Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 1 Capitole

Présentée et soutenue par

Tristan GARREC

Le 11 juillet 2019

Sur les jeux dynamiques : jeux stochastiques, recherche-dissimulation et transmission d'information

---

Ecole doctorale : **EDMITT - Ecole Doctorale Mathématiques, Informatique et Télécommunications de Toulouse**

Spécialité : **Mathématiques et Applications**

Unité de recherche :
UMR 5314 -TSE-R

Thèse dirigée par
Jérôme RENAULT

Jury

M. Steve ALPERN, Rapporteur
M. János FLESCH, Rapporteur
Mme Catherine RAINER, Examinatrice
M. Sylvain SORIN, Examinateur
M. Guillaume VIGERAL, Examinateur
M. Jérôme RENAULT, Directeur de thèse

À ma mère,
à la mémoire de mon père.

Remerciements

Je tiens à remercier vivement Jérôme Renault, qui a encadré cette thèse avec rigueur scientifique et bienveillance. Effectuer mes premiers pas dans le monde de la recherche sous sa direction a été une expérience très enrichissante.

It is an honour to have benefited from the expertise of Steve Apern and János Flesch as rapporteurs of this manuscript. They also allowed me to present my work at workshops and I sincerely thank them.

Je suis également honoré par la présence de chacun des membres du jury lors de ma soutenance. Je salue tout particulièrement Sylvain Sorin, qui a éveillé mon attrait pour la théorie des jeux lors de son cours de master à l'Université Pierre-et-Marie-Curie.

Depuis mes début en master, et tout au long de cette thèse, j'ai eu le plaisir de côtoyer à l'occasion de conférences, séminaires, et dans les laboratoires où j'ai travaillé, la conviviale communauté française de théorie des jeux. Je souhaite ici remercier tous ses membres pour leur accueil et pour les échanges que j'ai eus avec eux.

Ma première collaboration scientifique s'est faite avec Marco Scarsini, que je remercie cordialement. J'ai également grâce à lui présenté mes travaux à divers séminaires et conférences, et bénéficié de conditions de travail exceptionnelles lors de mes séjours à la *Libera Università Internazionale degli Studi Sociali*, à Rome.

J'adresse toute ma reconnaissance aux membres des administrations, notamment de l'école d'économie de Toulouse et de l'université Paris Dauphine, qui font fonctionner les départements de recherche et d'enseignement.

Je salue chaleureusement mes amis. Ils ont pu constater pendant ces années la corrélation de ma bonne humeur à l'avancement de cette thèse. Je leur souhaite succès dans les projets qu'ils entreprennent.

Une rencontre inattendue a empli de bonheur ces derniers mois, merci à Aurélie.

Ces dernières lignes vont à ma famille, que je remercie de tout mon cœur. À François, la force tranquille. À ma sœur, qui m'a appris par l'exemple. Et à ma mère, "qui ne m'a pas suffisamment poussé étant petit".

Table of contents

1 Introduction (en français) **7**

1.1 Généralités sur les jeux à somme nulle 7
1.2 Jeux stochastiques 10
1.3 Jeux de recherche-dissimulation 23
1.4 Jeux répétés à information incomplète 34

2 Introduction (in English) **40**

2.1 General information on zero-sum games 40
2.2 Stochastic games 43
2.3 Search games 55
2.4 Repeated games with incomplete information 65

3 Communicating zero-sum product stochastic games **72**

3.1 Introduction 72
3.2 Preliminaries on zero-sum product stochastic games 75
3.3 Main results 78
3.4 Proof of Theorem 3.5 80
3.5 Proof of Theorem 3.6 92

4 Continuous patrolling and hiding games **102**

4.1 Introduction 102
4.2 The models 105
4.3 Patrolling games 107
4.4 Hiding games 114
4.5 Properties of the value function 116

5 When Sally found Harry: A Stochastic search game **124**

5.1 Introduction 124
5.2 The model 127
5.3 Deterministic search games 129
5.4 Value of the game 132
5.5 Dynamic programming 133
5.6 Stochastic search games on trees 133
5.7 Stochastic search games on Eulerian graphs 140

6 Dynamic control of information **149**
6.1 Introduction 149
6.2 The model 150
6.3 The value of the game 152
6.4 The greedy strategy 152
6.5 The two-state case 155
6.6 More than two states 158

Bibliography **163**

This manuscript contains six chapters. The first two are introductions in French and in English. The following four are based on four articles published or intended for publication:

1. Chapter 3, *Communicating Zero-Sum Product Stochastic Games*, Journal of Mathematical Analysis and Applications 477 (2019) 60-84;

2. Chapter 4, *Continuous Patrolling and Hiding Games*, European Journal of Operational Research 277 (2019) 42-51;

3. Chapter 5, *When Sally Found Harry: A Stochastic Search Game*, in collaboration with Pr. Marco Scarsini, arXiv:1904.12852, submitted;

4. Chapter 6, *Dynamic Control of Information with Observed Return on Investment*, in collaboration with Pr. Jérôme Renault, in preparation.

Chapitre 1

Introduction (en français)

Cette thèse porte sur l'étude des *jeux dynamiques*, qui modélisent des processus de décisions prises par des agents rationnels en interactions stratégiques et dont la situation évolue au cours du temps.

Outre la présente introduction, ainsi que son pendant anglais, ce manuscrit comporte quatre chapitres, chacun d'entre eux s'intéressant à un ou plusieurs aspects dynamiques d'interactions stratégiques. Le Chapitre 3 s'intéresse aux jeux stochastiques, les Chapitres 4 et 5 étudient des modèles de jeux de recherche-dissimulation. Enfin, le Chapitre 6 traite de transmission dynamique d'information.

L'objectif de la présente introduction est double. D'une part, nous replaçons les sujet abordés dans leur contexte en présentant brièvement l'état de la littérature sur ces sujets. D'autre part, nous présentons les contributions apportées par cette thèse en en donnant les résultats, ainsi qu'en expliquant intuitivement les démarches suivies et les idées développées. Après quelques généralités sur les jeux à somme nulle, cette introduction comporte trois autres parties dans lesquelles s'inscrivent les quatre chapitres de la thèse.

1.1 Généralités sur les jeux à somme nulle

La théorie des jeux étudie les interactions dites stratégiques entre entités rationnelles appelées *joueurs*. Dans un cadre non coopératif, ceux-ci ont le choix d'une stratégie qui, associée aux stratégies des autres joueurs, induit un paiement individuel. Dans cette thèse nous nous plaçons dans le cadre des jeux à deux joueurs et à somme nulle, pour lesquels le gain d'un joueur est l'opposé de celui de son adversaire. Ainsi, les intérêts des deux protagonistes sont totalement opposés.

1.1.1 Modèle

Un *jeu sous forme stratégique à somme nulle* est donné par un triplet $\Gamma = (S, T, g)$ pour lequel S est l'ensemble des stratégies du joueur 1, T est l'ensemble des stratégies du joueur 2 et $g : S \times T \longrightarrow \mathbb{R}$ est la fonction de paiement. On se place dans le cadre où S et T sont des espaces métriques non vides, munis de leur tribu borélienne. Le jeu se déroule de la manière suivante. Simultanément, le joueur 1 choisit une stratégie $s \in S$ et le joueur 2 choisit une stratégie $t \in T$. Le joueur 1 reçoit le paiement $g(s,t)$ tandis que le joueur 2 reçoit le paiement $-g(s,t)$. Ainsi, le joueur 1 cherche à maximiser $g(s,t)$ tandis que le joueur 2 cherche à minimiser cette quantité.

1.1.2 Valeur

Quel peut être le paiement issu de l'interaction entre deux joueurs rationnels ? Pour répondre à cette question on introduit la notion de *valeur* d'un jeu.

Définition 1.1. Le jeu $\Gamma = (S, T, g)$ admet une valeur v si

$$v = \sup_{s \in S} \inf_{t \in T} g(s,t) = \inf_{t \in T} \sup_{s \in S} g(s,t). \tag{1.1.1}$$

Dans la Définition 1.1 ci-dessus, le paiement $\sup_{s \in S} \inf_{t \in T} g(s,t)$ correspond à la situation dans laquelle le joueur 1 choisit sa stratégie et l'annonce au joueur 2 qui à son tour fait son choix en fonction de l'annonce du joueur 1. À l'inverse, le paiement $\inf_{t \in T} \sup_{s \in S} g(s,t)$ correspond à la situation dans laquelle le joueur 2 choisit sa stratégie et l'annonce au joueur 1 qui à son tour fait son choix en fonction de l'annonce du joueur 2. Dès lors, lorsque le jeu considéré a une valeur, celle-ci correspond au paiement obtenu par des joueurs rationnels. On définit également le concept de *stratégies* ε*-optimales*, qui permettent aux joueurs de garantir un paiement proche de la valeur.

Définition 1.2. Soit $\varepsilon \geq 0$. Une stratégie $s \in S$ du joueur 1 est dite ε-optimale si

$$\inf_{t \in T} g(s,t) \geq v - \varepsilon. \tag{1.1.2}$$

De même, une stratégie $t \in T$ du joueur 2 est dite ε-optimale si

$$\sup_{s \in S} g(s,t) \leq v + \varepsilon. \tag{1.1.3}$$

Lorsque ε vaut 0 dans la Définition 1.2 ci-dessus on parle simplement de *stratégie optimale*.

1.1.3 Théorèmes de minmax

Considérons le jeu de pile ou face suivant, appelé *matching pennies* en anglais. Chaque joueur a deux stratégies : pile et face. Si les côtés de pièces choisis par les joueurs sont assortis, le joueur 2 donne sa pièce au joueur 1. Dans le cas contraire c'est le joueur 1 qui donne sa pièce au joueur 2. On représente cette situation de la manière suivante, voir Fig. 1.1.

Un jeu tel que le jeu de pile ou face pour lequel les ensembles de stratégies sont finis est dit *jeu matriciel*, car on peut représenter le jeu sous la forme d'une

	pile	face
pile	1	-1
face	-1	1

Figure 1.1: Un jeu de pile ou face

matrice dans laquelle le joueur 1 joue les lignes et le joueur 2 les colonnes. On parle alors d'*actions*, plutôt que de stratégies.

On s'aperçoit alors que ce jeu n'admet pas de valeur. En effet on a $\sup_{s\in S}\inf_{t\in T} g(s,t) = -1$ tandis que $\inf_{t\in T}\sup_{s\in S} g(s,t) = 1$. Afin de contrecarrer cette déconvenue, on autorise les joueurs à introduire de l'aléa dans leur choix. On parle de *stratégie mixte*. Formellement, on définit l'*extension mixte* d'un jeu de la manière suivante.

Soit X un espace métrique, alors $\Delta(X)$ désigne l'ensemble des mesures de probabilité boréliennes sur X.

Définition 1.3. Soit un jeu $\Gamma = (S, T, g)$. On suppose que le théorème de Fubini s'applique pour tout $(\sigma, \tau) \in \Delta(S) \times \Delta(T)$ à l'intégrale de g sur $S \times T$. L'extension mixte de Γ est le jeu $(g, \Delta(S), \Delta(T))$, où g est étendu linéairement, c'est-à-dire que si $\sigma \in \Delta(S)$ et $\tau \in \Delta(T)$ alors

$$g(\sigma,\tau) = \int_{S\times T} g(s,t)d(\sigma\times\tau)(s,t).$$

On doit le théorème suivant à von Neumann (1928).

Théorème 1.1. *Toute extension mixte d'un jeu fini admet une valeur ainsi que des stratégies optimales.*

Ainsi, l'extension mixte du jeu de pile ou face présenté plus haut a bien une valeur. Celle-ci vaut 0 et les stratégies optimales des joueurs consistent à jouer pile avec probabilité $1/2$ et face avec cette même probabilité.

Il existe de nombreuses généralisations du théorème de Von Neumann à des ensembles d'actions plus généraux. On cite les deux théorèmes suivants, voir (Mertens et al., 2015).

Théorème 1.2. *Soit un jeu $\Gamma = (S, T, g)$. On suppose que*

1. *S est compact ;*
2. *pour tout $t \in T$, $g(\cdot, t)$ est semi-continue supérieurement.*

Alors l'extension mixte de Γ admet une valeur, le joueur 1 a une stratégie optimale mixte et le joueur 2 a une stratégie ε-optimale mixte à support fini.

Théorème 1.3. *Soit un jeu $\Gamma = (S, T, g)$. On suppose que*

1. *S et T sont compacts ;*

2. *pour tout $t \in T$, $g(\cdot, t)$ est semi-continue supérieurement, et pour tout $s \in S$, $g(s, \cdot)$ est semi-continue inférieurement ;*

3. *g est bornée et mesurable par rapport à la tribu borélienne produit.*

Alors l'extension mixte de Γ admet une valeur, et chaque joueur a une stratégie optimale mixte.

1.2 Jeux stochastiques

Il est vraisemblable qu'une interaction stratégique entre deux joueurs ayant lieu se soit déjà déroulée dans le passé ou ait à nouveau lieu dans le futur. Il est donc tout naturel d'introduire la notion de *jeu répété*. Or si l'on répète un même jeu à somme nulle tel que présenté dans la partie précédente, il est optimal pour les joueurs de jouer à chaque étape une stratégie optimale du jeu en un coup, la répétition n'a donc pas d'influence d'un point de vue stratégique.

On se place dans la situation où le jeu courant dépend d'un l'*état de la nature*. Ce dernier évolue d'une étape à la suivante de manière aléatoire en fonction de l'état de la nature courant ainsi que des actions des joueurs. Si l'on suppose que les joueurs observent l'état courant ainsi que les actions passées, on se trouve alors face à un *jeu stochastique*. Ces derniers ont été introduits par Shapley (1953), nous en présentons dans cette partie les grands enjeux.

1.2.1 Modèle

Un jeu stochastique est la donnée d'un quintuplet $\Gamma = (\Omega, A, B, g, \rho)$ où,

- Ω est l'ensemble d'états ;
- A et B sont les ensembles d'actions des joueurs 1 et 2 respectivement ;
- $g : \Omega \times A \times B \to \mathbb{R}$ est la fonction de paiement ;
- $\rho : \Omega \times A \times B \to \Delta(\Omega)$ est la probabilité de transition.

Dans cette thèse on se placera le plus souvent sous les hypothèses suivantes.

Hypothèses 1.

- l'espace d'états Ω est non vide et fini, muni de la tribu discrète ;
- les ensembles d'actions A et B sont non vides et métriques compacts, munis de leur tribu borélienne ;
- les fonctions de paiement et de transition sont continues séparément en les actions des joueurs.

Soit $\omega_1 \in \Omega$ l'état initial du jeu stochastique. Le jeu $\Gamma(\omega_1)$ se déroule de la manière suivante.

- L'état initial $\omega_1 \in \Omega$ est connu des deux joueurs.
- À chaque étape $t \geq 1$, connaissant l'histoire
$$h_t = (\omega_1, a_1, b_1, \cdots, \omega_{t-1}, a_{t-1}, b_{t-1}, \omega_t),$$
le joueur 1 et le joueur 2 choisissent simultanément $a_t \in A$ et $b_t \in B$ respectivement.
- Ceci génère un paiement d'étape $g_t = g(\omega_t, a_t, b_t)$.
- L'état suivant ω_{t+1} est tiré selon $\rho(\cdot|\omega_t, a_t, b_t)$ et observé par les joueurs, tout comme les actions a_t et b_t.

Commençons par un exemple. Le jeu suivant, appelé *Big Match*, a été introduit par Gillette (1957). L'espace d'états est $\Omega = \{\omega, 1^*, 0^*\}$, les ensembles d'actions des joueurs 1 et 2 sont $A = \{H, B\}$ et $B = \{G, D\}$ respectivement. Les états 1^* et 0^* sont *absorbants*, c'est-à-dire que lorsqu'ils sont atteints, ils ne peuvent plus être quittés. Le paiement dans ces états est respectivement 1 et 0. Seul l'état ω n'est pas absorbant. On dit que le *Big Match* est un *jeu absorbant*. Les transitions et paiements dans l'état ω sont représentés Fig. 1.2. Ainsi, si par exemple les actions

	G	D
H	1^*	0^*
B	0	1

Figure 1.2: Le *Big Match* dans l'état ω

H et G sont jouées, alors le paiement d'étape est 1, l'état devient 1^*, et le paiement sera donc 1 à chaque étape. Si les actions B et G sont jouées, alors le paiement d'étape est 0, et l'état reste ω.

Dans le jeu $\Gamma(\omega)$ en une étape, la valeur du jeu vaut $1/2$ et le choix uniforme de chaque action est optimal pour les deux joueurs. En revanche, si le jeu est répété un plus grand nombre de fois, et que l'on considère la moyenne des paiements sur ces étapes, le joueur 1 doit moduler son choix de l'action H. En effet, jouer l'action H avec probabilité ε à chaque étape implique, si le joueur 2 joue D à chaque étape, que l'état est absorbé en 0^* en $1/\varepsilon$ étapes en moyenne. Dès lors, si le nombre de répétitions est grand devant $1/\varepsilon$, cette stratégie du joueur 1 paraît peu satisfaisante. Toutefois, ne jamais jouer H serait également une mauvaise stratégie, puisque le joueur 2 pourrait jouer G à chaque étape, ce qui induirait un paiement nul.

Apparaît donc un problème d'arbitrage typique des jeux stochastiques. En effet le joueur 1 doit à la fois obtenir un bon paiement d'étape, mais également se placer dans une bonne situation sur le plus long terme.

