbf{S}$, is a polytope. Theorem 7.30 (with $\mathbf{D} = \Delta(I)$) implies that the solution, denoted by $B(\mu)$ for every $\mu \in \mathcal{S}$, of the parameterized program

$$\max_{x \in \Delta(I)} \min_{y \in \mathbf{s}^{-1}(\mu)} \rho(x, y)$$

is polytopial constant (and we denote by $\{K(l), l \in L\}$ a corresponding polytopial complex). It is clear that if B is constant on $K(l)$, then it is obviously also constant on $\widehat{K}(l) = \Pi_{\mathbf{S}}^{-1}(K(l))$, which is a finite union of polytopes. □

Proof of Lemma 7.11

For every $l \in L$, there exist a finite family $\left\{ c_t \in \mathbb{R}^d, b_t \in \mathbb{R}, t \in T_l \right\}$ such that :

$$P(l) = \left\{ \mu \in \mathbb{R}^d, \langle \mu, c_t \rangle \leq b_t, \forall t \in T_l \right\}.$$

For every $\varepsilon \geq 0$, we define

$$P_\varepsilon(l) = \left\{ \mu \in \mathbb{R}^d, \langle \mu, c_t \rangle \leq b_t + \varepsilon, \forall t \in T_l \right\}.$$

Equation (7.4) can be rephrased as : if x belongs to $P_\varepsilon(l)$ then $d(x, P(l))$ is smaller than $M_P \varepsilon$. By convexity of $d(\cdot, P(l))$, it is enough to prove this property for each vertex v_ε of $P_\varepsilon(l)$. Since it is a vertex, there exists $\{t_1, \ldots, t_d\}$, a finite subset of T_l, such that

$$v_\varepsilon = \bigcap_{k=1}^{d} \left\{ \mu \in K^1, \langle \mu, c_{t_k} \rangle = b_{t_k} + \varepsilon \right\}$$

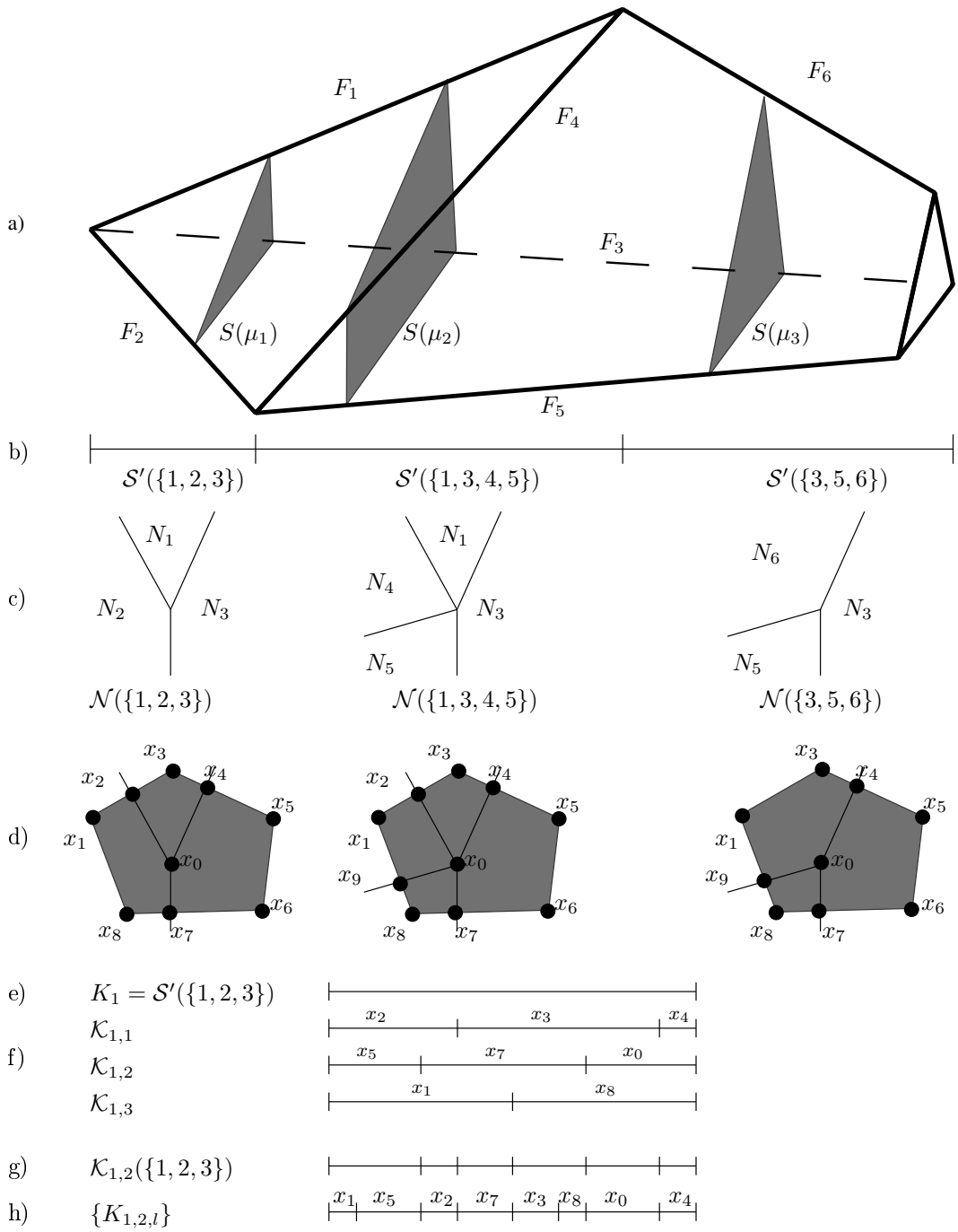

FIGURE 7.1: Steps of the proof of Theorem 7.30. From top to bottom : a) $\mathbf{S}$ (with $\mathcal{M} \subset \mathbb{R}$) and some S_μ in shaded. b) The polytopial complex of $\mathcal{M}$ with constant normal cones. c) Each polytopial complex induced by normal cones. d) The fans and their vertices x_k. f) Every polytopial complex of K_1 induced by the normal cones (the vertex that is a maximizer is indicated above a polytope). g) The intersection of polytopial complexes of K_1. h) The final polytopial complex of K_1

and $\{c_{t_1}, \ldots, c_{t_d}\}$ is a basis of $\mathbb{R}^d$. Let v be the corresponding vertex of $P(l)$, *i.e.* the vertex defined by $v = \bigcap_{k=1}^{d} \left\{\mu \in \mathbb{R}^d, \langle \mu, c_{t_k} \rangle = b_{t_k}\right\}$,

With this notation, for every $k \in \{1, \ldots, d\}$, $\langle v_\varepsilon - v, c_{t_k} \rangle = \varepsilon$ and there exists a unique decomposition $v_\varepsilon - v = \sum_{k=1}^{d} \alpha_k c_{t_k}$. Define the symmetric $d \times d$ matrix Q_l by $Q_l^{kk'} = \langle c_{t_k}, c_{t_{k'}} \rangle$ and $\alpha = (\alpha_1, \ldots, \alpha_d)$. Then following classical properties hold :

1) $\|v_\varepsilon - v\|^2 = \alpha^T Q_l \alpha$ and there exists a $d \times d$ matrix P and a diagonal matrix $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$ with $0 < \lambda_1 \leq \cdots \leq \lambda_d$ such that $P^{-1} = P^T$ and $Q_l = P^T D P$;
2) $Q\alpha = \underline{\varepsilon} = (\varepsilon, \ldots, \varepsilon)$ therefore $\alpha = Q_l^{-1} \underline{\varepsilon}$;
3) $\|\overline{v}_\varepsilon - v\|^2 = (Q_l^{-1} \underline{\varepsilon})^T Q_l (Q_l^{-1} \underline{\varepsilon}) = \underline{\varepsilon}^T P^T D^{-1} P \underline{\varepsilon} \leq \varepsilon^2 d_1 \lambda_1^{-1}$.

Therefore, $\|\mu - \Pi_l(\mu)\| \leq \|v_\varepsilon - v\| \leq \varepsilon . \sqrt{d} \sqrt{\lambda_1}^{-1}$ and the results follow from the fact that L is finite. The constant M_P in Lemma 7.11 is the square root of the inverse of the smallest eigenvalue of all Q_l times $2\sqrt{d}$ (and therefore depends on the set of scalar products $\langle c_t, c_{t'} \rangle$ and on the dimension of $\mathcal{M}$).

Acknowledgments : I am very thankful to my advisor Sylvain Sorin for both his help and comments. I also acknowledge helpful remarks from Gilles Stoltz and Michel Pocchiola.

8

Approachability of Convex Sets

Ce chapitre est dédié à l'étude de l'approchabilité d'ensembles convexes avec observations partielles. On exhibe notamment une condition nécessaire et suffisante d'approchabilité, qui généralise celle de Blackwell avec observations complètes. Par contre, la plupart des résultats connus, comme par exemple, la dualité entre approchabilité et repoussabilité, ne s'étendent pas.
Ce chapitre est issu de l'article Approachability of Convex Sets in Games with Partial Monitoring

Sommaire

8.1 Approachability **136**
Full monitoring case 137
Partial monitoring case 139
8.2 Internal regret with partial monitoring **140**
8.3 Proofs of the main results **142**
Proof of Theorem 8.8 142
Proof of proposition 8.9 145
Remarks on the counterexample 147
8.4 Repeated game with incomplete information **148**

Blackwell [17] introduced the notion of approachability in a two-person (infinitely) repeated game with vector payoffs. A player can approach a given set E, if he can insure that, after some stage and with a great probability, the average payoff will always remain close to E. When both players observe their opponent's moves, Blackwell [17] proved that if E satisfies some geometric condition (E is then called a B-set), then Player 1 can approach it. He also deduced that given a convex set C either Player 1 can approach it or Player 2 can exclude it, *i.e.* he can approach the complement of a neighborhood of C.

We consider the partial monitoring framework, where players do not observe their opponent's moves but receive random signals. We give a necessary and sufficient condition for the approachability of a convex set and the construction of an approachability strategy derived from the construction (provided by Perchet [92]) of a strategy that has no internal regret, in the partial monitoring case (see Definition 9 in Lehrer and Solan [74], which is an extension of the notion of external regret introduced by Rustichini [100]).

Three classical results that hold in the full monitoring case do not extend to the partial monitoring framework. Indeed, in a specific game introduced in section **??**, there exists a convex set C that is neither approachable by Player 1 nor excludable by Player 2 (see Theorem 3 in Blackwell [17]). Moreover, C is not approachable by Player 1 while every half-space that contains it is approachable by Plater 1 (see Corollary 2 in Blackwell [17]). Finally, C is neither weakly-approachable nor weakly-excludable (see Vieille [113]). We recall that weak-approachability is a weaker notion than approachability, also introduced by Blackwell [17], in finitely repeated games (see Definition 8.2 in section 8.1).

The notion of approachability was also used by Kohlberg [67] to construct optimal strategies for the uninformed player, in the class of zero-sum repeated games with incomplete information on one side (introduced by Aumann and Maschler [9]). Our result can be used in this framework to provide a simple proof of the existence of a value in the infinitely repeated game through the construction of an ϵ-optimal strategy for Player 2.

8.1 Approachability

Consider a two-person game Γ repeated in discrete time. At stage $n \in \mathbb{N}_*$, Player 1 (resp. Player 2) chooses an action $i_n \in I$ (resp. $j_n \in J$), where both sets I and J are finite. This generates a vector payoff $\rho_n = \rho(i_n, j_n)$ where ρ is a mapping from $I \times J$ to $\mathbb{R}^d$. Player 1 does not observe j_n nor ρ_n but receives a random signal $s_n \in S$ whose law is $s(i_n, j_n)$ where s is a mapping from $I \times J$ to $\Delta(S)$ (the set of probabilities over the finite set S). Player 2 observes i_n, j_n and s_n. The choices of i_n and j_n depend only on the past observations of the players and may be random.

Explicitly, a strategy σ of Player 1 (resp. a strategy τ of Player 2) is a function from the set of finite histories of Player 1 $H^1 = \bigcup_{n\in\mathbb{N}} (I \times S)^n$ to $\Delta(I)$ (resp. from $H^2 = \bigcup_{n\in\mathbb{N}} (I \times S \times J)^n$ to $\Delta(J)$), where $\sigma(h_n^1)$ (resp. $\tau(h_n^2)$) is the law of i_{n+1} (resp. j_{n+1}) given $h_n^1 \in (I \times S)^n$ (resp. $h_n^2 \in (I \times S \times J)^n$). A couple of strategies (σ, τ) generates a probability, denoted by $\mathbb{P}_{\sigma,\tau}$, over $\mathcal{H} = (I \times S \times J)^{\mathbb{N}}$, the set of plays of the game embedded with the cylinder σ-field.

The two functions ρ and s are extended multilinearly to $\Delta(I) \times \Delta(J)$ by $\rho(x, y) = \mathbb{E}_{x,y}[\rho(i, j)] \in \mathbb{R}^d$ and $s(x, y) = \mathbb{E}_{x,y}[s(i, j)] \in \Delta(S)$.

The following notations will be used : for any sequence $a = \{a_m \in \mathbb{R}^d\}_{m\in\mathbb{N}}$, the average of a up to stage n is denoted by $\overline{a}_n = \sum_{m=1}^n a_m/n$ and for any set $E \subset \mathbb{R}^d$,

the distance to E is denoted by $d_E(z) = \inf_{e \in E} \|z - e\|$, where $\|\cdot\|$ is the euclidian norm.

Definition 8.1 (*Blackwell [17]*)

i) A closed set $E \subset \mathbb{R}^d$ is approachable by Player 1 if for every $\varepsilon > 0$, there exist a strategy σ of Player 1 and $N \in \mathbb{N}$ such that for every strategy τ of Player 2 and every $n \geq N$:

$$\mathbb{E}_{\sigma,\tau}\left[d_E(\overline{\rho}_n)\right] \leq \varepsilon \quad \text{and} \quad \mathbb{P}_{\sigma,\tau}\left(\sup_{n \geq N} d_E(\overline{\rho}_n) \geq \varepsilon\right) \leq \varepsilon.$$

Such a strategy independent of ε is called an approachability strategy of E.

ii) A set E is excludable by Player 2, if there exists $\delta > 0$ such that the complement of E^δ is approachable by Player 2, where $E^\delta = \{z \in \mathbb{R}^d, d_E(z) \leq \delta\}$.

In words, a set $E \subset \mathbb{R}^d$ is approachable by Player 1, if he has a strategy such that the average payoff converges almost surely to E, uniformly with respect to the strategies of Player 2. Obviously, a set E cannot be both approachable by Player 1 and excludable by Player 2.

Definition 8.2

i) A closed set $E \subset \mathbb{R}^d$ is weakly-approachable by Player 1 if for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for every $n \geq N$, there is some strategy σ_n of Player 1 such that for every strategy τ of Player 2 :

$$\mathbb{E}_{\sigma_n,\tau}\left[d_E(\overline{\rho}_n)\right] \leq \varepsilon.$$

ii) A set E is weakly-excludable by Player 2, if there exists $\delta > 0$ such that the complement of E^δ is weakly-approachable by Player 2.

We emphasize the fact that in the definition of weak-approachability, the strategy of Player 1 depends on n, the length of the game, which was not the case in the definition of approachability.

Full monitoring case

A game satisfies full monitoring if Player 1 observes the moves of Player 2 (hence if $S = J$ and $s(i,j) = j$). Blackwell [17] gave a sufficient geometric condition for a closed set E to be approachable by Player 1 and a full characterization if E is a convex set. Stating his condition requires the following notations : $\Pi_E(z) = \{e \in E, d_E(z) = \|z - e\|\}$ is the set of closest points to $z \in \mathbb{R}^d$ in E, and $P^1(x) =$

$\{\rho(x,y), y \in \Delta(J)\}$ (resp. $P^2(y) = \{\rho(x,y), x \in \Delta(I)\}$) is the set of expected payoffs compatible with $x \in \Delta(I)$ (resp. $y \in \Delta(J)$).

Definition 8.3

A closed subset E of $\mathbb{R}^d$ is a B-set, if for every $z \in \mathbb{R}^d$, there exist $p \in \Pi_E(z)$ and $x\,(= x(z)) \in \Delta(I)$ such that the hyperplane through p and perpendicular to $z - p$ separates z from $P^1(x)$, or formally :

$$\forall z \in \mathbb{R}^d, \exists p \in \Pi_E(z), \exists x \in \Delta(I), \langle \rho(x,y) - p, z - p \rangle \leq 0, \quad \forall y \in \Delta(J). \tag{8.1}$$

Condition (8.1) and therefore Theorem 8.4 do not require that Player 1 observes Player 2's moves, but only his own payoffs.

Theorem 8.4 (*Blackwell [17]*)

If E is a B-set, then E is approachable by Player 1. Moreover, the strategy σ of Player 1 defined by $\sigma(h_n) = x(\overline{\rho}_n)$ is such that, for every strategy τ of Player 2 and every $\eta > 0$:

$$\mathbb{E}_{\sigma,\tau}[d_E^2(\overline{\rho}_n)] \leq \frac{4B}{n} \quad \text{and} \quad \mathbb{P}_{\sigma,\tau}\left(\sup_{n \geq N} d_E(\overline{\rho}_n) \geq \eta\right) \leq \frac{8B}{\eta^2 N}, \tag{8.2}$$

with $B = \sup_{i,j} \|\rho(i,j)\|^2$.

For a closed convex set C, a full characterization is available :

Corollary 8.5 (*Blackwell [17]*)

A closed convex set $C \subset \mathbb{R}^d$ is approachable by Player 1 if and only if :

$$P^2(y) \cap C \neq \emptyset, \quad \forall y \in \Delta(J). \tag{8.3}$$

Using a minmax argument, Blackwell [17] proved that condition (8.3) implies condition (8.1), therefore the B-set C is approachable by Player 1. This characterization implies the following properties on convex sets :

Corollary 8.6 (*Blackwell [17]*)

1. *A closed convex set C is either approachable by Player 1 or excludable by Player 2.*
2. *A closed convex set C is approachable by Player 1 if and only if every half-space that contains C is approachable by Player 1.*

If condition (8.3) is not fulfilled for some $y_0 \in \Delta(J)$, then (by the law of large numbers) Player 2 just has to play accordingly to y_0 at each stage to exclude C. If every half-space that contains C is approachable, then C is a B-set and conversely any set that contains an approachable set is approachable.

Blackwell also conjectured the following result on weak-approachability, proved by Vieille :

Theorem 8.7 (*Vieille [113]*)

A closed set is either weakly-approachable by Player 1 or weakly-excludable by Player 2.

Vieille [113] construct a differential game $\mathcal{D}$ (in continuous time and with finite length) such that the finite repetitions of Γ can be seen as a discretization of $\mathcal{D}$. The existence of the value for $\mathcal{D}$ implies the result.

Partial monitoring case

The main objective of this section is to provide a simple necessary and sufficient condition for a convex set C to be approachable in the partial monitoring case, *i.e.* where Player 1 observes random signals.

Before stating it, we introduce the following notations : the vector of probabilities over S defined by $\mathbf{s}(y) = (s(i,y))_{i\in I} \in \Delta(S)^I$ is called the flag generated by $y \in \Delta(J)$. This flag is not observed by Player 1 since, given a probability distribution $x \in \Delta(I)$, Player 1 actually plays its realization $i \in I$ and he only observes the realization of one chosen component of $\mathbf{s}(y)$. However, it is theoretically the maximal information available to him about $y \in \Delta(J)$, *i.e.* Player 2's choice of probability distribution over his actions.

We denote by $\mathcal{S}$ the range of $\mathbf{s}$ and, given a flag $\mu \in \mathcal{S}$, $\mathbf{s}^{-1}(\mu) = \{y \in \Delta(J), \mathbf{s}(y) = \mu\}$ is the set of mixed actions of Player 2 compatible with μ. $P(x,\mu) = \{\rho(x,y), y \in \mathbf{s}^{-1}(\mu)\}$ is the set of expected payoffs compatible with $x \in \Delta(I)$ and $\mu \in \mathcal{S}$.

Our main result is :

Theorem 8.8

A closed convex set $C \subset \mathbb{R}^d$ is approachable by Player 1 if and only if :

$$\forall \mu \in \mathcal{S}, \exists x \in \Delta(I), P(x,\mu) \subset C. \tag{8.4}$$

$P(x,\cdot)$ can be extended to $\Delta(S)^I$ (without changing condition (8.4)) by defining either $P(x,\mu) = \emptyset$ or $P(x,\mu) = P(x, \Pi_{\mathcal{S}}(\mu))$, where $\Pi_{\mathcal{S}}(\cdot)$ is the projection onto $\mathcal{S}$ if $\mu \notin \mathcal{S}$.

In the full monitoring case, condition (8.4) is exactly condition (8.3). Indeed, if Player 1 observes Player 2's action then $S = J$, $\mathcal{S} = \{(y, \ldots, y) \in \Delta(J)^I, y \in \Delta(J)\}$ and given $\mathbf{y} = (y, \ldots, y) \in \mathcal{S}$, $P(x, \mathbf{y}) = \{\rho(x, y)\}$. Condition (8.4) implies that for every $y \in \Delta(J)$ there exists $x \in \Delta(I)$ such that $\rho(x, y) \in C$, or equivalently $P^2(y) \cap C \neq \emptyset$.

An other main result is that Corollary 8.6 and Theorem 8.7 do not extend :

Proposition 8.9

1. *There exists a closed convex set that is neither approachable by Player 1 nor excludable by Player 2.*
2. *An half-space is either approachable by Player 1 or excludable by Player 2.*
3. *There exists a closed convex set that is not approachable by Player 1 while every half-space that contains it is approachable by Player 1.*

As said in the introduction, the proof of Theorem 8.8 relies on the construction of a strategy that has no internal regret in an auxiliary game with partial monitoring. Point 2) of Proposition 8.9 should be compared with Corollary 8.6.

8.2 Internal regret with partial monitoring

Consider the following two-person repeated game $\mathcal{G}$ with random signals where, at stage $n \in \mathbb{N}$, Player 1 chooses $x_n \in \Delta(I)$, the law of his action $i_n \in I$, and Player 2 chooses $y_n \in \Delta(J)$, the law of his action $j_n \in J$. Player 1 observes a signal s_n whose law is the i_n-th coordinate of $\mu_n = \mathbf{s}(j_n)$.

