Mémoire de Master : Les jeux de révision

Benoit Duvocelle

17 juillet 2017

Remerciements

Je voudrais tout d'abord remercier mon directeur de mémoire, Tristan Tomala, pour ses conseils avisés, son aide patiente et ses qualités pédagogiques. Il a été un guide éclairant tout au long de cette recherche.

Je remercie infiniment ma sœur Agnès, mon frère Florian, ma mère Reine, ma tante Ginette, ma voisine Maryse et mes colocataires Maxime, Florian, Jérémy, Edouard et Thibault qui ont eu la gentillesse et la patience de lire ce mémoire afin d'en améliorer la qualité grâce à leurs questions et à leurs suggestions pertinentes.

Je remercie enfin toute ma famille pour son soutien continu et bienveillant.

Table des matières

Introduction

La théorie des jeux est une branche des mathématiques ayant de nombreuses applications en économie, mais également en physique et en biologie. Elle s'intéresse aux interactions entre des individus, appelés "joueurs", qui sont amenés à effectuer des choix dans un ensemble de circonstances, appelé "jeu".

Dans son ouvrage de 1938, *Applications aux Jeux de Hasard* (Borel, 1938), Émile Borel développe un théorème du minmax pour les jeux à somme nulle à deux joueurs, c'est-à-dire les jeux dans lesquels ce que gagne l'un est perdu par l'autre. C'est le début de la théorie des jeux.

Cette théorie des jeux devient un champ de recherche à part entière avec la publication de *Theory of Games and Economic Behavior* (1944) (Théorie des jeux et du comportement économique) par John von Neumann et Oskar Morgenstern. Cet ouvrage fondateur détaille la méthode de résolution des jeux à somme nulle.

Vers 1950, John Forbes Nash développe la notion d'équilibre de Nash.

En 1950, Loyd Shapley, futur prix Nobel d'économie, introduit les jeux stochastiques, c'est-à-dire les jeux répétés une infinité de fois dans lesquels les paiements courants influent sur les suivants.

En 1994, John Nash, Reinhard Selten et John Harsanyi reçoivent le prix Nobel d'économie pour leurs travaux sur la théorie des jeux. Ce choix témoigne de l'importance prise par la théorie des jeux dans l'analyse économique.

Les jeux de révision qui font l'objet de ce mémoire ont quant à eux été introduits en 2008 par Yuichiro Kamada et Michihiro Kandori. Ce sont des jeux qui se déroulent en temps continu et dans lesquels les joueurs reçoivent aléatoirement des opportunités de changer leur choix, appelé "révision". L'étude de ces jeux a des applications directes en économie de marché, plus particulièrement dans les ventes aux enchères dans un laps de temps fini. C'est dans le but de mieux comprendre ce sujet que j'ai choisi ce thème de mémoire sous la direction de Tristan Tomala, spécialiste de la théorie des jeux.

Chapitre 1

Jeux sous forme normale

1.1 Premières définitions

Définition 1.

On appelle jeu sous forme normale la donnée d'un ensemble $\mathcal{N}$ de joueurs, d'une famille d'ensembles de stratégies (ou d'actions) $(A_i)_{i\in\mathcal{N}}$ et d'une famille de fonctions de paiements $(g_i)_{i\in\mathcal{N}}$ avec $g_i : A := \prod_{i\in\mathcal{N}} A_i \to \mathbb{R}$. L'ensemble des joueurs sera toujours supposé fini et non-vide. Les ensembles d'actions seront toujours supposés non vides et on parlera de jeu fini lorsque A_i est fini pour tout i dans $\mathcal{N}$. Un jeu sous forme normale représente une interaction entre joueurs rationnels : chaque joueur $i \in \mathcal{N}$ choisit une action $a_i \in A_i$, les choix étant simultanés, et si l'élément $a = (a_i)_{i\in\mathcal{N}}$ de A est le profil d'actions choisi, le joueur i reçoit le paiement $g_i(a)$. Tous les joueurs connaissent le jeu et le but de chaque joueur est d'obtenir le paiement le plus grand possible, sans se soucier de savoir s'il sera plus grand ou non que celui des autres joueurs.

Exemple 1. *Le Dilemme du Prisonnier*

Deux criminels sont arrêtés et interrogés dans des pièces séparées. Ils ont le choix entre dénoncer leur complice (action D) ou se taire et donc coopérer avec leur complice (action C). Un criminel dénoncé par son complice se verra infliger une lourde peine (quatre ans, que l'on représente par le paiement -4) et une peine légère dans le cas contraire (un an, que l'on représente par le paiement -1). De plus le fait de dénoncer l'autre permet d'obtenir une remise de peine (un an de moins, traduit par +1 dans son paiement), que l'on soit soi-même dénoncé ou pas. Chaque joueur classe les issues du jeu par préférence décroissante selon l'ordre suivant : (ne pas être dénoncé et dénoncer), (ne pas être dénoncé et ne pas dénoncer), (être dénoncé et dénoncer), (être dénoncé et ne pas dénoncer). Le jeu peut donc se représenter ainsi :

	C	D
C	$-1,-1$	$-4,0$
D	$0,-4$	$-3,-3$

Commençons par donner quelques notions simples de bonne stratégie. Nous adopterons les notations suivantes. Pour tout joueur i, $-i$ désigne l'ensemble des autres joueurs $\mathcal{N}\backslash\{i\}$. Si $(E^i)_{i\in\mathcal{N}}$ est une famille d'ensembles indexée par $\mathcal{N}$, nous notons $E := \prod_{i\in\mathcal{N}} E_i$, $E^{-i} := \prod_{j\neq i} E^j$. Un élément e de E pourra se noter $e = (e_i)_{i\in\mathcal{N}} = (e_i, e_{-i})$ cette dernière notation étant utilisée lorsque l'on veut séparer le joueur i des autres.

Définition 2. *Une stratégie $a_i \in A_i$ du joueur i est dominée si*

$$\exists b_i \in A_i\backslash\{a_i\}, \forall a_{-i} \in A_{-i}, g_i(a_i, a_{-i}) \leq g_i(b_i, a_{-i}).$$

Une stratégie $a_i \in A_i$ du joueur i est strictement dominée si

$$\exists b_i \in A_i\backslash\{a_i\}, \forall a_{-i} \in A_{-i}, g_i(a_i, a_{-i}) < g_i(b_i, a_{-i}).$$

Une stratégie $a_i \in A_i$ du joueur i est dominante si

$$\forall b_i \in A_i\backslash\{a_i\}, \forall a_{-i} \in A_{-i}, g_i(a_i, a_{-i}) \geq g_i(b_i, a_{-i}).$$

Une stratégie $a_i \in A_i$ du joueur i est strictement dominante si

$$\forall b_i \in A_i\backslash\{a_i\}, \forall a_{-i} \in A_{-i}, g_i(a_i, a_{-i}) > g_i(b_i, a_{-i}).$$

Un joueur rationnel ne jouera jamais de stratégie strictement dominée, il jouera à coup sûr une stratégie strictement dominante si elle existe (et alors elle est unique) et ne perd rien à jouer une stratégie dominante. On peut remarquer que dans le jeu du Dilemme du Prisonnier, la stratégie D est strictement dominante pour chaque joueur, l'issue rationnelle du jeu est donc (D, D).

1.2 Les jeux à somme nulle

Un jeu à somme nulle est un jeu à deux joueurs $G = (A_1, A_2, g_1, g_2)$ tel que pour tout $a \in A$, $g_1(a) = -g_2(a)$. Pour cette partie posons, $A_1 = S$, $A_2 = T$, et $g_1 = g$. Ainsi, un jeu à somme nulle est déterminé par une application $g : S \times T \to \mathbb{R}$. Dans la suite, par souci de simplicité, nous supposerons g bornée.

Définition 3. *Le joueur 1 garantit le paiement $d \in \mathbb{R}$ si :*

$$\exists s \in S, \forall t \in T, g(s,t) \geq d.$$

Le joueur 1 défend le paiement $d \in \mathbb{R}$ si :

$$\forall t \in T, \exists s \in S, g(s,t) \geq d.$$

Le joueur 2 garantit le paiement $d \in \mathbb{R}$ si :

$$\exists t \in T, \forall s \in S, g(s,t) \leq d.$$

Le joueur 2 défend le paiement $d \in \mathbb{R}$ si :

$$\forall s \in S, \exists t \in T, g(s,t) \leq d.$$

Les propriétés suivantes sont immédiates :

Proposition 1. *Posons*

$$\underline{v}(g) = \sup_{s \in S} \inf_{t \in T} g(s,t) \text{ et } \bar{v}(g) = \inf_{t \in T} \sup_{s \in S} g(s,t).$$

On a $\underline{v}(g) \leq \bar{v}(g)$ et

$$\underline{v}(g) = \sup \{d | 1 \text{ garantit } d\} = \inf \{d | 2 \text{ défend } d\},$$

$$\bar{v}(g) = \sup \{d | 1 \text{ défend } d\} = \inf \{d | 2 \text{ garantit } d\},$$

Définition 4. *On dit que le jeu (S, T, g) a une valeur lorsque $\underline{v}(g) = \bar{v}(g)$ et on note $v(g)$ cette valeur.*

Proposition 2. *Pour toutes applications $f, g : S \times T \to \mathbb{R}$:*

$$|\underline{v}(g) - \underline{v}(g)| \leq ||g - f|| \text{ et } |\bar{v}(g) - \bar{v}(g)| \leq ||g - f||.$$

Preuve.
Si $a > 0$ et $b \in \mathbb{R}$ et si l'on pose $af + b$ l'application $(s,t) \mapsto af(s,t) + b$, on a : $\underline{\mathrm{v}}(af + b) = a\underline{\mathrm{v}}(f) + b$. De plus, si pour tout (s,t), $f(s,t) \leq g(s,t)$, alors $\underline{\mathrm{v}}(f) \leq \underline{\mathrm{v}}(g)$. Comme $f(s,t) \leq g(s,t) + ||g - f||$, on a $\underline{\mathrm{v}}(f) \leq \underline{\mathrm{v}}(g) + ||g - f||$, d'où le résultat.

1.3 Jeux à somme non nulle

La notion centrale de solution pour les jeux sous forme normale est l'équilibre de Nash (1950). Soit $G = (\mathcal{N}, (A^i)_{i \in \mathcal{N}}, (g^i)_{i \in \mathcal{N}})$ un jeu.

Définition 5. *Un équilibre de Nash du jeu G est un profil de stratégies $a = (a^i)_{i \in \mathcal{N}}$ tel que :*

$$\forall i \in \mathcal{N}, \forall b^i \in A^i, g^i(b^i, a^{-i}) \leq g^i(a).$$

Cette définition peut se reformuler différemment. Par exemple, étant donné un profil de stratégies a , on dit que le joueur i a une déviation profitable par rapport à a, s'il existe b^i telle que $g^i(b^i, a^{-i}) > g^i(a)$. Un équilibre de Nash est un profil de stratégies pour lequel il n'existe pas de déviation profitable. C'est donc un point tel que, si tous les joueurs savent qu'on va jouer a, alors chacun a effectivement intérêt à le jouer. La reformulation ci-dessous va suggérer une méthode de calcul et de démonstration d'existence.

Définition 6. *Pour chaque joueur i et profil d'action de ses adversaires a^{-i}, on dit que a^i est meilleure réponse contre a^{-i} si :*

$$\forall b^i \in A^i, g^i(b^i, a^{-i}) \leq g^i(a^i, a^{-i}).$$

On appelle correspondance de meilleure réponse du joueur i, l'application

$$\begin{aligned} MR_i : A^{-i} &\to \mathcal{P}(A^i) \\ a^{-i} &\mapsto \arg\max_{b^i \in A^i} g(b^i, a^{-i}). \end{aligned}$$

On notera parfois $MR_i: \ A^{-i} \Rightarrow A^i$.
On appelle correspondance de meilleure réponse du jeu G, l'application

$$\begin{aligned} MR : A &\to \mathcal{P}(A) \\ a &\mapsto \prod_{i\in\mathcal{N}} MR^i(a^{-i}). \end{aligned}$$

On voit alors que a est un équilibre de Nash de G si et seulement si $a \in MR(a)$. On dit que a est un point fixe de la correspondance de meilleure réponse. La procédure de calcul des équilibres est donc la suivante : tracer le graphe de la correspondance de meilleure réponse de chaque joueur, et chercher l'intersection des graphes. Cette dernière reformulation nous sera très utile pour démontrer un théorème d'existence d'équilibre.

On rappelle le théorème de Brouwer :

Théorème 1. ***(de Brouwer).*** *Soit C un convexe compact non-vide de $\mathbb{R}^k$ et $f : C \to C$ une fonction continue, alors f admet un point fixe : il existe $c \in C$ tel que $c = f(c)$.*

Le théorème de Brouwer a été généralisé aux correspondances par Kakutani (1941) dans les espaces vectoriels de dimension finie, et par Glicksberg (1952) dans les espaces vectoriels topologiques.

Théorème 2. ***(de Kakutani-Glicksberg).*** *Soit C un convexe compact non-vide d'un espace vectoriel topologique localement convexe séparé, et F une correspondance de C dans ses parties telle que*

1. *pour tout $c \in C$, $F(c)$ est un convexe compact non vide,*

2. *le graphe de F, à savoir $\{(c,d) \in C \times C | c \in F(d)\}$ est fermé.*

Alors, il existe $c \in C$ tel que $c \in F(c)$.

On peut aussi formuler ce théorème de la façon suivante : Soit X un sous-ensemble fermé de $C \times C$ tel que pour tout $c \in C$, la section de X au-dessus de C (i.e. $\{d \in C | (c,d) \in X\}$) est un convexe compact non-vide. Alors, X coupe la diagonale de $C \times C$.

Preuve. voir Kakutani (1941) et Glicksberg (1952).

Définition 7. *Soit $f : X \to \mathbb{R}$ où X est un sous-ensemble convexe d'un espace vectoriel, on dit que f est quasi-concave si pour tout réel c, l'ensemble $\{x | f(x) \geq c\}$ est convexe.*

Théorème 3. ***(de Glicksberg-Nash)*** *Soit $G = (\mathcal{N}, (A^i)_{i\in\mathcal{N}}, (g^i)_{i\in\mathcal{N}})$ un jeu tel que : pour tout $i \in \mathcal{N}$, A^i est un convexe compact non-vide dans un espace vectoriel topologique localement convexe, $g^i : A \to \mathbb{R}$ est continue et pour tout $a^{-i} \in A^{-i}$, l'application partielle $a^i \mapsto g^i(a^i, a^{-i})$ est quasi-concave.*

Alors, l'ensemble des équilibres de Nash de G est un fermé non-vide.

Preuve. On peut appliquer le théorème de Kakutani-Glicksberg. Les hypothèses de compacité, de continuité et de quasi-concavité des fonctions des paiements assurent que $MR_i(a^{-i})$ est convexe fermé non-vide ($\forall i, \forall a^{-i}$), $MR(a)$ est donc bien convexe compact non-vide ($\forall a$). Le graphe de la correspondance MR s'écrit :

$$\{(a,b) \in A \times A | \forall i \in \mathcal{N}, \forall c^i \in A^i, g^i(a^i, b^{-i}) \geq g^i(c^i, b^{-i})\}.$$

Les fonctions de paiements étant continues,

$$\{(a,b) \in A \times A | g^i(a^i, b^{-i}) \geq g^i(c^i, b^{-i})\}$$

est fermé ($\forall i \in \mathcal{N}, \forall c^i \in A^i$). Le graphe de la correspondance MR est donc une intersection de fermés. L'ensemble des équilibres de Nash est donc un fermé non-vide.