Désormais on donne formellement la notion de stratégie dans un jeu stochastique. On note $H_t = \Omega \times (\Omega \times A \times B)^{t-1}$ l'ensemble des histoires possibles à l'étape

$t \geq 1$. On munit H_t de la tribu produit notée $\mathcal{H}_t$. On note $H_\infty = (\Omega \times A \times B)^{\mathbb{N}^*}$ l'ensemble des histoires infinies, que l'on munit de la tribu produit engendrée par $\bigcup_{t\geq 1} \mathcal{H}_t$.

Définition 1.4. Une *stratégie de comportement* du joueur 1 est une suite d'applications $\sigma = (\sigma_t)_{t\geq 1}$ telle que pour tout $t \geq 1$, σ_t est une application mesurable de $(H_t, \mathcal{H}_t)$ dans $\Delta(A)$. On définit de manière analogue les stratégies de comportement du joueur 2.

On note $\mathcal{S}$ et $\mathcal{T}$ les ensembles de stratégies de comportement des joueurs 1 et 2 respectivement. Une stratégie de comportement est dite *pure* si pour chaque histoire finie $h_t \in H_t$, on a $\sigma_t(h_t) \in A$. Une *stratégie mixte* est une distribution de probabilité sur l'ensemble des stratégies pures muni de la tribu produit. On a deux définitions analogues pour le joueur 2.

De manière générale, une stratégie prend donc en compte à une étape $t \geq 1$ toute l'histoire h_t pour sélectionner une action. On est amené à considérer des stratégies nécessitant moins de mémoire.

Définition 1.5. Une stratégie est dite *markovienne* si à chaque étape le choix d'une action ne dépend que de l'étape et de l'état courant.

Une stratégie markovienne peut donc être vue comme une application de $\mathbb{N}^* \times \Omega$ à valeur dans $\Delta(A)$, ou dans $\Delta(B)$.

Définition 1.6. Une stratégie est dite *stationnaire* si à chaque étape le choix d'une action ne dépend que de l'état courant.

Une stratégie stationnaire peut donc être vue comme une application de Ω à valeur dans $\Delta(A)$, ou dans $\Delta(B)$.

Un état initial ω_1 ainsi qu'un couple de stratégies (σ, τ), de comportement ou mixtes, induisent naturellement sur l'ensemble des histoires finies $\bigcup_{t\geq 1} H_t$ une mesure de probabilité. Le théorème d'extension de Kolmogorov nous indique que cette mesure s'étend de manière unique à H_∞. On note cette mesure de probabilité $\mathbb{P}^{\omega_1}_{\sigma,\tau}$, et $\mathbb{E}^{\omega_1}_{\sigma,\tau}$ l'espérance correspondante.

Les stratégies de comportement se révèlent souvent plus aisées à manipuler que les stratégies mixtes. Le théorème de Kuhn, voir (Sorin, 2002, Appendice D) et (Aumann, 1964), établit l'équivalence entre les stratégies de comportement et les stratégies mixtes au sens suivant.

Théorème 1.4 (Théorème de Kuhn). *Pour toute stratégie de comportement σ du joueur 1, il existe une stratégie mixte $\tilde{\sigma}$ du joueur 1 telle que pour toute stratégie du joueur 2 (de comportement ou mixte) τ, on ait*

$$\mathbb{P}^{\omega_1}_{\sigma,\tau} = \mathbb{P}^{\omega_1}_{\tilde{\sigma},\tau}.$$

Réciproquement, pour toute stratégie mixte σ du joueur 1, il existe une stratégie de comportement $\tilde{\sigma}$ du joueur 1 telle que pour toute stratégie du joueur 2 (de comportement ou mixte) τ, on ait

$$\mathbb{P}^{\omega_1}_{\sigma,\tau} = \mathbb{P}^{\omega_1}_{\tilde{\sigma},\tau}.$$

Le Théorème 1.4 reste vrai en échangeant les rôles des joueurs 1 et 2. Il existe principalement deux façons d'agréger les paiements d'étapes, à savoir au travers des moyennes d'Abel et de Césaro.

Soit $\lambda \in]0,1]$, le *jeu escompté au taux* λ, noté $\Gamma_\lambda(\omega_1)$, est le jeu dont les ensembles de stratégies sont $\mathcal{S}$ et $\mathcal{T}$ et dont le paiement est $\gamma_\lambda^{\omega_1} : \mathcal{S} \times \mathcal{T} \to \mathbb{R}$ défini par

$$\gamma_\lambda^{\omega_1}(\sigma,\tau) = \mathbb{E}_{\sigma,\tau}^{\omega_1}\left(\sum_{t\geq 1} \lambda(1-\lambda)^{t-1} g_t\right).$$

Soit $N \in \mathbb{N}^*$, le *jeu répété* N *fois*, noté $\Gamma_N(\omega_1)$ est le jeu dont les ensembles de stratégies sont $\mathcal{S}$ et $\mathcal{T}$ et dont le paiement est $\gamma_N^{\omega_1} : \mathcal{S} \times \mathcal{T} \to \mathbb{R}$ défini par

$$\gamma_N^{\omega_1}(\sigma,\tau) = \mathbb{E}_{\sigma,\tau}^{\omega_1}\left(\frac{1}{N}\sum_{t=1}^{N} g_t\right).$$

Sous les Hypothèses 1, le théorème du minmax Théorème 1.3 nous assure l'existence de la valeur dans les jeux $\Gamma_\lambda(\omega_1)$ et $\Gamma_N(\omega_1)$, ainsi que celle de stratégies optimales. On note

$$\begin{aligned} v_\lambda(\omega_1) &= \max_{\sigma\in\mathcal{S}} \min_{\tau\in\mathcal{T}} \gamma_\lambda^{\omega_1}(\sigma,\tau) = \min_{\tau\in\mathcal{T}} \max_{\sigma\in\mathcal{S}} \gamma_\lambda^{\omega_1}(\sigma,\tau), \\ v_N(\omega_1) &= \max_{\sigma\in\mathcal{S}} \min_{\tau\in\mathcal{T}} \gamma_\lambda^{\omega_1}(\sigma,\tau) = \min_{\tau\in\mathcal{T}} \max_{\sigma\in\mathcal{S}} \gamma_N^{\omega_1}(\sigma,\tau). \end{aligned}$$

Plus généralement, on peut également agréger les paiements d'étapes selon une mesure de probabilité sur $\mathbb{N}^*$. Soit $\theta \in \Delta(\mathbb{N}^*)$. Le jeu $\Gamma_\theta = (S, T, g_\theta)$ est le *jeu répété pondéré par* θ où $g_\theta : S \times T \to \mathbb{R}$ est défini par

$$\gamma_\theta(\sigma,\tau) = \mathbb{E}_{\sigma,\tau}\left(\sum_{t\geq 1} \theta_t g_t\right).$$

Quelques classes de jeux stochastiques se distinguent dans la littérature, soit pour leur intérêt propre, soit dans l'espoir de généraliser certains résultats. Un jeu stochastique est *absorbant* s'il ne possède qu'un état non absorbant, un *état absorbant* étant un état qui une fois atteint ne peut être quitté. Le *Big Match* présenté plus haut en est un exemple. Un jeu stochastique est *récursif* si le paiement dans tout état non absorbant est nul. Un *processus de décision markovien* est un jeu stochastique à un seul joueur.

1.2.2 Opérateur et équations de Shapley

Soit $\mathcal{F}$ l'ensemble des fonctions bornées de Ω dans $\mathbb{R}$. À tout élément $f \in \mathcal{F}$ et tout état $\omega \in \Omega$ on associe un jeu $\Gamma(f)(\omega)$ avec espaces d'actions A et B et de paiement

$$(a,b) \mapsto g(\omega,a,b) + \sum_{\omega'\in\Omega} \rho(\omega'|\omega,a,b) f(\omega').$$

Sous les Hypothèses 1, ce jeu à une valeur en stratégies mixtes, que l'on note $\Psi(f)(\omega)$. On appelle Ψ l'*opérateur de Shapley* du jeu stochastique Γ.

Théorème 1.5. *Les valeurs des jeux escompté et répété N fois sont caractérisées par les* équations de Shapley *suivantes.*

$$\begin{aligned} v_\lambda(\omega_1) &= \max_{\mu\in\Delta(A)} \min_{\nu\in\Delta(B)} \left[\lambda g(\omega_1,\mu,\nu) + (1-\lambda)\mathbb{E}^{\omega_1}_{\mu,\nu}(v_\lambda)\right] \\ &= \min_{\nu\in\Delta(B)} \max_{\mu\in\Delta(A)} \left[\lambda g(\omega_1,\mu,\nu) + (1-\lambda)\mathbb{E}^{\omega_1}_{\mu,\nu}(v_\lambda)\right], \end{aligned}$$

et

$$\begin{aligned} v_{N+1}(\omega_1) &= \max_{\mu\in\Delta(A)} \min_{\nu\in\Delta(B)} \left[\frac{1}{N+1} g(\omega_1,\mu,\nu) + \frac{N}{N+1}\mathbb{E}^{\omega_1}_{\mu,\nu}(v_N)\right] \\ &= \min_{\nu\in\Delta(B)} \max_{\mu\in\Delta(A)} \left[\frac{1}{N+1} g(\omega_1,\mu,\nu) + \frac{N}{N+1}\mathbb{E}^{\omega_1}_{\mu,\nu}(v_N)\right], \end{aligned}$$

avec

$$\mathbb{E}^{\omega_1}_{\mu,\nu}(v_N) = \sum_{\omega\in\Omega} v_N(\omega) \int_{A\times B} \rho(\omega|\omega_1,a,b)d\mu(a)d\nu(b).$$

On a donc

$$v_\lambda = \lambda\Psi\left(\frac{1-\lambda}{\lambda}v_\lambda\right) \text{ et } v_{N+1} = \frac{1}{N+1}\Psi\left(Nv_N\right).$$

Les équations de Shapley ont une interprétation intuitive : si l'état initial est ω_1, les joueurs savent qu'ils pourront jouer optimalement à l'étape suivante dans le jeu de continuation. Une conséquence importante est la suivante.

Corollaire 1.7.
- *Les joueurs 1 et 2 ont des stratégies stationnaires optimales dans le jeu $\Gamma_\lambda(\omega_1)$, quel que soit ω_1.*
- *Les joueurs 1 et 2 ont des stratégies markoviennes optimales dans le jeu $\Gamma_N(\omega_1)$, quel que soit ω_1.*

Une question fondamentale de l'étude des jeux répétés est celle du comportement asymptotique des valeurs (v_λ) et (v_N). On présente trois approches de cette question : l'approche asymptotique, l'approche uniforme ainsi que le jeu infini.

1.2.3 Valeur asymptotique

L'*approche asymptotique* des jeux stochastiques consiste en l'étude de la convergence de (v_λ) et (v_N) lorsque λ tend vers 0 et N tend vers $+\infty$, c'est-à-dire à mesure que les joueurs deviennent de plus en plus patients.

Définition 1.8. Un jeu stochastique admet une *valeur asymptotique* si (v_λ) et (v_N) convergent vers la même limite lorsque λ tend vers 0 et N tend vers $+\infty$.

Le résultat suivant sur les jeux stochastiques avec espace d'états et ensembles d'actions finis est un classique de la littérature, on le doit à Bewley and Kohleberg (1978).

Théorème 1.6. *Tout jeu stochastique avec espace d'états et ensembles d'actions finis admet une valeur asymptotique.*

De plus (v_λ) admet un développement en série de Puiseux : il existe $\lambda_0 > 0$, $M \in \mathbb{N}^*$ et $(r_t) \in (\mathbb{R}^\Omega)^\mathbb{N}$ tels que pour tout $\lambda \in]0, \lambda_0[$ et tout $\omega \in \Omega$

$$v_\lambda(\omega) = \sum_{t \geq 0} r_t(\omega) \lambda^{\frac{t}{M}}.$$

Citons également le théorème taubérien de Ziliotto (2016a) qui sous des hypothèses faibles, et vérifiées par les Hypothèses 1 assure que (v_λ) converge si et seulement si (v_N) converge, et dans le cas où l'une converge les deux limites sont égales.

L'idée longtemps répandue selon laquelle la valeur asymptotique existe dans les jeux à espace d'états fini et ensembles d'ations compacts s'est vue réfutée par un contre-exemple de Vigeral (2013). On présente ci-dessous un contre-exemple simple de Renault (2019). Celui-ci est une légère variation d'un contre-exemple de Ziliotto (2016b), on le retrouve également chez Sorin and Vigeral (2015). Tous ces contre-exemples ont en commun d'avoir des transitions ou des ensembles d'actions qui ne sont pas semi-algébriques. C'est le point central qui permet de faire osciller la valeur. Dans le Contre-exemple 1.1 ci-dessous, c'est l'ensemble d'actions du joueur 1 qui n'est pas semi-algébrique.

Contre-exemple 1.1. L'espace d'états est $\Omega = \{0, 1, 0^*, 1^*\}$. Les espaces d'actions des joueurs 1 et 2 sont A et B respectivement, que l'on précise ci-après. Le joueur 1 joue $\alpha \in A$ dans l'état 0 et le joueur 2 joue $\beta \in B$ dans l'état 1. Les états 0^* et 1^* sont absorbants. Le paiement dans les états 0 et 0^* vaut 0. Il vaut 1 dans les états 1 et 1^*. On représente le jeu Fig. 1.3, les probabilités de transitions sont représentées par des flèches entre les états. On montre alors que qu'en choisissant

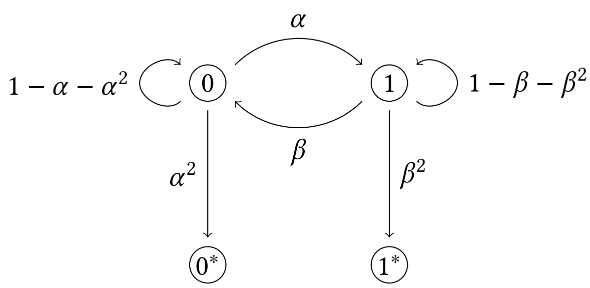

Figure 1.3: Un contre-exemple simple

$A = \{0\} \cup \left\{1/2^{2n} \mid n \geq 1\right\}$ et $B = [0, 1/4]$, la valeur du jeu escompté (v_λ) ne converge pas. L'idée est la suivante. Dans le jeu $\Gamma_\lambda(0)$ avec ensemble d'action

$A = [0, 1/4]$, le joueur 1 jouerait approximativement $\sqrt{\lambda}$. En choisissant A comme plus haut, on fait en sorte que le joueur 1 puisse jouer optimalement comme dans le jeu non contraint pour certains taux d'escompte, mais pas pour d'autres, ce qui induit un paiement plus faible et fait donc osciller la valeur. Nous reviendrons plus en détail sur ce contre-exemple au Chapitre 3.

La valeur asymptotique, lorsqu'elle existe, correspond donc approximativement au paiement obtenu par les joueurs s'ils jouent suffisamment longtemps, c'est-à-dire lorsque λ tend vers 0 et N tend vers $+\infty$. Toutefois les stratégies optimales des joueurs dans ces jeux de plus en plus longs peuvent être fonctions du taux d'escompte λ ou du nombre d'étapes N. Il est donc tout naturel de se demander s'il existe des stratégies qui soient optimales dans tout jeu suffisamment long, indépendamment de la longueur du jeu. Pour répondre à cette question on se place du point de vue de l'*approche uniforme*.

1.2.4 Valeur uniforme

Fixons un état initial ω_1. On dit que le joueur 1 *garantit uniformément* la quantité v_∞ s'il possède une stratégie qui garantit cette quantité, à ε près, contre n'importe quelle stratégie du joueur 2, dans tout jeu répété N fois $\Gamma_N(\omega_1)$, pourvu que N soit assez grand. Formellement,

$$(\forall \varepsilon > 0)\ (\exists \sigma \in \mathcal{S})\ (\exists M \in \mathbb{N}^*)\ (\forall \tau \in \mathcal{T})\ (\forall N \geq M)\ \gamma_N^{\omega_1}(\sigma, \tau) \geq v_\infty - \varepsilon.$$

Il en va de même pour le joueur 2.

Définition 1.9. S'il existe $v_\infty \in \mathbb{R}^\Omega$ telle que les deux joueurs garantissent $v_\infty(\omega_1)$ dans le jeu $\Gamma(\omega_1)$, alors le jeu Γ admet une valeur uniforme.

Soit $\varepsilon \geq 0$. Une stratégie $\sigma \in \mathcal{S}$ est dite (uniformément) ε-optimale si

$$(\exists M \in \mathbb{N}^*)\ (\forall \tau \in \mathcal{T})\ (\forall N \geq M)\ \gamma_N^{\omega_1}(\sigma, \tau) \geq v_\infty - \varepsilon.$$

Il en va également de même pour le joueur 2.

Il est à noter que si un joueur garantit uniformément une certaine quantité, il peut également garantir cette quantité dans tout jeu escompté, pourvu que le taux d'escompte soit suffisamment petit. Pour cette raison on s'intéresse dans l'approche uniforme au jeu répété N fois. De plus, si un jeu admet une valeur uniforme, alors (v_λ) et (v_N) convergent toutes deux vers la valeur uniforme. La réciproque n'est pas toujours vraie.

Le résultat suivant sur les jeux stochastiques avec espaces d'états et ensembles d'actions finis est un classique de la littérature, on le doit à Mertens and Neyman (1981).

Théorème 1.7. *Tout jeu stochastique avec espace d'états et ensembles d'actions finis admet une valeur uniforme.*

De nombreux autres résultats positifs quant à l'existence de la valeur uniforme existent dans la littérature. À titre d'exemples on cite, dans le cadre d'un espace d'états fini et d'ensembles d'actions compacts, les travaux de Mertens et al. (2009) pour les jeux absorbants, Li and Sorin (2016) pour les jeux récursifs, dont les démonstrations utilisent l'approche opérateur de Rosenberg and Sorin (2001) qui repose sur l'opérateur de Shapley, et qui contient entièrement la dynamique du jeu. Toujours pour un espace d'états fini, Bolte et al. (2014) ont montré que les jeux aux transitions et ensembles d'actions semi-algébriques (ou plus généralement définissables) ont une valeur uniforme. Enfin, Renault (2010) a montré l'existence de la valeur uniforme dans les processus de décision markoviens à espace d'états fini et ensemble d'actions quelconque.