Payoffs are unobserved, however given a flag $\mu \in \Delta(S)^I$ and $x \in \Delta(I)$, Player 1 evaluates his payoff through $G(x, \mu)$ where G is a map from $\Delta(I) \times \Delta(S)^I$ into $\mathbb{R}$, not necessarily linear.

In the full monitoring case, Foster and Vohra [42] defined internally consistent strategies (or strategies that have no internal regret) as follows : Player 1 has asymptotically no internal regret if for every $i \in I$, either the action i is a best response to his opponent's empirical distribution of actions on the set of stages where he actually played i, or the density of this set (also called the frequency of the action i) converges to zero.

In our framework, G is not linear so every action $i \in I$ (or the Dirac mass on i) might never be a best response. Hence if we want to define internal regret, we cannot distinguish the stages as a function of the actions actually played (*i.e.* $i_n \in I$) but we must do it as a function of the laws of the actions (*i.e.* $x_n \in \Delta(I)$) since best responses are elements of $\Delta(I)$.

We consider strategies described as follows : at stage n, given his history, Player 1 chooses (at random) a law $x(l_n)$ in a finite set $\{x(l) \in \Delta(I), l \in L\}$ and given that

choice, i_n is drawn accordingly to $x(l_n)$; l_n is called the type of the stage n. The set L is assumed to be finite, otherwise there exist trivial strategies such that the frequency of every $x(l)$ converges to zero.

We make the following regularity assumption of G :

Assumption 8.10

For every $\varepsilon > 0$, there exist $\eta, \delta > 0$ and two finite families $\{\mu(l) \in \Delta(S)^I\}_{l\in L}$ and $\{x(l) \in \Delta(I)\}_{l\in L}$ such that :

1) $\Delta(S)^I \subset \bigcup_{l\in L} B(\mu(l), \delta)$;

2) For every $l \in L$, if $\|x - x(l)\| \leq \eta$ and $\|\mu - \mu(l)\| \leq \delta$, then $x \in BR_\varepsilon(\mu)$;

where for every $\delta \geq 0$ and $\mu \in \Delta(S)^I$, $B(\mu, \delta) = \{\mu' \in \Delta(S)^I, \|\mu' - \mu\| \leq \delta\}$ and for every $\varepsilon \geq 0$ $BR_\varepsilon(\mu) = \{x \in \Delta(I) : G(x, \mu) \geq \sup_{z\in\Delta(I)} G(z, \mu) - \varepsilon\}$.

In words, Assumption 8.10 implies that G is regular with respect to μ and with respect to x in the following sense : given $\varepsilon > 0$, the set of flags can be covered by a finite number of balls centered in $\{\mu(l), l \in L\}$, such that $x(l)$ is an ε-best response to any μ in this ball. In addition, if x is close enough to $x(l)$, then x is an ε-best response to any μ close to $\mu(l)$.

We denote by $N_n(l) = \{1 \leq m \leq n, l_m = l\}$ the set of stages (before the n-th) of type l and for any sequence $a = \{a_m \in \mathbb{R}^d\}_{m\in\mathbb{N}}$, $\overline{a}_n(l) = \sum_{m\in N_n(l)} a_m/|N_n(l)|$ is the average of a on $N_n(l)$.

Definition 8.11 (*Lehrer and Solan [74]*)

For every $n \in \mathbb{N}$ and every $l \in L$, the internal regret of type $l \in L$ at stage n is

$$\mathcal{R}_n(l) = \sup_{x\in\Delta(I)} \left[G(x, \overline{\mu}_n(l)) - G(\overline{\imath}_n(l), \overline{\mu}_n(l))\right],$$

where $\overline{\imath}_n(l)$ is the empirical distribution of actions of Player 1 on $N_n(l)$ and $\overline{\mu}_n(l)$ is the average flag on $N_n(l)$.

A strategy σ of Player 1 is (L, ε)-internally consistent if for every strategy τ of Player 2 :

$$\limsup_{n\to+\infty} \frac{|N_n(l)|}{n}\Big(\mathcal{R}_n(l) - \varepsilon\Big) \leq 0, \quad \forall l \in L, \quad \mathbb{P}_{\sigma,\tau}\text{−as.}$$

Note that any (L, ε)-internally consistent strategy is such that, asymptotically and for every $l \in L$, $\overline{\imath}_n(l)$ and $x(l)$ belong to $BR_\varepsilon(\overline{\mu}_n(l))$, as soon as $|N_n(l)|/n$ (the frequency of the type l) is not too small. In words, either $x(l)$ is an ε-best response to $\overline{\mu}_n(l)$ the unobserved average flag on the set of stages where $x(l)$ was the law of Player 1's action, or this set has a very small density.

Theorem 8.12 (*Lehrer and Solan [74]*)

Under Assumption 8.10, for every $\varepsilon > 0$, there exist a finite set L and a (L, ε)-internally consistent strategy σ such that for every strategy τ of Player 2 :

$$\mathbb{E}_{\sigma,\tau}\left[\sup_{l\in L}\frac{|N_n(l)|}{n}\Big(\mathcal{R}_n(l)-\varepsilon\Big)\right] = O\left(\frac{1}{\sqrt{n}}\right) \quad \text{and}$$

$$\forall\eta>0, \mathbb{P}_{\sigma,\tau}\left(\exists n\geq N, l\in L, \frac{|N_n(l)|}{n}\Big(\mathcal{R}_n(l)-\varepsilon\Big)>\eta\right)\leq O\left(\frac{1}{\eta^2 N}\right).$$

The first proof of the existence of such strategies is due to Lehrer and Solan [74].

8.3 Proofs of the main results

This section is devoted to the proof of the theorems stated in the previous section.

Proof of Theorem 8.8

Proof of Theorem 8.8 : Sufficiency : Let C be a convex set such that for every $\mu \in \Delta(S)^I$ there exists $x_\mu \in \Delta(I)$ such that $P(x_\mu, \mu) \subset C$ and let $\varepsilon > 0$ be fixed. The construction of an approachability strategy in Γ relies on the construction of an (L, ϵ)-internally consistent strategy in an auxiliary game $\mathcal{G}$ with partial monitoring. A strategy in $\mathcal{G}$ defines a strategy in Γ as follows :

At stage $n \in \mathbb{N}$, Player 1 chooses the type $l_n \in L$ of that stage in the game $\mathcal{G}$; given that type, the same action i_n in $\mathcal{G}$ and Γ is drawn accordingly to $x(l_n)$. Independently, Player 2 chooses $j_n \in J$ which generates the flag $\mu_n = \mathbf{s}(j_n)$ and the signal s_n (whose law is $\mu_n^{i_n}$). As defined in section 8.1, the (unobserved) vector payoff in Γ is $\rho_n = \rho(i_n, j_n)$ while the payoff function $G : \Delta(I) \times \Delta(S)^I \to \mathbb{R}$ is defined in $\mathcal{G}$ by

$$G(x,\mu) = -\sup_{y\in\mathbf{s}^{-1}(\mu)} d_C\left(\rho(x,y)\right)$$

if $\mu \in \mathcal{S}$. If $\mu \notin \mathcal{S}$, then $G(x, \mu) = G\left(x, \Pi_{\mathcal{S}}(\mu)\right)$ where $\Pi_{\mathcal{S}}$ is the projection onto $\mathcal{S}$. G fulfills Assumption 8.10 since it is uniformly Lipschitz (see Lugosi, Mannor and Stoltz [78], section 3 and Appendix A, or Perchet [92] Corollary 2).

The main ideas of the proof are quite simple : given $\varepsilon > 0$, consider the finite family $\{x(l), l \in L\}$ given by Assumption 8.10 and σ a (L, ε)-internally consistent strategy of Player 1. Then for every $l \in L$, either $|N_n(l)|/n$ is very small, or $\mathcal{R}_n(l) \leq \varepsilon$ and in that case, the definition of G implies that $\overline{\rho}_n(l)$ is ε-close to C. Since

$$\overline{\rho}_n = \sum_{l\in L}\frac{|N_n(l)|}{n}\overline{\rho}_n(l), \tag{8.5}$$

$\overline{\rho}_n$ is almost a convex combination of terms that are ε-close to C and since C is convex, $\overline{\rho}_n$ is also close to C.

Formally, let σ be a (L,ε)-internally consistent strategy of Player 1 given by Theorem 8.12. For every $\theta > 0$, there exists $N^1 \in \mathbb{N}$ such that for any strategy τ of Player 2 :

$$\mathbb{P}_{\sigma,\tau}\left(\forall n \geq N^1, \sup_{l\in L} \frac{|N_n(l)|}{n}\Big(\mathcal{R}_n(l) - \varepsilon\Big) \leq \theta\right) \geq 1-\theta. \tag{8.6}$$

Recall that for any $\mu \in \Delta(S)^I$ there exists $x_\mu \in \Delta(I)$ such that $P(x_\mu,\mu) \subset C$, therefore $\sup_{z\in\Delta(I)} G(z,\mu) = G(x_\mu,\mu) = 0$ and

$$\mathcal{R}_n(l) = \sup_{y\in \mathbf{s}^{-1}(\overline{\mu}_n(l))} d_C\left(\rho(\overline{\imath}_n(l), y)\right) \geq d_C\Big(\rho\big(\overline{\imath}_n(l), \overline{\jmath}_n(l)\big)\Big),$$

because $\mathbf{s}(\overline{\jmath}_n(l)) = \mu_n(l)$ by linearity of $\mathbf{s}$. The choices of l_n and j_n are independent and so are the choices of i_n and j_n given l_n, so Hoeffding-Azuma's inequality for sums of bounded martingale differences (see Azuma [11] and Hoeffding [61]) implies that $\rho\left(\overline{\imath}_n(l), \overline{\jmath}_n(l)\right)$ is close to $\overline{\rho}_n(l)$. Explicitly, for every $\theta > 0$, there exists $N^2 \in \mathbb{N}$ (independent of σ and τ) such that :

$$\mathbb{P}_{\sigma,\tau}\left(\forall n \geq N^2, \exists l \in L, \frac{|N_n(l)|}{n}\big|\overline{\rho}_n(l) - \rho\left(\overline{\imath}_n(l), \overline{\jmath}_n(l)\right)\big| \leq \theta\right) \geq 1-\theta. \tag{8.7}$$

Equations (8.6) and (8.7) imply that for every $n \geq N = \max\{N^1, N^2\}$ and every $l \in L$, with probability at least $1-2\theta$:

$$\frac{|N_n(l)|}{n}\big(d_C\left(\overline{\rho}_n(l)\right) - \varepsilon\big) \leq 2\theta.$$

Since C is a convex set, $d_C(\cdot)$ is convex, therefore for any strategy τ of Player 2, with $\mathbb{P}_{\sigma,\tau}$-probability at least $1-2\theta$, for every $n \geq N$:

$$d_C(\overline{\rho}_n) \leq \sum_{l\in L} \frac{|N_n(l)|}{n} d_C(\overline{\rho}_n(l)) \leq 2L\theta + \varepsilon,$$

and C is approachable by Player 1.

Necessity : Conversely, assume that there exists $\mu_0 \in \Delta(S)^I$ such that for all $x \in \Delta(I)$, there is some $y(= y(x)) \in \mathbf{s}^{-1}(\mu_0)$ such that $d_C\left(\rho(x,y)\right) > 0$. Since $\Delta(I)$ is compact, we can assume that there exists $\delta > 0$ such that $d_C\left(\rho(x,y(x))\right) \geq \delta$.

Let $\mathcal{T}_0$ be the subset of strategies of Player 2 that generate at any stage the same flag μ_0 (explicitly, a strategy τ belongs to $\mathcal{T}_0$ if for every possible finite history h_n^2, $\tau(h_n^2) \in \mathbf{s}^{-1}(\mu_0)$). Recall that a strategy σ of Player 1 depends only on his past actions and on the signals he received. Since at any stage, two strategies τ and τ' in $\mathcal{T}_0$ induce the same laws of signals, the couples (σ,τ) and (σ,τ') generate the same probability

on the infinite sequences of moves of Player 1, and $\mathbb{E}_{\sigma,\tau}[\overline{\imath}_n] = \mathbb{E}_{\sigma,\tau'}[\overline{\imath}_n] := \overline{x}_n$ is independent of the strategy in $\mathcal{T}_0$.

For every $n \in \mathbb{N}$, define the strategy τ_n in $\mathcal{T}_0$ by $\tau_n(h) = y(\overline{x}_n)$, for all finite history h. Since $d_C(\cdot)$ is convex, by Jensen's inequality

$$\mathbb{E}_{\sigma,\tau_n}[d_C(\overline{\rho}_n)] \geq d_C(\mathbb{E}_{\sigma,\tau_n}[\overline{\rho}_n]).$$

Since j_m is independent of the history h_{m-1} :

$$\mathbb{E}_{\sigma,\tau_n}\left[\rho(i_m, j_m)\middle|h_{m-1}\right] = \mathbb{E}_{\sigma,\tau_n}\left[\rho(i_m, y(\overline{x}_n))\middle|h_{m-1}\right] = \rho\left(\mathbb{E}_{\sigma,\tau_n}\left[i_m\middle|h_{m-1}\right], y(\overline{x}_n)\right),$$

therefore $\mathbb{E}_{\sigma,\tau_n}[\overline{\rho}_n] = \rho(\overline{x}_n, y(\overline{x}_n))$. So

$$\mathbb{E}_{\sigma,\tau_n}[d_C(\overline{\rho}_n)] \geq d_C(\mathbb{E}_{\sigma,\tau_n}[\overline{\rho}_n]) \geq d_C(\rho(\overline{x}_n, y(\overline{x}_n))) \geq \delta$$

and for any strategy σ of Player 1 and any stage $n \in \mathbb{N}$, Player 2 has a strategy such that the expected average payoff is at a distance greater than $\delta > 0$ from C, so C is not approachable by Player 1. □

Remark : The fact that C is a convex set is crucial in both parts of the proof. Otherwise in the sufficient part, it would be possible that $\overline{\rho}_n(l) \in C$ for every $l \in L$, while $\overline{\rho}_n \notin C$, and in the necessary part, the counterpart could happen : $d_C(\mathbb{E}[\overline{\rho}_n]) \geq \delta$ while $\mathbb{E}[d_C(\overline{\rho}_n)] = 0$.

Remark : Since the approachability strategy relies on a (L, ε)-internally consistent strategy, one can easily show that :

$$\mathbb{E}_{\sigma,\tau}[d_C(\overline{\rho}_n)] = \varepsilon + O\left(\frac{1}{\sqrt{n}}\right) \quad \text{and}$$

$$\mathbb{P}_{\sigma,\tau}(\exists n \geq N, d_C(\overline{\rho}_n) - \varepsilon > \eta) \leq O\left(\frac{1}{\eta^2 N}\right).$$

Corollary 8.13

If C satisfies Condition (8.4), there exists σ a strategy of Player 1 such that for every $\eta > 0$, there is some $N \in \mathbb{N}$ such that for every strategy τ of Player 2 and $n \geq N$, $\mathbb{E}_{\sigma,\tau}[d_C(\overline{\rho}_n)] \leq \eta$.

Proof. The proof relies uniquely on the use of a doubling-trick (see *e.g.* Sorin [106], Proposition 3.2 p. 56), recalled below. We assume that payoffs are bounded by 1/2 (so that the distance between any convex combination of payoffs and C is at most 1). Denote by σ_p the strategy constructed in the proof of Theorem 8.8 for $\varepsilon_p = 1/2^{p+3}$. Define the strategy σ by blocks as follows : during N_1 stages, play accordingly to σ_1, then during N_2 stages accordingly to σ_2 and so on. Formally, for n such that

$\sum_{k=1}^{p-1} N_k \leq n \leq \sum_{k=1}^{p} N_k$, $\sigma(h_n) = \sigma_p(h_n^p)$ where $h_n^p = (i_m, l_m, s_m)_{m \in \{\sum_{k=1}^{p-1} N_k, \ldots, n\}}$ is the partial history on the last block.

The main issue is to choose wisely every N_p so that the expected distance between $\overline{\rho}_n$ and C is always smaller than $1/2^{p-2}$ on the p-th block and smaller than $1/2^p$ on its last stage. Remark 8.3 implies that for every $p \in \mathbb{N}$ there exists $M_p \in \mathbb{N}$ such that $\mathbb{E}_{\sigma_p,\tau}\left[d_C\left(\overline{\rho}_n\right)\right] \leq 1/2^{p+1}$ for every $n \geq M_p$. Define $N_p = \max\left(2^p M_{p+1}, 4\sum_{k=1}^{p-1} N_k\right)$ so that during the first M_{p+1} stages of the block $p+1$:

$$\mathbb{E}_{\sigma,\tau}\left[d_C(\overline{\rho}_n)\right] \leq \frac{M_{p+1}}{n} + \frac{\sum_{k=1}^{p} N_k}{n}\frac{1}{2^p} \leq \frac{1}{2^{p-1}}. \tag{8.8}$$

After this stage, the definition of M_{p+1} implies that

$$\mathbb{E}_{\sigma,\tau}\left[d_C(\overline{\rho}_n)\right] \leq \frac{n - \sum_{k=1}^{p} N_k}{n}\frac{1}{2^{p+2}} + \frac{\sum_{k=1}^{p} N_k}{n}\frac{1}{2^p} \leq \frac{1}{2^{p-1}} \tag{8.9}$$

and for $n = \sum_{k=1}^{p+1} N_k$:

$$\mathbb{E}_{\sigma,\tau}\left[d_C\left(\overline{\rho}_n\right)\right] \leq \frac{1}{2^{p+2}} + \frac{\sum_{k=1}^{p} N_k}{N_{p+1}}\frac{1}{2^p} \leq \frac{1}{2^{p+1}}. \tag{8.10}$$

The result follows from equations (8.8), (8.9) and (8.10). □

Remark : In the full monitoring case, condition (8.4) is condition (8.3) : if Player 1 observes Player 2's action then $S = J$, $\mathcal{S} = \{(y, \ldots, y) \in \Delta(J)^I, y \in \Delta(J)\}$ and given $\mathbf{y} = (y, \ldots, y) \in \mathcal{S}$, $P(x, \mathbf{y}) = \{\rho(x, y)\}$. Condition (8.4) implies that for every $y \in \Delta(J)$ there exists $x \in \Delta(I)$ such that $\rho(x, y) \in C$, or equivalently $P^2(y) \cap C \neq \emptyset$.

Proof of proposition 8.9

In the proof of Theorem 8.8, we have shown that if a convex set is not approachable by Player 1 then for any strategy σ of Player 1 and any $n \in \mathbb{N}$, Player 2 has a strategy τ_n such that $\overline{\rho}_n$ is at a distance at least δ from C. This does not imply that C is excludable by Player 2, since this would require that τ_n does not depend on σ nor n. The proof of Proposition 8.9 relies mainly on the study of the following example (with one-dimensional payoffs), pointed out to me by J. Renault and G. Stoltz.

Proof of Proposition 8.9 : Consider the following matrix two-person repeated game where Player 1 (the row player) receives no signal and his one-dimensional payoffs are defined by :

	L	R
T	0	1
B	-1	0

$\mathbf{C := [0; 1/2]}$ is neither approachable nor excludable : The closed convex set $C := [0; 1/2]$ is obviously not approachable by Player 1 (otherwise Theorem 8.8

implies that there exists $x \in \Delta(I)$ such that $\rho(x,y) \in [0,1/2]$ for every $y \in \Delta(J)$). More precisely, given a strategy σ of Player 1, we define τ_n as follows : if $\overline{x}_n$ (the expected frequency of T up to stage $n \in \mathbb{N}$ — it does not depend on Player 2's strategy) is smaller than $1/4$, then τ_n is the strategy that always plays L, otherwise that always plays R. Then the law of large numbers implies that, for n big enough, $\mathbb{E}_{\sigma,\tau_n}\left[d_C\left(\overline{\rho}_n\right)\right]$ is arbitrarily close to $1/4$.

It remains to show that Player 2 cannot exclude C. We prove this by constructing a strategy σ of Player 1 such that the average payoff is infinitely often close to 0 : σ is played in blocks and the length of the p-th block is p^{2p+1}. On odd blocks, Player 1 plays T while on even blocks he plays B. At the end of the block p, the average payoff is at most $1/p$ if it is an odd block and at least $-1/p$ otherwise. Hence on two consecutive blocks (the p-th and the $p+1$-th) there is at least one stage such that the average payoff is at a distance smaller than $1/p$ to $\{0\}$. Therefore $\{0\}$ and C (since it contains $\{0\}$) is not excludable by Player 2.

An half-space is either approachable by Player 1 or excludable : Let E be an half-space not approachable by Player 1. Then there exists $\mu_0 \in \Delta(S)^I$ such that, for every $x \in \Delta(I)$, $P(x,\mu_0) \not\subset E$. This implies that there exists $\delta > 0$ such that $\inf_{x\in\Delta(I)} \sup_{y\in\mathbf{s}^{-1}(\mu_0)} d_E\left(\rho(x,y)\right) \geq \delta > 0$ and therefore for every $x \in \Delta(I)$, there exists $y \in \Delta(J)$ such that $\rho(x,y)$ is in the complement of E^δ which is convex, since E is an half-space. Blackwell's result applies for Player 2 (since we assumed he has full monitoring), so he can approach the complement of E^δ and exclude E.

C is not approachable by Player 1 while every half-space that contains it is : An half-space that contains C contains either $(-\infty,0]$ or $[0,+\infty)$ which are approachable by, respectively, always playing T or always playing B.

C is neither weakly-approachable by Player 1 nor weakly excludable by Player 2 : we proved that for every strategy σ of Player 1 and every $n \in \mathbb{N}$ big enough, Player 2 has a strategy τ_n such that $\mathbb{E}_{\sigma,\tau_n}\left[d_C(\overline{\rho}_n)\right] = 1/2$. Hence C is not weakly approachable.