1.4 Jeux finis et stratégies mixtes

Définition 8. *On dit qu'un jeu $G = (\mathcal{N}, (A^i)_{i\in\mathcal{N}}, (g^i)_{i\in\mathcal{N}})$ est fini lorsque les ensembles de stratégies sont toutes des ensembles finis.*

Soit E un ensemble fini. On note $\Delta(E)$ l'ensemble des mesures de probabilités sur E :

$$\Delta(E) = \left\{ p \in \mathbb{R}^E | \forall e \in E, p(e) \geq 0, \sum_{e\in E} p(e) = 1 \right\}.$$

Cet ensemble est un convexe compact.

Définition 9. *Soit $G = (\mathcal{N}, (A^i)_{i\in\mathcal{N}}, (g^i)_{i\in\mathcal{N}})$ un jeu fini.*

On appelle stratégie mixte du joueur i, une probabilité $\sigma^i \in \Delta(A^i)$ et stratégie pure une action $a^i \in A^i$. L'ensemble des stratégies mixtes du joueur i est $\Delta(A^i)$.

Etant donné un profil de stratégie $\sigma = (\sigma^i)_{i\in\mathcal{N}} \in \prod_{i\in\mathcal{N}} \Delta(A^i)$ que l'on note encore g^i, on appelle paiement espéré du joueur i la quantité :

$$g^i(\sigma) = \sum_{a\in A} \left(\prod_{i\in\mathcal{N}} \sigma^i(a^i) \right) g^i(a).$$

Ceci définit une extension de l'application g^i de $\prod_{i\in\mathcal{N}} A^i$ à $\prod_{i\in\mathcal{N}} \Delta(A^i)$ que l'on note encore g^i.

On appelle extension mixte de G le jeu $(\mathcal{N}, (\Delta(A^i))_{i\in\mathcal{N}}, (g^i)_{i\in\mathcal{N}})$.

Un équilibre de Nash de l'extension mixte de G s'appellera équilibre de G en stratégies mixtes.

Théorème 4. ***(du MinMax, Von Neumann, 1928)***

Tout jeu fini (A^1, A^2, g) admet une valeur v et les deux joueurs ont des stratégies optimales. De plus,

$$v = \max_{\sigma^1\in\Delta(A^1)} \min_{a^2\in A^2} g(\sigma^1, a^2) = \min_{\sigma^2\in\Delta(A^2)} \max_{a^1\in A^1} g(a^1, \sigma^2)$$

Preuve. Admise.

Théorème 5. ***(de Nash, 1950)*** *Tout jeu fini $G = (\mathcal{N}, (A^i)_{i\in\mathcal{N}}, (g^i)_{i\in\mathcal{N}})$ admet un équilibre de Nash en stratégies mixtes.*

Preuve. Admise.

Théorème 6. *Soit $G = (\mathcal{N}, (A^i)_{i\in\mathcal{N}}, (g^i)_{i\in\mathcal{N}})$ un jeu fini. Un profil de stratégies $\sigma = (\sigma^i)_{i\in\mathcal{N}}$ est un équilibre de Nash en stratégies mixtes de G si et seulement si :*

$$\forall i \in \mathcal{N}, \forall a^i \in A^i, (\sigma^i(a^i) > 0 \Rightarrow a^i \in MR^i(\sigma^{-i}))$$

Preuve. Admise.

1.5 Conclusion

Ces premiers résultats, intéressants par eux-mêmes, vont permettre dans la suite d'étudier les jeux de révision, qui sont des jeux en temps continus et qui s'articulent sur la base des jeux sous forme normale, dans lesquels les joueurs peuvent changer leur action.

Le matériel introduit jusqu'à présent n'est que la base de la théorie des jeux qui nous servira à étudier les jeux de révision. Le lecteur intéressé pourra se pencher sur le livre *Bases de la théorie des jeux* (1992) de Tomala, Laraki et Renault.

Chapitre 2

Jeux de révision

2.1 Présentation et généralités

Les jeux de révision sont des jeux qui se déroulent en temps continu, où les joueurs peuvent avoir, à certains temps inconnus et tirés aléatoirement, des opportunités de changer d'action. On introduit tout d'abord un modèle très général qui est celui des jeux de révision stochastiques avec des ensembles d'états et d'actions finis avant d'expliquer leur déroulement et de donner de premiers résultats.

Définition 10.

Un jeu de révision stochastique est défini par les éléments suivants, dont nous supposerons tous les ensembles finis et non vides.

— Un ensemble $\mathcal{N}$ de joueurs.

— Un ensemble K d'états.

— Un état initial $k_0 \in K$.

— Pour chaque état $k \in K$ et pour chaque joueur $i \in \mathcal{N}$, le joueur i peut choisir une action parmi l'ensemble $A_i(k)$ à l'état k. On note $A(k) := \prod_{k \in K} A_i(k)$ l'ensemble des actions possibles à l'instant k, et $A := \cup_{k \in K} A(k)$
.

— Une fonction de gain $r : K \to \mathbb{R}^{\mathcal{N}}$. Pour tout $i \in \mathcal{N}$ on note r_i la fonction de gain du joueur i.

— Une fonction de transition $q : K \times A \to \Delta(K)$.

— A chaque état k on associe la fréquence d'opportunité $\lambda_k > 0$.

— Un intervalle de temps $[-T, 0]$.

Le jeu se déroule de la façon suivante :

— Le jeu commence à l'instant $-T$ à l'état k_0. On note ce jeu $\Gamma(-T, k_0)$.

— Au cours du temps, une cloche sonne de façon aléatoire. A chaque temps $t \in [-T, 0)$, le temps $t' := t + \tau$ du prochain son de cloche est tiré aléatoirement en suivant la loi de Poisson $\mathcal{E}(\lambda_k)$.

— Si $t' > 0$, le jeu s'arrête, et si l'état final est k, les joueurs reçoivent le paiement $r(k)$.

— Si $t' < 0$, on dit que t' est un temps de révision. Si l'état du jeu à t' est k, à t' chaque joueur i choisit une action $a_i \in A_i(k)$, ces choix se faisant simultanément. Un nouvel état k' est alors tiré selon la probabilité $q(k'|k, a)$ et on commence le nouveau jeu $\Gamma(t', k')$.

Remarque 1.

On peut modéliser le fait qu'un joueur ait plus de probabilité qu'un autre d'avoir une opportunité dans un certain état à l'aide de la fonction de transition q. Par exemple si $q(., k)$ ne dépend que de i, alors à l'état k seul le joueur i peut profiter des opportunités. Notons cependant que la fonction de transition est supposée indépendante du temps.

Nous étudierons par la suite plus particulièrement le cas des jeux de révision stochastiques où au plus un joueur peut avoir une opportunité à un temps quelconque.

Définition 11. *Un jeu de révision stochastique est asynchrone si pour tout état $k \in K$, il existe au plus un joueur $i \in \mathcal{N}$ tel que $A_i(k)$ n'est pas réduit à un élément : $\forall j \neq i$, $|A_j(k)| = 1$ et $|A_i(k)| > 1$.*

Remarque 2.

On peut voir un jeu de révision stochastique comme un jeu stochastique à temps discret (i.e. un jeu que l'on répète une infinité de fois et dont les paiements futurs dépendent des paiements présents) mais dont l'ensemble des états est infini et compact. Un état du jeu stochastique associé serait de la forme $\omega = (t, k)$, où k est l'état courant dans le jeu de révision stochastique et t est le temps restant. La loi de transition de ω à ω' du jeu stochastique associé serait une probabilité à densité donnée par une loi de Poisson et la fonction de transition $q(k'|k, a)$.

Stratégies et équilibres

En toute généralité, les stratégies des joueurs dépendent de toutes les informations qu'ils ont reçues depuis le début du jeu. Une histoire est constituée des temps de révision passés, des états et des profils d'actions associés à ces temps de révision ainsi que le temps courant (ou le temps restant, c'est égal). Ce qui s'écrit :

$$h = (t_1, k_1, a_1, ..., t_n, k_n, a_n, t_{n+1}).$$

Notons qu'on ne note pas k_0 car c'est un paramètre du jeu, de même qu'on ne note pas $-T$ au début. La suite $(t_k)_{1 \leq k \leq n+1}$ est strictement croissante et telle que $t_n < 0$. Si $t_{n+1} > 0$ alors l'histoire est finie, l'état final est k_n et le paiement est $r(k_n)$. Sinon, les joueurs jouent le jeu $\Gamma(t_{n+1}, k_n)$.

Une stratégie (de comportement) pour le joueur i est une application mesurable qui à une histoire non finie associe une probabilité sur l'ensemble des actions. Plus précisément, on note H_n l'ensemble des histoires non finies de la forme $h = (t_1, k_1, a_1, ..., t_n, k_n, a_n, t_{n+1})$ telles que $-T < t_1 < ... < t_{n+1} < 0$ et telles que si $k_m \in K$, alors $a_m \in A(k_m)$ pour $m = 1, ..., n$. On munit $[-T, 0]$ de la tribu des boréliens et $[-T, 0] \times (K \times A \times [-T, 0])^n$ de la tribu produit. L'ensemble H_n est alors un sous-ensemble mesurable de $[-T, 0] \times (K \times A \times [-T, 0])^n$. Une stratégie pour le joueur i est alors une suite d'applications mesurables $(\sigma_{i,n})_{n \geq 1}$ telle que pour tout n et pour tout $h \in H_n$, $\sigma_{i,n}(h) \in \Delta(A_i(k_n))$. Un profil σ de stratégies induit une distribution de probabilité $\mathbb{P}_\sigma$ sur l'ensemble des histoires. On note $\mathbb{E}_\sigma$ l'espérance associée à $\mathbb{P}_\sigma$.

On s'aperçoit qu'une histoire infinie n'arrive presque sûrement jamais. En effet comme le temps d'attente entre deux temps de révision est distribué exponentiellement avec un paramètre compris entre $\min_k \lambda_k$ et $\max_k \lambda_k$, alors le n-ième temps de révision tend vers l'infini après probabilité 1, sous n'importe quel profil de stratégies des joueurs. Dans ce qui suit nous ignorons le cas des histoires infinies.

Pour toute histoire non finie h et toute stratégie σ_i, on note $\sigma_i(.|h)$ la continuité de la stratégie σ_i après l'histoire h, qui correspond avec $\sigma_i(.)$ pour toute histoire non finie h' commençant par h. On peut alors introduire la notion d'équilibre de Nash et d'équilibre en sous-jeux parfait.

Définition 12. *1. Un équilibre de Nash du jeu de révision stochastique est un profil de stratégie σ tel que pour tout joueur i et pour toute stratégie σ'_i,*

$$\mathbb{E}_\sigma[r(k_0)] \geq \mathbb{E}_{\sigma'_i, \sigma_{-i}}[r(k_0)].$$

2. Un équilibre en sous-jeux parfait est un profil de stratégies σ tel que pour toute histoire non finie $h = (t_1, k_1, a_1, ..., t_n, k_n, a_n, t_{n+1})$, la continuité du profil de stratégies $\sigma(.|h)$ est un équilibre de Nash du jeu $\Gamma(t_{n+1}, k_n)$. C'est-à-dire, pour tout joueur i et toute stratégie σ'_i,

$$\mathbb{E}_{\sigma(.|h)}[r(k_0)] \geq \mathbb{E}_{\sigma'_i, \sigma_{-i}(.|h)}[r(k_0)].$$

Cela s'interprète de la façon suivante : σ est un équilibre de Nash (resp. un équilibre en sous-jeux parfait) si c'est un profil de stratégies tel que (resp. tel qu'après chaque histoire non finie), si chaque joueur sait que les autres joueurs vont jouer ce profil de stratégies, alors il trouve son intérêt à le jouer.

Remarque 3. *Ces profils de stratégies sont utiles car ils permettent une bonne anticipation du jeu dans certains cas. Cependant ils ne sont en général pas uniques, n'apportent pas tous les mêmes espérances de gain et sont difficiles à trouver compte tenu de la complexité des stratégies.*

Dans la section suivante nous allons montrer qu'il existe toujours un tel équilibre.

Equilibres markoviens parfaits

Cette section aborde un résultat important d'existence d'équilibre markovien parfait, c'est-à-dire un équilibre en sous-jeux parfait qui ne dépend que du temps restant et de l'état courant. Autrement dit c'est un profil de stratégies stable et qui n'utilise pas la mémoire des joueurs sur ce qui s'est passé avant l'instant présent, d'où le terme "markovien".

Définition 13. *Une stratégie markovienne pour le joueur i est une application borélienne $\sigma_i : [-T, 0] \times K \to \Delta(A_i(K))$ telle que pour tout $(t, k) \in [-T, 0] \times K$, $\sigma_i(t, k) \in \Delta(A_i(k))$.*

Ici cela signifie qu'après toute histoire $h = (t_1, k_1, a_1, ..., t_n, k_n, t_{n+1})$, avec $k_n = k$ et $t_{n+1} = t$, la stratégie du joueur i est de choisir l'action $\sigma_i(t, k)$ qui ne dépend que de l'état courant et du temps restant.

Définition 14. *Un équilibre parfait markovien est un profil de stratégies markoviennes qui est un équilibre en sous-jeux parfait. C'est-à-dire, pour tout $(t, k) \in [-T, 0] \times K$ et après toute histoire $h = (t_1, k_1, a_1, ..., t_n, k_n, a_n, t_{n+1})$ telle que $t_{n+1} = t$ et $k_n = k$, le profil de stratégie $\sigma(t, k)$ est un équilibre de Nash du jeu $\Gamma(t, k)$.*

On présente une définition équivalente. Soit σ un profil de stratégies. On considère l'espérance de gain du joueur i étant donné l'état k et le temps t, qu'on note $\mathbb{E}_\sigma^{t,k}$:

$$U_i^\sigma(t, k) = \mathbb{E}_\sigma^{t,k}[r_i(k)],$$

où $U_i^\sigma(t, k)$ est le paiement de continuité du jeu de révision $\Gamma(t, k)$ associé à σ, c'est-à-dire le paiement espéré par le joueur i si le profil de stratégies des joueurs est σ. L'application $U^\sigma : t \mapsto U^\sigma(t) = (U_i^\sigma(t, k))_{i \in \mathcal{N}, k \in K}$ est l'application de paiement de continuité associée à σ. Puisque la probabilité qu'il y ait une opportunité entre t et $t + dt$ tend vers 0 quand dt tend vers 0, il vient que U^σ est continue en t. On verra qu'elle est même lipschitzienne en t.

Considérons maintenant un temps de révision t à l'état k. Les joueurs choisissent simultanément leur action dans $A(k)$. Ainsi au temps t si la stratégie σ est appliquée dans le futur, les joueurs agissent comme s'ils jouaient dans le jeu sous forme normale dont la fonction paiement est

$$u_i^\sigma(t, k, a) := \sum_{l \in K} q(l|k, a) U_i^\sigma(t, l).$$

On l'appelle fonction de gain de continuité.

Lemme 1. *Un profil de stratégies σ est un équilibre parfait markovien si et seulement si pour tout k et tout $t \in [-T, 0]$, $\sigma(t, k)$ est un équilibre de Nash de $u^\sigma(t, k, .)$.*

Cette propriété est analogue au principe de déviation en un coup valide pour les équilibres en sous-jeux parfaits des jeux dynamiques. Ainsi, pour vérifier qu'un profil de stratégies est un équilibre parfait markovien, il suffit de vérifier qu'il n'existe pas de stratégie qui a une déviation profitable pour un joueur après une histoire donnée.

Preuve. Soit un équilibre parfait markovien σ. En particulier c'est un équilibre en sous-jeux parfait. Alors pour tout t et k, aucun joueur n'a envie de dévier, étant donné qu'aucun joueur n'a envie de dévier dans le futur. Donc $\sigma(t, k)$ est un équilibre de Nash de $u^\sigma(t, k, .)$.

Réciproquement, prenons un profil de stratégies σ qui ne soit pas un équilibre en sous-jeux parfait. On va montrer qu'il existe un ensemble mesurable de mesure non nulle sur lequel il existe un état et un joueur qui a un déviation profitable.