1.2.5 Jeu infini

Nous présentons ici une dernière approche des jeux stochastiques en temps long. Il s'agit ici de définir une fonction de paiement directement sur les histoires infinies et d'étudier le jeu sous forme stratégique résultant. On considère usuellement les deux fonctions $\underline{\gamma_\infty}$ et $\overline{\gamma_\infty}$ définies sur $\Omega \times \mathcal{S} \times \mathcal{T}$ par

$$\underline{\gamma_\infty}^{\omega_1}(\sigma, \tau) = \mathbb{E}^{\omega_1}_{\sigma,\tau}\left(\liminf_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} g_t\right) \text{ et } \overline{\gamma_\infty}^{\omega_1}(\sigma, \tau) = \mathbb{E}^{\omega_1}_{\sigma,\tau}\left(\limsup_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} g_t\right).$$

Le résultat suivant est également dû à Mertens and Neyman (1981). Il unifie les approches uniforme et infinie des jeux stochastiques.

Théorème 1.8. *Les jeux $(\mathcal{S}, \mathcal{T}, \underline{\gamma_\infty})$ et $(\mathcal{S}, \mathcal{T}, \overline{\gamma_\infty})$ avec espace d'états et ensembles d'actions finis admettent une valeur, qui est égale à la valeur uniforme.*

1.2.6 Contribution de la thèse : jeux stochastiques produits communicants à somme nulle

Cette partie fait écho au Chapitre 3 intitulé *Communicating zero-sum product stochastic games*. La contribution de cette thèse en matière de jeux stochastiques porte sur l'étude de propriétés de communication entre les états, lorsque l'espace d'états Ω est sous la forme d'un produit $X \times Y$, et que les joueurs 1 et 2 contrôlent la dynamique sur leur composante de l'état X et Y, respectivement.

On introduit une nouvelle classe de stratégies appelées *stratégies markoviennes N-périodique*. Ces dernières sont les stratégies markoviennes dont la dépendance en l'étape est considérée modulo la période N. On peut donc les voir comme des applications de $\{1, \ldots, N\} \times \Omega$ dans $\Delta(A)$, ou $\Delta(B)$.

Afin d'insister sur la différence entre les stratégies dépendant de l'histoire des deux joueurs, et les stratégies ne dépendant que de la composante (état et action) d'un joueur, à l'instar d'un processus de décision markovien, on nomme les secondes des *politiques*.

Les deux propriétés de communication sur les composantes des espaces d'états que l'on considère sont les suivantes. On les appelle *propriété de communication forte* et *propriété de communication faible*.

Un joueur a la propriété de communication forte s'il existe un temps T tel qu'indépendamment de son choix de politique (ne dépendant que de sa composante de l'histoire), il ait une probabilité strictement positive d'aller de n'importe quel état à n'importe quel autre état de sa composante en T étapes. Formellement on donne la définition ci-dessous.

Définition 1.10. Le joueur 1 a la propriété de communication forte s'il existe $T \in \mathbb{N}^*$ tel que pour toute politique σ et tous états $x, x' \in X$, on a $\mathbb{P}^x_\sigma(X_T = x') > 0$.

On donne une définition similaire pour le joueur 2, et l'on dit qu'un jeu est fortement communicant d'un côté si l'un des joueurs a la propriété de communication forte.

Un joueur a la propriété de communication faible s'il existe un temps T et une politique (ne dépendant que de sa composante de l'histoire), tels qu'il ait une probabilité strictement positive d'aller de n'importe quel état à n'importe quel autre état de sa composante en T étapes. Formellement on donne la définition ci-dessous.

Definition 1.11. Le joueur 1 a la propriété de communication faible s'il existe $T \in \mathbb{N}^*$ et une politique σ tels que pour tous états $x, x' \in X$, on a $\mathbb{P}^x_\sigma(X_T = x') > 0$.

On donne une définition similaire pour le joueur 2, et l'on dit qu'un jeu est faiblement communicant des deux côtés si les deux joueurs ont la propriété de communication faible.

Il est à noter que la propriété de communication forte implique la faible. L'Exemple 1.1 ci-dessous illustre ces deux propriétés.

Exemple 1.1. On représente Fig. 1.4 l'espace d'états et les transitions du joueur 1. L'espace d'états est $X = \{x, y, z\}$. Dans l'état x le joueur 1 choisit $\alpha \in A$. Si l'ensemble d'actions A est l'intervalle $[0, 1]$, alors le joueur 1 a la propriété de communication faible mais pas la propriété de communication forte. En revanche, si A vaut $[\varepsilon, 1]$ avec $\varepsilon > 0$, le joueur 1 a la propriété de communication forte.

On démontre les deux théorèmes suivants, qui sont les principaux résultats de cette thèse en matière de jeux stochastiques.

Théorème 1.9. *Tout jeu fortement communicant d'un côté a une valeur uniforme.*

De plus, en supposant que le joueur 1 a la propriété de communication forte, la valeur uniforme ne dépend que de l'état initial du joueur 2, et pour tout $\varepsilon > 0$ le joueur 1 a une stratégie markovienne périodique ε-optimale.

Théorème 1.10. *Il existe un jeu faiblement communicant des deux côtés qui n'admet pas de valeur asymptotique.*

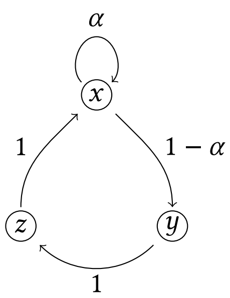

Figure 1.4: Communications faible et forte

On remarque qu'il est suffisant que l'un des deux joueurs ait la propriété de communication forte pour que la valeur uniforme existe, tandis que même si les deux joueurs ont la propriété de communication faible, la valeur asymptotique peut ne pas exister.

On détaille désormais l'intuition derrière les démonstrations des Théorèmes 1.9 et 1.10.

Démonstration du Théorème 1.9

On suppose ici que le joueur 1 a la propriété de communication forte. On classifie l'espace d'états Y via les classes de récurrence induites par les politiques stationnaires sur Y, d'une manière similaire à celle de Ross and Varadarajan (1991).

Définition 1.12. Un sous-ensemble C de Y est un *ensemble maximalement communicant* si

1. Il existe une politique stationnaire sur Y telle que C soit une classe de récurrence pour la chaîne de Markov induite sur Y ;

2. C est maximal, c'est-à-dire que s'il existe un sous-ensemble C' de Y vérifiant 1 et tel que $C \subseteq C'$, alors $C' = C$.

On note $C_1, \ldots, C_L$ les ensembles maximalement communicants, et D celui des états qui sont transitoires sous toute politique stationnaire. On montre alors la proposition suivante.

Proposition 1.13. $\{C_1, \ldots, C_L, D\}$ *est une partition de* Y.

L'Exemple 1.2 ci-dessous montre la manière dont l'espace d'états du joueur 2 se décompose en ensembles maximalement communicants. Il permet également de souligner un comportement du processus d'état qui peut se produire lorsque l'ensemble d'actions est compact, mais pas lorsqu'il est fini.

Exemple 1.2. Comme représenté sur la Fig. 1.5, l'espace d'états du joueur 2 est $Y = \{x, y, z\}$ et l'ensemble d'actions $B = [0, 1/2]$. Les probabilités de transitions

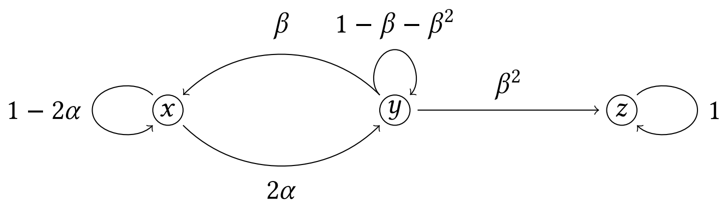

Figure 1.5: États, actions et transitions du joueur 2

sont représentées par des flèches. Les ensembles maximalement communicants sont $\{x\}$, $\{y\}$ et $\{z\}$. Il est à noter que si l'état initial est x, en jouant $\alpha = 1/2$ dans l'état x et $\beta = \frac{1}{2n}$ dans l'état y à l'étape $n \geq 1$, le joueur 2 a une probabilité strictement positive de commuter infiniment souvent entre les ensembles maximalement communicants $\{x\}$ et $\{y\}$.

Ceci ne peut pas se produire lorsque l'ensemble d'action est fini. Dans ce cas, pour toute politique du joueur 2, après un nombre fini d'étapes le processus d'état $(Y_n)_{n\geq 1}$ restera à tout jamais dans l'un des ensembles maximalement communicants avec probabilité 1, voir (Ross and Varadarajan, 1991, lemma 2 et proposition 2).

Pour chacun des ensembles maximalement communicants C_i, on introduit un jeu stochastique auxiliaire Γ_i comme suit. Pour tout $i \in \{1, \ldots, L\}$, si $y \in C_i$, on définit l'ensemble des actions du joueur 2 dans l'état y telles que l'état reste dans C_i avec probabilité 1,

$$B_y = \{b \in B \mid q(C_i|y, b) = 1\}.$$

On a alors la proposition suivante.

Proposition 1.14. *Pour tout $i \in \{1, \ldots, L\}$ le jeu Γ_i a une valeur uniforme v^i_∞, qui est constante sur $X \times C_i$.*

De plus, pour tout $\varepsilon > 0$ il existe $N_0 \in \mathbb{N}^*$ tel que dans chaque jeu Γ_i, les deux joueurs aient une stratégie markovienne N_0-périodique qui soit ε-optimale.

On considère alors le processus de décision markovien $\mathcal{G} = (Y, B, q, g)$, dans lequel seul le joueur 2 joue et vise à minimiser

$$\begin{array}{rcll} g & : & Y & \longrightarrow & [0,1] \\ & & y & \longmapsto & \begin{cases} v^i_\infty \text{ s'il existe } i \in \{1,\ldots,L\} \text{ tel que } y \in C_i \\ 1/2 \text{ si } y \in D. \end{cases} \end{array}$$

$\mathcal{G}$ a une valeur uniforme $w_\infty \in [0,1]^Y$, et pour tout $\varepsilon > 0$, le joueur 2 a une politique stationnaire ε-optimale, voir (Sorin, 2002).

On donne l'interprétation suivante. L'objectif du joueur 2 est d'atteindre en espérance l'ensemble maximalement communicant C_i tel que le jeu auxilaire correspondant Γ_i ait la plus petite valeur uniforme v^i_∞ possible, puis reste dans C_i. Le paiement de 1/2 dans D est arbitraire et ne change pas la valeur de w_∞.

On démontre que la valeur uniforme w_∞ du processus de décision markovien $\mathcal{G}$ est également la valeur uniforme du jeu initial.

L'idée pour le joueur 2 est de jouer tout d'abord optimalement dans le processus de décision markovien $\mathcal{G}$ (indépendamment du joueur 1) puis de changer pour une stratégie optimale dans Γ_i une fois l'ensemble maximalement communicant C_i atteint.

Montrer que le joueur 1 garantit uniformément w_∞ constitue l'une des principales difficultés. Il est naturel pour le joueur 1 de jouer optimalement dans chaque jeu Γ_i. Or le joueur 1 ne contrôle pas les transitions du joueur 2 d'un C_i à l'autre, qui on l'a vu dans l'Exemple 1.2, peuvent se produire infiniment souvent. On résout ce problème en laissant le joueur 2 jouer une stratégie de meilleure réponse markovienne périodique, ce qui empêche le comportement problématique. On conclut en montrant que le paiement ainsi obtenu peut également être généré par le joueur 2 comme paiement limite dans $\mathcal{G}$.

Démonstration du Théorème 1.10

La démonstration du Théorème 1.10 repose sur le Contre-exemple 1.1 introduit plus tôt. Ce contre-exemple n'a pas d'espace d'états produit et ne vérifie pas la propriété de communication faible. Nous construisons un contre-exemple qui vérifie ces propriétés et qui en reproduit la dynamique.

L'espace d'états du joueur 1 est $X = \{x, y\} \times C_8$ où $C_8 = \mathbb{Z}/8\mathbb{Z}$. Soit $I = \{0\} \cup \left\{1/2^{2n} \mid n \geq 1\right\}$. Soit $A = I \times \{-1, +1\} \cup \{0, 1\} \times \{0\}$ l'ensemble d'actions du joueur 1.

Pour $i \in \{x, y\}$ on note $-i$ l'élément de $\{x, y\} \backslash \{i\}$. Dans l'état $(i, k) \in X$ si le joueur 1 joue $(\alpha, p) \in I \times \{-1, +1\}$ alors avec probabilité $1 - \alpha - \alpha^2$ le nouvel état est $(i, k + p)$, avec probabilité α le nouvel état est $(-i, k + p)$, et avec probabilité α^2 le nouvel état est $(i, k - p)$ (voir Fig. 1.6).

Toujours dans l'état $(i, k) \in X$, si le joueur 1 joue $(\alpha, 0) \in \{0, 1\} \times \{0\}$, alors avec probabilité $1 - \alpha$ l'état reste en (i, k) et avec probabilité α le nouvel état est $(-i, k)$ (voir Fig. 1.7).

L'espace d'états et les transitions du joueur 2 sont une copie de ceux du joueur 1. En revanche l'ensemble d'actions du joueur 2 est $B = J \times \{-1, +1\} \cup \{0, 1\} \times \{0\}$, où $J = [0, 1/4]$.

Le fait que I et donc A ne soient pas semi-algébriques est essentiel. En effet puisque X et Y sont finis, et que les transitions sont polynomiales, si A et B étaient définissables dans une structure o-minimale, le jeu aurait une valeur uniforme, voir (Bolte et al., 2014, Theorem 4).

Ainsi en jouant $(0, p)$, le joueur 1 contrôle totalement la dynamique sur C_8, et en jouant $(\alpha, 0)$ avec α égal à 0 ou 1 il contrôle totalement la dynamique sur $\{x, y\}$. Il en va de même pour le joueur 2.

La fonction de paiement est définie comme suit. Si la distance entre les joueurs sur le cercle est supérieure à 3 le paiement vaut 1. Si la distance entre les joueurs sur le cercle est inférieure à 1 le paiement vaut 0. Enfin, si la distance entre les

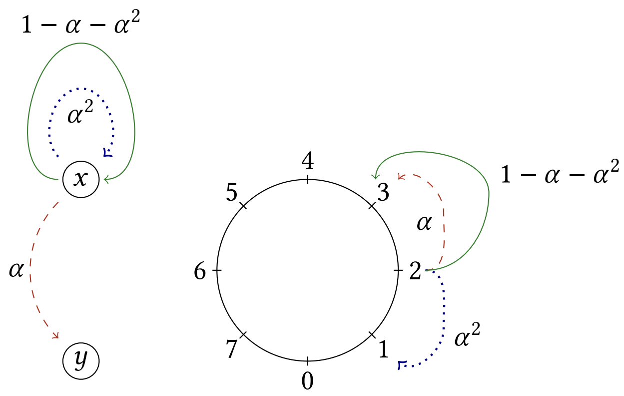

Figure 1.6: Transitions en jouant $(\alpha, +1)$, $\alpha \in I$ dans l'état $(x, 2)$

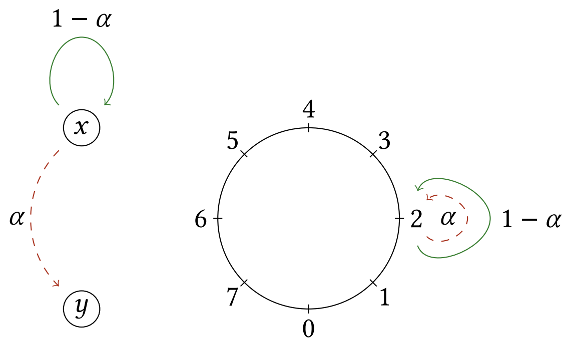

Figure 1.7: Transitions en jouant $(\alpha, 0)$, $\alpha \in \{0, 1\}$ dans l'état $(x, 2)$

joueurs sur le cercle vaut 2, alors la paiement vaut 1 si les joueurs 1 et 2 se trouvent dans le même état de $\{x, y\}$, et 0 s'ils se trouvent dans un état différent.

Le jeu a une interprétation en terme de jeu de poursuite-évasion escompté qui est la suivante. Le joueur 1 souhaite maximiser sa distance au joueur 2, qui lui souhaite minimiser sa distance au joueur 1. Si la distance entre les joueurs sur le cercle est inférieure à 1 ou supérieure à 3, alors leur position sur $\{x, y\}$ n'importe pas. En revanche s'ils sont à distance 2 sur le cercle, le joueur 1 ne veut pas avoir la même position que le joueur 2 sur $\{x, y\}$.

Il est notable que si la distance entre les joueurs sur le cercle est supérieure à 3, alors le joueur 1 peut jouer de manière à ce qu'elle reste supérieure à 3 pour toujours. De même si la distance entre les joueurs sur le cercle est inférieure à 1, alors le joueur 2 peut jouer de manière à ce qu'elle reste inférieure à 1 pour toujours. Ainsi ces états joints agissent comme des états absorbants avec paiements 1 et 0 respectivement.

On montre alors que le jeu construit a les mêmes équations de Shapley que celles du Contre-exemple 1.1.