Conversely, let τ be a strategy of Player 2 in the game repeated $2n$ times (where n is large enough) and $M \in \mathbb{N}$ be any integer. Consider the strategy σ of Player 1 that consists in playing T during the first n stages. Since $\overline{\rho}_n$, the average payoff after those n stages, belongs to $[0;1]$, there exists an integer $k_1 \in \{1,\ldots,M\}$ such that $\overline{\rho}_n$ belongs to $[\frac{k_1-1}{M};\frac{k_1}{M}]$ with $\mathbb{P}_{\sigma,\tau}$-probability at least $\frac{1}{M}$. Note that, given τ , Player 1 can compute this k.

Assume that, from stage $n+1$ on, the strategy σ dictates to play i.i.d action B with probability $\frac{k_1}{M}$ and action T with probability $1-\frac{k_1}{M}$. If n is large enough, the probability that the average payoff between stages $n+1$ and $2n$ belongs to $[-\frac{k_1}{M}-\frac{1}{M};1-\frac{k_1}{M}+\frac{1}{M}]$ is close to one (say bigger than $1/2$, this is again a direct consequence of the law of large number). Therefore, this strategy σ ensures that with $\mathbb{P}_{\sigma,\tau}$-probability at least $\frac{1}{2M}$ the average payoff over the $2n$ stages belongs to $[-\frac{1}{M};\frac{1}{2}+\frac{1}{2M}]$.

Denote by $\left(C^{2/M}\right)^c$ the complement of the $\frac{2}{M}$-neighborhood of C. Given a strategy τ of Player 2 and an integer n big enough, the strategy σ we described ensures that $\mathbb{E}_{\sigma,\tau}\left[d_{\left(C^{2/M}\right)^c}\left(\overline{\rho}_{2n}\right)\right] \geq \frac{1}{2M^2}$. Therefore, for every $M \in \mathbb{N}$, $\left(C^{2/M}\right)^c$ is not weakly-approachable by Player 2 hence C is not weakly-excludable.

The strategy σ we described can be quite easily made independent of the strategy of Player 2 by, for example, choosing $k_1 \in \{1, \ldots, M\}$ at random ; indeed, this would simply imply that $\mathbb{E}_{\sigma,\tau}\left[d_{\left(C^{2/M}\right)^c}\left(\overline{\rho}_{2n}\right)\right] \geq \frac{1}{2M^3}$. □

These results hold if one chooses $C_3 := [0; 1/3]$ instead of $[0; 1/2]$. In fact, it only remains to prove that C_3 is not weakly-excludable by Player 2. Consider the game repeated $3n$ times and the strategy σ, defined by block of size n, that plays on the first block always T, on the second block i.i.d. action B with probability $\frac{k_1}{M}$. The average payoff on those two block belongs to a small neighborhood of $[0; 1/2]$, hence to some $[\frac{k_2-1}{M}, \frac{k_2}{M}]$ (where $k_2 \leq \frac{M}{2}$) with probability at least $\frac{1}{M}$. Assume that on the third block Player 1 plays i.i.d action B with probability $\frac{2k_2}{M}$ then the average payoff over the three blocks belongs to a small neighborhood of $[0; 1/3]$ with probability at least $\frac{1}{(2M)^2}$. Therefore C_3 is not weakly excludable.

Since this proof can be generalized to any set $C_k = [0; \frac{1}{k}]$, even the singleton $\{0\}$ is neither weakly-approachable nor weakly-excludable ; we recall that in the full monitoring framework all those convex sets are approachable by Player 1.

Remarks on the counterexample

Following Mertens, Sorin and Zamir's notations [84] (see Definition 1.2 p. 149), we say that in a zero-sum repeated game Γ_∞, Player 1 can guarantee $\underline{v}$ if

$$\forall \varepsilon > 0, \exists \sigma_\varepsilon, \exists N \in \mathbb{N}, \mathbb{E}_{\sigma_\varepsilon,\tau}\left[\overline{\rho}_n\right] \geq \underline{v} - \varepsilon, \forall \tau, \forall n \geq N,$$

where σ_ε is a strategy of Player 1, and τ any strategy of Player 2. Player 2 can defend $\underline{v}$ if :

$$\forall \varepsilon > 0, \forall \sigma_\varepsilon, \exists \tau, \exists N \in \mathbb{N}, \mathbb{E}_{\sigma_\varepsilon,\tau}\left[\overline{\rho}_n\right] \leq \underline{v} + \varepsilon, \forall n \geq N.$$

If Player 1 can guarantee $\underline{v}$ and Player 2 defend $\underline{v}$, then $\underline{v}$ is the maxmin of Γ_∞. The minmax $\overline{v}$ is defined in a dual way and Γ_∞ has a value if $\underline{v} = \overline{v}$.

These definitions can be extended to the vector payoff framework : Player 1 can guarantee a set E if he can approach E :

$$\forall \varepsilon > 0, \exists \sigma_\varepsilon, \exists N \in \mathbb{N}, \mathbb{E}_{\sigma_\varepsilon,\tau}\left[d_E\left(\overline{\rho}_n\right)\right] \leq \varepsilon, \forall \tau, \forall n \geq N.$$

In the counterexample of the proof of Proposition 8.9, Player 1 cannot guarantee the convex set $C = \{0\}$ and Player 2 cannot defend it since :

$$\exists \sigma, \forall \varepsilon > 0, \forall \tau, \forall N \in \mathbb{N}, \exists n \geq N, \mathbb{E}_{\sigma,\tau}\left[d_C\left(\overline{\rho}_n\right)\right] \leq \varepsilon.$$

To keep the notations of zero-sum repeated game, one could say that the game we constructed *has no maxmin.*

Blackwell [17] also gave an example of a game (with vector payoff) without maxmin in the full monitoring case. The main differences between the two examples are that in the partial monitoring case the set can be convex (which is impossible in the full monitoring framework) and that Player 1 has a strategy such that the average payoff is infinitely often close to C but, unlike in Blackwell's example, he does not know at which stages.

8.4 Repeated game with incomplete information

Aumann and Maschler [9] introduced the class of two-person zero-sum games with incomplete information on one side. Those games are described as follows : Nature chooses a state k_0 from a finite set K of states according to some known probability $p \in \Delta(K)$. Player 1 (the maximizer) is informed about which k is chosen but not Player 2. At stage $m \in \mathbb{N}$, Player 1 (resp. Player 2) chooses $i_m \in I$ (resp. $j_m \in J$) and the payoff is $\rho_m^k = \rho^k(i_m, j_m)$. Player 1 observes j_m and Player 2 does not observe i_m nor ρ_m but receives a signal s_m whose law is $s^k(i_m, j_m) \in \Delta(S)$. As in the previous sections, we define $\mathbf{s}^k(x) = \left(s^k(x,j)\right)_{j\in K}$, for every $x \in \Delta(I)$.

A strategy σ (resp. τ) of Player 1 (resp. Player 2) is a function :

$$\begin{array}{rcl} \sigma : K \times \bigcup_{\in\mathbb{N}} (I\times J)^m & \to & \Delta(I) \\ (k, h_m^1) & \mapsto & \sigma^k(h_m^1) \end{array} \quad \text{and} \quad \begin{array}{rcl} \tau : \bigcup_{m\in\mathbb{N}} (J\times S)^m & \to & \Delta(J) \\ h_n^2 & \mapsto & \tau(h_m^2). \end{array}$$

We define Γ_1 the one-shot game with expected payoff $\sum_{k\in K} p^k \rho^k(x^k, y)$ and $\Gamma_\infty(p)$ the infinitely repeated game. We denote by $v_\infty(p)$ its value, if it exists (*i.e.* if both Player 1 and Player 2 can guarantee it). Aumann and Maschler [9] (Theorem C, p. 191) proved that $\Gamma_\infty(p)$ has a value and characterized it.

Let us first introduce the operator **Cav** and the non-revealing game $D(p)$: For any function f from $\Delta(I)\times\Delta(J)$ to $\mathbb{R}$, $\mathbf{Cav}(f)(\cdot)$ is the smallest (pointwise) concave function greater than f.

A profile of mixed actions $x = (x^k)_{k\in K} \in \Delta(I)^K$ is non-revealing at $p \in \Delta(K)$ (and induces the flag $\mu \in \Delta(S)^J$) if the flag induced by x is independent of the state :

$$NR(p,\mu) = \left\{ x = (x^1, \ldots, x^K) \in \Delta(I)^K \,\middle|\, \mathbf{s}^k(x^k) = \mu, \forall k \text{ st } p^k > 0 \right\}.$$

The set of non-revealing strategies is denoted by $NR(p) = \bigcup_{\mu\in\Delta(S)^J} NR(p,\mu)$. For every $\mu \in \Delta(S)^J$, $D(p,\mu)$ (resp. $D(p)$) is the one-stage game Γ_1 where Player 1 is restricted to $NR(p,\mu)$ (resp. $NR(p)$) and its value is denoted by $u(p,\mu)$ (resp. $u(p)$), with $u(p,\mu) = -\infty$ if $NR(p,\mu) = \emptyset$ (resp. $u(p) = -\infty$ if $NR(p) = \emptyset$).

Theorem 8.14 (*Aumann Maschler [9]*)

Γ_∞ has a value defined by $v_\infty(p) = \mathbf{Cav}(u)(p)$.

Proof. Player 1 can guarantee $u(p)$: indeed if $NR(p) \neq \emptyset$, he just has to play i.i.d. an optimal strategy in $NR(p)$ otherwise $u(p) = -\infty$. Therefore, using the splitting procedure (see Lemma 5.2 p. 25 in [9]), he can guarantee $\mathbf{Cav}(u)(p)$.

It remains to prove that Player 2 can also guarantee $\mathbf{Cav}(u)(p)$. The function $\mathbf{Cav}(u)(\cdot)$ is concave and continuous, so there exists $\mathbf{m} = (\mathbf{m}^1, \ldots, \mathbf{m}^k) \in \mathbb{R}^K$ such that $\mathbf{Cav}(u)(p) = \langle \mathbf{m}, p \rangle$ and $u(q) \leq \mathbf{Cav}(u)(q) \leq \langle \mathbf{m}, q \rangle$. Instead of constructing a strategy for Player 2 that minimizes the expected payoff $\sum_{k \in K} p^k \overline{\rho}_n^k$, it is enough to construct a strategy such that each $\overline{\rho}_n^k$ is smaller than $\mathbf{m}^k$, for every state k that has a positive probability accordingly to Player 2's posterior.

Therefore, we consider an auxiliary repeated two-person game with vector payoff where at stage $n \in \mathbb{N}$, Player 2 (resp. Player 1) chooses j_n accordingly to $y_n \in \Delta(J)$ (resp. $i_n = (i_n^1, \ldots, i_n^K)$ accordingly to $x_n = (x_n^1, \ldots, x_n^K) \in \Delta(I)^K$). Player 2 receives a signal s_n whose law is $s^{k_0}(i_n^{k_0}, j_n)$ where k_0 is the true state. As before, we denote by $\mu_n = \mathbf{s}^{k_0}(x_n^{k_0})$ the expected flag of stage n. The k-th component of the vector payoff ρ_n is defined by $\rho^k(i_n^k, j_n)$ if μ_n belongs to $\mathcal{S}^k$, the range of $\mathbf{s}^k$ and $-A := -\max_{k \in K} \|\rho^k\|_\infty$ otherwise. We use this notation, because if μ_n is not in the range of $\mathbf{s}^k$, then Player 2 knows that the true state is not k, and therefore does not need to minimize the k-th component of the payoff vector.

Conversely, the set of compatible payoffs given a flag $\mu \in \Delta(S)^J$, $y \in \Delta(J)$ and a state k, is defined by :

$$P^k(\mu, y) = \left\{ \rho^k(x^k, y) \,\middle|\, \mathbf{s}^k(x^k) = \mu \right\} \text{ if } \mu \in \mathcal{S}^k, \text{ otherwise } P^k(\mu, y) = \{-A\},$$

and the set of compatible vector payoffs is defined by $P(\mu, y) = \Pi_{k \in K} P^k(\mu, y) \subset \mathbb{R}^K$. If Player 2 can approach the convex set $M = \{m \in \mathbb{R}^K, m^k \leq \mathbf{m}^k, \forall k \in K\}$ (the set M is an orthant since it equals $\mathbf{m} + \mathbb{R}_-^K$), then he can guarantee $\mathbf{Cav}(u)(p)$. Theorem 8.8 implies that M is approachable if and only if, for every $\mu \in \Delta(S)^I$ there exists $y \in \Delta(J)$ such that $P(\mu, y) \subset M$.

Hence it is enough to prove that this property holds. Assume the converse : there exists $\mu_0 \in \Delta(S)^I$ such that for every $y \in \Delta(J)$, $P(\mu_0, y)$ is not included in M.

We denote by $K(\mu_0) = \left\{ k \in K, \mu_0 \in \mathcal{S}^k \right\}$ the set of states that are compatible with μ_0 : if Player 2 observes μ_0, then he knows that the true state is in $K(\mu_0)$. For every $y \in \Delta(J)$ and $k \in K(\mu_0)$, $\omega_0^k(y) = \sup_{\mathbf{s}^k(x^k) = \mu_0} \rho^k(x^k, y)$ is the worst payoff for Player 2 in state k. Since $P(\mu_0, y)$ is not included in M, then $\omega_0(y) = (\omega_0^k(y))_{k \in K(\mu_0)}$ does not belong to $M_0 = \{m \in \mathbb{R}^{K(\mu_0)}, m^k \leq \mathbf{m}^k, \forall k \in K(\mu_0)\}$. Define the convex set :

$$W_0 = \{\omega_0(y), y \in \Delta(J)\} + \mathbb{R}_+^{K(\mu_0)} \bigcap B(0, A),$$

with $B(0,A)$ the closed ball of radius A. Obviously $W_0 \cap M_0 = \emptyset$ and, by linearity of each ρ^k, W_0 is a compact convex set. So there exists a strongly separating hyperplane $H_0 = \{\omega \in \mathbb{R}^{K(\mu_0)}, \langle \omega, q_0 \rangle = b\}$ such that $\sup_{m \in M_0} \langle m, q_0 \rangle < \inf_{\omega \in W_0} \langle \omega, q_0 \rangle$. Every component of q_0 must be non-negative (since M_0 is negatively comprehensive), therefore up to a normalization, we can assume that q_0 belongs to $\Delta(K(\mu_0))$.

Define $W = W_0 \times \mathbb{R}^{K \setminus K(\mu_0)}$ and $q \in \Delta(K)$ by $q(k) = q_0(k)$ if $k \in K(\mu_0)$ and 0 otherwise. Then, $H = \{\omega \in \mathbb{R}^K, \langle \omega, q \rangle = b\}$ strongly separates M and W, therefore :

$$\langle \mathbf{m}, q \rangle < \min_{\omega \in W_0} \langle \omega, q \rangle = \min_{y \in \Delta(J)} \max_{x \in NR(q,\mu_0)} \sum_{k \in K} q^k \rho^k(x^k, y) = u(q, \mu_0) \leq u(q) \leq \langle \mathbf{m}, q \rangle.$$

Therefore M is approachable by Player 2, he can guarantee $\mathbf{Cav}(u)(p)$ in $\Gamma_\infty(p)$ and $v_\infty(p) = \mathbf{Cav}(u)(p)$. $\square$

Acknowledgments : I am grateful to my advisor Sylvain Sorin for his help, comments and time. I also want to thank Jérôme Renault and Gilles Stoltz for sharing ideas and pointing out to me the counter-examples.

9

Purely Informative Game.

Dans ce chapitre, on associe à un jeu répété un jeu purement informatif *abstrait où le paiement d'un joueur à une étape est son information maximale, représentée comme une mesure de probabilité dans l'espace L_2 de Wasserstein. Dans ce dernier, on utilise une notion de normale proximale à un ensemble ce qui permet d'étendre la notion de B-set aux ensembles de mesures. Ce chapitre est issu de l'article* Purely Informative Game. Approachability in Wasserstein Space *écrit avec M. Quincampoix.*

Sommaire

9.1 Approachability in the purely informative game **152**
Game with full monitoring 152
Game with partial monitoring 153
Purely informative game 154
Links between approachability in these games 155
9.2 Characterization of approachable sets **160**
Preliminaries on the space of probability measures 161
Approachability in the purely informative game 164
9.3 Characterization of convex approachable sets **168**
9.4 Convex games **170**

REPEATED games can be studied by considering sequences of payoffs and constructing, stage by stage, strategies with the requirement that the outcome at the next stage will have good properties given the past ones. Perhaps, the most revealing examples of this claim are Shapley's [104] operator approach that describes the value of stochastic zero-sum games (repeated a finite number of times), the *exponential weight algorithm* for predictions with expert advices (see *e.g.* Cesa-Bianchi and Lugosi [29], Chapter 6) or Blackwell's [17] approachability theory.

We recall that in a two-person repeated game with vector payoffs in some Euclidian space $\mathbb{R}^k$, a player can approach a given set $E \subset \mathbb{R}^k$, if he can insure that, after

some stage and with a great probability, the average payoff will always remain close to E. When both players observe their opponent's moves (or at least his payoff), Blackwell [17] proved that if E satisfies some geometrical condition with respect to the game (E is then called a B-set), then Player 1 can approach it. He also deduced that either Player 1 can approach a convex set C or Player 2 can exclude it, *i.e.* he can approach the complement of a neighborhood of C.

In the partial monitoring framework, when players do not observe their opponent's moves but receive random signals (whose laws may depend on the actions played), working on the space of payoffs might not be sufficient (except for specific games such as prediction games).

Attempts were made to circumvent this issue, notably by Aumann and Maschler [9] and Kohlberg [67] in the framework of repeated games with incomplete information. Lehrer and Solan [74] also considered strategies that are defined, not as a function of the unknown past payoffs, but as a function of the past signals, and they proved the existence of strategies that satisfy an extension of the consistency property (as defined in the full monitoring framework). Perchet [91] also used this approach to provide a complete characterization of approachable convex sets (that extends Blackwell's one in the full monitoring case).

We formalize and generalize this insight by introducing in section 9.1 a game $\widetilde{\Gamma}$ (that we call *Purely Informative Game*). The outcome of this game is the information available to each player — seen as a probability measure. His objective is that the sequence of informations satisfies a property equivalent to Blackwell's approachability in the space of measures — called Wasserstein Space.

The remaining of this paper is organized as follows. In section 9.1, we recall Blackwell's definition of approachability, we extend it to $\widetilde{\Gamma}$ and we link these two notions. In section 9.2, we define a $\widetilde{B}$-set, the extension of a B-sets. A set in $\widetilde{\Gamma}$ is approachable if and only if it contains a $\widetilde{B}$-set. We provide a full characterization of approachable convex sets in section 9.3. The last section is devoted to a specific class of games, called *convex games*, in which there exist explicit and optimal bounds of convergence.

9.1 Approachability in the purely informative game

Game with full monitoring

We consider a two player repeated game Γ^f where at stage $n \in \mathbb{N}$, Player 1 (resp. Player 2) chooses an action i_n in a finite set I (resp. $j_n \in J$). This generates a vector payoff $\rho_n = \rho(i_n, j_n) \in \mathbb{R}^k$ where ρ is a mapping from $I \times J$ to $\mathbb{R}^k$, extended to $\Delta(I) \times \Delta(J)$ by $\rho(x, y) = \mathbb{E}_{x,y}[\rho(i, j)]$, where $\Delta(I)$ (resp. $\Delta(J)$) is the set of probability measures over I (resp. J).

As usual, a behavioral strategy σ of Player 1 (resp. τ of Player 2) is a mapping from the set of finite histories $H := \bigcup_{n\in\mathbb{N}} (I \times J)^n$ into $\Delta(I)$ (resp. from H to $\Delta(J)$) and a couple of strategies (σ, τ) generates a probability, denoted by $\mathbb{P}_{\sigma,\tau}$, over the set of infinite histories $H^\infty := (I \times J)^{\mathbb{N}}$, endowed with the cylinder topology.

Given $E \subset \mathbb{R}^k$ and $\delta \geq 0$, we denote by $d(z, E) = \inf_{e\in E} \|z - e\|$, with $\|\cdot\|$ the Euclidian norm, the distance to E, by $E^\delta = \{\omega \in \mathbb{R}^k, d(\omega, E) < \delta\}$ the δ-neighborhood of E and by $\Pi_E(z) = \{e \in E, d(z, E) = \|z - e\|\}$ the set of projections of z to E. Given $\{a_m\}_{m\in\mathbb{N}}$, $\overline{a}_n = \frac{1}{n}\sum_{m=1}^n a_m$ is its average up to the n-th term.

Definition 9.1

i) A closed set $E \subset \mathbb{R}^k$ is approachable by Player 1 if for every $\varepsilon > 0$, there exist a strategy σ of Player 1 and $N \in \mathbb{N}$, such that for every strategy τ of Player 2 and every $n \geq N$:

$$\mathbf{E}_{\sigma,\tau}\left[d(\overline{\rho}_n, E)\right] \leq \varepsilon.$$

ii) A set E is excludable by Player 2, if there exists $\delta > 0$ such that the complement of E^δ is approachable by Player 2.

In words, a set $E \subset \mathbb{R}^k$ is approachable by Player 1, if he has a strategy such that the expected average payoff converges to E, uniformly with respect to the strategies of Player 2.

Game with partial monitoring

We consider Γ^p a game with the same action spaces and payoff function than Γ^f but where Player 1 has partial monitoring. At stage n, he does not observe Player 2's action j_n, nor his payoff ρ_n, but he receives a random signal s_n whose law is $s(i_n, j_n) \in \Delta(S)$, where s is a mapping from $I \times J$ to $\Delta(S)$, extended to $\Delta(I) \times \Delta(J)$ by $s(x, y) = \mathbb{E}_{x,y}[s(i, j)] \in \Delta(S)$.

We define the mapping $\mathbf{s}$ from $\Delta(J)$ to $\Delta(S)^I$ by $\mathbf{s}(y) = (s(i, y))_{i\in I}$. Its range $\mathcal{S} \subset \Delta(S)^I$ is a polytope (*i.e.* the convex hull of a finite number of points) and any of its element is called a *flag*. Whatever being his move, Player 1 cannot distinguish between two actions $y_0, y_1 \in \Delta(J)$ that generate the same flag $\mu \in \mathcal{S}$ (*i.e.* such that $\mathbf{s}(y_0) = \mathbf{s}(y_1) = \mu$) thus $\mathbf{s}(y)$ (although not observed) is the maximal information available to Player 1, given $y \in \Delta(J)$. Note that with full monitoring, a flag is simply the law of the action of Player 2.