Comme σ n'est pas un équilibre en sous-jeux parfait, il existe une déviation profitable pour un joueur après une certaine histoire. Il existe un $\epsilon > 0$, un temps $s < 0$, un état k, un joueur i et une stratégie τ_i tels que

$$\mathbb{E}_{\tau_i, \sigma_{-i}}^{s,k}[r_i(k)] > \mathbb{E}_\sigma^{s,k}[r(k)] + \epsilon.$$

Analysée ainsi, la déviation τ_i peut différer de σ_i après un nombre infini d'histoires. Pour tout entier n on note τ_i^n la stratégie qui coïncide avec τ_i les n pour les n opportunités, et avec σ_i pour les suivantes. La probabilité d'avoir plus de n révisions est majorée par $\mathbb{P}(X > n)$ où X est une variable aléatoire de loi Poisson de paramètre $\lambda^* = \max \lambda_k$. Donc cette probabilité tend vers 0 quand n tend vers l'infini. On en déduit que pour n assez grand,

$$\left| \mathbb{E}_{\tau_i, \sigma_{-i}}^{s,k}[r_i(k) - \mathbb{E}_{\tau_i^n, \sigma_{-i}}^{s,k}[r_i(k)] \right| \leq \epsilon/2,$$

et donc,

$$\mathbb{E}_{\tau_i^n, \sigma_{-i}}^{s,k}[r_i(k)] > \mathbb{E}_\sigma^{s,k}[r_i(k)] + \epsilon/2.$$

On en déduit ainsi l'existence d'une déviation profitable (de $\epsilon/2$) qui se différencie de σ seulement les n premières révisions. Supposons désormais que,

$$\mathbb{E}_{\tau_i^{n-1}, \sigma_{-i}}^{s,k}[r_i(k)] < \mathbb{E}_{\tau_i^n, \sigma_{-i}}^{s,k}[r_i(k)],$$

cela signifierait que dévier les n premières révisions fait gagner davantage que de ne dévier que les $n - 1$ premières. En particulier, cela implique que la probabilité d'avoir au moins n révisions est strictement positive. On considère l'ensemble des $(l,t) \in K \times [s,0]$ tels que t est le n-ième temps de révision dans $\Gamma(s,k)$ et $k_t = l$. D'après l'inégalité précédente, la probabilité de cet ensemble est strictement positive sous $\mathbb{P}_{\tau_i^{n-1},\sigma_{-i}}$. Cela implique qu'il existe un ensemble de temps t de mesure strictement positive sur lequel il existe un état l et un joueur i qui a une déviation profitable sur un coup.

Si l'inégalité inverse est vérifiée,

$$\mathbb{E}_{\tau_i^{n-1},\sigma_{-i}}[r_i(k)] \geq \mathbb{E}_{\tau_i^n,\sigma_{-i}}[r_i(k)] > \mathbb{E}_\sigma^{s,k}[r_i(k)] + \epsilon/2,$$

on déduit l'existence d'une déviation profitable qui dévie de σ seulement les $n-1$ premières révisions. On conclut par récurrence.

En se basant sur le lemme précédent, on peut définir un équilibre markovien parfait comme un profil de stratégies tel qu'à chaque temps, les actions mixtes courantes forment un équilibre de Nash pour le paiement continu induit par cette stratégie. On utilise cette idée pour définir les équilibres corrélés. Un équilibre corrélé parfait markovien est tel qu'à chaque instant, les distributions des actions corrélées forment un équilibre corrélé pour les paiements continus. Cela assure que, pour tout temps de révision, tous les joueurs ont intérêt à jouer les actions recommandées par le système de corrélation.

Définition 15. *1. Une stratégie corrélée markovienne est une application borélienne σ définie sur $[-T,0] \times K$ telle que pour tout $(t,k) \in [-T,0] \times K$, $\sigma(t,k) \in \Delta(\prod_i A_i(k))$.*

2. Une stratégie corrélée markovienne σ est un équilibre corrélé parfait markovien si pour tout état k et tout $t \in [-T,0]$, $\sigma(t,k)$ est un équilibre corrélé de $u^\sigma(t,k,.)$.

3. Une stratégie corrélée markovienne σ est un équilibre corrélé public parfait markovien si pour tout état k et tout $t \in [-T,0]$, $\sigma(t,k)$ est une combinaison convexe des équilibres de Nash de $u^\sigma(t,k,.)$.

Un équilibre corrélé publique est tel que tous les joueurs reçoivent un signal avant de choisir leur action. Pour les jeux sous forme normale, cela revient à choisir une combinaison convexe d'équilibres de Nash.

Existence d'un équilibre parfait markovien

On a introduit les outils qui nous permettent à présent de démontrer le théorème important suivant.

Théorème 7.

-Tout jeu stochastique admet un équilibre corrélé publique parfait markovien.

-Tout jeu de révision stochastique asynchrone admet un équilibre parfait markovien.

On constate que si le jeu est asynchrone, alors dans chaque état il y a un seul joueur qui a un ensemble d'action non réduit à un élément. Ainsi, le paiement continu $u^\sigma(t,k,.)$ est un jeu à un joueur à chaque opportunité de révision. Par conséquent un équilibre corrélé est simplement un équilibre de Nash en actions mixtes, et le second cas du théorème est une conséquence du premier.

Le plan de la preuve est le suivant. D'une part, on associe à une stratégie σ une application U^σ. D'autre part, à chaque fonction de paiement continu V on associe l'ensemble des profils de stratégies que joue un équilibre corrélé publique pour le paiement continu donné par V en presque tout t. Le but est de montrer que cette correspondance admet un point fixe en utilisant le théorème de Kakutani-Glicksberg (introduit au chapitre 1). Il nous suffit donc de vérifier que les hypothèses du théorème sont bien satisfaites. Techniquement le plus difficile est de choisir les bonnes topologies pour les stratégies et les paiements et de montrer les hypothèses de continuité requises pour appliquer le théorème de Kakutani-Fan-Glicksberg. La topologie appropriée pour les stratégies est la topologie faible, c'est-à-dire de la convergence étroite. Une étape importante est de montrer que l'ensemble des applications de paiement de continuité n'est pas seulement faiblement compact, mais qu'il est aussi fortement compact. Il est essentiel de montrer que la correspondance de meilleure réponse a un graphe fermé. Il est à noter que la topologie faible identifie les fonctions mesurables à un ensemble de mesure nulle près. La preuve sera complétée à l'aide d'un théorème de sélection mesurable pour définir un équilibre en tout point.

Preuve du théorème d'existence

Définition 16. *Une stratégie corrélée markovienne σ est un équilibre corrélé parfait markovien faible si pour tout k et pour presque tout $t \in [-T,0]$, $\sigma(t,k)$ est un équilibre corrélé de $u^\sigma(t,k,.)$.*

Topologie des profils de stratégies markoviennes

Une stratégie corrélée markovienne σ définit pour tout $t \in [-T, 0]$, $k \in K$ et $a \in A(k)$, la probabilité $\sigma(t)[a|k]$ du profil d'action a au temps t étant donné un état k. On considère le sous-ensemble convexe et compact de $\mathbb{R}^{K\times A}$ suivant :

$$X = \left\{ x = (x_{ka})_{ka} \in \mathbb{R}^{K\times A} : \forall (k,a), x_{ka} \geq 0, \forall k, \sum_{a\in A(k)} x_{ka} = 1 \right\}.$$

L'ensemble Σ des stratégies corrélées markoviennes est l'ensemble des applications boréliennes de $[-T, 0]$ dans X. Puisque X est convexe, Σ est aussi convexe. On munit Σ de la topologie faible-* des fonctions mesurables bornées, ce qui rend Σ compact pour cette topologie. (conséquence du théorème de Banach-Alaoglu-Bourbaki). C'est ce que l'on démontre dans le lemme suivant :

Lemme 2 (). *Soit $X \subset \mathbb{R}^D$ un ensemble convexe et compact de dimension finie. Soit $\mathcal{F}$ l'ensemble des fonctions boréliennes de $[-T, 0]$ sur X. On dit que la suite $(f_n)_{n\in\mathbb{N}} \subset \mathcal{F}$ converge faiblement vers $f \in \mathcal{F}$ si pour toute fonction Lebesgue-intégrable g et pour toute coordonnée $d = 1, ..., D$,*

$$\int_{-T}^{0} (f_n)_d(t)g(t)dt \underset{n\to\infty}{\longrightarrow} \int_{-T}^{0} f_d(t)g(t)dt.$$

Alors, $\mathcal{F}$ est convexe et compact pour cette convergence faible.

Preuve. On montre tout d'abord que pour des fonctions uniformément bornées, la topologie faible-* et la topologie faible de L^2 coïncident. Puisque toute fonction bornée sur $[-T, 0]$ est intégrable et de carré intégrable, il est suffisant de démontrer le fait suivant :

Fait 1. *Soit $f_n : [-T, 0] \to \mathbb{R}$ une suite de fonctions mesurables uniformément bornées, c'est-à-dire qu'il existe $M > 0$ tel que $|f_n(t)| \leq M$ pour tout n et tout t. Alors les deux conditions suivantes sont équivalentes :*

1. *$\int f_n(t)g(t)dt \to \int f(t)g(t)dt$ pour tout g Lebesgue intégrable,*

2. *$\int f_n(t)g(t)dt \to \int f(t)g(t)dt$ pour toute fonction mesurable bornée g.*

Preuve du fait. Puisque toute fonction mesurable bornée sur $[-T, 0]$ est Lebesgue intégrable, le sens $\Leftarrow$ de la preuve est évident. Supposons donc que $\int f_n g \to \int fg$ pour toute fonction mesurable bornée g et fixons une fonction Lebesgue intégrable h. Pour tout entier p, soit $h_p = h.\mathbb{1}_{\{|h|\leq p\}}$, où $\mathbb{1}_{(.)}$ est la fonction indicatrice. Pour tout entier p, la fonction h_p est bornée, et donc $\int f_n h_p \to \int fh_p$. On écrit alors :

$$\left|\int f_n h - \int fh\right| \leq \left|\int f_n h - \int f_n h_p\right| + \left|\int f_n h_p - \int fh_p\right| + \left|\int fh_p - \int fh\right|.$$

Fixons $\epsilon > 0$ et considérons le membre de droite. Puisque tous les f_n sont bornées par M, le premier terme est borné par $M \left|\int h_p - \int h\right|$. Par le théorème de convergence dominée, ce terme tend vers 0 quand $p > p_0$ assez grand. Il est donc inférieur à $\epsilon/4$ pour p assez grand. Toujours avec le théorème de convergence dominée on peut montrer que le troisième terme tend vers 0 quand p tend vers l'infini, et est donc plus petit que $\epsilon/4$ pour $p > p_1$ assez grand. Fixons $p \geq \max\{p_0, p_1\}$. Le second terme est plus petit que $\epsilon/2$ pour n assez grand. La preuve est terminée.

On revient à présent à la preuve du lemme. On applique le théorème de Banach-Alaoglu (voir annexe). L'ensemble $\mathcal{F}$ est clairement borné et convexe. C'est facile à voir car comme X est supposé convexe et compact, alors $\mathcal{F}$ est fermé pour la topologie forte sur les fonctions bornées. A présent, un ensemble convexe qui est fortement fermé est aussi faiblement fermé. Par le fait précédent, sur un ensemble de fonctions uniformément bornées, la topologie faible (c'est-à-dire L^2) coïncide avec la topologie faible-*, donc $\mathcal{F}$ est faible-* fermé, ce qui conclut la preuve du lemme.

Corollaire 1. *L'ensemble Σ des stratégies corrélées markoviennes est un ensemble convexe qui est compact pour la topologie faible.*

Transitions et paiements

On se fixe une stratégie corrélée markovienne $\sigma : [-T, 0] \times K \to \Delta(A)$ qui est une application mesurable telle que $\sigma(t, k) \in \Delta(A(k))$ pour tout t et pour tout k. On définit également une fonction $\bar{\sigma} : [0, 1] \to (\Delta(K \times A))^K$ comme suit :

$$\bar{\sigma}(t)[k', a|k] = \sigma[a|k]q(k'|k, a).$$

C'est la probabilité que si à l'instant t l'état est k et que la stratégie (corrélée) des joueurs est σ, alors l'action jouée est a et l'état suivant est k'.

On note $P^\sigma_{kj}(s,t) = P_\sigma(k_t = j, k_s = k)$, et donc $P^\sigma(s,t) = (P^\sigma_{kj}(s,t))_{kj}$ est la matrice de transition du temps s au temps t. On note $M^\sigma_{kj}(t) = \lambda_k \bar{\sigma}(t)[j|k] - \lambda_k \mathbb{1}_{\{k=j\}}$.

Lemme 3. *La fonction $(s,t) \mapsto P^\sigma(s,t)$ est solution de l'équation différentielle suivante pour tout $t \in (-T, 0]$ et pour presque tout $s \in [-T, t)$:*

$$\forall t \in (-T,0], \frac{d}{ds}P(s,t) = -M^\sigma(s)P(s,t), \ \ s \text{ presque sûrement}, P(t,t) = I$$

où I est la matrice identité.

On reconnaît une équation rétrograde.

Preuve. On fixe $s < t$ et $h > 0$ et on écrit :

$$\mathbb{P}_\sigma(k_t = j | k_{s-h} = k) = \sum_l \mathbb{P}_\sigma(k_t = j, k_s = l)\mathbb{P}_\sigma(k_s = l, k_{s-h} = k).$$

Pour un processus de Poisson la probabilité d'avoir deux opportunités dans un intervalle de longueur h est $o(h) = h\epsilon(h)$ avec $\epsilon(h) \to 0$ quand h tend vers 0. Par conséquent, pour $l \neq k$,

$$\mathbb{P}_\sigma(k_s = l, k_{s-h} = k) = \int_0^h \lambda_k \exp(-\lambda_k z)\bar{\sigma}(s-h+z)[l|k]dz + o(h).$$

Pour $l = k$,

$$\mathbb{P}_\sigma(k_s = k, k_{s-h} = k) = \int_0^h \lambda_k \exp(-\lambda z)\bar{\sigma}(s-h+z)[k|k]dz + \exp(-\lambda_k h) + o(h).$$

Donc,

$$\frac{P^\sigma_{kj}(s-h,t) - P^\sigma_{kj}(s,t)}{h} =$$

$$\sum_l P^\sigma_{lj}(s,t)\frac{1}{h}\int_0^h \lambda_k \exp(-\lambda_k z)\bar{\sigma}(s-h+z)[l|k]dz - P^\sigma_{kj}(s,t)\frac{1-\exp(-\lambda_k h)}{h} + \epsilon(h).$$

Par le théorème de convergence dominée, pour presque tout $s \in [-T, t)$,

$$-\frac{d}{ds}P^\sigma_{kj}(s,t) = \lim_{h\to 0}\frac{P^\sigma_{kj}(s-h,t) - P^\sigma_{kj}(s,t)}{h} = \sum_l P^\sigma_{lj}(s,t)\lambda_k\bar{\sigma}(s)[l|k] - P^\sigma_{kj}(s,t)\lambda_k$$

ce que l'on voulait.

On déduit du lemme précédent que l'application $\sigma \mapsto P^\sigma$ est continue pour la topologie faible sur l'ensemble des stratégies.