1.3 Jeux de recherche-dissimulation

Dans un second temps on s'intéresse à une autre forme d'interaction dynamique. Dans un *jeu de recherche-dissimulation*, deux joueurs interagissent sur un *espace de recherche*. Le premier est appelé *le chercheur* (ou Sally, *the Searcher* en anglais), et le second est appelé *le dissimulateur* (ou Harry, *the Hider* en anglais). Harry choisit un endroit où se dissimuler dans l'espace de recherche. Les deux paradigmes que considère Sally sont alors typiquement

1. soit de minimiser le temps nécessaire pour trouver Harry ;

2. soit de maximiser la probabilité de trouver Harry en un temps imparti.

L'objectif de Harry est opposé. La dynamique présente dans l'interaction ne tient donc non pas d'un état de la nature changeant, comme c'était le cas pour les jeux stochastiques introduits dans la Section 1.2, ou les jeux de transmission d'information que l'on présentera dans la Section 1.4, mais bien à la trajectoire de recherche de Sally. On verra toutefois au travers des jeux de recherche-dissimulation stochastiques, Section 1.3.4, que ces deux aspects dynamiques peuvent être considérés simultanément.

1.3.1 Modèle standard

On présente ici le modèle standard de jeux de recherche-dissimulation, introduit par Isaacs (1965). Celui-ci considère le paradigme 1 de minimisation du temps de découverte. Le jeu est donné par

- un espace de recherche Q, qui est un sous-ensemble compact de $\mathbb{R}^n$ muni d'une norme $\|\cdot\|$ qui induit une distance d ;

- un rayon de détection $r \geq 0$;

- une origine de la trajectoire de Sally $O \in Q$.

On pourrait également relâcher ce dernier point de telle sorte à ce que Sally choisisse son point de départ, ce qui sera le cas pour les jeux de patrouille présentés dans la Section 1.3.3.

L'ensemble de stratégies $\mathcal{S}$ de Sally est celui des trajectoires s qui sont 1-lispchitziennes de $\mathbb{R}_+$ dans Q, et telles que $s(0) = O$. Ainsi, Sally commence sa trajectoire en O. De plus $s(t) \in Q$ représente la position de Sally à l'instant t. Enfin, Sally se déplace à vitesse au plus 1. L'ensemble des stratégies de Harry est Q. On définit alors pour $(s, h) \in \mathcal{S} \times Q$ le paiement

$$g(s, h) = \inf\{t \in \mathbb{R}_+ \mid d(s(t), h) \leq r\},$$

où l'infimum sur l'ensemble vide est $+\infty$. Ainsi, la capture a lieu lorsque Sally et Harry se trouvent à distance inférieure au rayon de détection de Sally, et le paiement $g(s, h)$ représente le temps de capture lorsque Sally suit la trajectoire de recherche s et Harry se dissimule en h. Le jeu de recherche-dissimulation a une valeur en stratégies mixtes, que nous notons v.

On donne dans un premier temps un exemple simple de jeu de recherche-dissimulation. Considérons le jeu où l'espace de recherche Q est le cercle de rayon 1, l'origine O est un point quelconque de Q, et le rayon de détection r vaut 0. Soit la stratégie du chercheur qui consiste à parcourir à vitesse 1 le cercle avec probabilité $1/2$ dans les sens trigonométrique et anti-trigonométrique respectivement. Soit $h \in Q$ le point de dissimulation. Le paiement induit par ces stratégies est

$$\frac{1}{2}h + \frac{1}{2}(2\pi - h) = \pi,$$

et ne dépend donc pas du point de dissimulation.

Supposons que le dissimulateur joue la distribution uniforme sur le cercle. Une meilleure réponse du chercheur est de parcourir le cercle dans le sens trigonométrique à vitesse 1. Ceci induit le paiement

$$\frac{1}{2\pi}\int_0^{2\pi} t dt = \pi.$$

Ainsi la valeur du jeu est π, et l'on a exhibé un couple de stratégies optimales.

On considère le taux auquel Sally peut découvrir de nouveaux points de Q. La mesure de Lebesgue d'un sous-ensemble mesurable $B \subset \mathbb{R}^n$ est notée $\lambda(B)$. De plus, $B_r(0)$ désigne la boule fermée de rayon r et de centre 0. Enfin, $\lambda(B_r)$ désigne la mesure de Lebesgue de toute boule fermée de rayon r.

Définition 1.15. Le *taux de recherche maximal* ρ du chercheur est donné par

$$\rho = \sup_{s \in \mathcal{S}, t>0} \frac{\lambda(s([0,t]) + B_r(0)) - \lambda(B_r)}{t},$$

où $s([0,t]) = \{s(\tau) \mid \tau \in [0,t]\}$ est l'image de $[0,t]$ par s, et $s([0,t]) + B_r(0) = \{y \in \mathbb{R}^n \mid d(s([0,t],y)) \leq r\}$.

Ainsi, dans $\mathbb{R}^2$ et $\mathbb{R}^3$ munis de la norme euclidienne, et sous des hypothèses raisonnables, le taux de recherche maximal est $2r$ et πr^2 respectivement, c'est-à-dire la largeur de balayage de Sally.

La proposition suivante donne une borne inférieure sur la valeur, garantie par la *distribution uniforme* sur Q du dissimulateur. Intuitivement, si le dissimulateur joue la distribution uniforme, le mieux que puisse faire le chercheur est de couvrir tout Q le plus rapidement possible.

Proposition 1.16. *Soit Q un espace de recherche de mesure de Lebesgue strictement positive. La valeur du jeu de recherche-dissimulation vérifie*

$$v \geq \frac{(\lambda(Q) - \lambda(B_r))^2}{2\lambda(Q)\rho}.$$

1.3.2 Recherche-dissimulation sur les réseaux

Une part importante de la littérature sur les jeux de recherche-dissimulation est consacrée à l'étude des *réseaux*. Si ces espaces de recherche particuliers ont un intérêt naturel du point de vue des applications, leur étude se révèle également fructueuse du point de vue théorique.

De manière informelle, on entend par réseau un graphe continu. Ainsi Harry peut se dissimuler sur n'importe quel point d'une arête, et Sally se déplace continûment le long de ces dernières. On utilise un vocabulaire similaire à celui de la théorie des graphes, et l'on parlera par exemple d'arête, de sommet, de chemin, de cycle, d'arbre, etc. De plus, on définit naturellement une distance, ainsi qu'une mesure de Lebesgue sur les réseaux, également notée λ. Ces notions sont définies formellement au Chapitre 4.

Deux classes de réseaux ont une importance particulière au regard des jeux de recherche-dissimulation : les *réseaux eulériens* et les *arbres*. Ces classes représentent les cas limites en terme de valeur du jeu à longueur de réseau fixée.

On suppose ici que le rayon de détection r de Sally vaut 0, c'est-à-dire que Sally rencontre Harry s'ils sont au même point. Dans un réseau, le taux de recherche maximal ρ vaut 1.

Définition 1.17. On appelle *cycle minimal* ou encore *cycle du postier chinois* d'un réseau $\mathcal{N}$, un cycle visitant tous les points de $\mathcal{N}$ de longueur minimale. On note $\overline{\lambda}(\mathcal{N})$ la longueur d'un tel cycle.

Définition 1.18. Un réseau $\mathcal{N}$ est dit eulérien s'il possède un cycle minimal de longueur $\lambda(\mathcal{N})$. Un tel cycle est appelé *cycle eulérien*.

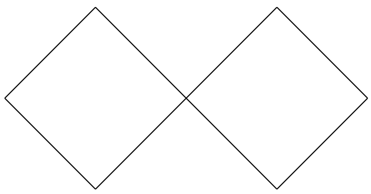

Figure 1.8: Un réseau eulérien

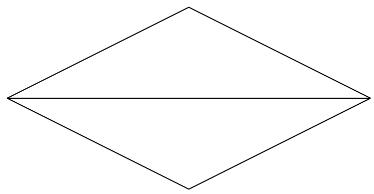

Figure 1.9: Un réseau non eulérien

La *stratégie du postier chinois uniforme* de Sally consiste à choisir uniformément un cycle minimal débutant en l'origine O. Dans un réseau eulérien on parle de *stratégie eulérienne uniforme*.

Dans un arbre, la *distribution égalisatrice aux embranchements* de Harry est la mesure de probabilité avec support sur les feuilles de l'arbre, telle que pour tout sommet v avec arêtes sortantes $e_1, \ldots, e_n$, la probabilité que Harry se cache dans le sous-arbre émanent de e_i est égale à la longueur de ce sous-arbre divisée par la longueur du sous-arbre de racine v. On illustre cette stratégie Fig. 1.10. La longueur des arêtes est indiquée le long de celles-ci, tandis que la probabilité affectée à chaque feuille est indiquée sous ces dernières.

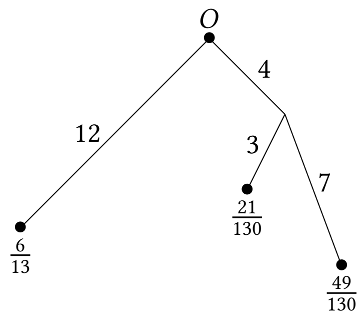

Figure 1.10: La distribution égalisatrice aux embranchements

Le Théorème 1.11 ci-dessous constitue l'un des principaux théorèmes liés à l'étude des jeux de recherche-dissimulation, voir (Alpern and Gal, 2003).

Théorème 1.11. *Soit un réseau $\mathcal{N}$. La valeur du jeu de recherche-dissimulation vérifie*

$$\frac{\lambda(\mathcal{N})}{2} \leq v \leq \frac{\bar{\lambda}(\mathcal{N})}{2} \leq \lambda(\mathcal{N}).$$

De plus la borne inférieure est atteinte si et seulement si $\mathcal{N}$ est eulérien, et la borne supérieure $\lambda(\mathcal{N})$ est atteinte si et seulement si $\mathcal{N}$ est un arbre.

Si $\mathcal{N}$ est eulérien, alors la distribution uniforme de Harry et la stratégie eulérienne uniforme de Sally forment un couple de stratégies optimales.

Si $\mathcal{N}$ est un arbre, alors la distribution égalisatrice aux embranchements de Harry et la stratégie du postier chinois uniforme de Sally forment un couple de stratégies optimales.

Une question naturelle est de savoir pour quels réseaux la stratégie du postier chinois uniforme est optimale. Autrement dit, quels sont les réseaux pour lesquels la borne $\bar{\lambda}(\mathcal{N})/2$ du Théorème 1.11 est atteinte ? La réponse est donnée par une classe de réseaux contenant à la fois les réseaux eulériens et les arbres. Un *réseau faiblement eulérien* est un réseau obtenu en remplaçant les sommets d'un arbre par un réseau eulérien.

Le théorème suivant dû à Gal (2000) répond à notre question.

Théorème 1.12. *La stratége du postier chinois uniforme est optimale si et seulement si $\mathcal{N}$ est faiblement eulérien, et l'on a*

- *si $\mathcal{N}$ est faiblement eulérien, alors $v = \frac{\bar{\lambda}(\mathcal{N})}{2}$;*
- *si $\mathcal{N}$ n'est pas faiblement eulérien, alors $v < \frac{\bar{\lambda}(\mathcal{N})}{2}$.*

1.3.3 Contribution de la thèse : jeux de patrouille et de dissimulation continus

Cette partie fait écho au Chapitre 4 intitulé *Continuous patrolling and hiding games.* Nous en présentons ici les principaux résultats. Le modèle étudié ici, dit *jeu de patrouille* considère le paradigme 2 mentionné plus haut, à savoir la maximisation de la probabilité de capture en un temps imparti.

Dans un jeu de patrouille, un *attaquant* choisit un temps et un lieu à attaquer dans l'espace de recherche Q, tandis qu'un *patrouilleur* y marche continûment à vitesse au plus 1, en ayant choisi son point de départ. Lorsque l'attaque survient, le patrouilleur a un délai $m \in \mathbb{R}_+$ pour être à distance au plus $r \in \mathbb{R}_+$ du point d'attaque, auquel cas l'attaque est désamorcée. Dans le cas contraire, l'attaque est un succès. Le paramètre m représente la durée nécessaire à la mise en place de l'attaque. On peut par exemple penser au temps nécessaire à l'amorçage d'une bombe.

Un jeu de patrouille est donc donné par un triplet (Q, m, r). On note $\mathcal{W}$ l'ensemble des stratégies du patrouilleur. C'est l'ensemble des fonctions 1-lipschitziennes de $\mathbb{R}_+$ à valeur dans Q. L'ensemble de stratégies de l'attaquant est $\mathcal{A} = Q \times \mathbb{R}_+$.

Le paiement du patrouilleur $g_{m,r} : \mathcal{W} \times \mathcal{A} \to \{0, 1\}$ est donné par

$$g_{m,r}(w, (y, t)) = \begin{cases} 1 & \text{si } d(y, w([t, t + m])) \leq r \\ 0 & \text{sinon,} \end{cases}$$

où $w([t, t + m]) = \{w(\tau) \mid \tau \in [t, t + m]\}$.

Il est à noter que les jeux de patrouilles ont été introduits, dans le cas discret lorsque l'espace de recherche est un graphe, par Alpern et al. (2011), voir également (Alpern et al., 2016a, 2019). Dans le cas continu, citons (Alpern et al., 2016b).

Il résulte du Théorème 1.2 ainsi que du théorème d'Ascoli que le jeu de patrouille (Q, m, r) joué en stratégies mixtes a une valeur, notée $V_Q(m, r)$. On donne une borne supérieure sur la valeur qui, comme pour les jeux de recherche-dissimulation, repose sur le taux de recherche maximal (Définition 1.15) et sur la *stratégie uniforme* de l'attaquant, qui consiste à attaquer uniformément l'espace de recherche au temps 0. Intuitivement, une meilleure réponse du patrouilleur est de couvrir autant de points de Q que possible entre les temps 0 et m.

Proposition 1.19. *Soit Q un espace de recherche tel que $\lambda(Q) > 0$. Alors la valeur du jeu de patrouille (Q, m, r) vérifie*

$$V_Q(m, r) \leq \frac{m\rho + \lambda(B_r)}{\lambda(Q)}.$$

On donne également le résultat de décomposition suivant. On considère un espace de recherche Q comme l'union d'espaces de recherche $Q_1, \ldots, Q_n$, pour lesquels la valeur du jeu de patrouille correspondant est éventuellement connue. La valeur du jeu de patrouille sur Q est alors bornée inférieurement en fonction des valeurs de jeux de patrouilles impliqués dans la décomposition.

Proposition 1.20. *Soient Q et $Q_1, \ldots, Q_n$ des espaces de recherche tels que $Q = \cup_{i=1}^{n} Q_i$. Alors pour tous $m, r \in \mathbb{R}_+$*

$$V_Q(m,r) \geq \frac{1}{\sum_{i=1}^{n} V_{Q_i}(m,r)^{-1}}.$$

Concernant les jeux de patrouille sur les réseaux, on montre le résultat suivant donnant valeur et stratégies optimales lorsque le réseau est eulérien. La *stratégie uniforme* du patrouilleur sur un réseau eulérien consiste à fixer un cycle eulérien et à choisir uniformément un point de départ.

Théorème 1.13. *Si $\mathcal{N}$ est un réseau eulérien, alors*

$$V_{\mathcal{N}}(m,0) = \min\left(\frac{m}{\lambda(\mathcal{N})}, 1\right).$$

De plus les stratégies uniformes de l'attaquant et du patrouilleur sont optimales.

Un espace de recherche simple est un espace de recherche dans $\mathbb{R}^2$ dont la frontière vérifie des hypothèses faibles de régularité. Pour ces espaces de recherche on démontre le résultat asymptotique suivant.

Théorème 1.14. *Si Q est un espace de recherche simple muni de la norme euclidienne, alors*

$$V_Q(m,r) \sim \frac{2rm}{\lambda(Q)},$$

lorsque r tend vers 0.

Supposons que la durée nécessaire au succès de l'attaque m vaille 0. Dans ce cas on montre que le jeu revient à un jeu statique appelé *jeu de dissimulation*, introduit par Gale and Glassey (1974) dans le cas où l'espace de recherche est un disque, voir également (Ruckle, 1983), dont l'ensemble de stratégies des deux joueurs est Q et dont le paiement est

$$h_r : (x,y) \mapsto \begin{cases} 1 & \text{si } \|x-y\| \leq r \\ 0 & \text{sinon.} \end{cases}$$

C'est-à-dire que les deux joueurs choisissent un point de Q, s'ils sont à distance inférieure à r, c'est un succès pour le chercheur, dans le cas contraire c'est un échec.

Dans un jeu de dissimulation, les deux joueurs ont donc le même ensemble de stratégies. De plus la fonction de paiement est symétrique au sens où, si $\mu \in \Delta(Q)$ et $y \in Q$, alors $h_r(\mu, y) = h_r(y, \mu) = \mu(B_r(y) \cap Q)$.

On dit que la stratégie d'un joueur est *égalisatrice*, si le paiement lorsque ce joueur joue cette stratégie ne dépend pas de l'action de l'autre joueur. On a alors la proposition suivante.

Proposition 1.21. *Soit $\mu \in \Delta(Q)$. Alors μ est une stratégie égalisatrice (de paiement constant c) si et seulement si μ est une stratégie optimale pour chacun des joueurs (et dans ce cas $V_Q(r) = c$).*

On démontre le théorème suivant, qui donne asymptotiquement la valeur d'un jeu de dissimulation sur un espace de recherche de mesure de Lebesgue strictement positive.

Théorème 1.15. *Soit Q un espace de recherche tel que $\lambda(Q) > 0$, alors*

$$V_Q(r) \sim \frac{\lambda(B_r)}{\lambda(Q)},$$

lorsque r tend vers 0.