Given $x \in \Delta(I)$ and $\mu \in \Delta(S)^I$, the set of compatible payoffs is defined by :

$$P(x, \mu) = \left\{\rho(x, y), y \in \mathbf{s}^{-1}(\mu)\right\}.$$

With partial monitoring, a strategy σ of Player 1 (resp. τ of Player 2) is a mapping from the set of past finite observations $H^1 := \bigcup_{n\in\mathbb{N}}(I \times S)^n$ into $\Delta(I)$ (resp.

from $H^2 := \bigcup_{n\in\mathbb{N}}(I \times S \times J)^n$ into $\Delta(J)$). A couple of strategies (σ, τ) generates a probability, also denoted by $\mathbb{P}_{\sigma,\tau}$ on $H^\infty := (I \times S \times J)^\infty$ embedded with the cylinder topology. With this notation, the definition of approachability with partial monitoring is exactly the same as the one with full monitoring.

In order to treat simultaneously Γ^f and Γ^p define X and $\mathcal{X}$ — the *informative actions spaces* — and P the multivalued application from $X \times \mathcal{X}$ to $\mathbb{R}^k$ by the following. In both games, $X = \Delta(I)$ and :

In Γ^f : $\mathcal{X} = \Delta(J)$ and $P(x, \xi) = \{\rho(x, \xi)\}$ for every $x \in X$ and $\xi \in \mathcal{X}$.

In Γ^p : $\mathcal{X} = \mathcal{S}$ and $P(x, \xi) = \{\rho(x, y); y \in \mathbf{s}^{-1}(\xi)\}$ for every $x \in X$ and $\xi \in \mathcal{X}$.

Purely informative game

The purely informative game $\widetilde{\Gamma}$ is an abstract game defined as follows : at stage $n \in \mathbb{N}$, Player 1 (resp. Player 2) chooses $\mathbf{x}_n \in \Delta(X)$ (resp. $\boldsymbol{\xi}_n \in \Delta(\mathcal{X})$), the set of probability measures over X, a compact convex subset of some Euclidian space endowed with the weak-$\star$ topology. Those choices generate the outcome (the term payoff will only be used in Γ^p or Γ^f) :

$$\theta_n = \theta(\mathbf{x}_n, \boldsymbol{\xi}_n) = \mathbf{x}_n \otimes \boldsymbol{\xi}_n \in \Delta(X \times \mathcal{X}).$$

A strategy σ of Player 1 is a mapping from $\bigcup_{n\in\mathbb{N}} (\Delta(X) \times \Delta(\mathcal{X}))^n$ to $\Delta(X)$. Similarly, a strategy τ of Player 2 is defined as a mapping from $\bigcup_{n\in\mathbb{N}} (\Delta(X) \times \Delta(\mathcal{X}))^n$ to $\Delta(\mathcal{X})$. A couple of strategies (σ, τ) induces a unique sequence $(\mathbf{x}_n, \boldsymbol{\xi}_n)_{n\in\mathbb{N}}$ in $(\Delta(X) \times \Delta(\mathcal{X}))^{\mathbb{N}}$.

Let us briefly recall the definition of the (quadratic) Wasserstein distance W_2 which is a metric compatible with the weak convergence of probability measures on a compact set, see *e.g.* Dudley [35], Chapter 11.8. Its definition relies on the following proposition called Kantorovitch duality.

Consider μ and ν in $\Delta^2\left(\mathbb{R}^N\right)$, the set of measures with a finite moment of order 2 in some Euclidian space $\mathbb{R}^N$. We define for every $\gamma \in \Delta\left(\mathbb{R}^N \times \mathbb{R}^N\right)$ and couple of fonctions $(\phi, \psi) \in L^1_\mu\left(\mathbb{R}^N, \mathbb{R}\right) \times L^1_\nu\left(\mathbb{R}^N, \mathbb{R}\right)$:

$$I[\gamma] = \int_{\mathbb{R}^N\times\mathbb{R}^N} \|x - y\|^2 \mathrm{d}\gamma(x, y), \text{ and } J(\phi, \psi) = \int_{\mathbb{R}^N} \phi \mathrm{d}\mu + \int_{\mathbb{R}^N} \psi \mathrm{d}\nu.$$

Denote by $\Pi(\mu, \nu)$ the set of probability measures γ on $\mathbb{R}^N \times \mathbb{R}^N$ with marginals μ and ν and by Ξ the set of measurable functions $(\phi, \psi) \in L^1_\mu(\mathbb{R}^N, \mathbb{R}) \times L^1_\nu(\mathbb{R}^N, \mathbb{R})$ satisfying $\phi(x) + \psi(y) \leq \|x - y\|^2, \quad \mu \otimes \nu$-as.

Proposition 9.2

For every μ and ν in $\Delta^2\left(\mathbb{R}^N\right)$:

$$W_2^2(\mu, \nu) := \inf_{\gamma\in\Pi(\mu,\nu)} I[\gamma] = \sup_{(\phi,\psi)\in\Xi} J(\phi, \psi). \tag{9.1}$$

There exists an other equivalent formulation of W_2 — referred later as the probabilistic interpretation — implied by the Skorokhod's representation Theorem ; it yields that

$$\inf_{\gamma\in\Pi(\mu,\nu)} I[\gamma] = \inf_{U\sim\mu;V\sim\nu} \mathbb{E}\left[\|U-V\|^2\right],$$

where $U \sim \mu$ means that the law of the random variable U is μ.

This metric allows us to define approachability in the space of measures. We denote by $\overline{\theta}_n$ the average outcome up to stage n in the following sense : for every Borel subset of $F \subset X \times \mathcal{X}$, $\overline{\theta}_n(F) = \frac{1}{n}\sum_{m=1}^{n}\theta_m(F)$.

Definition 9.3

A closed set $\widetilde{E} \subset \Delta(X \times \mathcal{X})$ is approachable by Player 1 if for every $\varepsilon > 0$ there exist a strategy σ of Player 1 and $N \in \mathbb{N}$ such that for every strategy τ of Player 2 :

$$\forall n \geq N, W_2\left(\overline{\theta}_n, \widetilde{E}\right) := \inf_{\theta\in\widetilde{E}} W_2(\overline{\theta}_n, \theta) \leq \varepsilon.$$

In this definition the choice of the quadratic distance is not compulsory ; we can indeed define approachability with respect of any distance W_r, for $r \geq 1$ (see *e.g.* Dudley [35], Chapter 11.8). We stress out that $\widetilde{\Gamma}$ is independent of Γ^f and Γ^p and can be defined for any compact convex subsets X and $\mathcal{X}$.

Links between approachability in these games

We define the set of outcomes compatible with $\theta \in \Delta(X \times \mathcal{X})$ by :

$$\rho(\theta) = \int_{X\times\mathcal{X}} P(x,\xi)\mathrm{d}\theta \subset \mathbb{R}^k$$

where the integral is in Aumann's sense of integration of correspondence (see *e.g.* Aubin and Frankowska [5]) : it is the set of all integrals of measurable selection of P. For every subset $\widetilde{E} \subset \Delta(X \times \mathcal{X})$, the set of compatible outcomes $\rho(\widetilde{E}) \subset \mathbb{R}^k$ is defined by :

$$\rho(\widetilde{E}) = \left\{\rho(\theta), \theta \in \widetilde{E}\right\} \subset \mathbb{R}^k.$$

Reciprocally, we define for every subset $E \subset \mathbb{R}^k$, the set of compatible measures $\widetilde{\rho}(E) \subset \Delta(X \times \mathcal{X})$ by :

$$\widetilde{\rho}(E) = \{\theta \in \Delta(X \times \mathcal{X}), \rho(\theta) \subset E\}.$$

The signal mapping $\mathbf{s}$ does not appear in the description of $\widetilde{\Gamma}$ but only in the definition of two mappings ρ and $\widetilde{\rho}$ that link Γ^p and $\widetilde{\Gamma}$. Given a set $E \subset \mathbb{R}^k$ in Γ^p,

we have defined through $\widetilde{\rho}$ a compatible set $\widetilde{E} = \widetilde{\rho}(E) \subset \Delta(X \times \mathcal{X})$ in $\widetilde{E}$; it is quite intuitive that E is approachable if and only $\widetilde{E}$ is (see Theorem 9.4 below).

Although the outcomes are less complex in Γ^p than in $\widetilde{\Gamma}$, the description of approachable sets in the latter is a lot simpler, because it is independent of ρ and $\mathbf{s}$. This is the reason why we introduced the purely informative game.

Theorem 9.4

i) A set $E \subset \mathbb{R}^k$ is approachable in Γ^f (resp. in Γ^p) if and only if the set $\widetilde{\rho}(E) \subset \Delta(X \times \mathcal{X})$ is approachable in $\widetilde{\Gamma}$;

ii) If a set $\widetilde{E} \subset \Delta(X \times \mathcal{X})$ is approachable in $\widetilde{\Gamma}$ then the set $\rho(\widetilde{E}) \subset \mathbb{R}^k$ is approachable in Γ^f (resp. in Γ^p);

iii) For every convex set $C \subset \mathbb{R}^k$, $\widetilde{\rho}(C) \subset \Delta(X \times \mathcal{X})$ is a (possibly empty) convex set and for every convex set $\widetilde{C} \subset \Delta(X \times \mathcal{X})$, $\rho(\widetilde{C})$ is a convex set.

Proof. The third point is obvious, so we only need to prove that if $\widetilde{E} \subset \Delta(X \times \mathcal{X})$ is approachable in $\widetilde{\Gamma}$ then $\rho(\widetilde{E}) \subset \mathbb{R}^k$ is approachable in Γ^p (see **part 1**) and that if $E \subset \mathbb{R}^k$ is approachable in Γ^p then $\widetilde{\rho}(E) \subset \Delta(X \times \mathcal{X})$ is also approachable (see **part 2**). The remaining easily follows from the fact that $\rho\big(\widetilde{\rho}(E)\big) \subset E$.

part 1 : The proof consist in two steps. First, we link the Wasserstein distance between two probability measures $\overline{\theta}, \underline{\theta} \in \Delta(X \times \mathcal{X})$ and the distance between the two sets $\rho(\overline{\theta}) \subset \mathbb{R}^k$ and $\rho(\underline{\theta}) \subset \mathbb{R}^k$ (see Lemmas 9.6 and 9.7). We will prove this step with the use of the 1-Wasserstein distance (defined below). In the second step, we transform a strategy in $\widetilde{\Gamma}$ into a strategy in Γ^p.

Step 1 : The 1-Wasserstein distance between μ and ν in $\Delta(X)$ is defined by (see *e.g.* Dudley [35], Chapter 11.8) :

$$W_1(\overline{\theta}, \underline{\theta}) = \inf_{\gamma \in \Pi(\overline{\theta}, \underline{\theta})} \int_X \|x - y\| \mathrm{d}\gamma(x, y) = \sup_{\phi \in \mathrm{Lip}_1(X, \mathbb{R})} \int_X \phi \, \mathrm{d}\overline{\theta} - \mathrm{d}\underline{\theta} = \inf_{U \sim \overline{\theta}, V \sim \underline{\theta}} \mathbb{E}[\|U - V\|],$$

where $\mathrm{Lip}_1(X, \mathbb{R})$ is the set of 1-Lipschitzian functions from X to $\mathbb{R}$. Since by Jensen's inequality, for any random variable U and V, $\mathbb{E}\left[\|U - V\|\right]^2 \leq \mathbb{E}\left[\|U - V\|^2\right]$, the probabilistic interpretation implies that $W_1(\overline{\theta}, \underline{\theta}) \leq W_2(\overline{\theta}, \underline{\theta})$.

Definition 9.5

The multivalued map $P : X \times \mathcal{X} \rightrightarrows \mathbb{R}^k$ is a L-Lipschitizian convex hull, with $L > 0$, if there exists a family $\{p_\kappa : X \times \mathcal{X} \to \mathbb{R}^k; \kappa \in \mathcal{K}\}$ of L-Lipchitzian functions such that $P(x, \xi) = \mathrm{co}\left\{p_\kappa(x, \xi); \kappa \in \mathcal{K}\right\}$.

Lemma 9.6

Let $\widetilde{E}$ be a compact subset of $\Delta(X\times\mathcal{X})$ and $\overline{\theta}$ such that $W_1\left(\overline{\theta},\widetilde{E}\right)\leq\varepsilon$. Assume that P is a L-Lipschitzian convex hull, then $\sup_{z\in\rho(\overline{\theta})}d(z,\rho(\widetilde{E}))\leq\sqrt{k}L\varepsilon$.

Proof. the result is almost trivial in Γ^f. Indeed, $\rho : X\times\mathcal{X}$ is L-Lipschitzian with $L=\|\rho\|_\infty$ and if we denote $\rho=(\rho^1,\ldots,\rho^k)$ then each ρ^k/L is 1-Lipschitzian. Let us denote by $\underline{\theta}$ any projection of $\overline{\theta}$ on $\widetilde{E}$, then the definition of W_1 implies that :

$$\begin{aligned}d(\rho(\overline{\theta}),\rho(\widetilde{E}))&\leq d\left(\int_{X\times\mathcal{X}}\rho(x,\xi)\mathrm{d}\overline{\theta},\int_{X\times\mathcal{X}}\rho(x,\xi)\mathrm{d}\underline{\theta}\right)\\&\leq\sqrt{k}L\sup_{t\leq k}\left|\int_{X\times\mathcal{X}}\frac{\rho^t(x,\xi)}{L}\mathrm{d}\overline{\theta}-\int_{X\times\mathcal{X}}\frac{\rho^t(x,\xi)}{L}\mathrm{d}\underline{\theta}\right|\leq\sqrt{k}L\varepsilon.\end{aligned}$$

In Γ^p, $P(\cdot)$ is not a singleton, therefore any of its selection (in the definition of the Aumann integral) may not be lipschitzian, so the argument in the proof for Γ^f cannot be directly used. However, since $P(x,\xi)=\mathrm{co}\left\{p_\kappa(x,\xi);\kappa\in\mathcal{K}\right\}$ where every p_κ is L-Lipschitzian, for every $\overline{\theta}\in\Delta(X\times\mathcal{X})$, by convexity of the integral see *e.g.* Klein and Thompson [66], Theorem 18.1.19 :

$$\int_{X\times\mathcal{X}}P(x,\xi)\mathrm{d}\overline{\theta}=\int_{X\times\mathcal{X}}\mathrm{co}\left\{p_\kappa(x,\xi),\kappa\in\mathcal{K}\right\}\mathrm{d}\overline{\theta}=\mathrm{co}\left\{\int_{X\times\mathcal{X}}p_\kappa(x,\xi)\mathrm{d}\overline{\theta},\kappa\in\mathcal{K}\right\}. \tag{9.2}$$

Since p_κ is L-Lipschitzian, $d\left(\int_{X\times\mathcal{X}}p_\kappa(x,\mu)\mathrm{d}\overline{\theta},\rho(\underline{\theta})\right)\leq\sqrt{k}L\varepsilon$ and since $\rho(\underline{\theta})$ is a convex set $d\left(\rho(\overline{\theta}),\rho(\underline{\theta})\right)\leq\sqrt{k}L\varepsilon$. □

Lemma 9.7

If the functions s and ρ are linear, then $(x,\mu)\mapsto P(x,\mu)=\{\rho(x,y),y\in\mathbf{s}^{-1}(\mu)\}$ is a L-Lipschitzian convex hull.

Proof. Since the graph of $\mathbf{s}^{-1}$ is a polytope of $\mathbb{R}^{SI}\times\mathbb{R}^J$, there exists a finite family of (so called) *extreme points* functions $\{y_\kappa(\cdot),\kappa\in\mathcal{K}\}$ from $\mathcal{S}$ into $\Delta(J)$ that are all piecewise linear and continuous (thus Lipschitzian) such that $\mathbf{s}^{-1}(\mu)=\mathrm{co}\left\{y_\xi(\mu),\kappa\in\mathcal{K}\right\}$, for every $\mu\in\mathcal{S}$. Since $\rho(x,\cdot)$ is linear on $\Delta(J)$:

$$\begin{aligned}P(x,\mu)&=\left\{\rho(x,y);y\in\mathbf{s}^{-1}(\mu)\right\}=\left\{\rho(x,y);y\in\mathrm{co}\left\{y_\kappa(\mu),\kappa\in\mathcal{K}\right\}\right\}\\&=\mathrm{co}\left\{\rho(x,y_\kappa(\mu)),\kappa\in\mathcal{K}\right\}.\end{aligned}$$

Therefore P is indeed a L-Lipschitzian convex hull. □

Step 2 : This step will use the following lemma known as Hoeffding-Azuma's [61, 11] inequality for sums of martingale differences.

Definition 9.8

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ be a filtration. A stochastic process $U_n \in \mathbb{R}$ is a sequence of bounded (by $B > 0$) martingale differences if for every $n \in \mathbb{N}$ U_n is adapted to $\mathcal{F}_n$, $|U_n| < B$ almost surely and $\mathbb{E}[X_n|\mathcal{F}_{n-1}] = 0$.

Lemma 9.9 (*Hoeffding-Azuma's [61, 11] inequality*)

Let U_n be a sequence of bounded (by $B > 0$) martingale differences in $(\Omega, \mathcal{F}, \mathcal{P})$ adapted to $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$. Then for every $n \in \mathbb{N}$ and every $\varepsilon > 0$:

$$\mathbb{P}\left(\left|\overline{U}_n\right| \geq \varepsilon\right) \leq 2\exp\left(\frac{-n\varepsilon^2}{2K^2}\right).$$

In particular, for every $\varepsilon > 0$, there exists $N > 0$ such that $\mathbb{E}\left[\sup_{n\geq N}|\overline{U}_n|\right] \leq \varepsilon$.

Let $\widetilde{\sigma}$ be a strategy of Player 1 that approaches (up to $\varepsilon > 0$) a set $\widetilde{E} \subset \Delta(X \times \mathcal{X})$. This strategy cannot be directly played in Γ^p in order to approach $\rho(\widetilde{E})$ for two reasons :

1) in Γ^p, Player 1 chooses an action $i \in I$ and not some $\mathbf{x} \in \Delta(\Delta(I))$;
2) at stage n in Γ^p, he does not observe the flag $\mu_n = \mathbf{s}(i_n, j_n)$ but receives a signal s_n whose law is $s(i_n, j_n)$.

We will use classic statistical tools in order to transform a strategy $\widetilde{\sigma}$ in $\widetilde{\Gamma}$ into a strategy σ in Γ^p. We decompose $\mathbb{N}$ into blocks of length N (where N is large enough and will be defined later) ; roughly speaking, the n-th block in Γ^p will correspond to the n-th stage in $\widetilde{\Gamma}$.

The strategy σ is defined inductively as follows. Denote by $\mathbf{x}_1 = \widetilde{\sigma}(\emptyset)$ the action Player 1 should play in $\widetilde{\Gamma}$ at the first stage accordingly to $\widetilde{\sigma}$ and let $x_1 = \mathbb{E}_{\mathbf{x}_1}[x] \in \Delta(I)$ be its expectation. At every stage of the first block, Player 1 will play in Γ^p accordingly to an η-perturbation of x_1, defined by $(1-\eta)x_1 + \eta u$ (where u is the uniform distribution over I). If the first block is long enough, the empirical distribution of Player 1's action on this block (denoted by $x(1)$) will be 2η-close to x_1.

The η-perturbation (which was introduced by Auer, Cesa-Bianchi, Freund, and Schapire [6]) allows to construct c_n an unbiased estimator of μ_n by defining :

$$c_n = \frac{\mathbb{1}_{(i,s)=(i_n,s_n)}}{x_1[i_n]} \in \mathbb{R}^{SI}$$

where (i_n, s_n) is the couple "action actually played-signal received by Player 1 at stage n" and $x_1[i_n] > 0$ is the probability that Player 1 chose the action i_n. The vector c_n is indeed an unbiased estimator of μ_n since $\mathbb{E}_{\sigma,\tau}[c_n] = \mu_n$, where $\Delta(S)^I$ is seen as a subset of $\mathbb{R}^{SI}$. If the first block is long enough, the average of every c_n on this block (denoted by $c(1)$) will be close to the average flag denoted by $\mu(1)$.

Finally, we denote by $\rho(1)$ the average payoff on the first block and $y(1)$ the empirical distribution of Player 2's action.