Lemme 4. *Si σ_n converge faiblement vers σ, alors pour tout $0 \leq s < t \leq 1$, $P^{\sigma_n}(s,t)$ converge vers $P^\sigma(s,t)$.*

Preuve. Clairement, M^{σ_n} converge faiblement vers $M := M^\sigma$. Par le lemme précédent, pour tout $s < t$,

$$P^{\sigma_n}(s,t) = I + \int_s^t M^{\sigma_n}(z)P^{\sigma_n}(z,t)dz.$$

Fixons t. La suite de fonctions $g^n(z) = M^{\sigma_n}P^{\sigma_n}(z,t)$ est uniformément bornée (car M^{σ_n} l'est, la suite vient en appliquant le lemme de Gronwall à P^{σ_n}), et donc admet une sous-suite qui converge faiblement et une limite au sens faible g. On veut montrer que $Q(s) = P^\sigma(s,t)$. Pour cela, on va montrer que $g(z) = M(z)Q(z)$ pour presque tout $z \in [s,t]$, ce qui revient à montrer que pour toute fonction ϕ bornée,

$$\int_s^t \phi(z)g(z)dz = \int_s^t \phi(z)M(z)Q(z)dz.$$

Pour toute fonction ϕ bornée, $\int_s^t \phi(z)g^n(z)dz$ converge vers $\int_s^t \phi(z)g(z)dz$. Maintenant,

$$\int_s^t \phi(z)M^{\sigma_n}(z)P^{\sigma_n}(z,t)dz = \int_s^t \phi(z)M^{\sigma_n}(z)Q(z)dz + \int_s^t \phi(z)M^{\sigma_n}(z)(P^{\sigma_n}(z,t) - Q(z))dz.$$

Par convergence faible, le premier terme à droite converge vers $\int_s^t \phi(z)M(z)Q(z)dz$. Pour le second terme,

$$\left|\int_s^t \phi(z)M^{\sigma_n}(z)(P^{\sigma_n}(z,t) - Q(z))dz\right| \leq C\int_s^t |\phi(z)|.|P^{\sigma_n}(z,t) - Q(z)|dz$$

où on a utilisé que $||M^{\sigma_n}||$ est bornée par $C = \max_k \lambda_k$. Puisque $P^{\sigma_n}(z,t)$ converge vers $Q(z)$ pour tout z, le membre de droite tend vers 0 par le théorème de convergence dominée.

On conclut que $g(z) = M(z)Q(z)$ pour presque tout $z \in [s,t]$, et donc $I + \int_s^t M(z)Q(z)dz = Q(s)$ pour tout s. Par unicité des solutions du problème de Cauchy, $Q(s) = P^{\sigma}(s,t)$ pour tout s.

De plus, nous avons vu que si σ_n converge faiblement vers σ alors P^{σ_n} converge en tout point vers P^{σ} le long d'une sous-suite. Si P^{σ_n} ne converge pas en tout point vers P^{σ}, alors on peut trouver $s < t$ et une sous-suite $\psi(n)$ telle que $P^{\psi(n)}(s,t)$ converge vers $R(s,t) \neq P^{\sigma}(s,t)$. Puisque $\sigma_{\psi_{(n)}}$ converge faiblement vers σ, d'après le raisonnement précédent, on déduit que $P^{\psi(n)}$ converge en tout point vers P^{σ} en suivant une sous-sous-suite, ce qui est une contradiction.

Une première conséquence est que les paiements de continuité sont continus pour la topologie faible.

Corollaire 2. *Si σ_n converge faiblement vers σ, alors pour chaque joueur i, pour chaque état k, et à chaque instant t de $[-T,0]$, le paiement de continuité $U_i^{\sigma_n}(t,k)$ converge vers $U_i^{\sigma}(t,k)$.*

Preuve. Le paiement final espéré peut s'écrire

$$U_i^{\sigma_n}(t,k) = \mathbb{E}_{\sigma_n}^{t,k}[r_i(k)] = \sum_l r_i(l) P_{kl}^{\sigma_n}(t,0)$$

La convergence suit directement du lemme précédent.

Une autre conséquence importante est que le paiement de continuité $U_i^{\sigma}(t,k)$ est une fonction lipschitzienne en temps. Plus précisément, si $P^{\sigma}(t,0) = I + \int_t^0 M^{\sigma}(z)P^{\sigma}(z,0)dz$, on peut écrire :

$$U_i^{\sigma}(t,k) = \sum_l P_{kl}^{\sigma}(t,0) r_i(l) = r_i(k) + \int_t^0 \langle M^{\sigma}(z)P^{\sigma}(z,0), r\rangle dz, \tag{2.1}$$

où $\langle M^{\sigma}(z)P^{\sigma}(z,0), r\rangle$ est un raccourci pour $\sum_l \sum_m M_{km}^{\sigma}(z)P_{ml}^{\sigma}(s,z)r(l)$. Si on note $||r|| = \max_{i,k}|r_i(k)|$ le plus grand paiement en valeur absolue dans le jeu et $C = \max_k \lambda_k \times ||r||$. On trouve,

$$\forall (s,t), \max_{i,k} |U_i^{\sigma}(t,k) - U_i^{\sigma}(s,k)| \leq C|t-s|. \tag{2.2}$$

Pour conclure cette sous-section, on doit préciser les topologies que l'on considère pour les paiements de continuité ;

Lemme 5. *Soit $\mathcal{V}$ l'ensemble des fonctions $v : [-T,0] \to [-||r||, ||r||]^{\mathcal{N}\times K}$ telles que*

$$\forall (s,t), \max_{i,k} |v_i(t,k) - v_i(s,k)| \leq C|t-s|.$$

Alors $\mathcal{V}$ est convexe et compact aussi bien pour la topologie faible que pour la topologie forte induite par la convergence uniforme. De plus, l'application de paiement continu $\sigma \mapsto U^{\sigma}$ envoie Σ sur $\mathcal{V}$.

Preuve. Clairement, $\mathcal{V}$ est un ensemble convexe d'après le Lemme 2, il est compact pour la topologie faible. De plus, les fonctions de $\mathcal{V}$ sont lipschitziennes avec la même constante. Cet ensemble est donc équicontinu et par le théorème d'Ascoli il est compact pour la topologie de la convergence uniforme. Par l'équation (2.2), c'est vrai pour tout $\sigma \in \Sigma$, $U^{\sigma} \in \mathcal{V}$.

Le point fixe

On prouve ici l'existence d'un équilibre corrélé parfait markovien faible. On définit deux correspondances $\varphi_{\mathcal{V}} : \Sigma \to \mathcal{V}$ et $\varphi_{\Sigma} : \mathcal{V} \to \Sigma$ telle qu'un point fixe de la correspondance

$$(\sigma, v) \mapsto \phi(\sigma, v) = \phi_{\Sigma}(v) \times \phi_{\mathcal{V}}(\sigma),$$

c'est-à-dire un couple $(\sigma, v) \in \Sigma \times \mathcal{V}$ tel que $(\sigma, v) \in \phi(\sigma, v)$, vérifie que σ un ECPMf de la fonction de paiement de continuité v.

La composante en $\mathcal{V}$ est une application $\sigma \mapsto U^{\sigma}$ qui à une stratégie σ associe sa fonction de paiement continu : $\phi_{\mathcal{V}}(\sigma) = \{U^{\sigma}\}$.

La composante en Σ est une application qui à une fonction de paiement continu v associe l'ensemble des stratégies qui sont des équilibres corrélés pour les jeux sous forme normale induits par v en presque tout temps. Précisément, une stratégie σ appartient à $\phi_{\Sigma}(v)$ si pour tout état k, pour tout joueur i, pour chaque paire d'action a_i, b_i et pour presque tout t,

$$\sum_{a_i,l} \sigma(t)[a_i, a_{-i}|k] q(l|a_i, a_{-i}, k) v_{i,l}(t) \geq \sum_{a_{-i},l} \sigma(t)[a_i, a_{-i}|k] q(l|b_i, a_{-i}, k) v_{i,l}(t). \tag{2.3}$$

Le membre de droite est le paiement espéré du joueur i quand il devrait jouer a_i et qu'il le joue effectivement, et le membre de droite est le paiement espéré par le joueur i s'il devrait jouer a_i et qu'il joue b_i à la place. En d'autres termes, $\sigma(t)[.|k]$ est un équilibre corrélé du jeu joué au temps de révision t à l'été k, et les gains sont déterminés par la fonction de transition et la fonction v.

Proposition 3. *La correspondance $\phi : \Sigma \times \mathcal{V} \to \phi \times \mathcal{V}$ définie par $\phi(\sigma, v) = \phi_\Sigma(v) \times \phi_\mathcal{V}(\sigma)$ admet un point fixe.*

Preuve. Elle consiste à appliquer le théorème de Kakutani-Glicksberg vu au chapitre 1. $\Sigma \times \mathcal{V}$ est convexe et faiblement compact. Pour tout t, l'ensemble des $\sigma(t)$ qui satisfont l'équation (2.3) pour tout k, i, a_i, b_i est non-vide et convexe, comme ensemble des équilibres corrélés d'un jeu fini. L'ensemble $\phi_\Sigma(v)$ est donc convexe pour tout v, il suit que les valeurs de ϕ sont des ensembles convexes.

Fait 2. *Pour tout $v \in \mathcal{V}$, $\phi_\sigma(v)$ est faiblement compact.*

Preuve. $\phi_\Sigma(v)$ est clairement fermé pour la topologie de la convergence uniforme. Comme il est convexe, il est aussi compact pour la topologie faible d'après le Lemme 2

La correspondance ϕ est donc à valeurs convexes et compactes. Maintenant, on prouve que ϕ a un graphe faiblement fermé. Puisque l'application $\sigma \mapsto U^\sigma$ est continue, il suffit de montrer le Lemme suivant :

Lemme 6. *ϕ_Σ a un graphe fermé.*

Preuve. Par le théorème d'Ascoli, il existe une sous-suite de (g_n) qui converge uniformément vers g. Soit ϕ Lebesgue-intégrable, on écrit,

$$\int \phi(t) f_n(t) g_n(t) dt = \int \phi(t) f_n(t) g(t) dt + \int \phi(t) f_n(t)(g_n(t) - g(t)) dt$$

Par convergence faible de f_n, le premier terme du côté droit converge vers $\int \phi(t) f(t) g(t) dt$. La valeur absolue du second terme est bornée par en-dessous par $\int |\phi(t)|.|g_n(t) - g(t)| dt$ à une constante multiplicative près, qui converge vers 0 par le théorème de convergence dominée.

Grâce à ce fait on peut extraire une sous-suite et on obtient par passage à la limite :

$$\int \sum_{a_{-i},l} \sigma(t)[a_i, a_{-i}|k] q(l|a_i, a_{-i}, k) v_{i,l}(t) f(t) dt \geq \int \sum_{a_{-i},l} \sigma(t)[a_i, a_{-i}|k] q(l|b_i, a_{-i}, k) v_{i,l}(t) f(t) dt$$

C'est vrai pour tout k, i, a_i, b_i, donc $\sigma \in \phi_\Sigma(v)$, ce que l'on voulait.

Finalement la correspondance ϕ admet un point fixe qui est un équilibre corrélé parfait markovien faible. On va enfin montrer l'existence d'un équilibre corrélé parfait markovien à partir de ce premier.

Fait 3. *Si σ est un équilibre corrélé parfait markovien faible, alors il existe un équilibre corrélé parfait $\hat{\sigma}$ tel que $\sigma(t) = \hat{\sigma}(t)$ pour presque tout $t \in [-T, 0]$.*

Preuve. Soit σ un équilibre parfait markovien faible et $U^\sigma(t, k)$ l'application de paiement continu associée. Par le Lemme 3 et l'équation (2.1), si $\sigma(t) = \hat{\sigma}(t)$ pour presque tout $t \in [-T, 0]$, les fonctions de paiement continu associées sont les mêmes : $U^\sigma = U^{\hat{\sigma}}$.

Soit $Z \subseteq [-T, 0]$ un ensemble mesurable de mesure nulle tel que pour tout k, pour tout $t \in [-T, 0] \backslash Z$, $\sigma(t, k)$ est un équilibre corrélé pour le jeu dont la matrice de gain est $u^\sigma(t, k, .)$.

On choisit alors $\hat{\sigma}$ une sélection de la correspondance qui à $t \in Z$ associe l'ensemble des équilibres corrélés des matrices de jeux $(u^\sigma(t, k, .))_{k \in K}$, et $t \in [-T, 0] \backslash Z$ au singleton $\{\sigma(t)\}$. Cette correspondance est mesurable et admet une sélection mesurable (admis).

Le paiement continu induit n'est pas modifié. Alors, $\hat{\sigma}$ satisfait les conditions d'équilibre pour tout $t \in [-T, 0]$.

2.2 Jeux de révision à deux joueurs

Modèle

On considère un jeu à deux joueurs $\big((X^i)_{i=1,2}, (u_i)_{i=1,2}\big)$ le jeu sous forme normale où X^i est l'ensemble supposé fini des actions du joueur i avec $|X^i| \geq 2$ et $u_i : X \to \mathbb{R}$ est la fonction de gain du joueur i. Ici, $X = X_1 \times X_2$ est l'ensemble des profils d'actions. Le modèle est le suivant : le temps est continu, $t \in [-T, 0]$, et le jeu sous forme normale est joué au temps $t = 0$. A l'instant $t = -T$, un profil d'action est donné, c'est un paramètre du jeu de révision. Dans l'intervalle $(-T, 0]$ les joueurs sont susceptibles d'obtenir, indépendamment, des opportunités de révision, c'est-à-dire qu'en certains temps ils peuvent changer leur action. En fonction de deux paramètres de Poisson p_1 et p_2 de paramètres respectifs $\lambda_1 > 0$ et $\lambda_2 > 0$. Comme ces processus de Poisson sont indépendants, la probabilité que les deux joueurs changent leur action en même temps est nulle.

Nous avons défini les notions de stratégies pour les jeux de révision dans la section précédente. On notera pour cette section $H_i^n(t)$ et $H_i^r(t)$ les sous-ensembles des histoires non finies pour le joueur i telles que ce dernier n'a pas d'opportunité de révision à l'instant t, et celles où il a une opportunité de révision à l'instant t, respectivement. L'ensemble des histoires possibles se note alors : $H_i := \cup_{t \in [-T,0]} H_i^n(t) \cup H_i^r(t)$.

Une histoire pour le joueur i au temps $t \in (-T, 0]$ peut prendre l'une des deux formes suivantes, selon si le joueur i reçoit une opportunité au temps t ou non.

$$h_i(t) = \left((t_1^k, x_1^k)_{k=0}^{k_1}, (t_2^k, x_2^k)_{k=0}^{k_2} \right) \in H_i^n(t)$$

si le joueur i n'a pas eu d'opportunité de révision au temps t, et

$$h_i(t) = \left((t_1^k, x_1^k)_{k=0}^{k_1}, (t_2^k, x_2^k)_{k=0}^{k_2}, t \right) \in H_i^r(t)$$

s'il en reçoit une au temps t, où $t_i^0 := -T$, k_i est un entier positif ou nul pour $i \in \{1, 2\}$, $-T < t_i^1 < ... < t_i^{k_i} < t$ pour $i \in \{1, 2\}$. Le profil d'action x_i^0 est celui donné au joueur i au début, k_i est le nombre d'opportunités de révision qu'a eu le joueur i avant le temps t, t_i^k est le temps où le joueur i a reçu sa k-ième opportunité de révision et x_i^k est l'action du joueur i préparée au temps t_i^k.

Le but de cette section est de démontrer l'existence d'un unique équilibre en sous-jeux parfaits dans une certaine classe de jeux.

La preuve de ce théorème s'appuie sur une idée de récurrence rétrograde (backward induction en anglais) en temps continu. Introduisons tout d'abord ce lemme qui nous servira par la suite :

Lemme 7. *On suppose que pour tout $t \in (-T, 0]$ il existe $\epsilon > 0$ tel que l'événement $A_{t'}$ est vrai pour tout $t' \in (t - \epsilon, t]$ si l'événement $A_{t''}$ est vrai pour tout $t'' > t$. Alors, pour tout $t \in (-T, 0]$, l'événement A_t est vrai.*

Preuve. voir Calcagno, (2013)

Jeux à intérêt commun

On s'intéresse dans cette sous-section à des jeux de révisions dont le jeu sous forme normale associé a un équilibre de Pareto strictement dominant.

Définition 17. *Un équilibre de Pareto strictement dominant est un profil d'action $x^* \in X = X_1 \times X_2$ tel que*

$$\forall i \in \{1, 2\}, \forall x \in X \backslash \{x^*\}, u_i(x^*) > u_i(x).$$

On dit qu'un jeu est à intérêt commun s'il admet un équilibre de Pareto strictement dominant. Il est à noter que si x^* est un équilibre de Pareto strictement dominant, alors c'est un équilibre de Nash strict.