Une conséquence du Théorème 1.15 est que pour un espace de recherche de mesure de Lebesgue strictement positive, on a $V_Q(r) \sim r^n \lambda(B_1)/\lambda(Q)$ lorsque r tend vers 0. On montre que pour un espace de recherche de mesure de Lebesgue nulle, il n'est pas toujours vrai que V_Q admette un équivalent de la forme Mr^α, avec α et M strictement positifs, lorsque r tend vers 0. Notre contre-exemple s'appuie sur un ensemble de Cantor.

1.3.4 Contribution de la thèse : jeux de recherche-dissimulation stochastiques

Cette partie fait écho au Chapitre 5 intitulé *When Sally found Harry: A Stochastic search game*. Nous en présentons ici les principaux résultats. On introduit un modèle de jeu de recherche-dissimulation sur un graphe qui évolue avec le temps. On mélange ainsi les aspects dynamiques liés à l'évolution du graphe à ceux liés à la trajectoire du chercheur. Ce modèle est appelé *jeu de recherche-dissimulation stochastique*.

Nous étudions un jeu de recherche-dissimultation dans lequel Harry se cache dans une arête d'un graphe et Sally parcourt le graphe à la recherche de Harry. Son but étant de le trouver le plus vite possible.

La nouveauté du modèle est qu'en raison de circonstances diverses, à un moment donné, certaines arêtes peuvent ne pas être disponibles, de sorte que le graphe évolue de façon aléatoire avec le temps. À chaque étape, chaque arête e du graphe est, indépendamment des autres, active avec probabilité p_e et inactive avec probabilité $1 - p_e$. Décrivons le modèle plus en détails.

Soit $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ un graphe connexe non orienté, avec $\mathcal{V}$ l'ensemble des sommets et $\mathcal{E}$ l'ensemble d'arêtes. Soit $O \in \mathcal{V}$ l'*origine* du graphe $\mathcal{G}$. Soit $\mathbb{G}$ l'ensemble des sous-graphes de $\mathcal{G}$. Pour tout $v \in \mathcal{V}$, on appelle $\mathcal{N}(\mathcal{G}, v)$ le *voisinage immédiat* de v dans $\mathcal{G}$:

$$\mathcal{N}(\mathcal{G}, v) = \{v\} \cup \{u \in \mathcal{V} \mid \{v, u\} \in \mathcal{E}\}.$$

Le graphe évolue en temps discret de la manière suivante. Soit $\boldsymbol{p} = (p_e)_{e \in \mathcal{E}} \in]0, 1]^{\mathcal{E}}$. À chaque étape $t \geq 1$, chacune des arêtes $e \in \mathcal{E}$ est active avec probabilité p_e ou inactive avec probabilité $1 - p_e$, indépendamment des autres arêtes. On a ainsi défini un processus aléatoire de graphes sur $\mathbb{G}$ que l'on note $(\mathcal{G}_t)_t = (\mathcal{V}, \mathcal{E}_t)_{t\geq 1}$, où $\mathcal{E}_t$ est l'ensemble des arêtes actives à l'étape t.

Le jeu se déroule de la manière suivante. À l'étape 0 les deux joueurs connaissent $\mathcal{G}_0 = \mathcal{G}$ et la position initiale de Sally $v_0 = O$. Harry choisit une arête $e \in \mathcal{E}$. Puis le graphe $\mathcal{G}_1$ est sélectionné et Sally choisit $v_1 \in \mathcal{N}(\mathcal{G}_1, v_0)$. Si $\{v_0, v_1\} = e$, alors le jeu se termine et le paiement de Harry est 1, sinon le graphe $\mathcal{G}_2$ est sélectionné et le jeu continue. Récursivement, à chaque étape $t \geq 1$, connaissant $h_t = (\mathcal{G}_0, v_0, \ldots, \mathcal{G}_{t-1}, v_{t-1}, \mathcal{G}_t)$, Sally choisit $v_t \in \mathcal{N}(\mathcal{G}_t, v_{t-1})$. Si $\{v_{t-1}, v_t\} = e$, alors le jeu se termine et le paiement de Harry vaut t, sinon le graphe $\mathcal{G}_{t+1}$ est sélectionné et le jeu continue.

On peut borner la valeur du jeu de recherche-dissimulation stochastique pour tout $\boldsymbol{p} \in]0,1]^{\mathcal{E}}$, notée $\mathrm{val}(\boldsymbol{p})$, en fonction de la valeur du *jeu de recherche-dissimulation déterministe*, c'est-à-dire lorsque les paramètres d'activations $\boldsymbol{p}$ valent tous 1.

Proposition 1.22. *Pour tout $\boldsymbol{p} \in]0,1]^{\mathcal{E}}$ la valeur du jeu de recherche-dissimulation stochastique vérifie*

$$\frac{\mathrm{val}(\mathbf{1})}{1 - (1 - \min_{e\in\mathcal{E}} p_e)^{\delta}} \leq \mathrm{val}(\boldsymbol{p}) \leq \frac{\mathrm{val}(\mathbf{1})}{\min_{e\in\mathcal{E}} p_e},$$

où δ est le degré maximal de $\mathcal{G}$. Ainsi,

$$\mathrm{val}(\boldsymbol{p}) \longrightarrow \mathrm{val}(\mathbf{1}), \quad \textit{lorsque } \min_{e\in\mathcal{E}} p_e \longrightarrow 1.$$

On retranscrit dans un premier temps les notions et résultats définis dans le cas des réseaux à celui des graphes dans le cadre déterministe. On montre en particulier le théorème suivant, qui est l'analogue du Théorème 1.11 pour les réseaux.

Théorème 1.16. *Pour tout graphe $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, la valeur du jeu de recherche-dissimulation déterministe $\Gamma = \langle \mathcal{G}, O, \mathbf{1}\rangle$ satisfait les inégalités*

$$\frac{\operatorname{card}\mathcal{E} + 1}{2} \leq \mathrm{val}(1) \leq \operatorname{card}\mathcal{E}.$$

De plus, si $\operatorname{card}\mathcal{E} > 1$, la borne supérieure est atteinte si et seulement si $\mathcal{G}$ est un arbre et la borne inférieure est atteinte si et seulement si $\mathcal{G}$ est eulérien.

Si $\mathcal{G}$ est un graphe eulérien, alors la distribution uniforme sur $\mathcal{E}$ et la stratégie eulérienne uniforme forment un couple de stratégies optimales.

Si $\mathcal{G}$ est un arbre, alors la distribution égalisatrice aux branchements et la stratégie du postier chinois uniforme forment un couple de stratégies optimales.

On voit ici également que les arbres et les graphes eulériens représentent les classes limites en terme de valeur du jeu de recherche-dissimulation déterministe à nombre d'arêtes fixé. Notre but est d'étudier ces deux classes dans le cadre stochastique. On se place dans le cas où $\boldsymbol{p}_e = p$ pour tout $e \in \mathcal{E}$, c'est-à-dire que le paramètre d'activation est le même pour chaque arête.

On définit une classe de stratégies de Sally dans les arbres dites *stratégies de descente en premier*. Ces dernières ont la propriété en chaque sommet de ne jamais revenir en arrière avant d'avoir visité tout le sous-arbre. Elles généralisent les cycles du postier chinois du cas déterministe.

Définition 1.23. Une stratégie de descente en premier sur un arbre est une stratégie de Sally qui, en arrivant à un sommet :

- si l'ensemble des arêtes sortantes actives et non cherchées est non vide, emprunte l'une de ces arêtes (éventuellement de manière aléatoire) ;
- si toutes les arêtes sortantes non cherchées sont inactives, attend ;
- si toutes les arêtes sortantes ont été cherchées et l'arête remontante est active, la prend ;
- si toutes les arêtes sortantes ont été cherchées et l'arête remontante est inactive, attend.

Définition 1.24. Une stratégie de descente en premier sur un arbre $\mathcal{T}$ induit un temps espéré de parcours depuis l'origine, couvrant tout l'arbre et revenant à l'origine. On l'appelle le *temps de cycle* de $\mathcal{T}$ et on le note $\tau(O)$.

On remarque que $\tau(O)$ dépend de p, mais est indépendant de la stratégie de descente en premier choisie. À titre d'exemple, dans l'arbre représenté Fig. 1.11, le temps de cycle est $\tau(O) = \frac{2}{1-(1-p)^2} + \frac{6}{p}$.

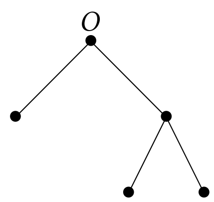

Figure 1.11: Un arbre simple

La *stratégie de descente en premier biaisée* est une stratégie de descente en premier qui en tout sommet v, lorsque des branches non cherchées sont actives simultanément, donne une certaine probabilité d'emprunter chacune d'entre elles qui dépend de p ainsi que de la géométrie du sous-arbre de racine v.

On généralise également la distribution égalisatrice aux embranchements au cadre stochastique. Le paramètre pertinent n'est plus la longueur ou le nombre d'arêtes des sous-arbres, mais le temps de cycle de ces derniers. La distribution égalisatrice aux embranchements de Harry est la mesure de probabilité avec support sur les arêtes feuilles de l'arbre, telle que pour tout sommet v avec arêtes sortantes $e_1, \ldots, e_n$, la probabilité que Harry se cache dans le sous-arbre émanent de e_i est égale au temps de cycle de ce sous-arbre divisé par la somme des temps de cycles des sous-arbres émanant des arêtes $e_1, \ldots, e_n$.

On démontre alors les théorèmes suivants pour les *arbres binaires*. La quantité Λ est définie par récurrence, elle dépend de p ainsi que de la géométrie de l'arbre.

Théorème 1.17. *Il existe $p_0 \in]0,1[$ tel que pour tout $p \geq p_0$, le temps espéré d'atteinte d'une arête feuille en utilisant la stratégie de descente en premier biaisée est $\frac{1}{2}\tau(O) + \Lambda(O)$. Ainsi, pour tout $p \geq p_0$, on a*

$$\mathrm{val}(p) \leq \frac{1}{2}\tau(O) + \Lambda(O).$$

Théorème 1.18. *La distribution égalisatrice aux embranchements de Harry donne un paiement constant contre toute stratégie de descente en premier, et ce paiement vaut*

$$\frac{1}{2}\tau(O) + \Lambda(O).$$

Les Théorèmes 1.17 et 1.18 impliquent que dans un arbre binaire si les stratégies de descente en premier sont des meilleures réponses à la distribution égalisatrice aux embranchements, alors pour tout p assez grand la valeur du jeu est $\frac{1}{2}\tau(O) + \Lambda(O)$. Et de plus les stratégies de descente en premier biaisée et la distribution égalisatrice aux embranchements sont optimales.

On montre toutefois l'existence d'un arbre binaire pour lequel les stratégies de descente en premier ne sont pas des meilleures réponses à la distribution égalisatrice aux embranchements. On a cependant le corollaire suivant. La *stratégie de descente en premier uniforme* est la stratégie de descente en premier qui choisit uniformément parmi les arêtes sortantes actives et non cherchées, elle est donc indépendante de p.

Corollaire 1.25. *Si le graphe $\mathcal{G}$ est une ligne composée de L arêtes, alors les stratégies de descente en premier sont des meilleures réponses à la distribution égalisatrice aux embranchements. La valeur vaut*

$$\mathrm{val}(p) = \frac{1}{2}\tau(O) + \Lambda(O) = \frac{L}{p} + \frac{1}{1-(1-p)^2} - \frac{1}{p},$$

pour tout $p \in]0,1]$, si l'origine O n'est pas l'un des sommets extrêmes.

De plus, la distribution égalisatrice aux embranchements et la stratégie de descente en premier uniforme forment un couple de stratégies optimales.

Pour les graphes eulériens, on généralise la stratégie eulérienne uniforme au cadre stochastique, et on appelle *stratégie eulérienne* une stratégie qui a chaque sommet choisit une arête active qui n'a pas encore été cherchée et telle que le chemin induit appartienne à un cycle eulérien. La stratégie eulérienne qui a chaque sommet choisit uniformément parmi les arêtes est la *stratégie eulérienne* *uniforme.* La stratégie eulérienne uniforme induit un temps espéré de parcours depuis l'origine couvrant tout le graphe $\mathcal{G}$. Il est noté $\theta(\mathcal{G})$.

Un *graphe eulérien parallèle* est composé d'un nombre pair de chemins parallèles qui joignent deux sommets, l'un de ces sommets étant l'origine O, voir la figure Fig. 1.12. Un tel graph est noté $\mathcal{P}_{2m}(\boldsymbol{\lambda})$, où $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{2m})$ est le vecteur des longueurs des chemins parallèles.

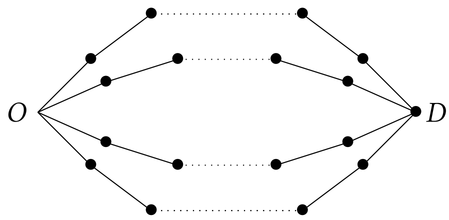

Figure 1.12: Un graphe eulérien parallèle

Pour les graphes eulériens parallèles on démontre les deux théorèmes suivants. La quantité Φ est définie par récurrence. Elle dépend de p ainsi que de du nombre de chemins parallèles.

Théorème 1.19. *Soit un graphe eulérien parallèle $\mathcal{P}_{2m}(\boldsymbol{\lambda})$. Pour tout $p \in]0,1]$, le temps espéré pour atteindre une arête en jouant la stratégie eulérienne uniforme est*

$$\frac{\theta(\mathcal{P}_{2m}(\boldsymbol{\lambda})) + 1/p}{2} + \Phi_m.$$

Ainsi, pour tout $p \in]0,1]$, on a

$$\mathrm{val}(p) \leq \frac{\theta(\mathcal{P}_{2m}(\boldsymbol{\lambda})) + 1/p}{2} + \Phi_m.$$

Théorème 1.20. *Soit un graphe eulérien parallèle $\mathcal{P}_{2m}(\boldsymbol{\lambda})$. Pour tout $p \in]0,1]$, la distribution uniforme de Harry donne un paiement constant contre toute stratégie eulérienne de Sally, qui est la quantité*

$$\frac{\theta(\mathcal{P}_{2m}(\boldsymbol{\lambda})) + 1/p}{2} + \Phi_m.$$

On montre que les stratégies eulériennes ne sont pas toujours meilleures réponses à la distribution uniforme. On a toutefois le corollaire suivant.

Corollaire 1.26. *La valeur du jeu joué sur le cercle composé de L arêtes est*

$$\mathrm{val}(p) = \frac{\theta(\mathcal{G}) + 1/p}{2} + \Phi_2 = \frac{1}{1-(1-p)^2} + \frac{L-1}{2p},$$

pour tout $p \in]0,1]$. De plus la distribution uniforme et la stratégie eulérienne uniforme sont optimales.

1.4 Jeux répétés à information incomplète

Jusqu'à présent, les modèles de jeux présentés avaient en commun la connaissance de la part des joueurs de l'état de la nature ainsi que des actions jouées précédemment. On se place désormais dans la situation où les joueurs reçoivent seulement des *signaux* sur l'état ainsi que sur les actions, qui peuvent ne pas révéler complètement ces derniers.

1.4.1 Modèle général de jeu répété avec signaux

Présentons dans un premier temps un modèle général de *jeu répété avec signaux*, voir (Mertens et al., 2015). Un jeu répété avec signaux est la donnée d'un septuplet $\Gamma = (\Omega, A, B, C, D, g, \rho)$ où

- Ω est l'ensemble d'états ;
- A et B sont les ensembles d'actions des joueurs 1 et 2 respectivement ;
- C et D sont les ensembles de signaux des joueurs 1 et 2 respectivement ;
- $g : \Omega \times A \times B \longrightarrow \mathbb{R}$ est la fonction de paiement ;
- $\rho : \Omega \times A \times B \longrightarrow \Delta(\Omega \times C \times D)$ est la probabilité de transition.

On suppose ici les ensembles Ω, A, B, C et D non vides et finis. Soit $p \in \Delta(\Omega \times C \times D)$ une distribution initiale. Le jeu $\Gamma(p)$ se déroule de la manière suivante.

- Avant le début du jeu, un triplet (ω_1, c_0, d_0) est tiré selon p. L'état initial est ω_1, les joueurs 1 et 2 reçoivent les signaux c_0 et d_0 respectivement.
- À chaque étape $t \geq 1$, les joueurs 1 et 2 choisissent simultanément des actions $a_t \in A$ et $b_t \in B$ respectivement. Un triplet (ω_{t+1}, c_t, d_t) est tiré selon $\rho(\cdot|\omega_t, a_t, b_t)$. Les joueurs 1 et 2 reçoivent les signaux c_t et d_t respectivement. Le paiement d'étape est $g_t = g(\omega_t, a_t, b_t)$. Le jeu va dans l'état ω_{t+1} et l'on passe à l'étape suivante.

On donne formellement la notion de stratégie dans un jeu répété avec signaux. On note $H_t = \Omega \times C \times D \times (\Omega \times A \times B \times C \times D)^{t-1}$ l'ensemble des histoires possibles à l'étape $t \geq 1$. On munit H_t de la tribu produit notée $\mathcal{H}_t$. On note $H_\infty = \Omega \times C \times D \times (\Omega \times A \times B \times C \times D)^{\mathbb{N}^*}$ l'ensemble des histoires infinies, que l'on munit de la tribu produit engendrée par $\bigcup_{t \geq 1} \mathcal{H}_t$.

L'ensemble des histoires privées du joueur 1 à l'étape t est $H_t^1 = C \times (A \times C)^{t-1}$. Il est muni de la tribu produit notée $\mathcal{H}_t^1$. L'ensemble des histoires privées du joueur 2 qui est défini de manière analogue.