Since $W_1\left(\delta_{x(1)} \otimes \delta_{c(1)}, \delta_{x_1} \otimes \delta_{\mu(1)}\right) = \left\| (x(1), c(1)) - (x_1, \mu(1)) \right\|$, $\mathbb{E}_{\sigma,\tau}[c_n] = \mu_n$, and $\mathbb{E}_{\sigma,\tau}[i_n] = (1-\eta)x_1 + \eta u$, Lemma 9.9 implies that for N large enough and η small enough

$$\mathbb{E}_{\sigma,\tau}\left[W_1\left(\delta_{x(1)} \otimes \delta_{c(1)}, \delta_{x_1} \otimes \delta_{\mu(1)}\right)\right] \leq \varepsilon.$$

Moreover, by definition of $\rho(1)$:

$$\begin{aligned}\rho(1) &= \frac{\sum_{m=1}^{N} \rho(i_m, j_m)}{N} = \frac{\sum_{m=1}^{N} \rho(i_m, j_m) - \rho(x_1, j_m)}{N} + \frac{\sum_{m=1}^{N} \rho(x_1, j_m)}{N} \\ &= \left| \frac{\sum_{m=1}^{N} \rho(i_m, j_m) - \rho(x_1, j_m)}{N} \right| + \rho(x_1, y(1)).\end{aligned}$$

Since $\mathbf{s}(y(1)) = \mu(1)$ and for every $m \leq N$, $\mathbb{E}[\rho(i_m, j_m)] = \rho((1-\eta)x_1 + \eta u, j_m)$, Lemma 9.9 implies that for N large enough and η small enough :

$$\mathbb{E}_{\sigma,\tau}\left[d\left(\rho(1), P[x_1, \mu(1)]\right)\right] = \mathbb{E}_{\sigma,\tau}\left[d\left(\rho(1), \rho(\delta_{x_1}, \delta_{\mu(1)})\right)\right] \leq \varepsilon.$$

Assume that σ is described for the K first blocks and denote by $c(k)$ (resp. $x(k)$, $\mu(k)$) the average of c_m (resp. i_m, μ_m) on the k-th block and by $\mathbf{x}_k \in \Delta(\Delta(I))$ the action specified by $\widetilde{\sigma}$ in $\widetilde{\Gamma}$. Then we define :

$$\mathbf{x}_{K+1} = \widetilde{\sigma}(\mathbf{x}_1, \ldots, \mathbf{x}_K, \delta_{c(1)}, \ldots, \delta_{c(K)}) \in \Delta(\Delta(I)), \text{ and } x_{K+1} = \mathbb{E}\left[\mathbf{x}_{K+1}\right] \in \Delta(I),$$

if $c(k)$ does not belong to $\int$ then consider instead its projection (which is closer to $\mu(k)$). Player 1 will play at every stage of the $K+1$-th block accordingly to η-perturbation of x_{K+1}. For every $k \in \mathbb{N}$ (and the proof is exactly the same as the one for $k = 1$) :

$$\mathbb{E}_{\sigma,\tau}\left[W_1\left(\delta_{x(k)} \otimes \delta_{c(k)}, \delta_{x_k} \otimes \delta_{\mu(k)}\right)\right] \leq \varepsilon \text{ and } \mathbb{E}_{\sigma,\tau}\left[d\left(\rho(k), \rho\left(\delta_{x_k}, \delta_{\mu(k)}\right)\right)\right] \leq \varepsilon.$$

Therefore if we define $\mathbf{t}_n = \delta_{x(n)} \otimes \delta_{c(n)}$ and $t_n = \delta_{x_n} \otimes \delta_{\mu(n)}$ then :

$$\mathbb{E}_{\sigma,\tau}\left[W_1(\overline{\mathbf{t}}_n, \overline{t}_n)\right] \leq \varepsilon \text{ and } \mathbb{E}_{\sigma,\tau}\left[d(\overline{\rho}(n), \rho(\overline{t}_n)\right] \leq \varepsilon.$$

Since $\widetilde{\sigma}$ is an approachability strategy of $\widetilde{E}$, if we denote by $\theta_n = \mathbf{x}_n \otimes \delta_{\mu(n)}$ then for n large enough (*i.e.* larger than $\overline{n}$) $W_1(\overline{\theta}_n, \widetilde{E}) \leq \varepsilon$. Finally, since $\rho(\cdot, y)$ is linear on $\Delta(I)$, then $\rho(\overline{\theta}_n) = \rho(\overline{t}_n)$ and since every block has the same length $\overline{\rho}_{Nn} = \overline{\rho}(n)$. Thus we have shown that

$$\mathbb{E}_{\sigma,\tau}\left[d\left(\overline{\rho}_{Nn}, \rho\left(\widetilde{E}\right)\right)\right] \leq (1 + \sqrt{k}L)\varepsilon \tag{9.3}$$

for every $n \geq \overline{n}$. Consider any integer $m \in \mathbb{N}$ such that $Nn \leq m \leq N(n+1)$, then $\|\overline{\rho}_m - \overline{\rho}_{Nn}\| \leq \frac{2\|\rho\|_\infty}{n}$, therefore $\rho(\widetilde{E})$ is approachable by Player 1.

Part 2 : Assume that $E \subset \mathbb{R}^k$ is approachable in Γ^p by Player 1. Consider the game where Player 1 observes in addition $\mu_n = \mathbf{s}(j_n)$ and his payoff is $\mathbb{E}_{x_n}[\rho(i_n, j_n)]$ where x_n is the law of i_n. This new game is *easier* for Player 1 because he has more information and actions, hence he can still approach $E \subset \mathbb{R}^k$. Since $\mathbf{s}^{-1}$ is convex on $\mathcal{S}$, allowing Player 2 to play any action in $\Delta(\mathcal{X})$ does not make the game harder for Player 1. Thus we can assume that at stage $n \in \mathbb{N}$, Player 1 observes $\boldsymbol{\xi}_n \in \Delta(\mathcal{X}) = \Delta(\mathcal{S})$, that he plays deterministically $\mathbf{x}_n \in \Delta(X) = \Delta(\Delta(I))$ and that his payoff is $\rho(\mathbf{x}_n \otimes \boldsymbol{\xi}_n)$. We call this new game by Γ^d.

The fact that E is approachable in Γ^d implies that for every ε there exists a strategy σ_ε in Γ^d and $N_\varepsilon \in \mathbb{N}$ such that for every $n \geq N_\varepsilon$ and strategy τ of Player 2 :

$$d\left(\frac{\sum_{m=1}^n \rho\left(\mathbf{x}_m \otimes \boldsymbol{\xi}_m\right)}{n}, E\right) := \sup\left\{d(z, E), z \in \frac{\sum_{m=1}^n \rho\left(\mathbf{x}_m \otimes \boldsymbol{\xi}_m\right)}{n}\right\} \leq \varepsilon. \quad (9.4)$$

If we denote as before $\theta_n = \mathbf{x}_n \otimes \boldsymbol{\xi}_n \in \Delta(X \times \mathcal{X})$ then equation (9.4) becomes :

$$\forall \varepsilon > 0, \exists N_\varepsilon \in \mathbb{N}, \forall n \geq N_\varepsilon, \overline{\theta}_n \in \widetilde{\rho}(E^\varepsilon).$$

Let us define similarly $\widetilde{\rho}(E)^\delta = \left\{\theta \in \Delta\left(X \times \mathcal{X}\right), W_1\left(\theta, \widetilde{E}\right) \leq \delta\right\}$. Since the sequence of compact sets $\widetilde{\rho}\left(E^\varepsilon\right)$ converges (as ε converges to zero) to $\widetilde{\rho}(E)$, for every $\delta > 0$, there exists $\underline{\varepsilon}$ such that for every $0 < \varepsilon \leq \underline{\varepsilon}$, $\widetilde{\rho}\left(E^\varepsilon\right) \subset \widetilde{\rho}(E)^\delta$. Therefore, for every $\delta > 0$, there exists $N \in \mathbb{N}$ such that for every $n \geq N$ and every strategy τ of Player 2, $\overline{\theta}_n$ belongs to $\widetilde{\rho}(E)^\delta$. Thus $\widetilde{\rho}(E)$ is approachable by Player 1. □

Remark : Point *ii*) cannot be a equivalence. Indeed consider a game with full monitoring where the payoff function is constant and equals 0. Define $\widetilde{E} = \{\mathbf{x_0} \otimes \mathbf{y_0}\}$ where $\mathbf{x_0}$ and $\mathbf{y_0}$ are chosen arbitrarily. Then $\rho(\widetilde{E}) = \{0\}$ is approachable and $\widetilde{E}$ is not, since Player 2 just has to play $\mathbf{y_1} \neq \mathbf{y_0}$ at each stage.

9.2 Characterization of approachable sets

In Γ^f, Blackwell [17] noticed that a closed set E that fulfills a geometrical condition with respect to the game, see Definition 9.10 below, is approachable by Player 1. In order to state it, we introduce $P^2(x) = \{\rho(x, y), y \in \Delta(J)\}$, the set of expected payoffs compatible with $x \in \Delta(I)$, and we define similarly $P^1(y)$.

Definition 9.10

A closed subset E of $\mathbb{R}^k$ is a B-set for Player 1 in Γ^f, if for every $z \in \mathbb{R}^k$, there exist $p \in \Pi_E(z)$ and $x\,(= x(z, p)) \in \Delta(I)$ such that the hyperplane through p and perpendicular to $z - p$ separates z from $P^2(x)$, or formally :

$$\forall z \in \mathbb{R}^k, \exists p \in \Pi_E(z), \exists x \in \Delta(I), \langle \rho(x, y) - p, z - p\rangle \leq 0, \quad \forall y \in \Delta(J). \quad (9.5)$$

An equivalent formulation using the set of proximal normals to E at q, denoted by $\mathrm{NC}_E(q)$ (see *e.g.* Bony [23] or Clarke [32]) is due to As Soulaimani, Quincampoix and Sorin [4]. E is a B-set if and only if

$$\forall p \in E, \forall q \in \mathrm{NC}_E(p), \exists x \in \Delta(I), \forall y \in \Delta(J) \langle \rho(x,y) - p, q \rangle \leq 0.$$

The main result with full monitoring is :

Proposition 9.11 (*Blackwell [17], Spinat [110]*)

E is approachable in Γ^f if and only if it contains a B-set.

We aim to extend the definition of a B-set to the space of measures. The definitions of proximal normals in this space, recalled in the following section (see Definition 9.13 and 9.14 below), require some knowledge on the Wasserstein Space.

Preliminaries on the space of probability measures

WASSERSTEIN SPACE

For any μ in $\Delta(\mathbb{R}^N)$ and $p \geq 1$ we denote by $L^p_\mu(\mathbb{R}^N, \mathbb{R})$ (resp. $L^p_\mu(\mathbb{R}^N, \mathbb{R}^N)$) the set of $\mu-$measurable maps $\phi : \mathbb{R}^N \to \mathbb{R}$ (resp. $\phi : \mathbb{R}^N \to \mathbb{R}^N$) such that

$$\|\phi\|_{L^p_\mu} := \int_{\mathbb{R}^N} \|\phi\|^p \mathrm{d}\mu < \infty.$$

For $\mu \in \Delta\left(\mathbb{R}^N\right)$ and $\psi : \mathbb{R}^N \to \mathbb{R}^N$, Borel measurable with at most a linear growth, we denote by $\psi\sharp\mu \in \Delta\left(\mathbb{R}^N\right)$ the push-forward of μ by ψ, which is also called the image probability measure of μ by ψ. It is defined by

$$\psi\sharp\mu(A) = \mu\left(\psi^{-1}(A)\right) \qquad \forall A \subset \mathbb{R}^N, \text{ Borel measurable}$$

or, equivalently by, for every Borel measurable bounded maps $f : \mathbb{R}^N \to \mathbb{R}$:

$$\int_{\mathbb{R}^N} f \mathrm{d}(\psi\sharp\mu) = \int_{\mathbb{R}^N} f(\psi(x)) \mathrm{d}\mu(x).$$

Recall the definition of W_2 given by the equation (9.1) :

$$W_2^2(\mu,\nu) := \inf_{\gamma \in \Pi(\mu,\nu)} I[\gamma] = \sup_{(\phi,\psi)\in\Xi} J(\phi,\psi) = \inf_{U\sim\mu; V\sim\nu} \mathbb{E}\left[\|U-V\|^2\right].$$

Assume that the support $K \subset \mathbb{R}^n$ of μ and ν is compact. Then since finding the infimum of $I[\gamma]$ over $\Pi(\mu,\nu)$ reduced to find the supremum of $\int_{\mathbb{R}^N\times\mathbb{R}^N} \langle x,y\rangle \mathrm{d}\gamma(x,y)$ over $\Pi(\mu,\nu)$, we can assume (as a consequence of Fenchel-Moreau-Rockafellar duality, see *e.g.* Brezis [25], Theorem 1.10 and 1.11 pages 10 and 11) that Ξ is reduced to the set of functions (ϕ,ϕ^*) such that

$$\phi^*(x) = \inf_{y\in K} \|x-y\|^2 - \phi(y), \;\; \phi = (\phi^*)^* \text{ and } \phi(x_0) = 0, \tag{9.6}$$

for a fixed $x_0 \in K$. The same argument also implies that the supremum in the dual definition is actually a minimum.

Since K is bounded every function in Ξ is at most $2\|K\|$-lipschitzian (where $\|K\|$ is the diameter of K) and Arzela-Ascoli's theorem imply that $(\Xi, \|\|_\infty)$ is compact.

We denote by $\Phi(\mu,\nu)$ the subset of Ξ that maximizes $J(\phi,\phi^*)$ and call its elements Kantorovitch potentials from μ to ν. Any probability measure $\gamma \in \Delta\left(\mathbb{R}^{2N}\right)$ that achieves the minimum is called an optimal plan from μ to ν.

PROXIMAL NORMALS

The usual definition of proximal normals to a set A can be extended to the Wasserstein space in two different ways, depending wether we use the primal definition of W_2 (*i.e.* $W_2^2(\mu,\nu) = \inf_{\gamma\in\Pi(\mu,\nu)} I[\gamma]$) or its dual definition (*i.e.* $W_2^2(\mu,\nu) = \sup_{(\phi,\psi)\in\Xi} J(\phi,\psi)$).

We recall the following result (which proof is a straightforward application of Riesz representation theorem).

Lemma 9.12 (***Cardaliaguet, Quincampoix [27]***)

Let $\mu,\nu \in \Delta(\mathbb{R}^N)$ and $\gamma \in \Pi(\mu,\nu)$. There are two maps $p \in L^2_\mu\left(\mathbb{R}^N,\mathbb{R}^N\right)$ and $q \in L^2_\nu\left(\mathbb{R}^N,\mathbb{R}^N\right)$ such that

$$\int_{\mathbb{R}^N} < \psi(x), p(x) > \mathrm{d}\mu(x) = \int_{\mathbb{R}^{2N}} < \psi(x), x-y > \mathrm{d}\gamma(x,y) \text{ and} \tag{9.7}$$

$$\int_{\mathbb{R}^N} < \psi(x), q(x) > \mathrm{d}\nu(x) = \int_{\mathbb{R}^{2N}} < \psi(y), x-y > \mathrm{d}\gamma(x,y) \tag{9.8}$$

for any Borel measurable map $\psi : \mathbb{R}^N \to \mathbb{R}^N$ with at most a linear growth.

We will denote by $\mathcal{P}(\gamma)$ the set of all $p \in L^2_\mu(\mathbb{R}^N,\mathbb{R}^N)$ satisfying (9.7) and by $\mathcal{Q}(\gamma)$ the set of all $q \in L^2_\mu(\mathbb{R}^N,\mathbb{R}^N)$ satisfying (9.8).

Definition 9.13 (***adapted from As Soulaimani [3]***)

Let $K \subset \mathbb{R}^N$ be a compact set and $\underline{\mu}$ be in A, a nonempty closed subset of $\Delta(K)$. A map $p \in L^2_{\underline{\mu}}(\mathbb{R}^N,\mathbb{R}^N)$ is a proximal gradient normal to A at $\underline{\mu}$ if there exists $\mu \notin A$ with projection $\underline{\mu}$ on A in the Wasserstein distance meaning, i.e.

$$W_2(\mu, A) = \inf_{\theta\in A} W_2(\mu,\theta) = W_2(\mu,\underline{\mu})$$

and γ an optimal plan from $\underline{\mu}$ to μ such that $p \in \mathcal{P}(\gamma)$. The set of proximal gradient normals to A is denoted by $NC^g_A\left(\underline{\mu}\right)$.

Definition 9.14

Let $\underline{\mu}$ be in A, a nonempty closed subset of $\Delta(K)$. A continuous function $\phi : K \to \mathbb{R}$ is a proximal potential normal to A at $\underline{\mu}$ if there exists $\mu \in \Delta(K)$ such that $\underline{\mu}$ is a projection of μ on A and ϕ is a Kantorovitch potential from $\underline{\mu}$ to μ. The set of proximal potential normals to A at $\underline{\mu}$ is denoted by $NC_A^p\left(\underline{\mu}\right)$.

The notation $\mu \ll_0 \boldsymbol{\lambda}$ means that the probability measure μ is absolutely continuous with respect to the Lebesgue measure $\boldsymbol{\lambda}$ and has a strictly positive density.

Proposition 9.15 (*Brenier [24]*)

Let μ and ν in $\Delta(K)$. If $\mu \ll_0 \boldsymbol{\lambda}$ then there exist a unique optimal plan γ and a unique Kantorovitch potential from μ to ν. They satisfy :

$$d\gamma(x,y) = d\mu(x)\delta_{\{x-\nabla\phi(x)\}} \text{ or equivalently } \gamma = (\mathrm{Id} \times (\mathrm{Id} - \nabla\phi))\,\sharp\mu.$$

Apart from the uniqueness of both the optimal plan and the Kantorovitch potential (we have assumed that its value at x_0 is normalized to 0), Proposition 9.15 implies that both definitions of proximal normals are, in some sense, equivalent :

Corollary 9.16

Let A be a compact subset of $\Delta(K)$ and $\underline{\mu} \in A$. If $\mu \ll_0 \boldsymbol{\lambda}$ then

$$\phi \in \mathrm{NC}_A^p(\underline{\mu}) \Longrightarrow \nabla\phi \in \mathrm{NC}_A^g(\underline{\mu}).$$

Finally, we will use the following technical lemmas.

Lemma 9.17

Let K_1 be any compact subset of $\Delta_0(K) = \{\mu \in \Delta(K), \mu \ll_0 \boldsymbol{\lambda}\}$, the set of probability measures absolutely continuous with respect to $\boldsymbol{\lambda}$. Then Φ is singlevalued and uniformly continuous on $K_1 \times \Delta(K)$.

Proof. Brenier's theorem implies that Φ is singlevalued on $\Delta_0(K) \times \Delta(K)$; it is also continuous since any accumulation point of (μ_n, ν_n) that converges to (μ, ν) belongs necessarily to $\Phi(\mu, \nu)$. □

Lemma 9.18 (*see e.g. Dudley [35]*)

Assume that $K \subset \mathbb{R}^N$ is a polytope with nonempty interior. For every $\varepsilon > 0$, there exists $\Delta_\varepsilon(K)$ a compact subset of $\Delta_0(K)$ such that for every $\mu \in \Delta(K)$, $W_2^2(\mu, \Delta_\varepsilon(K)) \leq \varepsilon$.

This lemma is well known and quite classic; its proof is a simple adaptation of Dudley [35], Lemma 11.8.4 (see also Exercice 11.8.7).

For instance, let $\{P_1, \ldots, P_M\}$ be a finite partition of K where each P_m is measurable, has non-empty interior and has a diameter smaller than $\varepsilon/(4\|K\|) > 0$. Define for every $m \leq M$, the probability measure $\boldsymbol{\lambda}_m$ by

$$\boldsymbol{\lambda}_m(A) = \frac{\boldsymbol{\lambda}(A \cap P_m)}{\boldsymbol{\lambda}(P_m)} \in \Delta_0(K), \text{ for every measurable subset } A.$$

then one can consider, for ε small enough, the set

$$\Delta_\varepsilon(K) := \left\{ \left(1 - \frac{\varepsilon}{\|K\|^2}\right) \sum_{m=1}^{M} \alpha_m \boldsymbol{\lambda}_m + \frac{\varepsilon}{\|K\|^2} \boldsymbol{\lambda}; \alpha_m \in [0,1], \sum_{m=1}^{M} \alpha_m = 1 \right\}.$$

Actually, the proof is valid for any compact set that is the closure of its interior. This is, from now on, assumed for X and $\mathcal{X}$.

Approachability in the purely informative game

It remains to describe approachable sets in $\widetilde{\Gamma}$ and to introduce the equivalent notion of a B-set, called $\widetilde{B}$-set :

Definition 9.19

A set $\widetilde{E} \subset \Delta(X \times \mathcal{X})$ is a $\widetilde{B}$-set if for every $\overline{\theta}$ not in $\widetilde{E}$ there exist $\underline{\theta} \in \Pi_{\widetilde{E}}(\overline{\theta})$, $\phi \in \Phi(\underline{\theta}, \overline{\theta})$ and $\mathbf{x}\left(= \mathbf{x}(\overline{\theta})\right) \in \Delta(X)$ such that :

$$\int_{X \times \mathcal{X}} \phi \, \mathrm{d}(\underline{\theta} - \mathbf{x} \otimes \boldsymbol{\xi}) \leq 0, \quad \forall \boldsymbol{\xi} \in \Delta(\mathcal{X}). \tag{9.9}$$

Or stated in terms of proximal normals :

$$\forall \underline{\theta} \in \widetilde{E}, \forall \phi \in \mathrm{NC}^p_{\widetilde{E}}(\underline{\theta}), \exists \mathbf{x} \in \Delta(X), \forall \boldsymbol{\xi} \in \Delta(\mathcal{X}), \int_{X \times \mathcal{X}} \phi \, \mathrm{d}(\underline{\theta} - \mathbf{x} \otimes \boldsymbol{\xi}) \leq 0.$$

The notion of $\widetilde{B}$-set is the extension of Blackwell's [17] definition of B-set to Wasserstein space because of the following main result which is the extension of Theorem 9.11 proved by Blackwell [17] and Spinat [110] with the use of geometrical arguments. We are able to adapt their proofs to our framework thanks to the Hilbertian structure induced by the proximal potential normals in the Wasserstein space.

Theorem 9.20

A set $\widetilde{E} \subset \Delta(X \times \mathcal{X})$ is approachable if and only if it contains a $\widetilde{B}$-set.

Proof. We first prove the sufficient part, *i.e.* a $\widetilde{B}$-set is approachable by Player 1 (and we adapt Blackwell [17]'s proof to our framework).

Let $\varepsilon > 0$ be fixed. For every probability distribution $\mu \in \Delta(X \times \mathcal{X})$, let μ^ε be the probability measure in $\Delta_\varepsilon(X \times \mathcal{X})$ as defined in Lemma 9.18 such that $W_2^2(\mu, \mu^\varepsilon) \leq \varepsilon$. Consider the strategy σ^ε of Player 1 where at stage $n \in \mathbb{N}$ he plays $\mathbf{x}(\bar{\theta}^\varepsilon_{n-1})$ given by the definition of a $\widetilde{B}$-set. Then, if we denote by $\underline{\theta}^\varepsilon_{n-1}$ the projection of $\bar{\theta}^\varepsilon_{n-1}$ over $\widetilde{E}$ and let $w_n = W_2^2\left(\bar{\theta}^\varepsilon_n, \widetilde{E}\right)$:

$$\begin{aligned}
w_n = W_2^2\left(\bar{\theta}^\varepsilon_n, \widetilde{E}\right) &\leq W_2^2\left(\bar{\theta}^\varepsilon_n, \underline{\theta}^\varepsilon_{n-1}\right) = W_2\left(\underline{\theta}^\varepsilon_{n-1}, \frac{n-1}{n}\bar{\theta}^\varepsilon_{n-1} + \frac{\theta^\varepsilon_n}{n}\right) \\
&= \sup_{\phi \in \Xi} \int_{X \times \mathcal{X}} \phi \, \mathrm{d}\underline{\theta}^\varepsilon_{n-1} + \int_{X \times \mathcal{X}} \phi^* \mathrm{d}\left(\frac{n-1}{n}\bar{\theta}^\varepsilon_{n-1} + \frac{\theta^\varepsilon_n}{n}\right) \\
&= \int_{X \times \mathcal{X}} \phi_n \, \mathrm{d}\underline{\theta}^\varepsilon_{n-1} + \int_{X \times \mathcal{X}} \phi^*_n \mathrm{d}\left(\frac{n-1}{n}\bar{\theta}^\varepsilon_{n-1} + \frac{\theta^\varepsilon_n}{n}\right) \\
&\leq \frac{n-1}{n} w_{n-1} + \frac{1}{n}\left(\int_{X \times \mathcal{X}} \phi_n \mathrm{d}\underline{\theta}^\varepsilon_{n-1} \int_{X \times \mathcal{X}} \phi^*_n \mathrm{d}\theta^\varepsilon_n\right)
\end{aligned}$$

where ϕ_n is the optimal Kantorovitch potential from $\underline{\theta}^\varepsilon_{n-1}$ to $\frac{n-1}{n}\bar{\theta}^\varepsilon_{n-1} + \frac{\theta^\varepsilon_n}{n}$. Let us denote by ϕ_0 the optimal Kantorovitch potential from $\underline{\theta}^\varepsilon_{n-1}$ to $\bar{\theta}^\varepsilon_{n-1}$ and by $\omega^\varepsilon(\cdot)$ the modulus of continuity of Φ restricted to the compact set $\widetilde{E} \times \Delta_\varepsilon(X \times \mathcal{X})$.