Théorème 8. *On considère un jeu sous forme standard à intérêt commun et x^* son équilibre de Pareto strictement dominant. Alors pour tout $\epsilon > 0$ il existe $T' > 0$ tel que pour tout $T > T'$, dans tout équilibre de révision, $x(0) = x^*$ avec probabilité $1 - \epsilon$.*

Preuve. voir Calcagno, (2013)

Plusieurs remarques :

— Premièrement, si les joueurs choisissent leur action à l'instant $-T$, on peut montrer qu'ils ont tout intérêt à choisir dès le départ de jouer x^*. Alors ils ne changent plus durant toute la durée du jeu.

— Deuxièmement, en temps long le jeu de révision se finit sur l'équilibre de Pareto strictement dominant même s'il y a des risques pour en changer en commençant dans un autre équilibre de Nash pas optimal.

(Le cas à n joueurs)

On peut également étudier les valeurs de révision d'un jeu à un nombre fini de joueurs avec un équilibre de Pareto strictement dominant x^*. Dans la suite on note I le nombre de joueurs, de cardinal fini, et λ_i le paramètre de la loi de Poisson qui donne les opportunités de révision du joueur i. On note aussi $r_i = \lambda_i / \sum_{j \in I} \lambda_j$. On suppose $\mathrm{Card}(I) = n \geq 2$.

Définition 18. *Un jeu à intérêt commun est un K-jeu coopératif si pour tout i, $j \in I$ et $x \in X$,*

$$\frac{u_i(x^*) - u_i(x)}{u_i(x^*) - \underline{u}_i} \leq K \frac{u_j(x^*) - u_j(x)}{u_j(x^*) - \underline{u}_j},$$

où $\underline{u}_i = \min_x u_i(x)$.

La valeur minimum de K est 1 quand le jeu est à coopération pure, où les joueurs ont exactement la même fonction de gain. Plus K est grand et plus les préférences des joueurs divergent.

On note $\alpha = \min_{i \in I} r_i$ et $\beta = \min_{i \in I, i \neq j^*} r_i \geq \alpha$ où j^* est un élément quelconque de $\arg\min_{i \in I} r_i$

Théorème 9. *Supposons qu'un jeu à intérêt commun est un K-jeu coopératif. Soit x^* son équilibre de Pareto strictement dominant. Si on a*

$$(1-\alpha-\beta)K < 1-\beta,$$

alors pour tout $\epsilon > 0$, il existe T' tel que pourtout $T > T'$, dans tout équilibre de révision, $x(0) = x^*$ avec une probabilité supérieure à $1-\epsilon$.

Autrement dit sous l'hypothèse de cette inégalité, quelle que soit la stratégie jouée par les autres joueurs, chaque joueur i a intérêt à jouer x_i^*.

A priori cette hypothèse ne semble pas importante et on pourrait croire que l'on peut s'en passer, mais il existe pourtant des jeux avec des stratégies stables telles que les joueurs individuellement n'ont pas envie de jouer l'équilibre.

Figure 2.1 – Contre-exemple

a_3 :

	P_2: a_2	b_2
P_1: a_1	$1,1,1$	$-1,-1,-1,$
b_1	$-1,-1,-1$	$0,0,-1$

b_3 :

	P_2: a_2	b_2
P_1: a_1	$-1,-1,-1$	$-1,0,0$
b_1	$0,-1,0$	$-1,-1,-1$

Dans la Figure 2.1 on prend $\lambda_1 = \lambda_2 = \lambda_3$. On considère le profil de stratégie suivant, qui est un équilibre en sous-jeux parfait markovien : chaque joueur prépare une action différente de son action actuelle si et seulement si le profil d'actions courant a exactement deux a (et par conséquent un seul b). Le jeu va s'avérer donner des valeurs de révisions différentes en fonction de la position de départ. Ainsi quand le profil d'actions courant est (a_i, a_j, b_k), le joueur i préfère jouer b_i. En effet s'il joue b_i il se retrouve dans un équilibre de Nash et garantit 0, tandis que s'il avait gardé a_i, ou bien la prochaine opportunité est pour lui, et il peut garantir 0, ou bien c'est le joueur j et le joueur i gagne -1, ou bien c'est le joueur k et le joueur i gagne 1, ou bien il n'y a plus d'opportunité et le joueur i gagne -1.

Plusieurs remarques s'imposent :

- On remarque tout d'abord avec cet exemple que la constante K donnée par l'inéquation du théorème précédent est optimale, dans le sens où le théorème n'est pas satisfait en cas d'égalité.
- Avec le contre-exemple, on voit également un exemple de jeu de révision ayant au moins deux équilibres en sous-jeux parfaits dont l'un est strictement moins bon que l'autre, pour tous les joueurs.

On verra plus tard des jeux de révision dans lesquels on peut calculer explicitement la valeur et les stratégies optimales associées.

Jeux d'intérêt opposé

Nous avons étudié des jeux dans lesquels un unique profil de stratégie était le meilleur pour les deux joueurs. Désormais nous nous concentrons sur des jeux à deux joueurs pour lesquels les joueurs ont plusieurs "meilleurs" profils d'action.

Figure 2.2 – Jeux à intérêt opposé

	G	D
H	$3,3$	$0,1$
B	$0,5$	$1,6$

	G	D
H	$2,1$	$0,0$
B	$0,0$	$1,2$

On va étudier les jeux à deux joueurs, à deux actions pour chaque joueur et ayant deux équilibres de Nash stricts (H, G) et (B, D) tels que :

$$u_1(H,G) > u_1(B,D) \text{ et } u_2(H,G) < u_2(B,D). \tag{2.4}$$

La première inégalité dit que le joueur 1 préfère l'équilibre de Nash en haut à gauche, et la seconde que le joueur 2 préfère au contraire l'équililibre de Nash en bas à droite.

	G	D
H	$u_1(H,G), u_2(H,G)$	$u_1(H,D), u_2(H,D)$
B	$u_1(B,G), u_2(B,G)$	$u_1(B,D), u_2(B,D)$

Soit

$$t_1^* = -\frac{1}{\lambda_1+\lambda_2}\ln\left(\frac{\lambda_1}{\lambda_2}\frac{u_1(B,D)-u_1(H,D)}{u_1(H,G)-u_1(B,D)} + \frac{u_1(H,G)-u_1(H,D)}{u_1(H,G)-u_1(B,D)}\right).$$

et

$$t_2^* = -\frac{1}{\lambda_1+\lambda_2}\ln\left(\frac{\lambda_2}{\lambda_1}\frac{u_2(H,G)-u_2(H,D)}{u_2(B,D)-u_2(H,G)} + \frac{u_2(B,D)-u_2(H,D)}{u_2(B,D)-u_2(H,G)}\right).$$

Théorème 10. *On suppose que le jeu sous forme normale satisfait la condition 2.4. Si $t_1^* \neq t_2^*$, alors il existe un unique équilibre de révision pour tout T. De plus quand $-T$ tend vers $-\infty$,*
— *si $t_1^* > t_2^*$, alors les paiements d'équilibre convergent vers $u_i(H, G)$.*

— *si $t_1^* < t_2^*$, alors les paiements d'équilibre convergent vers $u_i(B, D)$.*

Il est à noter que dans le cas où $t_1^* = t_2^*$, le jeu de révision a plusieurs équilibres.

Le Théorème 10 affirme que pour presque toutes les valeurs possibles des paramètres t_1^* et t_2^* il existe un unique équilibre de révision. L'équilibre tend vers un des deux équilibres de Nash. Celui qui est choisi dépend du ratio λ_1/λ_2 et de la fonction u. Dans la Figure 2.2, si $\lambda_1 = \lambda_2$, alors $t_1^* > t_2^*$ dans le jeu de gauche et $t_1^* = t_2^*$ dans celui de droite. Alors l'issue du jeu de gauche est (H, G) mais le théorème ne dit pas quelle est l'issue du jeu de droite. Si $\lambda_1 < \lambda_2$, alors le théorème s'applique sur le jeu de droite et l'issue du jeu est (H, G) tandis que si $\lambda_1 > \lambda_2$ le théorème s'applique encore mais l'issue du jeu est cette fois (B, D).

La première étape de la preuve du théorème 10 montre que quand t tend vers 0, chaque joueur préfère strictement préparer sa meilleure réponse à l'action courante de l'autre joueur dans le jeu sous forme normale en fonction de celle que le joueur adverse a. Ainsi, dans les jeux de la Figure 2.2, quand le temps est proche de zéro les joueurs s'écartent des positions (H, D) et (B, G) pour atteindre un des deux équilibres, et ils y restent jusqu'à la fin. Si t est assez loin de zéro, les deux joueurs préparent (H, D), dans l'espoir que l'autre joueur joue leur équilibre préféré. Vient un moment t^* où un des joueurs sera indifférent entre aller sur l'équilibre de Nash de l'autre joueur ou rester sur place, et à partir duquel il essaiera, peu importe le profil d'action, d'aller sur un équilibre de Nash. Ce joueur sera le joueur faible, l'autre sera appelé le joueur fort. Cela peut paraître surprenant, pourtant dans ces jeux à intérêt opposé, il est avantageux pour un joueur d'être lent, c'est-à-dire que u_i est une fonction décroissante en λ_i

Preuve. Par programmation dynamique. La preuve en détail sans outil de programmation dynamique est donnée dans Calcagno, (2013).

Les principes de programmation dynamique seront évoqués dans la section suivante. Ils permettent de résoudre parfaitement des jeux de révisions particuliers.

2.3 Jeux de révision à somme nulle

Il s'agit d'un cas particulier de jeu de révision à deux joueurs dans lequel $g_1 = -g_2$. Nous étudions certaines des propriétés de ces jeux avant de résoudre entièrement les jeux à somme nulle à deux actions par joueurs.

2.3.1 Jeux sous forme normale

Dans cette sous-section on notera X_1 et X_2 les ensembles (finis non vides) des actions des joueurs, et $X := X_1 \times X_2$ l'ensemble des profils d'action. Soit $U : X \to \mathbb{R}$ la fonction de gain du joueur 1. Celle du joueur 2 étant alors $-U$. On notera aussi U pour décrire l'ensemble de ses valeurs $(U(x))_{x \in X}$.

Pour chaque action pure $x_{-i} \in X_{-i}$ du joueur $-i$ on note encore $MR_i^U(x_{-i}) \subseteq X_i$ l'ensemble des meilleures réponses à x_{-i} dans le jeu sous forme normale de matrice de gain U.

Définition 19. *On dit que U est régulière si pour tout joueur i et pour toute action $x_{-i} \in X_{-i}$, $MR_i^U(x_{-i})$ est un singleton.*

On note V la valeur $\min\max$ de ce jeu en stratégies mixtes.

Définition 20. *Le paiement de Stackelberg S_i du joueur i est le gain maximal que peut remporter le joueur i s'il joue avant l'autre joueur :*

$$\max_{x_1} \min_{x_2} U(x_1, x_2) = S_1 \leq V \leq S_2 = \min_{x_2} \max_{x_1} U(x_1, x_2)$$

On remarque que $S_1 = S_2$ si et seulement si le jeu sous forme normale a un équilibre de Nash en stratégies pures.

Jeu de révision

On note encore λ_1 et λ_2 les paramètres des lois de Poisson des deux joueurs, $\lambda = \lambda_1 + \lambda_2$ et $q = \lambda_1/\lambda$. Le jeu de révision est cette fois-ci vu de cette façon : une cloche sonne en suivant un processus de Poisson de paramètre λ. Lorsque la cloche sonne, le joueur 1 a une probabilité q d'avoir une opportunité de révision. Sinon, c'est le joueur 2 qui en a une.

On garde les notations précédentes pour décrire les histoires et les stratégies, markoviennes ou non.

On définit la valeur $\min\max$ et la valeur $\max\min$ du jeu :

$$\bar{u}(t,x) = \inf_{\sigma_2}\sup_{\sigma_1} \mathbb{E}_\sigma[U(x(t))],\ \underline{u}(t,x) = \sup_{\sigma_1}\inf_{\sigma_2} \mathbb{E}_\sigma[U(x(t))].$$

Quand ces deux valeurs sont égales, on dit que le jeu a une valeur, et on note $u(t,x) = \bar{u}(t,x) = \underline{u}(t,x)$.

On notera par la suite $\Gamma(t,x)$ le jeu de révision sur l'intervalle $[t,0]$ et commençant en $x \in X$, $u(t)$ la matrice de gain $(u(t,x))_{x\in X}$, et $||u(t)|| := \max_{x\in X} |u(t,x)|$. Pour tout $t \geq 0$ et $x = (x_1, x_2)$, on note :

$$u^-(t,x) := \min_{y_2\in X_2} u(t,x_1,y_2) \text{ et } u^+(t,x) := \max_{y_1\in X_1} u(t,y_1,x_2).$$

On a clairement $u^- \leq u \leq u^+$.

Valeurs de révision

On commence par rappeler un résultat démontré dans le cas plus général des jeux de révisions stochastiques, et que nous reformulons ici dans le cadre des jeux de révisions à somme nulle :

Lemme 8. *Le jeu de révision admet un équilibre parfait markovien σ. Le jeu de révision $\Gamma(t,x)$ a une valeur donnée par $u(t,x)$. Cet équilibre et cette valeur sont tels que pour tout $x \in X$ et pour presque tout $t \in [-T,0]$,*

$$\sigma_i(t,x_{-i}) \in MR_i^{u(t)}(x_{-i}).$$

Pour interpréter cette équation, il faut voir le jeu de révision à chaque instant t comme un jeu dont la matrice de gain a les coefficients $u(t,x)$. Afin de maximiser son espérance de gain, le joueur i joue dans le jeu de révision comme s'il jouait dans le jeu de Stackelberg S_{-i} dans le jeu sous forme normale de matrice de gain $u(t)$. Ainsi, pour tout temps restant $|t| \geq 0$ avant la fin, le paiement espéré est $u(t,x)$ et ne dépend que de l'état actuel et du temps restant.

A partir de maintenant on appelle valeur de révision la fonction $(t,x) \mapsto u(t,x)$. Introduisons à présent quelques propriétés de la fonction valeur.

Lemme 9. *Pour tout $x \in X$, la fonction $u(.,x)$ est $2\lambda||U||$-lipschitzienne.*

Preuve. Soit $t' < t < 0$. Les histoires de $\Gamma(t,x)$ sont aussi des histoires de $\Gamma(t',x)$, donc les stratégies σ_i' de $\Gamma(t',x)$ induisent par restriction des stratégies dans $\Gamma(t,x)$. Inversement les stratégies σ_i de $\Gamma(t,x)$ peuvent être vues comme des restrictions de stratégies sur $\Gamma(t',x)$. Pour toute stratégie σ_i de $\Gamma(t,x)$, on définit la stratégie $\hat{\sigma}_i$ de $\Gamma(t',x)$ telle que le joueur i joue σ_i sur $[t,0]$ et ne révise pas son action sur $[t',t)$. Ainsi, on a

$$u(t,x) = \sup_{\hat{\sigma_1}}\inf_{\hat{\sigma_2}} \mathbb{E}_{\hat{\sigma}}[u(x(t))].$$

Pour un profil de stratégies $\sigma = (\sigma_1,\sigma_2)$ fixé, la probabilité d'avoir au moins un temps de révision dans l'intervalle $(t',t]$ est $(1-e^{-\lambda|t-t'|})$ alors $P_{\sigma'}(x(t) = x(t')) \geq e^{-\lambda|t-t'|}$, ce qui implique

$$\mathbb{E}_{\sigma'}[|u(x(t)) - u(x(t'))|] \leq ||U||(1-e^{-\lambda|t-t'|}) \leq 2||U||\lambda|t-t'|.$$

Le résultat suit.