Définition 1.27. Une *stratégie de comportement* du joueur 1 est suite d'applications $\sigma = (\sigma_t)_{t \geq 1}$ telle que pour tout $t \geq 1$, σ_t est une application mesurable de $(H_t^1, \mathcal{H}_t^1)$ dans $\Delta(A)$. On définit de manière analogue les stratégies de comportement du joueur 2.

On note $\mathcal{S}$ et $\mathcal{T}$ les ensembles de stratégies de comportement des joueurs 1 et 2 respectivement. Une stratégie de comportement est dite *pure* si pour chaque histoire finie $h^1_t \in H^1_t$, on a $\sigma_t(h^1_t) \in A$. Une *stratégie mixte* est une distribution de probabilité sur l'ensemble stratégies pures. On a deux définitions analogues pour le joueur 2.

Une probabilité initiale $p \in \Delta(\Omega \times C \times D)$ ainsi qu'un couple de stratégies (σ, τ), de comportement ou mixtes, induisent naturellement sur l'ensemble des histoires finies $\bigcup_{t \geq 1} H_t$ une mesure de probabilité. Le théorème d'extension de Kolmogorov nous indique que cette mesure s'étend de manière unique à H_∞. On note cette mesure de probabilité $\mathbb{P}^p_{\sigma,\tau}$, et $\mathbb{E}^p_{\sigma,\tau}$ l'espérance correspondante.

Tout comme pour les jeux stochastiques, le théorème de Kuhn, voir Théorème 1.4, s'applique dans ce contexte.

Il existe également deux façons principales d'agréger les paiements d'étapes, à savoir au travers des moyennes d'Abel et de Césaro.

Soit $\lambda \in]0,1]$, le *jeu escompté au taux* λ, noté $\Gamma_\lambda(p)$, est le jeu dont les ensembles de stratégies sont $\mathcal{S}$ et $\mathcal{T}$ et dont le paiement est $\gamma^p_\lambda : \mathcal{S} \times \mathcal{T} \to \mathbb{R}$ défini par

$$\gamma^p_\lambda(\sigma,\tau) = \mathbb{E}^p_{\sigma,\tau}\left(\sum_{t \geq 1} \lambda(1-\lambda)^{t-1} g_t\right).$$

Soit $N \in \mathbb{N}^*$, le *jeu répété* N *fois*, noté $\Gamma_N(p)$ est le jeu dont les ensembles de stratégies sont $\mathcal{S}$ et $\mathcal{T}$ et dont le paiement est $\gamma^p_N : \mathcal{S} \times \mathcal{T} \to \mathbb{R}$ défini par

$$\gamma^p_N(\sigma,\tau) = \mathbb{E}^p_{\sigma,\tau}\left(\frac{1}{N}\sum_{t=1}^{N} g_t\right).$$

Le théorème du Théorème 1.3 nous assure l'existence de la valeur dans les jeux $\Gamma_\lambda(p)$ et $\Gamma_N(p)$, ainsi que celle de stratégies optimales. On note

$$v_\lambda(p) = \max_{\sigma \in \mathcal{S}} \min_{\tau \in \mathcal{T}} \gamma^p_\lambda(\sigma,\tau) = \min_{\tau \in \mathcal{T}} \max_{\sigma \in \mathcal{S}} \gamma^p_\lambda(\sigma,\tau),$$
$$v_N(p) = \max_{\sigma \in \mathcal{S}} \min_{\tau \in \mathcal{T}} \gamma^p_\lambda(\sigma,\tau) = \min_{\tau \in \mathcal{T}} \max_{\sigma \in \mathcal{S}} \gamma^p_N(\sigma,\tau).$$

La valeur uniforme est définie de manière analogue à celle définie dans les jeux stochastiques.

Une littérature pléthorique s'est développée sur les jeux répétés à information incomplète. Citons, à titre d'exemples, les travaux de Cardaliaguet et al. (2012); Forges (1982); Gensbittel (2014); Gensbittel et al. (2014); Gensbittel and Renault (2015); Gimbert et al. (2016); Laraki (2001, 2002); Mertens and Zamir (1971); Neyman (2008); Renault (2006, 2012); Renault and Venel (2017); Rosenberg et al. (2003, 2004); Sorin (1984, 1985); Ziliotto (2016b). Pour une revue de la littérature, on renvoie également vers (Renault, 2018).

1.4.2 Jeux répétés à information incomplète d'un côté

On présente ici le modèle de *jeu répété avec information incomplète d'un côté* introduit par Aumann and Maschler (1995). Avant le début du jeu, la probabilité initiale $p \in \Delta(\Omega)$ sélectionne un état $\omega \in \Omega$ qui est fixé pendant toute la durée du jeu. Celui-ci est révélé au joueur 1 mais pas au joueur 2. Puis à chaque étape $t \geq 1$, les joueurs 1 et 2 sélectionnent simultanément des actions $a_t \in A$ et $b_t \in B$ respectivement. Celles-ci sont révélées aux deux joueurs avant l'étape suivante. La probabilité p représente donc la *croyance* initiale du joueur 2 sur l'état de la nature.

Voici comment le joueur 1 utilise son information. Ce dernier choisit sa première action (ou plus généralement un message ou signal k d'un ensemble fini K) en fonction de l'état de la nature ω sélectionné. Soit $x \in \Delta(K)^{\Omega}$ la probabilité de transition jouée par le joueur 1, à savoir que si l'état est ω, il envoie le signal k avec probabilité $x^{\omega}(k)$.

On note λ_k la probabilité totale que k soit sélectionné. Celle-ci vaut

$$\sum_{\omega \in \Omega} p(\omega) x^{\omega}(k),$$

de plus si $\lambda_k > 0$, alors la probabilité conditionnelle sur Ω sachant k, appelée également *croyance a posteriori* vaut

$$\hat{p}(x,k) = \left(\frac{p(\omega) x^{\omega}(k)}{\lambda_k} \right)_{\omega \in \Omega}.$$

Puisque $\sum_{k \in K} \lambda_k \hat{p}(x,k) = p$, les croyances a posteriori contiennent la *croyance a priori* p dans leur enveloppe convexe. Le lemme suivant, dit *lemme d'éclatement*, fondamental en jeux répétés à information incomplète, constitue en quelque sorte une réciproque : le joueur 1 peut induire toute croyance a posteriori contenant la probabilité initiale p dans leur enveloppe convexe.

Lemme 1.28 (Lemme d'éclatement). *Supposons que $p = \sum_{k \in K} \lambda_k p_k$ avec pour tout $k \in K$, $\lambda_k > 0$, $p_k \in \Delta(\Omega)$ et $\sum_{k \in K} \lambda_k = 1$.*

Alors il existe une probabilité de transition $x \in \Delta(K)^{\Omega}$ telle que pour tout $k \in K$

$$\lambda_k = \sum_{\omega \in \Omega} p(\omega) x^{\omega}(k) \text{ et } \hat{p}(x,k) = p_k.$$

Pour toute fonction f semi-continue supérieurement définie sur $\Delta(\Omega)$ à valeur dans $\mathbb{R}$, la plus petite fonction concave en tout point supérieure à f est notée $\operatorname{cav} f$. La fonction $\operatorname{cav} f$ est continue et pour tout $p \in \Delta(\Omega)$ on a

$$\begin{aligned} \operatorname{cav} f(p) = \max\{ \sum_{k \in K} \lambda_k f(p_k) \mid K \text{ fini}, \forall k \in K\ \lambda_k > 0,\ p_k \in \Delta(K), \\ \sum_{k \in K} \lambda_k = 1,\ \sum_{k \in K} \lambda_k p_k = p \}. \end{aligned}$$

Le lemme suivant est une conséquence du Lemme 1.28.

Lemme 1.29. *Dans tout jeu finiment ou infiniment répété, si le joueur 1 possède toute l'information, alors s'il garantit f il garantit également* $\operatorname{cav} f$.

Aumann and Maschler (1995) montrent le théorème suivant.

Théorème 1.21. *Soit $u(p)$ la valeur du jeu matriciel avec ensembles d'actions A et B et de paiement $g_p : A \times B \to \mathbb{R}$ défini pour tout $(a, b) \in A \times B$ par*

$$g_p(a, b) = \sum_{\omega \in \Omega} p(\omega) g(\omega, a, b).$$

Alors le jeu répété avec information incomplète d'un côté $\Gamma(p)$ a une valeur uniforme qui vaut $\operatorname{cav} u(p)$.

1.4.3 Contribution de la thèse : contrôle dynamique de l'information avec observation du retour sur investissement

Cette partie fait écho au Chapitre 6 intitulé *Dynamic control of information with observed return on investment.* Nous en présentons ici les principaux résultats. On y étudie un jeu dynamique avec information incomplète dans la lignée des modèles de persuasion bayésienne. Citons à cet égard les travaux de Kamenica and Gentzkow (2011) lorsque l'état de la nature est fixé, ainsi que ceux de Ely (2017). Renault et al. (2017) traitent un modèle de contrôle dynamique de l'information proche du nôtre, à cela près que les retours sur investissement ne sont pas observés, voir ci-après.

Les protagonistes sont un *conseiller* et un *investisseur.* À chaque étape n, le conseiller observe en privé la réalisation de l'état de la nature $\omega_n \in \Omega$ qui évolue avec le temps. Le conseiller décide des informations à révéler à l'investisseur, qui à son tour décide d'investir ou non. Si investissement il y a, l'investisseur observe si c'est un succès ou un échec.

À chaque investissement, le conseiller perçoit une commission fixée et normalisée à 1, et escompte ses gains selon le facteur $\delta < 1$. En cas d'investissement, l'investisseur paye un frais de 1 et gagne $M > 1$ avec probabilité $r(\omega_n)$ et 0 avec probabilité $1 - r(\omega_n)$, où $r : \Omega \to [0, 1]$.

On suppose que $(\omega_n)_{n\geq 0}$ suit une chaîne de Markov de transition $(\rho(\omega'|\omega))_{\omega',\omega\in\Omega}$. L'investisseur connaît la distribution de la suite $(\omega_n)_{n\geq 0}$. L'information supplémentaire reçue pendant le jeu vient du conseiller et des retours sur investissement observés. L'investisseur investit si et seulement si son paiement espéré est positif, c'est-à-dire que si la croyance courante est $p \in \Delta(\Omega)$, il investit si, en notant $c = 1/M$, $\langle p, r\rangle \geq c$. La *région d'investissement* est $I = \{p \in \Delta(\Omega),\ \langle p, r\rangle \geq c\}$ et la région de non investissement est $J = \Delta(\Omega) \setminus I$.

Le jeu se réduit à un processus de décision markovien que l'on note Γ, dans lequel le conseiller manipule les croyances a posteriori de l'investisseur afin de maximiser la fréquence escomptée espérée d'étapes avec investissement.

L'investisseur utilise sa connaissance de la distribution de $(\omega_n)_{n\geq 0}$ ainsi que les retour sur investissement qu'il observe afin de mettre à jour sa croyance a posteriori. On distingue ainsi trois croyances de l'investisseur à l'étape n. La *croyance a priori* $p_n \in \Delta(\Omega)$ est la croyance à l'étape n avant d'avoir reçu le message du conseiller. La *croyance intermédiaire* q_n est la croyance mise à jour après avoir reçu le message, en particulier l'investisseur investit si et seulement si $q_n \in I$. Enfin, la *croyance a posteriori* s_n est la croyance mise à jour de q_n après que le retour sur investissement, le cas échéant, ait été observé.

La croyance p_{n+1} diffère de s_n parce-que les états ne sont pas fixés. Pour tout $\omega' \in \Omega$ on a $p_{n+1}(\omega') = \sum_{\omega\in\Omega} s_n(\omega)\rho(\omega'|\omega)$. On définit en conséquence l'application linéaire $\phi : \Delta(\Omega) \to \Delta(\Omega)$ telle que $p_{n+1} = \phi(s_n)$.

La croyance q_n diffère de p_n du fait de l'information révélée. Soit une croyance $p \in \Delta(\Omega)$, soit $\mathcal{S}(p) \subset \Delta(\Delta(\Omega))$ l'ensemble des mesures de probabilité sur $\Delta(\Omega)$ d'espérance p. Les éléments de $\mathcal{S}(p)$ sont appelés *éclatements* de p. Une conséquence de la mise à jour bayésienne est que quelque soit la stratégie de dévoilement d'information, la distribution conditionnelle de q_n appartient à $\mathcal{S}(p_n)$. Réciproquement, le lemme d'éclatement Lemme 1.28, établit que pour une croyance $p \in \Delta(\Omega)$ donnée et pour un éclatement $\mu \in \mathcal{S}(p)$, le conseiller peut corréler ses messages avec l'état de la nature de telle sorte à ce que la distribution de la croyance mise à jour q soit μ.

Enfin, la croyance s_n diffère de q_n du fait de l'observation des retours sur investissement. Dans le cas où l'investisseur n'investit pas, la croyance est inchangée, c'est-à-dire que $q_n = s_n$. En cas d'investissement, avec probabilité $r(\omega_n)$ l'investisseur observe un succès et l'on a, pour tout $\omega \in \Omega$, $s_n(\omega) = \frac{q_n(\omega)r(\omega)}{\langle q_n, r\rangle}$. Et par ailleurs, avec probabilité $1 - r(\omega_n)$ l'investisseur observe un échec, et l'on a, pour tout $\omega \in \Omega$, $s_n(\omega) = \frac{q_n(\omega)(1-r(\omega))}{1-\langle q_n, r\rangle}$.

On définit les applications ψ^+ et ψ^- de $\Delta(\Omega)$ dans $\Delta(\Omega)$ ainsi : pour tout $p \in \Delta(\Omega)$, $\psi^+(p)(\omega) = \frac{p(\omega)r(\omega)}{\langle p, r\rangle}$ et $\psi^-(p)(\omega) = \frac{p(\omega)(1-r(\omega))}{1-\langle p, r\rangle}$. Ainsi en cas d'investissement à l'étape n, on a $s_n = \psi^+(q_n)$ en cas de succès, et $s_n = \psi^-(q_n)$ en cas d'échec.

Soit E l'ensemble des fonctions de $\Delta(\Omega)$ à valeur dans $[0, 1]$. L'opérateur de programmation dynamique T est défini de E dans E par

$$T(f) : p \mapsto \begin{cases} \delta f(\phi(p)) & \text{if } p \in J \\ 1 - \delta + \delta\langle p, r\rangle f(\phi(\psi^+(p))) + \delta(1 - \langle p, r\rangle) f(\phi(\psi^-(p))) & \text{if } p \in I. \end{cases}$$

On note $V_\delta(p)$ la valeur du problème d'optimisation dynamique $\Gamma(p)$ comme fonction de la croyance initiale p. C'est l'unique solution de l'équation fonctionnelle

$$\operatorname{cav} T(f) = f.$$

À chaque étape, si la croyance courante p n'est pas dans la zone d'investissement, le conseiller doit faire un compromis. Pour avoir un paiement strictement positif à l'étape courante, il doit dévoiler de l'information. Toutefois, ceci peut abaisser les paiements futurs à cause de la concavité de la fonction valeur. Pour faire face à ce problème, on introduit la stratégie gloutonne, qui minimise la quantité d'information révélée tout en maximisant le paiement d'étape.

La stratégie gloutonne, notée σ_* est telle que si $p \in I$, alors le conseiller ne dévoile aucune information. Si au contraire $p \in J$ alors on considère le problème suivant :

$$\begin{aligned} &\text{maximiser} && a_I \\ &\text{sous contraintes} && p = a_I q_I + a_J q_J, \\ &&& q_I \in I, \\ &&& a_I \ + a_J = 1,\ a_I,\ a_J \geq 0. \end{aligned} \tag{1.4.1}$$

Soit (a_I, a_J, q_I, q_J) une solution du problème 1.4.1, $\sigma_*(p) \in \mathcal{S}(p)$ est l'éclatement qui sélectionne q_I et q_J avec probabilités a_I et a_J respectivement.

Soit g la fonction de paiement induite par la stratégie gloutonne. Cette dernière vérifie

$$g : p \mapsto \begin{cases} a_I g(q_I) + a_J g(q_J) & \text{si } p \in J \\ 1 - \delta + \delta\langle p, r\rangle g(\phi(\psi^+(p))) + \delta(1 - \langle p, r\rangle)g(\phi(\psi^-(p))) & \text{si } p \in I. \end{cases}$$

La proposition ci-dessous, également valide dans le cadre étudié par Renault et al. (2017), donne une condition nécessaire et suffisante commode pour que la stratégie gloutonne soit optimale.

Proposition 1.30. *La stratégie gloutonne est optimale si et seulement si*

1. *g est concave, et ;*

2. *pour tout $p \in J$, on a $\delta g \circ \phi(p) \leq g(p)$.*

Grâce à la Proposition 1.30, on démontre le théorème suivant.

Théorème 1.22. *Si* $\mathrm{card}(\Omega) = 2$*, alors la stratégie gloutonne est optimale.*

S'il y a plus de trois états de la nature, la stratégie gloutonne peut ne pas être optimale. C'est l'objet du théorème ci-dessous qui s'appuie sur un contre-exemple de Renault et al. (2017).

Théorème 1.23. *Supposons que* $\mathrm{card}(\Omega) = 3$*. Alors la stratégie gloutonne n'est pas nécessairement optimale.*

Chapter 2

Introduction (in English)

This thesis focuses on the study of *dynamic games*, which model decision-making processes taken by rational agents in strategic interactions and whose situation changes over time.