The definition of W_2^2 implies that

$$W_2^2\left(\bar{\theta}^\varepsilon_{n-1}, \frac{n-1}{n}\bar{\theta}^\varepsilon_{n-1} + \frac{\theta^\varepsilon_n}{n}\right) \leq \frac{1}{n} W_2^2(\bar{\theta}^\varepsilon_{n-1}, \theta^\varepsilon_n) \leq \frac{4\|X\|^2}{n},$$

therefore $\|\phi_0 - \phi_n\|_\infty \leq \omega^\varepsilon(2\|X\|/\sqrt{n})$ and

$$w_n \leq \frac{n-1}{n} w_{n-1} + \frac{1}{n}\left(\int_{X \times \mathcal{X}} \phi_0 \mathrm{d}\underline{\theta}^\varepsilon_{n-1} + \int_{X \times \mathcal{X}} \phi^*_0 \mathrm{d}\theta^\varepsilon_n\right) + \frac{2}{n}\omega\left(\frac{2\|X\|}{\sqrt{n}}\right).$$

Recall that θ^ε_n is such that that $\int_{X \times \mathcal{X}} \phi_0 \mathrm{d}\theta_n + \int_{X \times \mathcal{X}} \phi^*_0 \mathrm{d}\theta^\varepsilon_n \leq W_2^2(\theta_n, \theta^\varepsilon_n,) \leq \varepsilon$, therefore

$$w_n \leq \frac{n-1}{n} w_{n-1} + \frac{1}{n}\left(\int_{X \times \mathcal{X}} \phi_0 \mathrm{d}\underline{\theta}^\varepsilon_{n-1} - \int_{X \times \mathcal{X}} \phi_0 \mathrm{d}\theta_n\right) + \frac{2}{n}\omega\left(\frac{2\|X\|}{\sqrt{n}}\right) + \frac{\varepsilon}{n}.$$

Since $\widetilde{E}$ is a $\widetilde{B}$-set and because of the choice of $\mathbf{x}_n \in \Delta(X)$, for every $\boldsymbol{\xi}_n \in \Delta(\mathcal{X})$, $\int_{X \times \mathcal{X}} \phi_0 \mathrm{d}\left(\underline{\theta}^\varepsilon_{n-1} - \mathbf{x}_n \otimes \boldsymbol{\xi}_n\right) \leq 0$, thus

$$w_n \leq \frac{n-1}{n} w_{n-1} + \frac{2}{n}\omega^\varepsilon\left(\frac{2\|X\|}{\sqrt{n}}\right) + \frac{\varepsilon}{n}$$

and this yields, by induction, that

$$W_2^2\left(\bar{\theta}^\varepsilon_n, \widetilde{E}\right) \leq \frac{1}{n+1} W_2^2\left(\overline{\theta^\varepsilon}_1, \widetilde{E}\right) + \frac{2}{n+1}\sum_{m=1}^{n} \omega^\varepsilon\left(\frac{2\|X\|}{\sqrt{k}}\right) + \varepsilon.$$

Hence w_n converges to ε (since $\omega^\varepsilon(2\|X\|/\sqrt{k})$ converges to 0 when k goes to infinity). Since $W_2\left(\overline{\theta}_n, \widetilde{E}\right) \leq W_2\left(\overline{\theta}_n^\varepsilon, \widetilde{E}\right) + \varepsilon$, then $\widetilde{E}$ is approachable by Player 1.

We now turn to the necessary part, *i.e.* any approachable set contains a $\widetilde{B}$-set.

Definition 9.21

A point $\theta \in \Delta(X \times \mathcal{X})$ is δ-secondary for $\widetilde{E}$ if there exists a corresponding couple : a point $\boldsymbol{\xi} \in \Delta(\mathcal{X})$ and a continuous function $\lambda : \Delta(X) \to (0,1]$ such that $\min_{\mathbf{x}\in\Delta(X)} W_2\left(\lambda(\mathbf{x})\theta + (1-\lambda(\mathbf{x}))\mathbf{x}\otimes\boldsymbol{\xi}, \widetilde{E}\right) \geq \delta$. A point θ is secondary to $\widetilde{E}$ if there exists $\delta > 0$ such that x is δ-secondary to $\widetilde{E}$.

We denote by $\mathcal{P}(\widetilde{E}) \subset \widetilde{E}$ the subset of primary point to $\widetilde{E}$ (i.e. points of $\widetilde{E}$ that are not secondary).

The fact that an approachable set contains a $\widetilde{B}$-set is a consequence of the following lemma, adapted from Spinat [110] :

Lemma 9.22 (*Spinat [110]*)

i) Any approachable compact set contains a minimal approachable set ;
ii) A minimal approachable set is a fixed point of $\mathcal{P}$;
iii) A fixed point of $\mathcal{P}$ is a $\widetilde{B}$-set.

Proof. *i)* Let $\mathcal{B} = \left\{ \widetilde{B} \subset \widetilde{E} \middle| \widetilde{B} \text{ is an approachable compact set} \right\}$ be a nonempty family ordered by inclusion. Every fully ordered subset of $\mathcal{B}$ has a minorant $\underline{B}$ (the intersection of every elements of the subset) that belongs to $\mathcal{B}$ since it is an approachable compact subset of $\widetilde{E}$. Thus Zorn's lemma yields that $\mathcal{B}$ contains at least one minimal element.

ii) We claim that if $\widetilde{E}$ is approachable then so is $\mathcal{P}(\widetilde{E})$, hence a minimal approachable set is necessarily a fixed point of $\mathcal{P}$. Indeed, if θ is δ-secondary there exists an open neighborhood V of θ such that every point of V is $\delta/2$-secondary, because of the continuity of W_2. Hence $\mathcal{P}(\widetilde{E})$ is a compact subset of $\widetilde{E}$.

Let θ_0 be a δ-secondary point of an approachable set $\widetilde{E}$ and $\boldsymbol{\xi}, \lambda$ the associated couple given in Definition 9.21. Let $\varepsilon < \delta/4$ and consider σ a strategy of Player 1 that ensures that $\overline{\theta}_n$ is, after some stage $N \in \mathbb{N}$, closer than ε to $\widetilde{E}$. We will show that $\overline{\theta}_n$ must be close to θ_0 only a finite number of times ; so Player 1 can approach $\widetilde{E}\backslash\{\theta\}$. Indeed, assume that there exists a stage $m \in \mathbb{N}$ such that $W_2(\overline{\theta}_m, \theta_0) \leq \delta/4$ and consider the strategy of Player 2 that consists in playing repeatedly $\boldsymbol{\xi}$ from this stage on. It is clear that (if m is big enough) after some stage $\overline{\theta}$ will be $\delta/2$-closed to $\lambda\left(\overline{\mathbf{x}}_{n,m}\right)\theta_0 + \left(1-\lambda\left(\overline{\mathbf{x}}_{n,m}\right)\right)\overline{\mathbf{x}}_{n,m}\otimes\boldsymbol{\xi}$ where $\overline{\mathbf{x}}_{n,m}$ is the average action played by Player 1 between stage m and $m+n$. Therefore $W_2\left(\overline{\theta}_n, \widetilde{E}\right) \geq \delta/2 > \varepsilon$ and since

$W_2(\overline{\theta}_n, \theta_0)$ can be bigger than $\delta/4$ only a finite number of times, the strategy of Player 1 approaches $\widetilde{E}\backslash\{\theta\}$. Since this is true for any secondary point, Player 1 can approach $\mathcal{P}(\widetilde{E})$.

iii) Assume that $\widetilde{E}$ is not a $\widetilde{B}$-set : there exists $\overline{\theta} \notin \widetilde{E}$ such that for any projection $\underline{\theta} \in \Pi_{\widetilde{E}}(\overline{\theta})$, any $\phi \in \Phi(\underline{\theta}, \overline{\theta})$ and any $\mathbf{x} \in \Delta(X)$, there exists $\boldsymbol{\xi} \in \Delta(\mathcal{X})$ such that $\int_{X\times\mathcal{X}} \phi \mathrm{d}(\underline{\theta} - \mathbf{x} \otimes \boldsymbol{\xi}) > 0$. This last expression is linear both in $\mathbf{x}$ and $\boldsymbol{\xi}$, so Von Neumann's minmax theorem imply that there exists $\boldsymbol{\xi}(= \boldsymbol{\xi}(\underline{\theta}, \phi))$ and δ such that $\int_{X\times\mathcal{X}} \phi \mathrm{d}(\underline{\theta} - \mathbf{x} \otimes \boldsymbol{\xi}) \geq \delta > 0$, for every $\mathbf{x} \in \Delta(X)$.

We can assume that $\overline{\theta} \ll_0 \boldsymbol{\lambda}$. Otherwise, let $(\overline{\theta}_n)_{n\in\mathbb{N}}$ be a sequence of such measures that converges to $\overline{\theta}$, $(\underline{\theta}_n)_{n\in\mathbb{N}}$ a sequence of projection of $\overline{\theta}_n$ onto $\widetilde{E}$ and $\phi_0^n \in \Phi(\overline{\theta}_n, \underline{\theta}_n)$. Up to two extractions, we can assume that $\underline{\theta}_n$ converges to $\underline{\theta}_0$ a projection of $\overline{\theta}$ and ϕ_0^n converges to $\phi_0 \in \Phi(\overline{\theta}, \underline{\theta}_0)$. Therefore, for n big enough and for every $\mathbf{x} \in \Delta(X)$,

$$0 < \frac{\delta}{2} \leq \int_{X\times\mathcal{X}} \phi_0^n \mathrm{d}\left(\underline{\theta}_n - \mathbf{x} \otimes \boldsymbol{\xi}(\underline{\theta}_0, \phi_0)\right)$$

since the right member converges to $\int_{X\times\mathcal{X}} \phi_0 \mathrm{d}\left(\underline{\theta}_0 - \mathbf{x} \otimes \boldsymbol{\xi}(\underline{\theta}_0, \phi_0)\right) \geq \delta$.

For every $\lambda \in [0,1]$ and $\mathbf{x} \in \Delta(X)$, we denote by $\phi_{\lambda,\mathbf{x}}$ the unique (since we assumed that $\overline{\theta} \ll_0 \boldsymbol{\lambda}$) Kantorovitch potential such that :

$$\begin{aligned} W_2^2\left((1-\lambda)\underline{\theta} + \lambda\mathbf{x} \otimes \boldsymbol{\xi}, \overline{\theta}\right) &= \int_{X\times\mathcal{X}} \phi_{\lambda,\mathbf{x}} \mathrm{d}\left((1-\lambda)\underline{\theta} + \lambda\mathbf{x} \otimes \boldsymbol{\xi}\right) + \int_{X\times\mathcal{X}} \phi^*_{\lambda,\mathbf{x}} \mathrm{d}\overline{\theta} \\ &= \int_{X\times\mathcal{X}} \phi_{\lambda,\mathbf{x}} \mathrm{d}\underline{\theta} + \int_{X\times\mathcal{X}} \phi^*_{\lambda,\mathbf{x}} \mathrm{d}\overline{\theta} - \lambda \int_{X\times\mathcal{X}} \phi_{\lambda,\mathbf{x}} \mathrm{d}(\underline{\theta} - \mathbf{x} \otimes \boldsymbol{\xi}). \end{aligned}$$

Since $(\lambda, \mathbf{x}) \mapsto \phi_{\lambda,\mathbf{x}}$ is continuous, $\phi_{\lambda,\mathbf{x}}$ converges to ϕ_0, for every $\mathbf{x} \in \Delta(X)$ which is compact. Hence there exists $\underline{\lambda} \in (0,1]$ such that :

$$\left| \int_{X\times\mathcal{X}} (\phi_{\lambda,\mathbf{x}} - \phi_0) \, \mathrm{d}(\underline{\theta} - \mathbf{x} \otimes \boldsymbol{\xi}) \right| \leq \delta/4, \quad \forall \lambda \leq \underline{\lambda}, \forall \mathbf{x} \in \Delta(X).$$

Therefore, one has $W_2^2\left(\overline{\theta}, (1-\underline{\lambda})\underline{\theta} + \underline{\lambda}\mathbf{x} \otimes \boldsymbol{\xi}\right) \leq W_2^2(\overline{\theta}, \underline{\theta}) - \underline{\lambda}\frac{\delta}{4}$ and

$$W_2\left(\overline{\theta}, (1-\underline{\lambda})\underline{\theta} + \underline{\lambda}\mathbf{x} \otimes \boldsymbol{\xi}\right) \leq W_2(\overline{\theta}, \underline{\theta}) - \frac{\underline{\lambda}\delta}{8W_2(\overline{\theta}, \underline{\theta})} := W_2(\overline{\theta}, \underline{\theta}) - \eta$$

which implies that $W_2\left((1-\underline{\lambda})\underline{\theta} + \underline{\lambda}\mathbf{x} \otimes \boldsymbol{\xi}, \widetilde{E}\right) \geq \eta$ and $\underline{\theta}$ is η-secondary to $\widetilde{E}$.

Consequently, a fixed point of $\mathcal{P}$, *i.e.* a set without any secondary point, is necessary a $\widetilde{B}$-set. □

This concludes the necessary part. Hence we have proved that a set is approachable in $\widetilde{\Gamma}$ if and only if it contains a $\widetilde{B}$-set. □

9.3 Characterization of convex approachable sets

In the case of a convex set C, Blackwell [17] provided a complete characterization in Γ^f. A convex set C is approachable by Player 1 if and only if for every $y \in \Delta(J)$, $P^1(y) \cap C \neq \emptyset$, or equivalently if and only if

$$\forall y \in \Delta(J), \exists x \in \Delta(I), \rho(x, y) \in C.$$

Perchet [91] provided a characterization of approachable convex sets in Γ^p. A set C is approachable if and only if

$$\forall \mu \in \Delta(S)^I, \exists x \in \Delta(I), P(x, \mu) \subset C.$$

Those two characterizations can be rephrased with the notations we introduced in :

Proposition 9.23 (*Blackwell [17], Perchet [91]*)

A convex set C is approachable by Player 1 if and only if :

$$\forall \xi \in \mathcal{X}, \exists x \in X, P(x, \xi) \subset C. \tag{9.10}$$

In $\widetilde{\Gamma}$, the characterization is the same :

Theorem 9.24

A convex set $\widetilde{C}$ is approachable if and only if for every $\boldsymbol{\xi}$ there exists $\mathbf{x}$ such that $\theta(\mathbf{x}, \boldsymbol{\xi}) \in \widetilde{C}$.

Proof. Once again, we will follow the idea of Blackwell. Assume that there exists $\boldsymbol{\xi}$ such that, for every $\mathbf{x} \in \Delta(X)$, $\theta(\mathbf{x}, \boldsymbol{\xi}) \notin \widetilde{C}$. The application $\mathbf{x} \mapsto W_2\left(\mathbf{x}, \widetilde{C}\right)$ is continuous on the compact set $\Delta(X)$, therefore there exists $\delta > 0$ such that $W_2^2\left(\theta(\mathbf{x}, \boldsymbol{\xi}), \widetilde{C}\right) \geq \delta$.

Consider the strategy of Player 2 that consists of playing $\boldsymbol{\xi}$ at every stages, then $\theta_n = \mathbf{x}_n \otimes \boldsymbol{\xi}$, $\overline{\theta}_n = \overline{\mathbf{x}}_n \otimes \boldsymbol{\xi} = \theta(\overline{\mathbf{x}}_n, \boldsymbol{\xi})$ and $W_2^2\left(\overline{\theta}_n, \widetilde{C}\right) \geq \delta > 0$. Therefore $\widetilde{C}$ is not approachable by Player 1.

Reciprocally, assume that for every $\boldsymbol{\xi} \in \Delta(\mathcal{X})$ there exists $\mathbf{x} \in \Delta(X)$ such that $\theta(\mathbf{x}, \boldsymbol{\xi}) \in \widetilde{C}$. We claim that this implies that $\widetilde{C}$ is a $\widetilde{B}$-set.

Let $\overline{\theta}$ be a probability measure that does not belong to $\widetilde{C}$ and assume (for the moment) that $\overline{\theta} \ll_0 \boldsymbol{\lambda}$. Denote by $\underline{\theta} \in \widetilde{C}$ any of its projection then, by definition of the projection and convexity of $\widetilde{C}$, $W_2(\overline{\theta}, \underline{\theta})$ is smaller than $W_2\left(\overline{\theta}, (1-\lambda)\underline{\theta} + \lambda \mathbf{x} \otimes \boldsymbol{\xi}\right)$

which is equal to :

$$\begin{aligned}
W_2^2\left((1-\lambda)\underline{\theta}+\lambda\mathbf{x}\otimes\boldsymbol{\xi},\overline{\theta}\right) &= \sup_{\phi\in\Xi}\int_{X\times\mathcal{X}}\phi\mathrm{d}\left((1-\lambda)\underline{\theta}+\lambda\mathbf{x}\otimes\boldsymbol{\xi}\right)+\int_{X\times\mathcal{X}}\phi^*\mathrm{d}\overline{\theta}\\
&= \sup_{\phi\in\Xi}\int_{X\times\mathcal{X}}\phi\mathrm{d}\underline{\theta}+\int_{X\times\mathcal{X}}\phi^*\mathrm{d}\overline{\theta}-\lambda\int_{X\times\mathcal{X}}\phi\mathrm{d}(\underline{\theta}-\mathbf{x}\otimes\boldsymbol{\xi})\\
&= \int_{X\times\mathcal{X}}\phi_\lambda\mathrm{d}\underline{\theta}+\int_{X\times\mathcal{X}}\phi_\lambda^*\mathrm{d}\overline{\theta}-\lambda\int_{X\times\mathcal{X}}\phi_\lambda\mathrm{d}(\underline{\theta}-\mathbf{x}\otimes\boldsymbol{\xi})\\
&\leq \int_{X\times\mathcal{X}}\phi_0\mathrm{d}\underline{\theta}+\int_{X\times\mathcal{X}}\phi_0^*\mathrm{d}\overline{\theta}-\lambda\int_{X\times\mathcal{X}}\phi_\lambda\mathrm{d}(\underline{\theta}-\mathbf{x}\otimes\boldsymbol{\xi})\\
&= W_2^2(\overline{\theta},\underline{\theta})-\lambda\int_{X\times\mathcal{X}}\phi_\lambda\mathrm{d}(\underline{\theta}-\mathbf{x}\otimes\boldsymbol{\xi})
\end{aligned}$$

where ϕ_λ (resp. ϕ_0) is the unique potential from $(1-\lambda)\underline{\theta}+\lambda\mathbf{x}\otimes\boldsymbol{\xi}$ (resp. $\underline{\theta}$) to $\overline{\theta}$. Therefore, for every $\lambda>0$, $\lambda\int_{X\times\mathcal{X}}\phi_\lambda\mathrm{d}(\underline{\theta}-\mathbf{x}\otimes\boldsymbol{\xi})\leq 0$. Dividing by $\lambda>0$ yields :

$$\int_{X\times\mathcal{X}}\phi_\lambda\mathrm{d}(\underline{\theta}-\mathbf{x}\otimes\boldsymbol{\xi})\leq 0,\quad\forall\lambda>0.$$

Since $(1-\lambda)\underline{\theta}+\lambda\mathbf{x}\otimes\boldsymbol{\xi}$ converges to $\underline{\theta}$, any accumulation point of $(\phi_\lambda)_{\lambda>0}$ has to belong (for every $\mathbf{x}$ and $\boldsymbol{\xi}$) to $\Phi(\underline{\theta},\overline{\theta})=\{\phi_0\}$. Stated differently, given $\phi_0\in\Phi(\underline{\theta},\overline{\theta})$, one has :

$$\max_{\boldsymbol{\xi}\in\Delta(\mathcal{X})}\min_{\mathbf{x}\in\Delta(X)}g_{\phi_0}(\mathbf{x},\boldsymbol{\xi}):=\max_{\boldsymbol{\xi}\in\Delta(\mathcal{X})}\min_{\mathbf{x}\in\Delta(X)}\int_{X\times\mathcal{X}}\phi_0\mathrm{d}(\underline{\theta}-\mathbf{x}\otimes\boldsymbol{\xi})\leq 0.$$

The function g_{ϕ_0} is linear in both of its variable, so Sion's Theorem implies that

$$\max_{\boldsymbol{\xi}\in\Delta(\mathcal{X})}\min_{\mathbf{x}\in\Delta(X)}g_{\phi_0}(\mathbf{x},\boldsymbol{\xi})=\min_{\mathbf{x}\in\Delta(X)}\max_{\boldsymbol{\xi}\in\Delta(\mathcal{X})}g_{\phi_0}(\mathbf{x},\boldsymbol{\xi})$$

hence $\widetilde{C}$ is a $\widetilde{B}$-set.