On va maintenant montrer que la fonction valeur est définie par une équation différentielle, à l'aide d'outils venant de la programmation dynamique.

Proposition 4. *Pour tout $t \leq 0$ et pour tout $x \in X$, la fonction valeur $u(t,x)$ du jeu de révision est l'unique solution du système d'équations différentielles suivant :*

$$\forall x \in X, \forall t \leq 0, \frac{\partial u(t,x)}{\partial t} = \lambda_1(u^+(t,x) - u(t,x)) + \lambda_2(u^-(t,x) - u(t,x)),\ u(0,x) = U(x).$$

Preuve. Montrons tout d'abord que la fonction valeur satisfait au principe de programmation dynamique suivant. Pour $s \leq 0$, on note $u^s(t,x)$ la valeur du jeu de révision commençant au temps $t \leq 0$ avec pour profil d'actions initial x et de jeu sous forme normale associé $u(s) = (u(s,x))_{x\in X}$. Alors $u^0(t,x) = u(t,x)$, $u^t(0,x) = u(t,x)$ et plus généralement on a le principe de programmation dynamique :

$$\forall x \in X, \forall t \leq 0, \forall s \in [t,0], u^s(t,x) = u(t+s,x).$$

On considère à présent le jeu commençant au temps $t < 0$, et soit s le temps en lequel arrive la première opportunité de révision. Si $s \leq 0$ et que le joueur 1 a une opportunité de révision, la valeur devient $u(s,\sigma_1(s,x),x_2) = u^+(s,x)$, et si le joueur 2 a une opportunité de révision la valeur devient $u(s,x_1,\sigma_2(s,x)) = u^-(s,x)$. Ainsi :

$$\begin{aligned} u(t,x) &= U(x)e^{\lambda t} + \int_{s=t}^0 \lambda e^{\lambda s}(qu^+(t-s,x) + (1-q)u^-(t-s,x))ds. \\ &= U(x)e^{\lambda t} + \int_{s=t}^0 e^{\lambda(t-s)}(\lambda_1 u^+(s,x) + \lambda_2 u^-(s,x))ds. \end{aligned}$$

Comme u^+ et u^- sont continues, u est de classe $\mathcal{C}^1$ en t, et la valeur du jeu est donnée par le système différentiel suivant :

$$\frac{\partial u(t,x)}{\partial t} = \lambda_1(u^+(t,x) - u(t,x)) + \lambda_2(u^-(t,x) - u(t,x)), \ u(0,x) = U(x).$$

L'application $u \mapsto \lambda_1 u^+ + \lambda_2 u^- - \lambda u$ est 2λ-lipscitzienne, donc par le théorème de Cauchy-Lipschitz, le système d'équation a une unique solution, donnée par la fonction valeur.

De cette caractérisation on déduit :

Corollaire 3. *Le jeu de révision a un équilibre parfait markovien en stratégies pures.*

Preuve. Comme vu ci-dessus, le jeu de révision $\Gamma(t,x)$ a un équilibre, donné par des stratégies qui garantissent la valeur, mais elles peuvent être mixtes. L'argument principal est qu'un joueur qui joue aléatoirement sur plusieurs actions peut en fait jouer avec des actions pures tout en gardant un gain optimal. Le fait qu'on soit dans un jeu à somme nulle est ici très important, dans un cas de jeu à somme non nulle, il est possible qu'un joueur soit indifférent. Nous étudierons un tel exemple avec le jeu de la balle en dernière section.

On détaille à présent la construction de cette stratégie. A cette fin on identifie X_i avec l'ensemble $\{1, ..., |X_i|\}$ de ses éléments ordonnés arbitrairement et dont on fixe l'ordre.

Soit σ^* une stratégie mixte qui garantit la valeur. On définit la stratégie $\sigma(1)$ du joueur 1 dans le jeu $\Gamma(t,x)$ de la manière suivante :
- Si la première opportunité de révision arrive au temps s et qu'elle est pour le joueur 1, alors ce dernier joue

$$x_1' = \min_{x_1 \in X_1} \left(\arg\max_{y_1 \in X_1} u(s, y_1, x_2) \right).$$

C'est-à-dire qu'au lieu de jouer la stratégie optimale (a priori mixte) à la première opportunité de révision, il joue une action pure qui maximise la valeur. Le minimum est pris sur l'ordre qu'on a mis sur X_1. Puis il joue ensuite σ^*.
- Sinon, le joueur 1 joue σ^*.

Contre toute stratégie du joueur 2, cette stratégie $\sigma(1)$ garantit le paiement :

$$U(x)e^{\lambda t} + \int_{s=t}^{0} \lambda e^{\lambda s}(qu^+(t-s,x) + (1-q)u^-(t-s,x))ds.$$

D'après l'équation de la programmation dynamique, c'est précisément la valeur $u(t,x)$, et donc $\sigma(1)$ est une stratégie optimale dans le jeu $\Gamma(t,x)$.

- Soit $n \geq 1$. Si pour un certain $m \leq n$, au m-ième son de cloche c'est le joueur 1 qui a une opportunité de révision, le joueur 1 joue $x_1' = \min_{x_1 \in X_1} \left(\arg\max_{y_1 \in X_1} u(s, y_1, x_2(s)) \right)$, où $x_2(s)$ est l'action courante du joueur 2 au temps s.
- Puis pour $m \geq n$, le joueur 1 suit σ^*.

Par récurrence faible sur n on voit que $\sigma(n)$ est une stratégie optimale pour le joueur 1 dans le jeu $\Gamma(t,x)$. Alors, pour tout t, la cloche va sonner un nombre fini de fois. Ainsi, en faisant tendre n vers l'infini, on voit que la stratégie $\sigma(\infty)$ qui consiste à jouer au temps s l'action pure

$$x_1' = \min_{x_1 \in X_1} \left(\arg\max_{y_1 \in X_1} u(s, y_1, x_2(s) \right)$$

est une stratégie optimale markovienne pure pour le joueur 1.
Par symétrie, le joueur 2 dispose d'une telle stratégie.

Donnons à présent quelques propriétés de régularité de la fonction valeur.

Lemme 10. *La valeur de révision $u(t,x)$ est 1-lipschitzienne en U, lipschitzienne en $\lambda \in (0, +\infty)$ et croissante en $q \in (0,1)$.*

Remarque : il y a là un grand contraste avec les jeux à somme non nulle, comme nous l'avons vu avec le jeu de la bataille des sexes Figure 2.2.

Preuve.
i) $u(t,x)$ **est 1-lipschitzienne en** U**.** On considère deux matrices de gain U et U'. Pour $-T \leq 0$ et $x \in X$, l'espace des stratégies des joueurs dans les jeux de révision $\Gamma(-T,x)$ et $\Gamma'(-T,x)$ est le même. Les espérances de gains données par un profil de stratégies σ sont $\mathbb{E}_\sigma[U]$ et $\mathbb{E}_\sigma[U']$, et ne peuvent différer

que d'au plus $||U-U'|| := \max_x |U(x)-U'(x)|$. Considérons une stratégie optimale dans le jeu dont la matrice de gain du jeu sous forme normale associé est U. La stratégie optimale du joueur 1 garantit un paiement d'au moins $u(-T,x)$, qu'importe ce que le joueur 2 joue. Si le joueur 1 utilise la même stratégie dans l'autre jeu, la différence sera d'au plus $||U-U'||$. Ainsi, $|u(-T,x)-u'(-T,x)| \leq ||U-U'||$ pour tout T et x.
ii) $u(t,x)$ est lipschitzienne en λ. Pour q fixé, augmenter la fréquence en fixant le temps revient à augmenter le temps en fixant la fréquence :

$$u(t,x)|_{\lambda'} = u(\frac{\lambda'}{\lambda}t,x)|_{\lambda}.$$

Comme u est lipschitzienne en t, elle l'est aussi en λ.
iii) Continuité en q. On fixe λ et U. Etant donné un profil de stratégie σ, on note $\mathbb{E}_\sigma$ l'espérance sous P_σ quand le paramètre est q, et $\mathbb{E}'_\sigma$ l'espérance sous P'_σ quand le paramètre est q'. On a la propriété de continuité uniforme suivante :

$$\forall \epsilon > 0, \exists \alpha > 0, |q-q'| \leq \alpha \Rightarrow \forall \sigma, |\mathbb{E}_\sigma - \mathbb{E}'_\sigma| \leq \epsilon.$$

Pour le voir, fixons une stratégie σ et comparons $\mathbb{E}_\sigma$ avec $\mathbb{E}'_\sigma$. Pour évaluer ces espérances, l'important est la distribution du nombre total de coups de cloches, et les alternances entre les joueurs. On affirme que cette distribution est continue par rapport à q. En effet pour tout $\epsilon > 0$ il existe un entier M assez grand pour lequel avec probabilité au moins $1-\epsilon$, le nombre total de coups de cloches est au plus M. Alors si on suppose que le nombre de coups de cloches est $m \leq M$, il y aura un nombre fini d'alternances possibles entre les joueurs. Pour un $m \leq M$ donné, la probabilité conditionnelle sur ces configurations est continue par rapport à q. Donc pour q' proche de q, la différence entre les deux probabilités est majorée par ϵ.

La fonction de gain du jeu de révision $\Gamma(t,x)$ avec paramètre q est alors proche de celle du jeu de révision $\Gamma'(t,x)$ avec paramètre q', uniformément par rapport aux stratégies. Les valeurs de ces deux jeux sont donc proches l'une de l'autre.
iv) Monotonie de q. Soit $q' < q$ et $u(t,x)|_q$ la valeur de révision pour q. On cherche à montrer que $u(t,x)|_{q'} \leq u(t,x)|_q$ pour tout t et x. A $t=0$, on a $u(0,x)|_{q'} = u(0,x)|_q = U(x)$. De l'équation de programmation dynamique, on obtient : $\frac{\partial u(0,x)|_{q'}}{\partial t} \leq \frac{\partial u(0,x)|_q}{\partial t}$ pour tout x. Par le lemme de Gronwall on obtient $u(t,x)|_{q'} \leq u(t,x)|_q$ pour tout t et x.

Valeur de révision limite

Dans cette section on analyse le comportement asymptotique de la valeur de révision quand le temps $-T$ tend vers $-\infty$, ou de façon équivalente quand la fréquence de révision λ tend vers $+\infty$ (à $-T$ fixé).

On note $\underline{R}(x)$ et $\bar{R}(x)$ la $\liminf$ et la $\limsup$ de la valeur de révision $u(t,x)$ quand t tend vers $-\infty$. Si $\underline{R}(x) = \bar{R}(x) = R$ est indépendant de x, on dit que R est la valeur de révision limite. On va prouver l'existence de cette valeur de révision et la comparer avec les solutions statiques du jeu : V, S_1 et S_2.

Proposition 5. *Pour tout jeu sous forme normale U et pour tout paramètre de révision $\lambda > 0$ et $q \in (0,1)$, il existe un nombre $R \in [S_1, S_2]$, tel que pour tout $t \geq 0$ et tout $x \in X$,*

$$|u(t,x) - R| \leq 2||U|| \exp(-\lambda(1-\max\{q, 1-q\})).$$

Il suit de cette propriété que la limite de révision $R = \lim_{t\to-\infty} u(t,x)$ existe et est indépendante de la position x initiale.

Ce nombre R satisfait $S_1 \leq R \leq S_2$ et donc si le jeu sous forme normale admet un équilibre de Nash en stratégies pures, alors $V = S_1 = S_2 = R$.

On remarque que la condition $0 < q < 1$ est indispensable, car si $q = 1$ seul le joueur 1 peut changer son action et le résultat dépend alors de la position initiale du joueur 2.

Preuve. On considère le maximum $M(t) = \max_{x\in X} u(t,x)$ de la valeur de révision du jeu démarrant au temps $t \leq 0$. Si $s < 0$, par le principe de programmation dynamique, on a pour tout x, $u(t,x) = u^s(t-s,x) \leq M(t-s)$, et par passage au $\max_{x\in X}$ à gauche on a $M(t) \leq M(t-s)$. Donc M est croissante en fonction du temps et donc converge vers $M_{-\infty}$ quand t tend vers $-\infty$. De même $m(t) = \min_{x\in X} u(t,x)$ est décroissante et converge vers $m_{-\infty}$ quand t tend vers $-\infty$. On a :

$$m = m(0) \leq m(t) \leq m_{-\infty} \leq M_{-\infty} \leq M(t) \leq M(0).$$

Pour un $t \leq 0$ donné, soit y tel que $m(t) = u(t,y)$ et x tel que $M(t) = u(t,x)$. Par le principe de programmation dynamique dans $\Gamma(t,x)$ on a :

$$M(t) = u(t,x) \leq ||U||e^{\lambda t} + \int_{s=t}^{0} \lambda e^{\lambda(t-s)}(qM(s) + (1-q)u(s,x_1,y_2)ds$$

puisque $u^+(s,x) \leq M(s)$ et $u^-(s,x) \leq u(s,x_1,y_2)$. De façon similaire,

$$m(t) = u(t,y) \geq -e^{\lambda t}||U|| + \int_{s=t}^{0} \lambda e^{\lambda(t-s)}((1-q)m(s) + qu(s,x_1y_2)ds.$$

On définit $w(t) = M(t) - m(t)$. Si $q \geq 1/2$, on écrit $qM(s) - (1-q)m(s) = qw(s) + (2q-1)m(s)$ et on déduit :

$$\begin{aligned} w(t) &\leq 2||U||e^{\lambda t} + q\int_t^0 \lambda e^{\lambda(t-s)}w(s)ds + \int_t^0 \lambda e^{\lambda(t-s)}(2q-1)(m(s) - u(s,x_1,y_2))ds \\ &\leq 2||U||e^{\lambda t} + q\int_t^0 \lambda e^{\lambda(t-s)}w(s)ds, \end{aligned}$$

où la deuxième inégalité vient du fait que $(2q-1)(m(s) - u(s,x_1,y_2)) \leq 0$.
De façon similaire, si $q \leq 1/2$, on écrit $qM(s) - (1-q)m(s) = (1-q)w(s) + (2q-1)M(s)$ et on obtient :

$$\begin{aligned} w(t) &\leq 2||U||e^{\lambda t} + (1-q)\int_t^0 \lambda e^{\lambda(t-s)}w(s)ds + \int_t^0 \lambda e^{\lambda(t-s)}(2q-1)(M(s) - u(s,x_1,y_2))ds \\ &\leq 2||U||e^{\lambda t} + (1-q)\int_t^0 \lambda e^{\lambda(t-s)}w(s)ds. \end{aligned}$$

On en déduit que pour tout $q \in (0,1)$, on a :

$$w(t) \leq 2||U||e^{\lambda t} + \max\{q, 1-q\}\int_t^0 \lambda e^{\lambda(t-s)}w(s)ds.$$

En notant $g(t) = e^{-\lambda t}w(t)$, on a :

$$g(t) \leq 2||U|| + \max\{q, 1-q\}\int_t^0 \lambda g(s)ds.$$

En appliquant le lemme de Gronwall à la fonction g on obtient :

$$g(t) \leq 2||U||e^{-\lambda t \max\{q,1-q\}}.$$

On en déduit que

$$w(t) \leq 2||U||e^{\lambda t(1-\max\{q,1-q\})} \underset{t\to-\infty}{\longrightarrow} 0.$$

Remarque : la preuve montre également que la valeur de révision tend exponentiellement vite vers la valeur de révision quand on remonte le temps.

Pour montrer que $S_1 \leq R \leq S_2$ il suffit de prouver que

$$S_1 \leq \liminf_{t\to-\infty} u(t,x) \leq \limsup_{t\to-\infty} u(t,x) \leq S_2.$$

On considère la stratégie du joueur 1 qui consiste à toujours jouer une stratégie optimale dans son jeu de Stackelberg, c'est-à-dire comme s'il savait que l'autre joueur allait jouer après lui. Comme la longueur de la phase de révision est arbitrairement longue, et que $\lambda q > 0$, la probabilité que le joueur 1 ait une opportunité est proche de 1. Ainsi cette stratégie garantit au joueur 1 à peu près S_1 quand T est grand et $\liminf_t u(t,x) \geq S_1$. Par symétrie, $\limsup_t u(t,x) \leq S_2$.