In addition to this introduction, as well as its French counterpart, this manuscript contains four chapters, each of which focuses on one or more dynamic aspects of strategic interactions. Chapter 3 focuses on stochastic games, in Chapters 4 and 5 we study models of search games. Finally, Chapter 6 deals with dynamic transmission of information.

The purpose of this introduction is twofold. On the one hand, we put the topics in context by briefly presenting the state of the literature on these topics. On the other hand, we present the contributions made by this thesis by giving the results, as well as intuitively explaining the approaches followed and the ideas developed. After some generalities on zero-sum games, this introduction is divided into three other parts corresponding to the four non-introductory chapters of the thesis.

2.1 General information on zero-sum games

Game theory studies the so-called strategic interactions between rational entities called *players*. In a non-cooperative setting, they have the choice of a strategy that, combined with the strategies of other players, induces an individual payoff. In this thesis we place ourselves in the context of two-player zero-sum games, where a player's payoff is the opposite of that of his opponent. Thus, the interests of the two protagonists are totally opposed.

2.1.1 The model

A *zero-sum game in strategic form* is given by a triplet $\Gamma = (S, T, g)$ for which S is the set of strategies of player 1, T is the set of strategies of player 2 and $g : S \times T \to \mathbb{R}$ is the payoff function. We place ourselves in the context where

S and T are non-empty metric spaces, endowed with their Borel σ-algebras. The game proceeds as follows. Simultaneously, player 1 chooses a strategy $s \in S$ and player 2 chooses a strategy $t \in T$. Player 1 receives payoff $g(s,t)$ while player 2 receives payoff $-g(s,t)$. Thus, player 1 seeks to maximize $g(s,t)$ while player 2 seeks to minimize this amount.

2.1.2 Value

What can be the payoff resulting from the interaction between two rational players? To answer this question, we introduce the notion of *value* of a game.

Definition 2.1. The game $\Gamma = (S, T, g)$ admits a value v if

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

In the above Definition 2.1, the payoff $\sup_{s \in S} \inf_{t \in T} g(s,t)$ corresponds to the situation in which player 1 chooses his strategy and announces it to player 2 who in turn makes his choice according to player 1's announcement. Conversely, the payoff $\inf_{t \in T} \sup_{s \in S} g(s,t)$ corresponds to the situation in which player 2 chooses his strategy and announces it to player 1 who in turn makes his choice based on player 2's announcement. Therefore, when the game considered has a value, it corresponds to the payment obtained by rational players. We also define the concept of *ε-optimal strategies*, which allow players to guarantee a payment close to value.

Definition 2.2. Let $\varepsilon \geq 0$. A strategy $s \in S$ of player 1 is called ε-optimal if

$$\inf_{t \in T} g(s,t) \geq v - \varepsilon.$$

Similarly, a strategy $t \in T$ of player 2 is said to be ε-optimal if

$$\sup_{s \in S} g(s,t) \leq v + \varepsilon.$$

When ε is 0 in Definition 2.2 above, we simply talk about an *optimal strategy*.

2.1.3 Minmax theorems

Consider the next game, called *matching pennies*. Each player has two strategies: heads and tails. If the sides of the pennies chosen by the players match, player 2 gives his penny to player 1, otherwise player 1 gives his penny to player 2. We represent this situation Fig. 2.1.

A game such as matching pennies, for which the strategy sets are finite, is called a *matrix game*, because one may represent the game as a matrix in which player 1 plays the rows and player 2 the columns. One may speak of *actions*, rather than strategies.

	heads	tails
heads	1	−1
tails	−1	1

Figure 2.1: *Matching Pennies*

We notice that this game does not admit any value. Indeed we have $\sup_{s\in S}\inf_{t\in T} g(s,t) = -1$ while $\inf_{t\in T}\sup_{s\in S} g(s,t) = 1$. In order to counteract this disappointment, players are allowed to introduce randomness into their choices. Such strategies are called *mixed strategies*. Formally, we define the *mixed extension* of a game as follows.

Let be X a metric space, then $\Delta(X)$ refers to the set of Borel probability measures on X.

Definition 2.3. Let $\Gamma = (S, T, g)$ be a game. We assume that Fubini's theorem applies for any $(\sigma,\tau) \in \Delta(S)\times\Delta(T)$ to the integral of g on $S\times T$. The mixed extension of Γ is the game $(g, \Delta(S), \Delta(T))$, where g is linearly extended, i.e., if $\sigma \in \Delta(S)$ and $\tau \in \Delta(T)$ then

$$g(\sigma,\tau) = \int_{S\times T} g(s,t)d(\sigma\times\tau)(s,t).$$

The following theorem is owed to von Neumann (1928).

Theorem 2.4. *Any mixed extension of a finite game admits a value as well as optimal strategies.*

Thus, the mixed extension of the game matching pennies presented above does have a value. It is equal to 0 and the optimal strategies of the players are to play heads and tails with equal probabilities.

There are many generalizations of Von Neumann's theorem to more general settings. We present the following two theorems, see (Mertens et al., 2015).

Theorem 2.5. *Let $\Gamma = (S, T, g)$ be a game. Suppose that*

1. *S is compact;*
2. *for all $t \in T$, $g(\cdot, t)$ is upper semi-continuous.*

Then the mixed extension of Γ admits a value, player 1 has an optimal mixed strategy and player 2 has an optimal mixed strategy with finite support.

Theorem 2.6. *Let $\Gamma = (S, T, g)$ be a game. Suppose that*

1. *S and T are compact;*
2. *for all $t \in T$, $g(\cdot, t)$ is lower semi-continuous, and for all $s \in S$, $g(s,\cdot)$ is upper semi-continue;*
3. *g is bounded and measurable with respect to the product Borel σ-algebra.*

Then the mixed extension of Γ admits a value, and each player has an optimal mixed strategy.

2.2 Stochastic games

It is likely that a strategic interaction between two players taking place has already taken place in the past or will take place again in the future. It is therefore only natural to introduce the notion of *repeated game*. However, if we repeat the same zero-sum game as presented in the previous section, it is optimal for the players to play at each stage an optimal strategy of the one-shot game, so the repetition has no strategic influence.

We place ourselves in the situation where the current game depends on a *state of nature*. The latter evolves from one stage to the next randomly according to the current state of nature, as well as the actions of the players. If we assume that players observe the current state as well as past actions, we are then faced with a *stochastic game*. These were introduced by Shapley (1953), we present in this section the main challenges.

2.2.1 The model

A stochastic game is given by a quintuplet $\Gamma = (\Omega, A, B, g, \rho)$ where,

- Ω is the state space;
- A and B are the action sets of players 1 and 2 respectively;
- $g : \Omega \times A \times B \longrightarrow \mathbb{R}$ is the payoff function;
- $\rho : \Omega \times A \times B \longrightarrow \Delta(\Omega)$ is the transition probability.

In this thesis, the following assumptions will most often be considered.

Assumption 2.7.
- the state space Ω is non-empty and finite, endowed with the finite σ-algebra;
- the action sets A and B are compact metric and non-empty, endowed with their Borel σ-algebras;
- the payoff and transition functions are separately continuous in the players' actions.

Let $\omega_1 \in \Omega$ be the initial state of the stochastic game. The game $\Gamma(\omega_1)$ is played as follows.

- The initial state $\omega_1 \in \Omega$ is known to both players.
- At each stage $t \geq 1$, knowing the history

$$h_t = (\omega_1, a_1, b_1, \cdots, \omega_{t-1}, a_{t-1}, b_{t-1}, \omega_t),$$

player 1 and player 2 simultaneously choose $a_t \in A$ and $b_t \in B$ respectively.

- This generates a stage payoff $g_t = g(\omega_t, a_t, b_t)$.
- The next state ω_{t+1} is drawn according to $\rho(\cdot|\omega_t, a_t, b_t)$ and observed by the players, as are the actions a_t and b_t.

Let us start with an example. The following game, called the *Big Match*, was introduced by Gillette (1957). The state space is $\Omega = \{\omega, 1^*, 0^*\}$, the action sets of players 1 and 2 are $A = \{T, B\}$ and $B = \{L, R\}$ respectively. The states 1^* and 0^* are *absorbing*, meaning that once they are reached, they can no longer be left. The payoffs in these states are 1 and 0 respectively. Only the state ω is not absorbing. The Big Match is said to be an *absorbing game*. Transitions and payoffs in state ω are represented Fig. 2.2. So, if for example the actions T and L are played, then

	L	R
T	1^*	0^*
B	0	1

Figure 2.2: The *Big Match* in state ω

the stage payoff is 1, the state becomes 1^*, and the payoff will be 1 at each stage. If the actions B and L are played, then the stage payoff is 0, and the state remains ω.

In the one-shot game $\Gamma(\omega)$, the value of the game is $1/2$ and the uniform choice of each action is optimal for both players. On the other hand, if the game is repeated a greater number of times, and we consider the average of the payoffs on these stages, player 1 must modulate his choice of the action T. Indeed, playing the action T with probability ε at each stage implies, if player 2 plays R at each stage, that the state is absorbed in 0^* in $1/\varepsilon$ stages on average. Therefore, if the number of repetitions is large in front of $1/\varepsilon$, this strategy of player 1 seems unsatisfactory. However, never playing T would also be a bad strategy, since player 2 could play L at each stage, which would result in a zero payoff.

This shows an arbitrage problem typical with stochastic games. Indeed, player 1 must obtain both a good stage payoff, but also place himself in a good situation over the longer term.

Now we formally define the notion of strategy in a stochastic game. We denote $H_t = \Omega \times (\Omega \times A \times B)^{t-1}$ the set off all possible histories at stage $t \geq 1$. We endow H_t with the product σ-algebra denoted $\mathcal{H}_t$. We denote $H_\infty = (\Omega \times A \times B)^{\mathbb{N}^*}$ the set of infinite histories, which is endowed with the product σ-algebra induced by $\bigcup_{t \geq 1} \mathcal{H}_t$.

Definition 2.8. A *behavioral strategy* of player 1 is a sequence of mappings $\sigma = (\sigma_t)_{t \geq 1}$ such that for every $t \geq 1$, σ_t is a measurable mapping from $(H_t, \mathcal{H}_t)$ to $\Delta(A)$. The behavior strategies of player 2 are defined in a similar way.

We denote $\mathcal{S}$ and $\mathcal{T}$ the behavior strategy sets of players 1 and 2 respectively. A behavior strategy is called *pure* if for each finite history $h_t \in H_t$, one has

$\sigma_t(h_t) \in A$. A *mixed strategy* is a probability distribution over all pure strategies endowed with the product σ-algebra. We have two similar definitions for player 2.

Generally speaking, a strategy therefore takes into account at one stage $t \geq 1$ the whole history h_t in order to select an action. We are led to consider strategies that require less memory.

Definition 2.9. A *Markov strategy* selects at each stage an action, as a function only on the stage and the current state.

A Markov strategy can therefore be seen as a mapping from $\mathbb{N}^* \times \Omega$ to $\Delta(A)$, or to $\Delta(B)$, where $\mathbb{N}^*$ denotes the set of positive integers.

Definition 2.10. A strategy is said to be *stationary* if at each stage the choice of an action depends only on the current state.

A stationary strategy can therefore be seen as a mapping from Ω to $\Delta(A)$, or to $\Delta(B)$.

An initial state ω_1 together with a pair of (behavior or mixed) strategies (σ, τ) naturally induce a probability measure on all finite histories $\bigcup_{t\geq 1} H_t$. The Kolmogorov extension theorem indicates that this measure extends uniquely to H_∞. We denote this probability measure $\mathbb{P}^{\omega_1}_{\sigma,\tau}$, and $\mathbb{E}^{\omega_1}_{\sigma,\tau}$ the corresponding expectation.

Behavior strategies are often easier to manipulate than mixed strategies. Kuhn's theorem, see (Sorin, 2002, Appendix D) and (Aumann, 1964), establishes the equivalence between behavior strategies and mixed strategies in the following sense.

Theorem 2.11 (Kuhn's theorem)**.** *For any behavior strategy σ of player 1, there exists a mixed strategy $\tilde{\sigma}$ of player 1 such that for any (behavior or mixed) strategy τ of player 2, one has*

$$\mathbb{P}^{\omega_1}_{\sigma,\tau} = \mathbb{P}^{\omega_1}_{\tilde{\sigma},\tau}.$$

Conversely, for any mixed strategy σ of player 1, there exists a behavior strategy $\tilde{\sigma}$ of player 1 such that for any (behavior or mixed) strategy τ of player 2, one has

$$\mathbb{P}^{\omega_1}_{\sigma,\tau} = \mathbb{P}^{\omega_1}_{\tilde{\sigma},\tau}.$$

Theorem 2.11 remains true by exchanging the roles of players 1 and 2. There are two main ways to aggregate stage payoffs, namely through Abel's and Caesaro's means.

Let $\lambda \in (0, 1]$, the λ*-discounted game*, denoted $\Gamma_\lambda(\omega_1)$, is the game whose strategy sets are $\mathcal{S}$ and $\mathcal{T}$ and whose payoff is $\gamma^{\omega_1}_\lambda : \mathcal{S} \times \mathcal{T} \to \mathbb{R}$ defined by

$$\gamma^{\omega_1}_\lambda(\sigma, \tau) = \mathbb{E}^{\omega_1}_{\sigma,\tau}\left(\sum_{t\geq 1} \lambda(1-\lambda)^{t-1} g_t\right).$$

Let be $N \in \mathbb{N}^*$, the N*-stage repeated game*, denoted $\Gamma_N(\omega_1)$ is the game whose strategy sets are $\mathcal{S}$ and $\mathcal{T}$ and whose payment is $\gamma_N^{\omega_1} : \mathcal{S} \times \mathcal{T} \to \mathbb{R}$ defined by

$$\gamma_N^{\omega_1}(\sigma, \tau) = \mathbb{E}_{\sigma,\tau}^{\omega_1}\left(\frac{1}{N}\sum_{t=1}^{N} g_t\right).$$

Under Assumption 2.7, the minmax theorem, Theorem 2.6, ensures the existence of the value in the games $\Gamma_\lambda(\omega_1)$ and $\Gamma_N(\omega_1)$, as well as optimal strategies. We denote

$$\begin{aligned} v_\lambda(\omega_1) &= \max_{\sigma\in\mathcal{S}}\min_{\tau\in\mathcal{T}} \gamma_\lambda^{\omega_1}(\sigma,\tau) = \min_{\tau\in\mathcal{T}}\max_{\sigma\in\mathcal{S}} \gamma_\lambda^{\omega_1}(\sigma,\tau), \\ v_N(\omega_1) &= \max_{\sigma\in\mathcal{S}}\min_{\tau\in\mathcal{T}} \gamma_\lambda^{\omega_1}(\sigma,\tau) = \min_{\tau\in\mathcal{T}}\max_{\sigma\in\mathcal{S}} \gamma_N^{\omega_1}(\sigma,\tau). \end{aligned}$$

More generally, we can also aggregate the stage payoffs according to a probability measure on $\mathbb{N}^*$. Let $\theta \in \Delta(\mathbb{N}^*)$. The game $\Gamma_\theta = (S, T, g_\theta)$ is the *repeated game weighted by* θ where $g_\theta : S \times T \to \mathbb{R}$ is defined by

$$\gamma_\theta(\sigma, \tau) = \mathbb{E}_{\sigma,\tau}\left(\sum_{t\geq 1} \theta_t g_t\right).$$

Some classes of stochastic games are distinguished in the literature, either for their own interest or in the hope of generalizing certain results. A stochastic game is *absorbing* if it has only one non-absorbing state, an *absorbing state* being a state that once reached cannot be left. The Big Match presented above is such a game. A stochastic game is *recursive* if the payoff in any non-absorbing state is zero. A Markov decision process is a stochastic game with only one player.

2.2.2 Shapley operator and Shapley equations

Let $\mathcal{F}$ be the set of bounded functions from Ω to $\mathbb{R}$. For any $f \in \mathcal{F}$ and any state $\omega \in \Omega$ let $\Gamma(f)(\omega)$ be the game with action sets A and B and with payoff function

$$(a, b) \mapsto g(\omega, a, b) + \sum_{\omega'\in\Omega} \rho(\omega'|\omega, a, b) f(\omega').$$

Under Assumption 2.7, this game has a value in mixed strategies, which is denoted $\Psi(f)(\omega)$. We call Ψ the *Shapley operator* of the stochastic game Γ.

Theorem 2.12. *The values of the N-stage repeated and λ-discounted games are characterized by the following* Shapley equations.

$$\begin{aligned} v_\lambda(\omega_1) &= \max_{\mu\in\Delta(A)}\min_{\nu\in\Delta(B)} \left[\lambda g(\omega_1,\mu,\nu) + (1-\lambda)\mathbb{E}_{\mu,\nu}^{\omega_1}(v_\lambda)\right] \\ &= \min_{\nu\in\Delta(B)}\max_{\mu\in\Delta(A)} \left[\lambda g(\omega_1,\mu,\nu) + (1-\lambda)\mathbb{E}_{\mu,\nu}^{\omega_1}(v_\lambda)\right], \end{aligned}$$

and

$$\begin{aligned} v_{N+1}(\omega_1) &= \max_{\mu \in \Delta(A)} \min_{\nu \in \Delta(B)} \left[\frac{1}{N+1} g(\omega_1, \mu, \nu) + \frac{N}{N+1} \mathbb{E}^{\omega_1}_{\mu,\nu}(v_N) \right] \\ &= \min_{\nu \in \Delta(B)} \max_{\mu \in \Delta(A)} \left[\frac{1}{N+1} g(\omega_1, \mu, \nu) + \frac{N}{N+1} \mathbb{E}^{\omega_1}_{\mu,\nu}(v_N) \right], \end{aligned}$$

with

$$\mathbb{E}^{\omega_1}_{\mu,\nu}(v_N) = \sum_{\omega \in \Omega} v_N(\omega) \int_{A \times B} \rho(\omega | \omega_1, a, b) d\mu(a) d\nu(b).$$

Hence one has

$$v_\lambda = \lambda \Psi \left(\frac{1-\lambda}{\lambda} v_\lambda \right) \text{ and } v_{N+1} = \frac{1}{N+1} \Psi \left(N v_N \right).$$

Shapley equations have an intuitive interpretation: if the initial state is ω_1, players know that they may play optimally at the next stage in the continuation game. An important consequence is the following.