Assume now that $\overline{\theta}$ is not absolutely continuous with respect to $\boldsymbol{\lambda}$ with positive density. Let $(\overline{\theta}_n)_{n\in\mathbb{N}}$ be a sequence of such measures that converges to $\overline{\theta}$, $(\underline{\theta}_n)_{n\in\mathbb{N}}$ a sequence of projections, and $\phi_n\in\Phi(\underline{\theta}_n,\overline{\theta}_n)$. Up to two extraction, we can assume that $\underline{\theta}_n$ and ϕ_n converge respectively to $\underline{\theta}$ and ϕ_0. Necessarily, $\underline{\theta}$ is a projection of $\overline{\theta}$ onto $\widetilde{C}$ and ϕ_0 belongs to $\Phi(\underline{\theta},\overline{\theta})$. Therefore :

$$\int_{X\times\mathcal{X}}\phi_0\mathrm{d}(\underline{\theta}-\mathbf{x}\otimes\boldsymbol{\xi})=\lim_{n\to\infty}\int_{X\times\mathcal{X}}\phi_n\mathrm{d}(\underline{\theta}-\mathbf{x}\otimes\boldsymbol{\xi})\leq 0$$

and $\widetilde{C}$ is a $\widetilde{B}$-set. □

Let us go back and quickly prove Proposition 9.23 which characterizes approachable convex sets in Γ^p as a consequence of Theorem 9.24 :

Proof of Proposition 9.23 : A convex subset C of $\mathbb{R}^k$ is approachable (see Theorem 9.4) if and only if the convex set $\widetilde{\rho}(C) \subset \Delta(X \times \mathcal{X})$ is approachable in $\widetilde{\Gamma}$, therefore (see Theorem 9.24) if and only if for every $\boldsymbol{\xi} \in \Delta(\mathcal{X})$, there exists $\mathbf{x} \in \Delta(X)$ such that $\mathbf{x} \otimes \boldsymbol{\xi} \in \widetilde{\rho}(C)$. For every $\mathbf{x} \in \Delta(X)$ and every $\boldsymbol{\xi} \in \Delta(\mathcal{X})$, let us denote by $\mathbb{E}_{\mathbf{x}}[x] \in \Delta(I)$ and $\mathbb{E}_{\boldsymbol{\xi}}[\xi] \in \mathcal{X}$ their expectations; then the definition of ρ and the convexity of P (in both of its variables) imply that :

$$\rho(\mathbf{x} \otimes \boldsymbol{\xi}) = \int_{X \times \mathcal{X}} P(x, \xi) \mathrm{d}\mathbf{x} \otimes \boldsymbol{\xi} \subset P\left(\mathbb{E}_{\mathbf{x}}[x], \mathbb{E}_{\boldsymbol{\xi}}[\xi]\right).$$

Therefore if for every $\boldsymbol{\xi} \in \Delta(\mathcal{X})$, there exists $\mathbf{x} \in \Delta(X)$ such that $\mathbf{x} \otimes \boldsymbol{\xi} \in \widetilde{\rho}(C)$ then for every $\mu \in \mathcal{S}$ (we recall that $\mathcal{X} = \mathcal{S}$), there exists $x \in \Delta(I)$ such that $P(x, \mu) \subset C$. In conclusion, a convex set $C\mathbb{R}^k$ is approachable in Γ^p if and only if

$$\forall \mu \in \mathcal{S}, \exists x \in \Delta(I), P(x, \mu) \subset C.$$

If μ does not belong to $\mathcal{S}$, then $P(x, \mu) = \emptyset$, therefore we can replace $\mathcal{S}$ by $\Delta(S)^I$ in the condition above. □

This also clearly explains why there exist convex sets that are neither approachable, nor excludable in Γ^p, see Perchet [91] which cannot occur in Γ^f, see Blackwell [17] : it simply due to the fact that $\widetilde{\rho}(C)$ can be empty.

9.4 Convex games

We restrict ourselves in this section to the particular class of games called *convex games* which have the following property : for every $q \in \Delta(X \times \mathcal{X})$:

$$\rho(q) = \int_{X \times \mathcal{X}} P(x, \xi) \mathrm{d}q(x, \xi) \subset P\left(\mathbb{E}_q\left[x\right], \mathbb{E}_x\left[\xi\right]\right).$$

This reduces in Γ^f to $\rho(q) = \rho\left(\mathbb{E}_q[x], \mathbb{E}_q[\xi]\right)$.

For example, the following game where the payoffs of player 1 are given by the matrix on the left and signals by the matrix on the right, is convex.

	L	C	R
T	(0,-1)	(1,-2)	(2,-4)
B	(1,0)	(2,-1)	(3,-3)

	L	C	R
T	a	a	b
B	a	a	b

In this game $I = \{T, B\}$, $J = \{L, C, R\}$ and $S = \{a, b\}$. If player 1 receives the signal a, he does not know whether Player 2 used the action L or C.

Let us introduce the notion of displacement interpolation (and also the displacement convexity) that will play the role of classic linear interpolation and convexity :

Definition 9.25 (***McCann [83]***)

Given μ, $\nu \in \Delta^2(\mathbb{R}^N)$ and $t \in [0,1]$, a displacement interpolation between μ and ν at time t is defined by $\widehat{\mu}_t = \sigma_t \sharp \gamma$, where $\gamma \in \Pi(\mu,\nu)$ is an optimal plan and $\sigma_t(x,y) = (1-t)x + ty$.

A set $\widehat{C}$ is displacement convex if for every $\mu, \nu \in \widehat{C}$, every $t \in [0,1]$ and every optimal plan $\gamma \in \Pi(\mu,\nu)$, $\sigma_t \sharp \gamma \in \widehat{C}$.

We define a new game $\widehat{\Gamma}$ as follows. At stage $n \in \mathbb{N}$, Player 1 (resp. Player 2) chooses $x_n \in X$ (resp. $y_n \in \mathcal{X}$) and the payoff is $\theta_n = \delta_{x_n} \otimes \delta_{y_n} = \delta_{(x_n,y_n)} \in \Delta(X \times \mathcal{X})$. We do not consider average payoffs in the usual sense (as in $\widetilde{\Gamma}$) but we define a sequence of recursive interpolation by :

$$\widehat{\theta}_{n+1} = \sigma_{\frac{1}{n+1}} \sharp \gamma_{n+1}, \text{ where } \gamma_{n+1} \in \Pi(\widehat{\theta}_n, \theta_{n+1}) \text{ is an optimal plan.}$$

By induction, this implies that $\widehat{\theta}_n = \delta_{\overline{x}_n} \otimes \delta_{\overline{y}_n}$. Indeed, $\widehat{\theta}_1 = \delta_{x_1} \otimes \delta_{y_1}$ and $\theta_2 = \delta_{x_2} \otimes \delta_{y_2}$ therefore :

$$\gamma_2 = (\delta_{x_1} \otimes \delta_{y_1}) \otimes (\delta_{x_1} \otimes \delta_{y_1}) \text{ and } \widehat{\theta}_2 = \sigma_{\frac{1}{2}} \sharp \gamma_2 = \delta_{\frac{x_1+x_2}{2}} \otimes \delta_{\frac{y_1+y_2}{2}}.$$

Definition 9.26

A closed set $\widehat{E} \subset \Delta(X \times \mathcal{X})$ is displacement approachable by Player 1 if for every $\varepsilon > 0$ there exist a strategy σ of Player 1 and $N \in \mathbb{N}$ such that for every strategy τ of Player 2 :

$$\forall n \geq N, W_2\left(\widehat{\theta}_n, \widehat{E}\right) \leq \varepsilon.$$

Consider any set $E \subset \mathbb{R}^d$ and assume that $\widetilde{\rho}(E)$ is displacement approachable by Player 1. Since $\widehat{\theta}_n = \delta_{\overline{x}_n} \otimes \delta_{\overline{y}_n}$, $\rho(\overline{\theta}_n) \subset \rho(\widehat{\theta}_n)$ and $\widetilde{\rho}(E)$ is also approachable (in $\widetilde{\Gamma}$). The use of displacement approachability provides explicit and optimal bounds (see Theorem 9.28 below). This is the reason we investigate this special case.

In this framework, we use proximal gradient normals to define a $\widehat{B}$-set.

Definition 9.27

A closed subset $\widehat{E} \subset \Delta(X \times \mathcal{X})$ is a $\widehat{B}$-set if for every θ not in $\widehat{E}$ there exist $\underline{\theta} \in \Pi_{\widehat{E}}(\mu)$, $\overline{p} \in NP^g_{\widehat{E}}(\underline{\theta})$ and $x = x(\theta) \in \Delta(X)$ such that for every $y \in \mathcal{X}$, there exists an optimal plan $\gamma(x,y) \in \Pi(\underline{\theta}, \delta_x \otimes \delta_y)$ and $p(x,y) \in \mathcal{P}(\gamma(x,y))$ such that :

$$\langle \overline{p}, p(x,y) \rangle_{L_2(\underline{\theta})} \leq 0.$$

This notion of $\widehat{B}$-set is the extension of Blackwell's one to $\widehat{\Gamma}$ because of the following Theorems 9.28 and 9.29.

Theorem 9.28

A set $\widehat{E}$ is approachable in $\widehat{\Gamma}$ if and only if it contains a $\widehat{B}$-set. Given a $\widehat{B}$-set, the strategy described by $x_{n+1} = x(\widehat{\theta}_n)$ ensures that there exists $K > 0$ such that $W_2\left(\widehat{\theta}_n, \widehat{E}\right) \leq K/\sqrt{n}$.

Proof. Assume that Player 1 plays, at stage n, $x_n = x(\widehat{\theta}_{n-1})$ and denote by $\theta_n = \delta_{x_n} \otimes \delta_{y_n}$ the outcome at stage n. Recall that, for every $n \in \mathbb{N}$, the displacement average outcome is $\widehat{\theta}_n = \delta_{\overline{x}_n} \otimes \delta_{\overline{y}_n} = \delta_{(\overline{x}_n, \overline{y}_n)}$.

If we denote by $\underline{\theta}_n \in \widehat{E}$ the projection of $\widehat{\theta}_n$ on $\widehat{E}$, then the optimal plan from $\underline{\theta}_n$ to $\widehat{\theta}_n$ is $\underline{\theta}_n \otimes \widehat{\theta}_n$. So the proximal normal $\overline{p}_n \in NP^g_{\widehat{E}}(\underline{\theta}_n)$ is defined by $\overline{p}_n(z) = z - (\overline{x}_n, \overline{y}_n)$. Similarly, we define $p_{n+1}(z) = z - (x_{n+1}, y_{n+1})$ so that the assumption that $\widehat{E}$ is a $\widehat{B}$-set (along with the choice of x_{n+1}) ensures that $\langle \overline{p}_n, p_{n+1} \rangle_{\underline{\theta}_n} \leq 0$.

As usual, we note that $W_2^2(\widehat{\theta}_{n+1}, \widehat{E}) \leq W_2^2(\widehat{\theta}_{n+1}, \underline{\theta}_n)$ which satisfies :

$$\begin{aligned}
W_2^2\left(\widehat{\theta}_{n+1}, \underline{\theta}_n\right) &= \int_{(X\times\mathcal{X})^2} \|x - z\|^2 \, \mathrm{d}\widehat{\theta}_{n+1} \otimes \underline{\theta}_n = \int_{X\times\mathcal{X}} \left\| \left(\overline{x}_{n+1}, \overline{y}_{n+1}\right) - z \right\|^2 \mathrm{d}\underline{\theta}_n(z) \\
&= \int_{X\times\mathcal{X}} \left\| \frac{n}{n+1} (\overline{x}_n, \overline{y}_n) + \frac{1}{n+1} (x_{n+1}, y_{n+1}) - z \right\|^2 \mathrm{d}\underline{\theta}_n(z) \\
&= \left(\frac{n}{n+1}\right)^2 \int_{X\times\mathcal{X}} \|(\overline{x}_n, \overline{y}_n) - z\|^2 \, \mathrm{d}\underline{\theta}_n(z) \\
&+ \left(\frac{1}{n+1}\right)^2 \int_{X\times\mathcal{X}} \|(x_{n+1}, y_{n+1}) - z\|^2 \, \mathrm{d}\underline{\theta}_n(z) \\
&+ 2\frac{n}{(n+1)^2} \int_{X\times\mathcal{X}} \langle (\overline{x}_n, \overline{y}_n) - z, (x_{n+1}, y_{n+1} - z) \rangle \, \mathrm{d}\underline{\theta}_n(z)
\end{aligned}$$

Therefore,

$$\begin{aligned}
W_2^2\left(\widehat{\theta}_{n+1}, \underline{\theta}_n\right) &= \left(\frac{n}{n+1}\right)^2 W_2^2\left(\widehat{\theta}_n, \underline{\theta}_n\right) + \left(\frac{1}{n+1}\right)^2 W_2^2\left(\theta_{n+1}, \underline{\theta}_n\right) \\
&+ 2\frac{n}{(n+1)^2} \langle \overline{p}_n, p_{n+1} \rangle_{\underline{\theta}_n} \leq \left(\frac{n}{n+1}\right)^2 W_2^2\left(\widehat{\theta}_n, \widehat{E}\right) + \left(\frac{K}{n+1}\right)^2.
\end{aligned}$$

We conclude by induction over $n \in \mathbb{N}$.

We sketch the proof of the necessary part. Conclusions of Lemma 9.22 hold in $\widehat{\Gamma}$ and the proof of the first two points are identical. Hence it remains to prove the third

point, *i.e.* that a set which is not a $\widehat{B}$-set has a secondary point. Let $\overline{\theta}$ be not in $\widehat{E}$, $\underline{\theta}$ one of its projection on $\widehat{E}$, and $\overline{p} \in \mathrm{NC}^g_{\widehat{E}}(\underline{\theta})$ the associated proximal normals such that :

$$\forall x \in X, \exists y \in \mathcal{X}, \langle \underline{p}, p(x,y)\rangle_{\underline{\theta}} = \int_{X\times\mathcal{X}} \langle \overline{p}(z), z-(x,y)\rangle \mathrm{d}\underline{\theta} > 0.$$

By linearity, there exists $\delta > 0$ and $y \in \mathcal{X}$ such that for every $x \in X$, $\langle \underline{p}, p(x,y)\rangle_{\underline{\theta}} \geq \delta$. If we denote by $\theta_\lambda = (\mathrm{Id}, \sigma_\lambda)\sharp\underline{\theta}_n \otimes \theta_{n+1}$ then using the same argument as in the proof of Lemma 9.30, we show that

$$W_2(\overline{\theta}, \theta_\lambda) \leq W_2(\overline{\theta}, \underline{\theta}) + \frac{K\lambda^2 - 2\lambda\delta}{2W_2(\overline{\theta}, \underline{\theta})} \leq W_2(\overline{\theta}, \underline{\theta}) - \frac{\lambda\delta}{2W_2(\overline{\theta}, \underline{\theta})},$$

for λ small enough. Hence, $\underline{\theta}$ is a secondary point. □

The following Theorem is the characterization of displacement convex approachable sets.

Theorem 9.29

A displacement convex set $\widehat{C}$ is approachable by Player 1 in $\widehat{\Gamma}$ if and only if for every $y \in \mathcal{X}$ there exists $x \in X$ such that $\delta_x \otimes \delta_y \in \widehat{C}$.

The proof is based on the following lemma :

Lemma 9.30

Let X be a compact subset of $\mathbb{R}^N$ and A be a displacement convex subset of $\Delta(X)$. Fix $\underline{\theta} \in A$. Then for all $\overline{p} \in \mathrm{NC}^g_A(\underline{\mu})$ and all $\theta_1 \in A$ we have

$$\forall p \in \mathcal{P}(\underline{\theta}, \theta_1), \quad \int_{\mathbb{R}^N} \langle \overline{p}(x), p(x)\rangle \mathrm{d}\underline{\theta}(x) := \langle \overline{p}, p\rangle_{\mathcal{L}^2_{\underline{\theta}}} \leq 0. \tag{9.11}$$

Proof. Let us consider $\underline{\theta}, \theta_0 \in A$ and $\overline{p} \in NP^g_A(\underline{\theta})$. From the definition of the proximal normal we know that there exist $\theta \notin A$ and $\gamma \in \Pi(\underline{\theta}, \theta)$ satisfying properties i) and ii) of Definition 9.13. Define $\gamma' = T\sharp\gamma$ where $T : (x,y) \mapsto (y,x)$. Then obviously γ' is an optimal plan from θ to $\underline{\theta}$.

Let $\widetilde{\gamma} \in \Pi(\underline{\theta}, \theta_0)$ be an optimal plan from $\underline{\theta}$ to θ_0 and for any $\lambda \in [0,1]$ we define $\theta_\lambda := \sigma_\lambda\sharp\widetilde{\gamma}$ and $\widetilde{\gamma}_\lambda = (\mathrm{Id}, \sigma_\lambda)\sharp\widetilde{\gamma}$ which belongs respectively to the displacement convex set A and to $\Pi(\underline{\theta}, \theta_\lambda)$.

By the disintegration of measure theorem for any $y \in \mathbb{R}^N$ there exists a probability measure $\widetilde{\gamma}_{\lambda,y}$ on $\mathbb{R}^N$ such that $\widetilde{\gamma}_\lambda = \int_{\mathbb{R}^N} (\delta_y \otimes \widetilde{\gamma}_{\lambda,y})\underline{\theta}(\mathrm{d}y)$ which means that for any continuous bounded function $u(y,z) : \mathbb{R}^{2N} \mapsto \mathbb{R}$

$$\int_{\mathbb{R}^{2N}} u(y,z)\widetilde{\gamma}_\lambda(\mathrm{d}y, \mathrm{d}z) = \int_{\mathbb{R}^N} \left[\int_{\mathbb{R}^N} u(y,z)\widetilde{\gamma}_{\lambda,y}(\mathrm{d}z)\right] \underline{\theta}(\mathrm{d}y)$$

We define $\widehat{\gamma} \in \Pi(\theta, \theta_1)$ by :

$$\forall \phi \in C_b, \int_{\mathbb{R}^{2N}} \phi \mathrm{d}\widehat{\gamma} = \int_{\mathbb{R}^{3N}} \phi(x,z)\widetilde{\gamma}_{\lambda,y}(\mathrm{d}z)\gamma'(\mathrm{d}x,\mathrm{d}y).$$

Since $\theta_\lambda \in A$ and $\widehat{\gamma} \in \Pi(\theta, \theta_\lambda)$, we obtain :

$$\begin{aligned}
W_2^2(\theta,\underline{\theta}) \leq W_2^2(\theta,\theta_\lambda) &\leq \int_{\mathbb{R}^{2N}} \|x-z\|^2 \mathrm{d}\widehat{\gamma} \\
&= \int_{\mathbb{R}^{2N}} \|x-z\|^2 \widetilde{\gamma}_{\lambda,y}(\mathrm{d}z)\gamma'(\mathrm{d}x,\mathrm{d}y) \\
&= \int_{\mathbb{R}^{3N}} \|x-y\|^2 \widetilde{\gamma}_{\lambda,y}(\mathrm{d}z)\gamma'(\mathrm{d}x,\mathrm{d}y) + 2\int_{\mathbb{R}^{3N}} \langle x-y, y-z\rangle \widetilde{\gamma}_{\lambda,y}(\mathrm{d}z)\gamma'(\mathrm{d}x,\mathrm{d}y) \\
&\qquad + \int_{\mathbb{R}^{3N}} |y-z|^2 \widetilde{\gamma}_{\lambda,y}(\mathrm{d}z)\gamma'(\mathrm{d}x,\mathrm{d}y) = a+b+c
\end{aligned}$$

where a, b and c denote respectively the three integral terms in the above equality. Now we will estimate a, b and c.

$$a = \int_{\mathbb{R}^{2N}} \|x-y\|^2 \gamma'(\mathrm{d}x,\mathrm{d}y) = W_2^2(\theta,\underline{\theta}).$$

$$\begin{aligned}
b &= 2\int_{\mathbb{R}^{2N}} \left\langle x-y, \int_{\mathbb{R}^N} (y-z)\widetilde{\gamma}_{\lambda,y}(\mathrm{d}z)\right\rangle \gamma'(\mathrm{d}x,\mathrm{d}y) \\
&= 2\int_{\mathbb{R}^{2N}} \left\langle y-x, \int_{\mathbb{R}^N} (x-z)\widetilde{\gamma}_{\lambda,x}(\mathrm{d}z)\right\rangle \gamma(\mathrm{d}x,\mathrm{d}y) \qquad \text{(by definition de } \gamma'\text{)} \\
&= -2\int_{\mathbb{R}^N} \left\langle \overline{p}(x), \int_{\mathbb{R}^N} (x-z)\widetilde{\gamma}_{\lambda,x}(\mathrm{d}z)\right\rangle \underline{\theta}(\mathrm{d}x) \qquad \text{(from the definition of } \overline{p}\text{)} \\
&= -2\int_{\mathbb{R}^{2N}} \langle \overline{p}(x), x-z\rangle\, \widetilde{\gamma}_{\lambda,x}(\mathrm{d}z)\underline{\theta}(\mathrm{d}x) \\
&= -2\int_{\mathbb{R}^{2N}} \langle \overline{p}(x), x-z\rangle\, \widetilde{\gamma}_{\lambda}(\mathrm{d}x,\mathrm{d}z) \qquad \text{(by the desintegration formula)}
\end{aligned}$$

Therefore

$$\begin{aligned}
b &= -2\int_{\mathbb{R}^{2N}} \langle \overline{p}(x), x-[(1-\lambda)x+\lambda z]\rangle\, \widetilde{\gamma}(\mathrm{d}x,\mathrm{d}z) \qquad \text{(by definition of } \widetilde{\gamma}_\lambda\text{)} \\
&= -2\lambda\int_{\mathbb{R}^{2N}} \langle \overline{p}(x), x-z\rangle\, \widetilde{\gamma}(\mathrm{d}x,\mathrm{d}z) = -2\lambda\int_{\mathbb{R}^{2N}} \langle \overline{p}(x), p(x)\rangle\, \underline{\theta}(\mathrm{d}x),
\end{aligned}$$

for any $p \in \mathcal{P}(\underline{\theta}, \theta_1)$.