De ce résultat on peut définir pour la matrice de gain U la valeur de révision limite $R(U) = \lim_{t\to-\infty} u(t,x)$, indépendante de x.

Montrons de la même façon que pour la valeur certaines propriétés de régularité de la fonction valeur de révision limite.

Lemme 11. *La valeur de révision limite R est 1-lipschitzienne en U, indépendante de $\lambda \in (0,\infty)$ et est continue et croissante en $q \in (0,1)$. De plus elle converge vers le paiement de Stackelberg du joueur 2 quand q tend vers 1, et vers le paiement de Stackelberg du joueur 1 quand q tend vers 0.*

Preuve.
i) R est 1-lipschitzienne en U. Soit deux matrices de gain U et U'. Par Le lemme 10, pour tout t et x, $|u(t,x) - u'(t,x)| \leq ||U - U'||$. En faisant tendre t vers $-\infty$ on obtient $|R(U) - R(U')| \leq ||U - U'||$.
ii) R ne dépend pas de λ. En effet, faire un changement sur λ revient à faire un changement de temps. De plus, la valeur est un passage à la limite dans le temps, R est indépendant du temps, et donc de la fréquence λ.
iii) R est continue en $q \in (0,1)$. Fixons $q \in (0,1)$ et prenons une suite (q_n) dans $(0,1)$ qui tend vers q. Pour $\epsilon > 0$, il existe t tel que $|t|$ est assez grand pour que $u(t,x) \in [R-\epsilon, R+\epsilon]$, pour tout $x \in X$.

Par le Lemme 10 $u(t,x)$ est continue en q, et pour n assez grand la valeur de révision $u_n(t,x)$ du jeu de paramètre q_n appartient à $[R-2\epsilon, R+2\epsilon]$, pour tout x. Ainsi, pour n assez large, la valeur de révision limite du jeu avec paramètre q_n appartient à $[R-2\epsilon, R+2\epsilon]$.
iv) R **est croissante en** q. Par le Lemme 10, $u(t,x)$ est continue en t et croissante en q. Donc, $R = \lim_{t\to-\infty} u(t,x)$ est croissante en q.
v) R **converge vers** S_2 **quand** q **tend vers 1.** Prenons t assez grand en valeur absolue pour qu'il y ait au moins une révision avec probabilité supérieure à $1-\epsilon$. A chaque fois que le joueur 1 est appelé à jouer, il peut choisir une meilleure réponse en action pure dans le jeu sous forme normale de matrice de gain U. Avec probabilité supérieure à $(1-\epsilon)q$ il y a au moins une opportunité de révision et le joueur 1 profitera de la dernière opportunité de révision. Alors pour tout x :

$$u(t,x) \geq (1-\epsilon)qS_2 - (1-(1-\epsilon)q)||U||.$$

En faisant tendre t vers $-\infty$ on obtient que la valeur de révision satisfait $R \geq (1-\epsilon)qS_2-(1-(1-\epsilon)q)||U||$, et comme $\epsilon > 0$ est arbitraire, en le faisant tendre vers 0 on obtient $R \geq qS_2-(1-q)||U||$. Comme $R \leq S_2$, R converge vers S_2 quand q tend vers 1.
De façon symétrique, R converge vers S_1 quand q tend vers 0.

2.3.2 Les jeux de révision à somme nulle à deux actions par joueur

Dans cette section on traite totalement le cas particulier des jeux de révision dont le jeu sous forme normale associé est à somme nulle et à deux actions par joueur. Pour cette classe on va caractériser les stratégies d'équilibre et la valeur de révision limite.

On considère un jeu de révision à somme nulle 2×2 et on note $X_1 = X_2 = \{\alpha,\beta\}$ l'ensemble d'actions commun aux deux joueurs. On se concentre sur le cas des jeux où U est régulière, c'est-à-dire que $MR_i^u(\alpha)$ et $MR_i^u(\beta)$ sont des singletons pour les deux joueurs. On remarque tout d'abord que dans ce cas il n'y a que trois scénarios possibles décrivant les meilleures réponses dans le jeu sous forme normale.

Définition 21.

— ***Scénario DD :*** *Chaque joueur a une action dominante. Par exemple :*

	α	β
α	0	-1
β	1	0

— ***Scénario DN :*** *Un joueur a une action dominante, la meilleure réponse de l'autre joueur varie en fonction de l'action du premier. Par exemple :*

	α	β
α	0	1
β	-1	-2

— ***Scénario NN :*** *La meilleure réponse de chaque joueur dépend de l'action courante de l'autre. Par exemple :*

	α	β
α	0	1
β	1	0

Le jeu a un équilibre de Nash en stratégie pure dans les scénarios DD et DN, et un équilibre de Nash en stratégies mixtes dans le scénario NN.

Quitte à remplacer U par $U - U(\alpha,\alpha)$ on peut supposer que $U(\alpha,\alpha) = 0$. On observe que si on a $U(\alpha,\beta)+U(\beta,\alpha)-U(\alpha,\alpha)-U(\beta,\beta) = 0$ chaque joueur a une action dominante et on est dans le cas DD.

Supposons à présent que nous soyons dans le cas $U(\alpha,\beta)+U(\beta,\alpha)-U(\alpha,\alpha)-U(\beta,\beta) \neq 0$. Alors quitte à normaliser on peut supposer $U(\alpha,\beta)+U(\beta,\alpha)-U(\alpha,\alpha)-U(\beta,\beta) = 1$. On adopte ainsi les notations suivantes :

	α	β
α	0	b
β	c	$b+c-1$

On considère un équilibre markovien du jeu de révision. Pour tout $t \geq 0$, on note la matrice de gain comme suit :

	α	β
α	$u(t,\alpha,\alpha)$	$u(t,\alpha,\beta)$
β	$u(t,\beta,\alpha)$	$u(t,\beta,\beta)$

où $u(0,x) = U(x)$. La proposition suivante montre comment caractériser l'équilibre du jeu de révision.

Proposition 6. *On considère une matrice de gain U de taille 2×2 régulière. L'équilibre du jeu de révision σ est le suivant :*

1. *Si le jeu sous forme normale associé est dans le scénario DD, alors chaque joueur prépare son action dominante du jeu sous forme normale à chaque instant $t \leq 0$. Il en résulte que $R = V$.*

2. *Si le jeu sous forme normale est dans le scénario DN, alors il y a un joueur i qui a une action dominante $\hat{x}_i$ et qui la prépare en tout $t \leq 0$:*

$$\sigma_i(t, x_{-i}) = \hat{x}_i, \forall t, \forall x_{-i} \in X_{-i}.$$

De plus il existe $t^* < 0$ tel que :
- pour $t > t^*$, le joueur $-i$ prépare sa meilleure réponse dans le jeu sous forme normale contre l'action courante du joueur i :

$$\sigma_{-i}(t, x_i) = B^U_{-i}(x_i);$$

- pour tout $t < t^*$, le joueur $-i$ prépare sa meilleure réponse dans le jeu sous forme normale contre l'action dominante du joueur i :

$$\sigma_{-i}(t, x_i) = B^U_{-i}(\hat{x}_i), \forall x_i \in X_i.$$

Il en résulte que $R = V$.

3. *Si le jeu sous forme normale est dans le scénario NN, alors il existe $t_1^* < t_2^* < 0$ et un joueur i tels que :*
- le joueur i prépare sa meilleure réponse contre l'action courante du joueur $-i$ si $t > t_1^*$, et prépare une action fixe $\hat{x}_i$ quand $t \geq t_1^*$:

$$\sigma_i(t, x_{-i}) = \begin{cases} MR^U_i(x_{-i}) & \text{pour } t > t_1^* \\ \hat{x}_i & \text{pour } t \leq t_1^*, \forall x_{-i}. \end{cases}$$

- le joueur $-i$ prépare sa meilleure réponse contre l'action courante du joueur i dans le jeu sous forme normale quand $t > t_2^*$. Quand $t < t_2^*$, le joueur $-i$ prépare sa meilleure réponse face à l'action fixe $\hat{x}_i$ du joueur i :

$$\sigma_{-i}(t, x_i) = \begin{cases} MR^U_{-i}(x_i) & \text{pour } t > t_2^* \\ MR^U_{-i}(\hat{x}_i) & \text{pour } t \leq t_2^*, \forall x_i \end{cases}$$

Il en résulte que $R \neq V$ en général pour les jeux de scénario NN.

Pour la démonstration on renvoie le lecteur à l'article de Tomala et Lovo (2015).

L'interprétation est la suivante :
- près de la deadline, si un joueur a une opportunité, il sait qu'il a une forte probabilité d'être le dernier à jouer, il va donc jouer en fonction de l'action courante de l'autre joueur.
- loin de la deadline, il ne se passe presque rien. Les joueurs jouent une action de sur-place et n'en changent pas avant d'être près de la deadline. Les joueurs commencent alors à agir près de la deadline où se passe l'essentiel des actions du jeu. Le détail des calculs et des cas pour des jeux particuliers sont détaillés dans l'article de Tomala et Lovo (2015) et dans l'article de Calcagno (2013).

Conclusion

Dans l'étude des jeux de révision à somme nulle nous avons pu remarquer que les valeurs avaient de nombreuses propriétés de régularité et que ces jeux de révision se résolvaient relativement bien avec la programmation dynamique. On a également vu que contrairement aux jeux de révision à somme non nulle, la rapidité d'un joueur est toujours profitable à ce dernier, contrairement aux jeux à intérêts opposés présentant deux équilibres de Nash.

2.4 Jeu de la balle

Le jeu de la balle est un jeu de révision stochastique à somme non nulle qui se déroule de la façon suivante. Deux joueurs se lancent une balle. Lorsque que la cloche sonne, celui qui a la balle peut choisir de la garder ou de la lancer. A la fin, un paiement est attribué à chacun des deux joueurs, ce paiement ne dépend que de la dernière action effectuée. L'action AA = "le joueur A garde la balle" donne le paiement (1,0). L'action AB = "le joueur A lance la balle" donne le paiement (1,2). L'action BA = "le joueur B lance la balle" donne le paiement (0,1). L'action BB = "le joueur B garde la balle" donne le paiement (2,1). Il peut se représenter sous les deux formes suivantes :

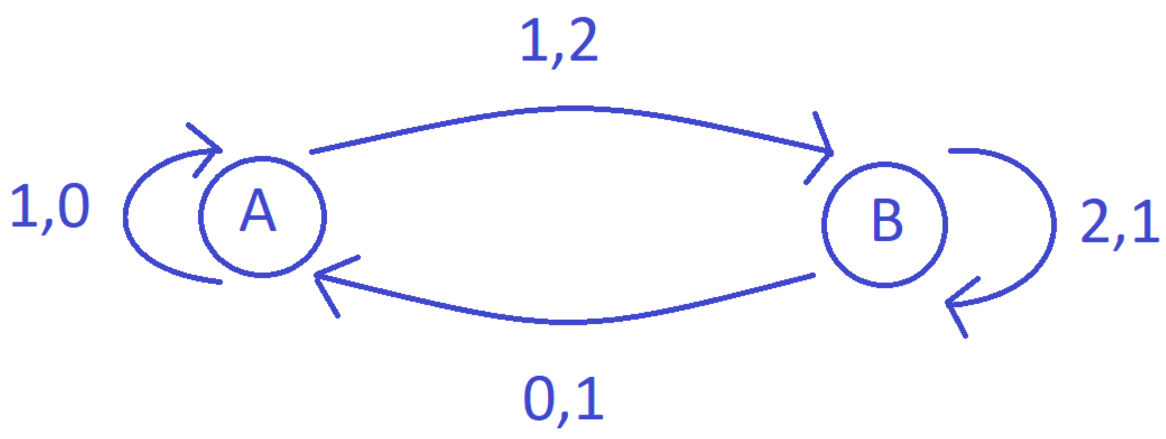

FIGURE 2.3 – Le jeu de la balle, première forme

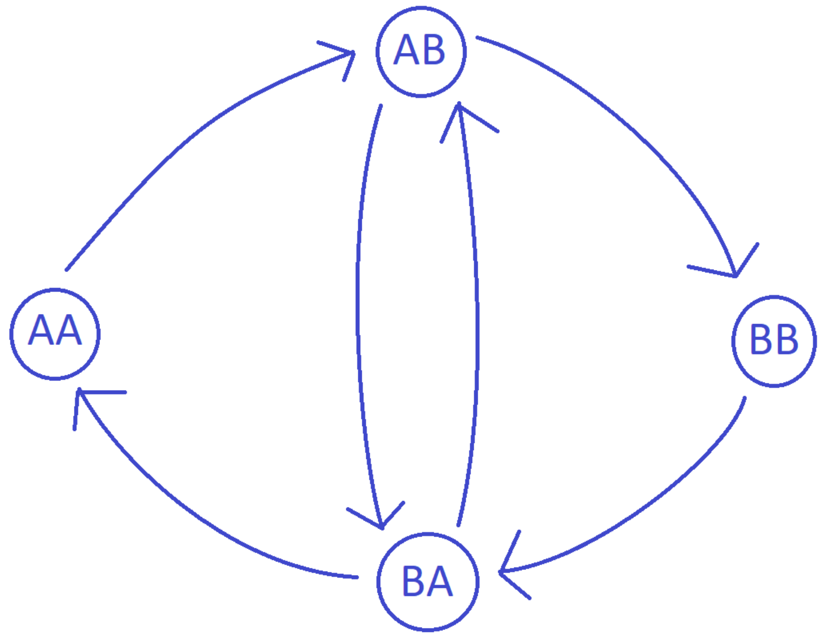

FIGURE 2.4 – Le jeu de la balle, deuxième forme

Dans la première forme, les états sont les joueurs, et les fonctions de paiement dépendent des deux derniers états, c'est pourquoi la deuxième forme est plus commode d'un point de vue rigoureux, même si la première forme permet de mieux comprendre le jeu. Nous travaillerons exclusivement avec la deuxième forme.

Ce jeu de révision stochastique a les paramètres suivants :

— Ensemble de joueurs $\mathcal{N} = \{A, B\}$,

— Ensemble des états et des actions (on identifie les deux par commodité) $K = \{AA, AB, BA, BB\}$,

— Ensemble d'actions des joueurs quand $k \in \{AA, BA\}$: $\{AA, AB\}$,

— Ensemble d'actions des joueurs quand $k \in \{AB, BB\}$: $\{BA, BB\}$,

— Fonction de gain $r : K \to \mathbb{R}^2$ vérifie $r(AA) = (1,0)$, $r(AB) = (1,2)$, $r(BA) = (0,1)$, $r(BB) = (2,1)$.

— Probabilité de transition $q : K \times K^2 \to \Delta(K)$ vérifie

$$q(k, a_1, a_2) = \begin{cases} a_1 \text{ si } k \in \{AA, BA\}, \\ a_2 \text{ si } k \in \{AB, BB\}. \end{cases}$$

— Une fréquence d'opportunité λ commune à tous les états,

— Un intervalle de temps $[-T, 0]$.

Enonçons d'abord quelques observations préliminaires.

On a naturellement envie de résoudre ce jeu avec le principe de la programmation dynamique comme l'ont fait Tomala et Lovo dans leur article sur les jeux de révision à somme nulle à deux actions par joueur. Le problème ici est que les joueurs sont indifférents quand on arrive près de la deadline : si le joueur B a la balle et une opportunité près de la fin, il garantit un paiement proche de 1 qu'il la donne ou qu'il la garde, idem pour le joueur A.