The disintegration of measure formula together with the definition of $\widetilde{\gamma}$ yield

$$c = \int_{\mathbb{R}^{2N}} \left\|y-[(1-\lambda)y+\lambda z]\right\|^2 \mathrm{d}\widetilde{\gamma}(y,z) = \lambda^2 \int_{\mathbb{R}^{2N}} \left\|y-z\right\|^2 \mathrm{d}\widetilde{\gamma}(y,z)$$

hence $c = \lambda^2 W_2^2(\underline{\theta}, \theta_1)$.

Summarizing our estimates, we have obtained

$$W_2^2(\theta, \underline{\theta}) \leq W_2^2(\theta, \underline{\theta}) - \lambda \int_{\mathbb{R}^{2N}} 2 < \overline{p}(x), p(x) > \underline{\theta}(\mathrm{d}x) + \lambda^2 W_2^2(\underline{\theta}, \theta_1).$$

Thus for any $\lambda \in (0, 1)$,

$$0 \leq \lambda^2 W(\underline{\theta}, \theta_1) - 2\lambda \int_{\mathbb{R}^{2N}} < \overline{p}(x), p(x) > \underline{\theta}(\mathrm{d}x).$$

Dividing firstly by $\lambda > 0$ and letting secondly λ tend to 0^+, this gives the wished conclusion. □

Proof of Theorem 9.29 Let θ be any measure not in $\widetilde{E}$ and denote by $\underline{\theta} \in \Pi_{\widehat{C}}(\theta)$ any of its projection and $\overline{p} \in NP^g_{\widehat{C}}(\underline{\theta})$, associated to some $\overline{\gamma} \in \Pi(\underline{\theta}, \theta)$, any proximal normal. For every $x \in X$ and $y \in \mathcal{X}$ the only optimal plan from $\underline{\theta}$ to $\delta_x \otimes \delta_y$ is $\widehat{\gamma} = \underline{\theta} \times (\delta_x \otimes \delta_y)$.

The function $h : X \times \mathcal{X}$ defined by

$$h(x, y) = \int_{(X \times \mathcal{X})^2} < \underline{p}(u), u - v > \mathrm{d}\widehat{\gamma}(x, y)$$

is affine in both of its variable since :

$$h(x, y) = \int_{X \times \mathcal{X}} \langle \overline{p}(u), u \rangle \, \mathrm{d}\underline{\theta}(u) - \left\langle \int_{X \times \mathcal{X}} \overline{p}(u) d\underline{\theta}(u), (x, y) \right\rangle = \overline{z} - \langle z, (x, y) \rangle,$$

$$\text{where } \overline{z} = \int_{X \times \mathcal{X}} \langle \overline{p}(u), u \rangle \mathrm{d}\underline{\theta}(u) \text{ and } z = \int_{X \times \mathcal{X}} \overline{p}(u) d\underline{\theta}(u).$$

Since for every $y \in \mathcal{X}$, there exists $x \in X$ such that $\delta_x \otimes \delta_y \in C$, Proposition 9.30 implies that for every $y \in \mathcal{X}$ there exists $x \in X$ such that $h(x, y) \leq 0$. X and $\mathcal{X}$ are compact sets, therefore Sion's theorem implies that there exists $x \in X$ such that $h(x, y)$ for every $y \in \mathcal{X}$. Hence $\widehat{C}$ is a $\widehat{B}$-set and is approachable by Player 1.

Reciprocally, assume that there exists $y \in \mathcal{X}$ such that $\delta_x \otimes \delta_y \notin \widehat{C}$ for every $x \in X$. Since X is compact, there exists $\eta > 0$ such that $\inf_{x \in X} W_2 \left(\delta_x \otimes \delta_y, \widehat{C} \right) \geq \eta$. The strategy of Player 2 that consists of playing at each stage δ_y ensures that $\widetilde{\theta_n} = \delta_{\overline{x}_n} \otimes \delta_y$ is always at, at least, $\delta > 0$ from $\widehat{C}$. Therefore it is not approachable by Player 1. □

Concluding remarks

Recall that the action spaces in $\widetilde{\Gamma}$ (resp. $\widehat{\Gamma}$) are $\Delta(X)$ and $\Delta(\mathcal{X})$ (resp. X and $\mathcal{X}$). Assume that in $\widetilde{\Gamma}$ players are restricted to X and $\mathcal{X}$; then a $\widetilde{B}$-set should satisfy :

$$\forall \underline{\theta} \in \widetilde{E}, \forall \phi \in \mathrm{NC}^p_{\widetilde{E}}(\underline{\theta}), \exists x \in X, \forall y \in \mathcal{X}, \int_{X\times\mathcal{X}} \phi \, \mathrm{d}\,(\underline{\theta} - \delta_x \otimes \delta_y) \leq 0.$$

The proof of the sufficient part of Theorem 9.20 does not change when we add this assumption, thus a $\widetilde{B}$-set is still approachable. However, both the proof of the necessary part of Theorem 9.20 and the proof of Theorem 9.24 are no longer valid (due to the lack of linearity).

Similarly, assume that in $\widehat{\Gamma}$ players can choose action in $\Delta(X)$ and $\Delta(\mathcal{X})$ and, at stage $n \in \mathbb{N}$, the outcome is $\theta_n = \mathbf{x}_n \otimes \boldsymbol{\xi}_n$. Strictly speaking, given such outcomes that might not be absolutely continuous with respect to $\boldsymbol{\lambda}$, the sequence of interpolation $\widehat{\theta}_n$ may not be unique. However, we can assume that the game begins at stage 2 and that $\theta_1 = \boldsymbol{\lambda}/\boldsymbol{\lambda}(X \times \mathcal{X})$; then, see *e.g.* Villani [114], Proposition 5.9, $\widehat{\theta}_2 \ll \boldsymbol{\lambda}$ and is unique. By induction, the sequence of $\widehat{\theta}_n$ is unique. Once again, using the same proof, we can show that a $\widehat{B}$-set is displacement approachable, but we cannot extends the necessary part nor the characterization of displacement approachable convex sets.

The proof of Theorem 9.4 relies on the fact that the graph of $\mathbf{s}^{-1}$ is a polytope and ρ is linear. More precisely, it is a consequence of the fact that there exists a family of L-lipschitzian function $(x, \theta) \mapsto p_\kappa(x, \theta)$ such that $P(x, \theta) = \mathrm{co}\{p_\kappa(x, \theta)\}$, where $\mathrm{co}(\cdot)$ stands for the convex hull. Thus, the linearity of s and ρ can replaced by this assumption.

We make two conjectures :

There must exist a strategy such that $W_2^2(\overline{\theta}_n, \widetilde{E}) \leq O\left(\frac{1}{n}\right)$ (the answer may relies in the study of the associated Monge-Ampere equation).

The results must hold even if the strategies of players are defined as functions from $\bigcup_{n\in\mathbb{N}} (X \times \mathcal{X})^n$ to X or $\mathcal{X}$.

Acknowledgments : I deeply thank my advisor Sylvain Sorin for all his acute remarks, I acknowledge useful comments of Filippo Santambrogio and I want to express my gratitude to Marc Quincampoix for his invitation to Brest, where this work begun.

Bibliographie

[1] N. Al Najjar, A. Sandroni, R. Smorodinsky, and J. Weinstein. Testing theories with learnable and predictive representations. *manuscript*, 2008.

[2] N. Al-Najjar and J. Weinstein. Comparative testing of experts. *Econometrica*, 76 :541–559, 2008.

[3] S. As Soulaimani. Viability with probabilistic knowledge of initial condition, application to optimal control. *Set-Valued Anal.*, 16 :1037–1060, 2008.

[4] 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.

[5] J.-P. Aubin and H. Frankowska. *Set-Valued Analysis*. Birkhäuser Boston Inc., 1990.

[6] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM J. Comput.*, 32 :48–77 (electronic), 2002/03.

[7] 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).

[8] R. J. Aumann. Subjectivity and correlation in randomized strategies. *J. Math. Econom.*, 1 :67–96, 1974.

[9] 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).

[10] F. Aurenhammer. A criterion for the affine equivalence of cell complexes in $\mathbb{R}^d$ and convex polyhedra in $\mathbb{R}^{d+1}$. *Discrete Comput. Geom.*, 2 :49–64, 1987.

[11] K. Azuma. Weighted sums of certain dependent random variables. *Tôhoku Math. J. (2)*, 19 :357–367, 1967.

[12] R. B. Bapat and T. E. S. Raghavan. *Nonnegative Matrices and Applications*, volume 64 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1997.

[13] K. Basu. The traveler's dilemma : paradoxes of rationality in game theory. *American Economic Review*, 84 :391 – 395, 1994.

[14] T. Becker, K. Carter, and J. Naeve. Experts playing the traveler's dilemma. *Discussion paper 252/2005*, 2005.

[15] M. Benaïm, J. Hofbauer, and S. Sorin. Stochastic approximations and differential inclusions. *SIAM J. Control Optim.*, 44 :328–348 (electronic), 2005.

[16] P. Billingsley. *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995.

[17] D. Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific J. Math.*, 6 :1–8, 1956.

[18] D. Blackwell. Controlled random walks. In *Proceedings of the International Congress of Mathematicians, 1954, Amsterdam, vol. III*, pages 336–338, 1956.

[19] D. Blackwell and L. Dubins. Merging of opinions with increasing information. *Ann. Math. Statist.*, 38 :882–886, 1962.

[20] D. Blackwell and M. A. Girshick. *Theory of Games and Statistical Decisions*. John Wiley and Sons, Inc., New York, 1954.

[21] A. Blum, E. Even-Dar, and K. Ligett. Routing without regret : on convergence to nash equilibria of regret-minimizing algorithms in routing games. In *PODC '06 : Proceedings of the twenty-fifth annual ACM symposium on Principles of distributed computing*, pages 45–52, New York, NY, USA, 2006. ACM.

[22] 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.

[23] J.-M. Bony. Principe du maximum, inégalite de Harnack et unicité du problème de Cauchy pour les opérateurs elliptiques dégénérés. *Ann. Inst. Fourier (Grenoble)*, 19(fasc. 1) :277–304 xii, 1969.

[24] Y. Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *C. R. Acad. Sci. Paris Sér. I Math.*, 305 :805–808, 1987.

[25] H. Brezis. *Analyse Fonctionnelle*. Collection Mathématiques Appliquées pour la Maîtrise. [Collection of Applied Mathematics for the Master's Degree]. Masson, Paris, 1983. Théorie et applications. [Theory and applications].

[26] R. C. Buck. Partition of space. *Amer. Math. Monthly*, 50 :541–544, 1943.

[27] P. Cardaliaguet and M. Quincampoix. Deterministic differential games under probability knowledge of initial condition. *International Game Theory Review (IGTR)*, 10 :1–16, 2008.

[28] N. Cesa-Bianchi and G. Lugosi. Potential-based algorithms in on-line prediction and game theory. In *Computational learning theory (Amsterdam, 2001)*, volume 2111 of *Lecture Notes in Comput. Sci.*, pages 48–64. Springer, Berlin, 2001.

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

[30] N. Cesa-Bianchi, G. Lugosi, and G. Stoltz. Minimizing regret with label efficient prediction. *IEEE Trans. Inform. Theory*, 51 :2152–2162, 2005.

[31] X. Chen and H. White. Laws of large numbers for Hilbert space-valued mixingales with applications. *Econometric Theory*, 12 :284–304, 1996.

[32] F. H. Clarke. *Optimization and Nonsmooth Analysis.* John Wiley & Sons Inc., New York, 1983.

[33] A. P. Dawid. The well-calibrated Bayesian. *J. Amer. Statist. Assoc.*, 77 :605–613, 1982.

[34] E. Dekel and Y. Feinberg. Non-Bayesian testing of a stochastic prediction. *Rev. Econom. Stud.*, 73 :893–906, 2006.

[35] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

[36] K. Fan. Minimax theorems. *Proc. Nat. Acad. Sci. U. S. A.*, 39 :42–47, 1953.

[37] 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.

[38] R. W. Farebrother. *Linear Least Squares Computations*, volume 91 of *Statistics : Textbooks and Monographs.* Marcel Dekker Inc., New York, 1988.

[39] Y. Feinberg and C. Stewart. Testing multiple forecasters. *Econometrica*, 76 :561–582, 2008.

[40] W. Feller. *An Introduction to Probability Theory and its Applications. Vol. I.* Third edition. John Wiley & Sons Inc., New York, 1968.

[41] D. Foster. A proof of calibration via blackwell's approachability theorem. *Games and Economic Behavior*, 29 :73 – 78, 1999.

[42] D. P. Foster and R. V. Vohra. Calibrated learning and correlated equilibrium. *Games Econom. Behav.*, 21 :40–55, 1997.

[43] D. P. Foster and R. V. Vohra. Asymptotic calibration. *Biometrika*, 85 :379–390, 1998.

[44] D. P. Foster and R. V. Vohra. Regret in the on-line decision problem. *Games Econom. Behav.*, 29 :7–35, 1999.

[45] D. P. Foster and H. P. Young. Learning, hypothesis testing, and Nash equilibrium. *Games Econom. Behav.*, 45 :73–96, 2003. First World Congress of the Game Theory Society (Bilbao, 2000).

[46] D. P. Foster and H. P. Young. Regret testing : learning to play nash equilibrium without knowing you have an opponent. *Theoretical Economics*, 1 :341–367, 2006.

[47] D. A. Freedman. On tail probabilities for martingales. *Ann. Probability*, 3 :100–118, 1975.

[48] D. Fudenberg and D. M. Kreps. Learning mixed equilibria. *Games Econom. Behav.*, 5 :320–367, 1993.

[49] 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.

[50] D. Fudenberg and D. K. Levine. *The theory of learning in games*, volume 2 of *MIT Press Series on Economic Learning and Social Evolution*. MIT Press, Cambridge, MA, 1998.

[51] D. Fudenberg and D. K. Levine. Conditional universal consistency. *Games Econom. Behav.*, 29 :104–130, 1999.

[52] F. Germano and G. Lugosi. Global Nash convergence of Foster and Young's regret testing. *Games Econom. Behav.*, 60 :135–154, 2007.

[53] L. G. Hačijan. Polynomial algorithms in linear programming. *Zh. Vychisl. Mat. i Mat. Fiz.*, 20 :51–68, 260, 1980.

[54] 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.

[55] J. Halpern and R. Pass. Iterated regret minimization : A new solution concept. *International Joint Conference on Artificial Intelligence*, 2009.

[56] 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.

[57] S. Hart and A. Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68 :1127–1150, 2000.

[58] S. Hart and A. Mas-Colell. A general class of adaptive strategies. *J. Econom. Theory*, 98 :26–54, 2001.

[59] 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.

[60] S. Hart and A. Mas-Colell. Uncoupled dynamics cannot lead to nash equilibrium. *The American Economic Review*, 93 :1830–1836, 2003.

[61] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58 :13–30, 1963.

[62] J Hofbauer and W. H. Sandholm. On the global convergence of stochastic fictitious play. *Econometrica*, 70 :2265–2294, 2002.

[63] 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.

[64] E. Kalai and E. Lehrer. Weak and strong merging of opinions. *J. Math. Econom.*, 23 :73–86, 1994.

[65] E. Kalai, E. Lehrer, and R. Smorodinsky. Calibrated forecasting and merging. *Games Econom. Behav.*, 29 :151–169, 1999. Learning in games : a symposium in honor of David Blackwell.

[66] E. Klein and A. C. Thompson. *Theory of correspondences.* Canadian Mathematical Society Series of Monographs and Advanced Texts. John Wiley & Sons Inc., New York, 1984.

[67] E. Kohlberg. Optimal strategies in repeated games with incomplete information. *Internat. J. Game Theory*, 4 :7–24, 1975.

[68] E. Lehrer. Any inspection is manipulable. *Econometrica*, 69 :1333–1347, 2001.

[69] E. Lehrer. Approachability in infinite dimensional spaces. *Internat. J. Game Theory*, 31 :253–268 (2003), 2002.

[70] E. Lehrer. A wide range no-regret theorem. *Games Econom. Behav.*, 42 :101–115, 2003.

[71] E. Lehrer and R. Smorodinsky. Compatible measures and merging. *Math. Oper. Res.*, 21 :697–706, 1996.

[72] E. Lehrer and R. Smorodinsky. Merging and learning. In *Statistics, probability and game theory*, volume 30 of *IMS Lecture Notes Monogr. Ser.*, pages 147–168. Inst. Math. Statist., Hayward, CA, 1996.

[73] E. Lehrer and E. Solan. Excludability and bounded computational capacity. *Math. Oper. Res.*, 31 :637–648, 2006.

[74] E. Lehrer and E. Solan. Learning to play partially-specified equilibrium. *manuscript*, 2007.

[75] E. Lehrer and E. Solan. Approachability with bounded memory. *Games Econom. Behav.*, 66 :995–1004, 2009.

[76] C. E. Lemke and J. T. Howson, Jr. Equilibrium points of bimatrix games. *J. Soc. Indust. Appl. Math.*, 12 :413–423, 1964.

[77] R. D. Luce and H. Raiffa. *Games and Decisions : Introduction and Critical Survey.* John Wiley & Sons Inc., New York, N. Y., 1957. A study of the Behavioral Models Project, Bureau of Applied Social Research, Columbia University ;.

[78] G. Lugosi, S. Mannor, and G. Stoltz. Strategies for prediction under imperfect monitoring. *Math. Oper. Res.*, 33 :513–528, 2008.

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

[80] S. Mannor and G. Stoltz. A geometric proof of calibration. *manuscript*, 2010.

[81] S. Mannor and J. N. Tsitsiklis. Approachability in repeated games : computational aspects and a Stackelberg variant. *Games Econom. Behav.*, 66 :315–325, 2009.

[82] D. A. Martin. The determinacy of Blackwell games. *J. Symbolic Logic*, 63 :1565–1581, 1998.

[83] R. J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128 :153–179, 1997.

[84] J.-F. Mertens, S. Sorin, and S. Zamir. *Repeated Games.* CORE discussion paper 9420–9422. 1994.

[85] J. Neveu. *Martingales à Temps Discret.* Masson et Cie, éditeurs, Paris, 1972.

[86] 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.

[87] W. Olszewski and A/ Sandroni. Manipulability of future-independent tests. *Econometrica*, 76 :1437–1466, 2008.

[88] W. Olszewski and A. Sandroni. Manipulability of comparative tests. *Proceedings of the National Academy of Science of the United States of America*, 106 :5029–5034, 2009.

[89] W. Olszewski and A. Sandroni. A nonmanipulable test. *Ann. Statist.*, 37 :1013–1039, 2009.

[90] W. Olszewski and A. Sandroni. Strategic manipulation of empirical tests. *Math. Oper. Res.*, 34 :57–70, 2009.

[91] V. Perchet. Approachability of convex sets in games with partial monitoring. *manuscript*.

[92] V. Perchet. Calibration and internal no regret with partial monitoring. *manuscript*.

[93] V. Perchet. No-regret with partial monitoring : Calibration-based optimal algorithms. *manuscript*.

[94] V. Perchet. Calibration and internal no-regret with random signals. *Proceedings of the 20th International Conference on Algorithmic Learning Theory*, pages 68–82, 2009.

[95] V. Perchet and M. Quincampoix. Purely informative game : Application to approachability with partial monitoring. *manuscript*.

[96] L. Renou and K. Schlag. Minimax regret and strategic uncertainty. *Journal of Economic Theory*, 145 :264 – 286, 2010.

[97] R. T. Rockafellar. *Convex Analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[98] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*, volume 317. Springer-Verlag, Berlin, 1998.

[99] W. Rudin. *Real and Complex Analysis*. New York, second edition, 1974. McGraw-Hill Series in Higher Mathematics.

[100] A. Rustichini. Minimizing regret : the general case. *Games Econom. Behav.*, 29 :224–243, 1999.

[101] A. Sandroni. The reproducible properties of correct forecasts. *Internat. J. Game Theory*, 32 :151–159, 2003. Special anniversary issue. Part 2.

[102] A. Sandroni, R. Smorodinsky, and R. V. Vohra. Calibration with many checking rules. *Math. Oper. Res.*, 28 :141–153, 2003.

[103] E. Seneta. *Nonnegative Matrices and Markov Chains*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 1981.

[104] L. S. Shapley. Stochastic games. *Proc. Nat. Acad. Sci. U. S. A.*, 39 :1095–1100, 1953.

[105] E. Shmaya. Many inspections are manipulable. *Theoretical Economics*, 3 :367–382, 2008.

[106] S. Sorin. Supergames. In *Game theory and applications (Columbus, OH, 1987)*, Econom. Theory Econometrics Math. Econom., pages 46–63. Academic Press, San Diego, CA, 1990.

[107] S. Sorin. *A First Course on Zero-Sum Repeated Games*. Springer-Verlag, 2002.

[108] S. Sorin. *Lectures on Dynamics in Games.* Unpublished Lecture Notes, 2008.

[109] S. Sorin. Exponential weight algorithm in continuous time. *Math. Program.*, 116 :513–528, 2009.

[110] X. Spinat. A necessary and sufficient condition for approachability. *Math. Oper. Res.*, 27 :31–44, 2002.

[111] G. Stoltz and G. Lugosi. Internal regret in on-line portfolio selection. *Mach. Learn.*, 59 :125–159, 2005.

[112] G. Stoltz and G. Lugosi. Learning correlated equilibria in games with compact sets of strategies. *Games Econom. Behav.*, 59 :187–208, 2007.

[113] N. Vieille. Weak approachability. *Math. Oper. Res.*, 17 :781–791, 1992.

[114] C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2003.

[115] J. Von Neumann. Zur theorie des gesellschaftsspiele. *Mathematische Annalen*, 100 :295–320, 1928.

[116] V. Vovk. A game of prediction with expert advice. *J. Comput. Syst. Sci.*, 56 :153–173, 1998.

[117] V. Vovk. Non-asymptotic calibration and resolution. *Theoret. Comput. Sci.*, 387 :77–89, 2007.

[118] 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.

[119] J. Wardrop. Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers, Part II*, 1 :352–362, 1952.

[120] V. Yurinskii. Exponential inequalities for sums of random vectors. *Journal of Multivariate Analysis*, 6 :473 – 499, 1976.

[121] A. Zapechelnyuk. Better-reply dynamics with bounded recall. *Math. Oper. Res.*, 33 :869–879, 2008.

[122] G. Ziegler. *Lectures on Polytopes*, volume 152 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 1995.