Une première idée serait de résoudre des jeux similaires avec des paiements approchés et de passer à la limite. Une étude (non présentée ici) des jeux où l'on a remplacé un gain de 1 par un gain de $1 + \epsilon$ ou $1 - \epsilon$ pour un seul des deux joueurs montre que les stratégies optimales données par programmation dynamique (et qui sont des équilibres en sous-jeux parfaits markoviens) sont des stratégies pures telles qu'il existe entre un et quatre intervalles par joueur sur lesquels ces derniers jouent toujours la même action. Le passage à la limite quand ϵ tend vers 0 (à $-T$ fixé) donne toujours la stratégie suivante :
- Le joueur B garde toujours la balle. Le joueur A garde lui aussi toujours la balle.
On verra que cette stratégie n'est pas optimale pour le joueur B s'il a la balle au début, et qu'elle n'est pas optimale pour le joueur A s'il a la balle au début. Le passage à la limite sur les gains se passe donc très mal.

Si le joueur A garde toujours la balle, il garantit le paiement $(1,0)$. S'il lance toujours la balle, en réponse le joueur B va lancer la balle sur un intervalle de fin pour garantir strictement plus que 1, selon si le dernier état est AB ou BA, tandis que le joueur A va garantir strictement moins de 1, ce qui est moins que s'il avait toujours garder la balle. Le joueur A ne lancera donc jamais la balle sur tout un intervalle de fin, c'est-à-dire de la forme $[t, 0]$.

Si le joueur B garde toujours la balle, il garantit $(2,1)$, tandis que s'il lance toujours la balle, en réponse le joueur A va toujours garder la balle et le joueur B va garantir 0, ce qui est moins que s'il avait toujours garder la balle. Le joueur B ne va donc pas lancer la balle sur un intervalle de fin.

La meilleure stratégie stationnaire pour le joueur B est qu'il garde toujours la balle. On verra plus loin qu'il peut mieux faire.
La meilleure stratégie stationnaire pour le joueur A est que le joueur B garde toujours la balle, ce qui permet au joueur A de garantir près de 2 en temps long, qui est le paiement le plus élevé qu'il puisse avoir. Pour cela il faut que le joueur B ait la balle au départ. La stratégie optimale du joueur A (non-markovienne) serait alors la suivante :

$$\sigma_A(h) = \begin{cases} AB \text{ si } h \in H_0, \\ AA \text{ sinon .} \end{cases}$$

Le joueur A ne lance donc la balle qu'au début si c'est lui qui l'a. S'il la reçoit du joueur B, il la garde jusqu'à la fin. Si le joueur A applique cette stratégie, le joueur B n'a jamais intérêt à lancer la balle au joueur A, donc il la garde. Ainsi en temps long, le joueur A garantit 2. Cette stratégie a deux défauts. D'une part elle utilise la mémoire du joueur A et n'est donc pas markovienne, ce qui n'est pas grave car elle reste simple à implémenter numériquement. Plus grave, si le joueur B lance quand même la balle, en suivant sa stratégie initiale le joueur A ne peut plus garantir plus de 1. Le joueur A préférera dans ce cas rejouer sa stratégie initiale sans tenir compte du passé pour garantir davantage. Sa stratégie n'est donc pas crédible et le joueur B ne s'y pliera pas. Elle n'a pas de chance d'être appliquée.

Les conclusions tirées jusqu'à présent montrent déjà la complexité de la résolution du jeu. Par ailleurs, si les deux joueurs utilisent des stratégies markoviennes, on voit qu'il n'existe pas d'intervalle de la forme $[t, 0]$ sur lequel un joueur aura une stratégie optimale stationnaire. S'il existe un équilibre en sous-jeux parfait markovien, ce dernier est donc soit à actions mixtes, soit tel que pour tout t et pour chaque joueur i il existe deux ensembles G_i, $L_i \subseteq [-T, 0]$ tels que le joueur i garde la balle sur G_i, lance la balle sur L_i, $\lambda([t, 0] \cap G_i) > 0$ et $\lambda([t, 0] \cap L_i) > 0$. Comme il s'agit d'un jeu de révision stochastique, on a vu qu'un tel équilibre existait. On va présenter un équilibre en sous-jeux parfait à actions mixtes, et montrer que ce n'est pas une stratégie optimale en présentant une stratégie non markovienne mais qui garantit plus en temps long pour chacun des deux joueurs. L'existence de cet équilibre ne garantit pas que ce soit le seul, ni la non-existence d'un équilibre en sous-jeux parfait markovien à actions pures. Néanmoins on énonce la conjecture de Tomala :

Conjecture (Tomala, 2016) Le jeu de la balle n'admet pas d'équilibre en sous-jeux parfait markovien à actions pures. C'est-à-dire que tout équilibre en sous-jeux parfait est soit non markovien, soit à actions mixtes et non toutes pures.

Présentons maintenant un équilibre en sous-jeux markovien à stratégies mixtes qui ne garantit pas la valeur.

Lemme 12. *La stratégie $\sigma_A(h) = \sigma_B(h)$ commune aux deux joueurs consistant à tirer au sort avec probabilité $(1/2, 1/2)$ entre garder la balle ou la lancer, à tout instant t et en tous $k \in K$ est un équilibre en sous-jeux markovien.*

Preuve. Supposons que le joueur A joue AA et AB avec probabilité $(1/2, 1/2)$ quand il a la balle. Supposons que le joueur B ait la balle. Montrons que le joueur B est indifférent entre garder la balle et la lancer par récurrence forte sur le nombre n d'opportunités restantes. S'il reste 0 opportunités, le joueur B n'a pas de choix, il est donc indifférent. S'il reste 1 opportunité, ou bien le joueur B garde la balle et il garantit 1, ou bien il la lance et il garantit 1. Il est indifférent. Supposons que pour tout $1 \leq k \leq n$ le joueur B soit indifférent entre lancer la balle et la garder s'il reste $k \leq n$ opportunités, et qu'il espère dans les deux cas le paiement 1. Supposons qu'il reste $n+1$ opportunités. Si le joueur B garde la balle la première fois qu'il a une opportunité, il est ramené au cas n et donc garantit un paiement de 1. S'il la lance, il reste n opportunités au joueur A. Avec probabilité $1/2^n$ le joueur A garde toujours la balle et B garantit le paiement 0. Avec probabilité $1/2^n$ le joueur A garde la balle $n-1$ fois avant de la lancer, le paiement du joueur B est alors 2. Sinon le joueur A lance la balle au joueur B au k-ième coup avec $k \leq n-2$ et B a de nouveau la balle avec au plus n opportunités. Par hypothèses de récurrence il est donc indifférent et espère le paiement 1. Au total, le joueur B espère $1/2^n \times 2 + 1/2^n \times 0 + 1 \times (1 - 1/2^n - 1/2^n) = 1$. Le joueur B est donc indifférent entre lancer la balle et la garder. Par récurrence forte, il est toujours indifférent s'il reste un nombre fini d'opportunités. Comme le nombre d'opportunités est fini presque sûrement, alors presque sûrement le joueur est B est indifférent entre garder la balle et la lancer.

De manière symétrique on montre que le joueur A est indifférent si le joueur B joue $(1/2, 1/2)$ quand il a la balle. Comme la stratégie ne dépend de rien, c'est un équilibre en sous-jeux parfait markovien.

Cette stratégie garantit le paiement $(1, 1)$ pour les deux joueurs. Nous introduisons maintenant une stratégie qui permet aux deux joueurs de garantir strictement plus de 1.

Les jeux de Stackelberg des deux joueurs

On a dit qu'avec des jeux approchés, les joueurs jouent avec des actions pures et ont des stratégies optimales pour lesquels ils changent d'action un nombre fini de fois. C'est pour cette raison que nous allons étudier les cas où les joueurs jouent avec des actions pures et changent d'avis un nombre fini de fois.

Jeu de Stackelberg du joueur A

Si le joueur A impose qu'il lance la balle sur un intervalle de fin, on a vu qu'en réponse le joueur B lui lançait la balle, et le joueur A garantit moins que s'il gardait la balle tout le temps. Cette stratégie n'étant pas intéressante car mauvaise, on s'intéresse au cas où le joueur A annonce qu'il va lancer la balle sur un intervalle $[x, 0]$ pour un certain $x < 0$:

Quelques explications. En réponse directe à cette annonce, le joueur B décide de garder la balle sur l'intervalle $[y, 0]$ avec $y \leq x$ car le joueur B ne veut pas lancer la balle s'il est sûr que le joueur A ne lui renverra pas. On a alors les conditions finales, on peut alors appliquer le principe de la programmation dynamique à partir de ces conditions. On en déduit l'existence des nombres $-\infty < -T < p < z < y < x$ tels que :
- Le joueur A garde la balle sur $[x, 0]$
- Le joueur B garde la balle sur $[y, 0]$
- Le joueur A lance la balle sur $[z, x]$
- Le joueur B lance la balle sur $[p, y]$
- Le joueur A garde la balle sur $[-T, z]$

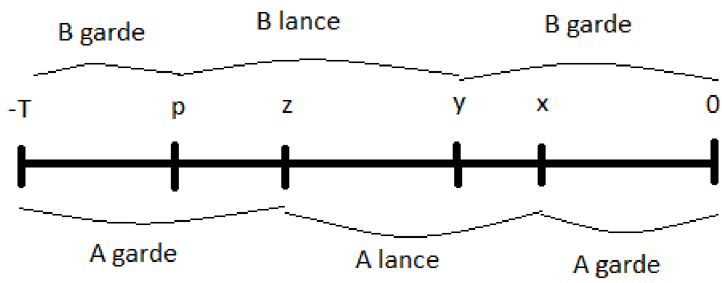

FIGURE 2.5 – Jeu de Stackelberg du joueur A

- Le joueur B garde la balle sur $[-T, p]$

Une fois cette stratégie construite on cherche à trouver le x qui permet au joueur A d'optimiser son gain en temps long, les autres paramètres dépendant exclusivement de x. On remarque alors que le x optimal vaut $-T$ si l'état initial est AB ou BB, ce qui signifie qu'on se retrouve dans la stratégie optimale où A ne rend jamais la balle au joueur B, et est telle que $y = -T$ si l'état initial est AA ou BA. Ainsi on voit que x tend vers $-\infty$ quand $-T$ tend vers $-\infty$. La stratégie du joueur A garantit un paiement proche de 2 en temps long, et le joueur B garantit un paiement proche de 1. Nous allons voir en étudiant le jeu de Stackelberg du joueur B que le joueur B va refuser de jouer dans le jeu de Stackelberg du joueur A pour privilégier le sien.

Jeu de Stackelberg du joueur B

On a vu que si le joueur B annonce qu'il lance la balle sur un intervalle de fin, le joueur A en réponse va garder la balle sur un intervalle de fin plus long. Le joueur B va ainsi garantir un paiement inférieur à 1, cette stratégie est donc mauvaise puisque moins bonne que la stratégie stationnaire "garder toujours la balle". Le joueur B va donc choisir un $y < 0$ tel qu'il va garder la balle sur l'intervalle $[y, 0]$:

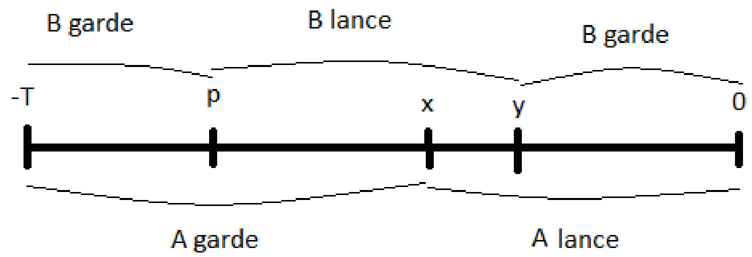

FIGURE 2.6 – Jeu de Stackelberg du joueur B

En réponse le joueur A a tout intérêt à lancer la balle sur un intervalle $[x, 0]$ avec $x \leq y$. Comme on sait ce que les joueurs font dans un intervalle comprenant 0, on peut donc appliquer le principe de programmation dynamique pour résoudre le jeu. On en déduit ainsi l'existence pour T assez grand des nombres $-T < p < x < y < 0$ tels que :
- Le joueur A lance la balle sur $[x, 0]$
- Le joueur B garde la balle sur $[y, 0]$
- Le joueur A garde la balle sur $[-T, x]$
- Le joueur B lance la balle sur $[p, y]$
- Le joueur B garde la balle sur $[-T, p]$

Ces nombres dépendant exclusivement de y, on cherche le y qui garantit un paiement v optimal pour le joueur B en temps long. On trouve $y^* = 2/\lambda$ si l'état initial est AA ou BA. Je n'ai pas réussi à finir les calculs pour les deux états initiaux BA et BB, ni pour le calculs des valeurs. Cette stratégie n'est pas stable, elle ne l'est que pour le joueur A qui a joué de façon optimale en fonction de ce qu'annonçait le joueur B. Le joueur B a cependant intérêt à dévier près de la deadline. Pour éviter cela, le joueur A utilise la stratégie de menace suivante. Si on note H l'ensemble des histoires non finies compatibles avec le jeu de Stackelberg du joueur B ci-dessus, alors la stratégie du joueur A est la suivante :

$$\sigma_A(h) = \begin{cases} AA \text{ si } -T \leq t < x & \text{et } h \in H \\ AB \text{ si } x \leq t < 0 & \text{et } h \in H \\ AA \text{ sinon .} \end{cases}$$

Le joueur B n'a alors aucune envie de dévier de la stratégie qu'il a annoncée, de peur d'être puni. Cette punition du joueur A est cette fois crédible, car si le joueur B le trahit, le joueur A n'a aucune raison de lui faire confiance en lui rendant la balle par la suite, et préfère garantir 1 en gardant la balle. La résolution du jeu n'est pas terminée. Nous ne savons pas s'il existe des équilibres en sous-jeux parfaits markovien, ni si le joueur A a une stratégie stable qui permettrait de gagner davantage que de jouer dans le jeu de Stackelberg du joueur B.

Chapitre 3

Conclusion

La résolution explicite d'un jeu de révision où le principe de la programmation dynamique ne s'applique pas est difficile. Nous n'avons pas réussi, sur un exemple, à trouver de méthode explicite pour résoudre le jeu de la balle. Le fait de voir beaucoup de stratégies optimales dans des jeux approchés donnent à la limite des stratégies mauvaises ou instables, alors qu'on peut trouver une meilleure stratégie stable et non markovienne pour les deux joueurs laisse perplexe. Une piste à exploiter semble la notion de joueur dominant : dans le jeu de la balle, on voit clairement que le joueur A a l'avantage car c'est lui qui a le dernier mot à coup sûr, et c'est la raison pour laquelle le jeu de Stackelberg du joueur B marche mieux. La résolution de ces jeux reste un problème ouvert.

Chapitre 4

Références

E. Borel, "La théorie du jeu et les équations intégrales à noyau symétrique gauche",C. R. Acad. Sci. Paris, 173 (1921), p. 1304–1308.

R. Calcagno, S. Lovo, Y. Kamada and T. Sugaya, "Asynchronicity and Coordination in Common and Opposing Interest Games", Theoretical Economics, (2014), 9(2), 409–434.

I. Glicksberg, "A further generalization of the Kakutani fixed point theorem, with applications to Nash equilibrium points", Proceedings of the American Mathematical Society, 3 (1952), p. 170–17.

S. Kakutani, "A generalization of Brouwer's fixed point theorem", Duke Mathematical Journal, 8 (1941), p. 416–427.

Y. Kamada et M. Kandori, "Revision Games", papier non publié, (2010).

S. Lovo et T. Tomala, "Markov Perfect Equilibria in Stochastic Revision Games", papier non publié, (2016).

S. Lovo et T. Tomala, "Zero-Sum revision games", papier non publié, (2015).

Morgenstern

L. S. Shapley, "Stochastic games", Proceedings of the national academy of sciences, (1953), 39(10), 1095–1100.

J. von Neumann, "Zur Theorie der Gesellschaftspiele", Mathematische Annalen, vol. 100,(1928), p. 295-320.