Corollary 2.13.
- *Players 1 and 2 have optimal stationary strategies in the game $\Gamma_\lambda(\omega_1)$, regardless of ω_1.*
- *Players 1 and 2 have optimal Markov strategies in the game $\Gamma_N(\omega_1)$, regardless of ω_1.*

A fundamental question in the study of repeated games is the asymptotic behavior of the values (v_λ) and (v_N). Three approaches to this question are presented: the asymptotic approach, the uniform approach and the infinite game.

2.2.3 Asymptotic value

The *asymptotic approach* of stochastic games is the study of the convergence of (v_λ) and (v_N) as λ goes to 0 and N goes to $+\infty$, i.e., as players become more and more patient.

Definition 2.14. A stochastic game admits an *asymptotic value* if (v_λ) and (v_N) converge to the same limit when λ goes to 0 and N goes to $+\infty$.

The following result of Bewley and Kohleberg (1978) on stochastic games with finite state space and action sets is a classic of the literature.

Theorem 2.15. *Any stochastic game with finite state space and finite action sets admits an asymptotic value.*

In addition, (v_λ) admits an expansion in Puiseux series: there exists $\lambda_0 > 0$, $M \in \mathbb{N}^*$ and $(r_t) \in (\mathbb{R}^\Omega)^\mathbb{N}$ such that for any $\lambda \in (0, \lambda_0)$ and any $\omega \in \Omega$

$$v_\lambda(\omega) = \sum_{t \geq 0} r_t(\omega) \lambda^{\frac{t}{M}}.$$

Let us also mention the Tauberian theorem of Ziliotto (2016a) which under weak assumptions, verified by Assumption 2.7, ensures that (v_λ) converges if and only if (v_N) converges, and if one converges the two limits are equal.

The long-standing idea that the asymptotic value exists in stochastic games with a finite state space and compact action sets has been refuted by a counterexample of Vigeral (2013). Below is a simple counterexample of Renault (2019). It is a slight variation of a counterexample of Ziliotto (2016b), we also find it in (Sorin and Vigeral, 2015). All these counterexamples have in common that they have transitions or action sets that are not semi-algebraic. This is the central point that allows the value to oscillate. In Counterexample 2.1 below, it is the action set of player 1 that is not semi-algebraic.

Counterexample 2.1. The state space is $\Omega = \{0, 1, 0^*, 1^*\}$. The action sets of players 1 and 2 are A and B respectively, which are specified below. Player 1 plays $\alpha \in A$ in state 0 and player 2 plays $\beta \in B$ in state 1. The states 0^* and 1^* are absorbing. The payoff in states 0 and 0^* is 0. It is 1 in states 1 and 1^*. We represent the game Fig. 2.3, the transition probabilities are represented by arrows between states.

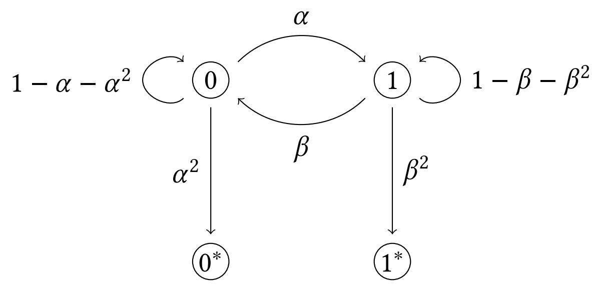

Figure 2.3: A simple counterexample

We then show that by choosing $A = \{0\} \cup \left\{1/2^{2n} \mid n \geq 1\right\}$ and $B = [0, 1/4]$, the value of the discounted game (v_λ) does not converge. The idea is as follows. In the game $\Gamma_\lambda(0)$ with action set $A = [0, 1/4]$, player 1 would play approximately $\sqrt{\lambda}$. By choosing A as above, we ensure that player 1 can play optimally as in the unconstrained game for some discount factors, but not for others, which induces a lower payment and therefore makes the value oscillate. We will come back in more detail on this counterexample in Chapter 3.

The asymptotic value, when it exists, therefore corresponds approximately to the payoff obtained by the players if they play long enough, i.e., when λ goes to 0 and N goes to $+\infty$. However, the optimal strategies of the players in these increasingly long games may depend on the discount factor λ or the number of stages N. It is therefore natural to ask whether there are optimal strategies in any game that is long enough, regardless of the length of the game. To answer this question, we consider the *uniform approach.*

2.2.4 Uniform value

Let ω_1 be an initial state. It is said that player 1 *uniformly guarantees* the quantity v_∞ if he has a strategy that guarantees this quantity, up to ε, against any strategy of player 2, in any N-repeated game $\Gamma_N(\omega_1)$, provided that N is large enough. Formally,

$$(\forall \varepsilon > 0)\,(\exists \sigma \in \mathcal{S})\,(\exists M \in \mathbb{N}^*)\,(\forall \tau \in \mathcal{T})\,(\forall N \geq M)\;\gamma_N^{\omega_1}(\sigma, \tau) \geq v_\infty - \varepsilon.$$

The same applies to player 2.

Definition 2.16. If there exists $v_\infty \in \mathbb{R}^\Omega$ such that both players guarantee $v_\infty(\omega_1)$ in the game $\Gamma(\omega_1)$, then the game Γ admits a *uniform value.*

Let $\varepsilon \geq 0$. A strategy $\sigma \in \mathcal{S}$ is said to be (uniformly) ε-optimal if

$$(\exists M \in \mathbb{N}^*)\,(\forall \tau \in \mathcal{T})\,(\forall N \geq M)\;\gamma_N^{\omega_1}(\sigma, \tau) \geq v_\infty - \varepsilon.$$

The same applies to player 2.

It should be noted that if a player uniformly guarantees a certain quantity, he can also guarantee that quantity in any discounted game, provided the discount factor is small enough. For this reason we are interested, in the uniform approach, in the N-stage repeated game. In addition, if a game admits a uniform value, then (v_λ) and (v_N) both converge to the uniform value. The converse is not always true.

The following result of Mertens and Neyman (1981) on stochastic games with finite state space and action sets is a classic of the literature.

Theorem 2.17. *Any stochastic game with finite state space and action sets admits a uniform value.*

Many other positive results regarding the existence of uniform value exist in the literature. For example, in the context of a finite state space and compact action sets, we cite the work of Mertens et al. (2009) for absorbing games, Li and Sorin (2016) for recursive games, whose demonstrations use the operator approach of Rosenberg and Sorin (2001) which is based on the Shapley operator, and which, recall, contains entirely the game's dynamics. Still for a finite state space, Bolte et al. (2014) have shown that games with semi-algebraic (or more generally definable) transitions and action sets have a uniform value. Finally, Renault (2010) has shown the existence of uniform value in Markov decision processes with finite state space and any action set.

2.2.5 Infinite game

We present here a last approach to long duration stochastic games. The aim here is to define a payoff function directly on infinite histories and to study the resulting game in strategic form. We usually consider the two payoff functions $\underline{\gamma_\infty}$ and $\overline{\gamma_\infty}$ defined on $\Omega \times \mathcal{S} \times \mathcal{T}$ by

$$\underline{\gamma_\infty}^{\omega_1}(\sigma,\tau) = \mathbb{E}^{\omega_1}_{\sigma,\tau}\left(\liminf_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} g_t\right) \text{ and } \overline{\gamma_\infty}^{\omega_1}(\sigma,\tau) = \mathbb{E}^{\omega_1}_{\sigma,\tau}\left(\limsup_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} g_t\right).$$

The following result is also due to Mertens and Neyman (1981). It unifies the uniform and infinite approaches of stochastic games.

Theorem 2.18. *The games $(\mathcal{S},\mathcal{T},\underline{\gamma_\infty})$ and $(\mathcal{S},\mathcal{T},\overline{\gamma_\infty})$ with finite state space and action sets admit a value, which is equal to the uniform value.*

2.2.6 Contribution of the thesis: zero-sum communicating product stochastic games

This part is in relation with Chapter 3 entitled *Communicating zero-sum product stochastic games*. The contribution of this thesis with regard to stochastic games is the study of communication properties between states, when the state space Ω is in the form of a product $X \times Y$, and players 1 and 2 control the dynamics on their state component X and Y, respectively.

A new class of strategies called *Markov N-periodic strategies* is introduced. These are Markov strategies whose dependence in the stage is considered modulo the period N. So we can see them as mappings from $\{1,\dots,N\} \times \Omega$ to $\Delta(A)$, or $\Delta(B)$.

In order to emphasize the difference between strategies that depend on the history of both players, and strategies that depend only on the component (state and action) of one player, as in Markov decision processes, we call the second ones *policies*.

The two communication properties on the components of the state spaces that are considered are as follows. They are called *strong communication property* and *weak communication property*.

A player has the strong communication property if there exists a time T such that regardless of his policy choice (depending only on his component of the history), he has a positive probability of going from any state to any other state of his component in T stages. Formally, the following definition is given.

Definition 2.19. Player 1 has the strong communication property if there exists $T \in \mathbb{N}^*$ such that for all policies σ and all states $x, x' \in X$, one has $\mathbb{P}^x_\sigma(X_T = x') > 0$.

A similar definition is given for player 2, and it is said that a game is strongly communicating on one side if one of the players has the strong communication property.

A player has the weak communication property if there exists a time T and a policy (depending only on his component of the history), such that he has a positive probability of going from any state to any other state of his component in T stages. Formally, the following definition is given.

Definition 2.20. Player 1 has the weak communication property if there exists $T \in \mathbb{N}^*$ and a policy σ such that for all states $x, x' \in X$, one has $\mathbb{P}^x_\sigma(X_T = x') > 0$.

A similar definition is given for player 2, and it is said that a game is weakly communicating on both sides if both players have the weak communication property.

It should be noted that the strong communication property implies the weak one. Example 2.1 below illustrates these two properties.

Example 2.1. We represent Fig. 2.4 the state space and transitions of player 1. The state space is $X = \{x, y, z\}$. In state x player 1 chooses $\alpha \in A$. If the action set A is the interval $[0, 1]$, then player 1 has the weak communication property but not the strong communication property. On the other hand, if A is $[\varepsilon, 1]$ with $\varepsilon > 0$, player 1 has the strong communication property.

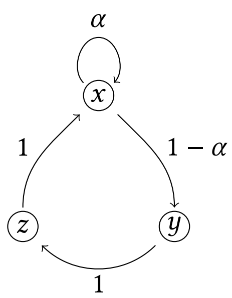

Figure 2.4: Weak and strong communication

We prove the two theorems below, these are the main results of this thesis on stochastic games.

Theorem 2.21. *Any strongly communicating on one side zero-sum product stochastic game has a uniform value.*

Moreover, assuming that player 1 has the strong communication property, the uniform value depends only on the initial state of player 2, and for every $\varepsilon > 0$, player 1 has an ε-optimal periodic Markov strategy.

Theorem 2.22. *There exists a weakly communicating on both sides zero-sum product stochastic game that does not admit an asymptotic value.*

Notice that it is sufficient that one of the two players has the strong communication property for the uniform value to exist, while even if both players have the weak communication property, the asymptotic value may not exist.

We now detail the intuition behind the demonstrations of Theorems 2.21 and 2.22.

Proof of Theorem 2.21

It is assumed here that player 1 has the strong communication property. The state space Y is classified via the recurrence classes induced by stationary policies on Y, in a similar way to that of Ross and Varadarajan (1991).

Definition 2.23. A subset C of Y is said to be a maximal communicating set if

1. there exists a stationary policy on Y such that C is a recurrent class of the induced Markov chain on Y. Such a policy is said to be a stationary policy associated to C;

2. C is maximal, i.e., if there exists C' a subset of Y such that 1 holds for C' and $C \subseteq C'$, then $C' = C$.

Let $C_1, \dots, C_L$ denote the maximal communicating sets. D denotes the set of transient states under every stationary policy. We have the following proposition.

Proposition 2.24. *$\{C_1, \dots, C_L, D\}$ is a partition of Y.*

Example 2.2 below shows how player 2's state space decomposes into maximal communicating sets. It also highlights a behavior of the state process that may occur when the action set is compact, but not when it is finite.

Example 2.2. As shown on Fig. 2.5, the state space of player 2 is $Y = \{x, y, z\}$ and his action set is $B = [0, 1/2]$. Transition probabilities are represented by arrows. The maximal communicating sets are $\{x\}$, $\{y\}$ and $\{z\}$. It should be noted that if

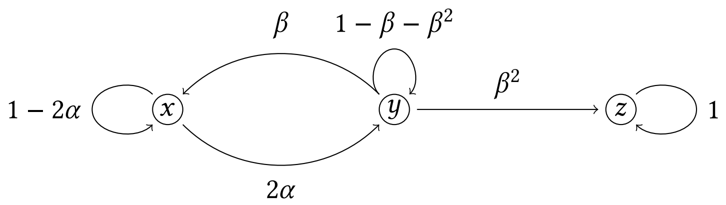

Figure 2.5: States, actions and transitions of player 2

the initial state is x, by playing $\alpha = 1/2$ in state x and $\beta = \frac{1}{2n}$ in state y at stage $n \geq 1$, player 2 has a positive probability of switching infinitely often between the maximal communicating sets $\{x\}$ and $\{y\}$.

This cannot happen when the action set is finite. In this case, for any policy of player 2, after a finite number of stages the state process $(Y_n)_{n\geq 1}$ will forever remain in one of the maximal communicating sets with probability 1, see (Ross and Varadarajan, 1991, lemma 2 and proposition 2).

For each of the maximal communicating sets C_i, an auxiliary stochastic game Γ_i is introduced as follows. For every $i \in \{1, \dots, L\}$, if $y \in C_i$, we define the set of actions of player 2 in state y such that the state remains in C_i with probability 1,

$$B_y = \{b \in B \mid \ q(C_i|y, b) = 1\}.$$

We then have the following proposition.

Proposition 2.25. *For all $i \in \{1,\dots,L\}$ the game Γ_i has a uniform value v^i_∞, which is constant over $X \times C_i$.*

Moreover, for all $\varepsilon > 0$ there exists $N_0 \in \mathbb{N}^*$ such that both players have an ε-optimal Markov N_0-periodic strategy in each Γ_i.

We then consider the Markov decision process $\mathcal{G} = (Y, B, q, g)$, in which only player 2 plays and aims to minimize

$$\begin{array}{rcl} g : Y & \to & [0,1] \\ y & \mapsto & \begin{cases} v^i_\infty \text{ if there exists } i \in \{1,\dots,L\} \text{ such that } y \in C_i \\ 1/2 \text{ if } y \in D. \end{cases} \end{array}$$

$\mathcal{G}$ has a uniform value $w_\infty \in [0,1]^Y$, and for all $\varepsilon > 0$, player 2 has an ε-optimal stationary policy, see (Sorin, 2002).

The following interpretation is given. Player 2's objective is to reach in expectation the maximal communicating set C_i such that the corresponding auxiliary game Γ_i has the smallest possible uniform value v^i_∞, then remains in C_i. The payoff $1/2$ in D is arbitrary and does not change the value of w_∞.

It is shown that the uniform value w_∞ of the Markov decision process $\mathcal{G}$ is also the uniform value of the initial game.

The idea for player 2 is to play optimally first in the Markov decision process $\mathcal{G}$ (regardless of player 1) and then to switch to an optimal strategy in Γ_i once the maximal communicating set C_i is reached.

Showing that player 1 uniformly guarantees w_∞ is one of the main difficulties. It is natural for player 1 to play optimally in each game Γ_i. However, player 1 does not control player 2's transitions from one C_i to another, which we have seen in Example 2.2, may occur infinitely often. This problem is solved by letting player 2 to play a best-response Markov periodic strategy, which prevents this problematic behavior. We conclude by showing that the payoff thus obtained can also be generated by player 2 as a limit payoff in $\mathcal{G}$.

Proof of Theorem 2.22

The proof of Theorem 2.22 is based on Counterexample 2.1 introduced earlier. This counterexample does not have a product state space and does not have the weak communication property. We build a counterexample that checks these properties and reproduces their dynamics.

The state space of player 1 is $X = \{x, y\} \times C_8$ where $C_8 = \mathbb{Z}/8\mathbb{Z}$. Let $I = \{0\} \cup \left\{1/2^{2n} \mid n \geq 1\right\}$. Let $A = I \times \{-1, +1\} \cup \{0,1\} \times \{0\}$ be the action set of player 1.

For $i \in \{x, y\}$ we denote $-i$ the element of $\{x,y\} \backslash \{i\}$. In state $(i,k) \in X$ if player 1 plays $(\alpha, p) \in I \times \{-1,+1\}$ then with probability $1 - \alpha - \alpha^2$ the new state is $(i, k+p)$, with probability α the new state is $(-i, k+p)$, and with probability α^2 the new state is $(i, k-p)$ (see Fig. 2.6).

Still in state $(i, k) \in X$, if player 1 plays $(\alpha, 0) \in \{0,1\} \times \{0\}$, then with probability $1-\alpha$ the state remains in (i, k) and with probability α the new state is $(-i, k)$ (see Fig. 2.7).

The state space and transitions of player 2 are a copy of those of player 1, but the action set of player 2 is $B = J \times \{-1, +1\} \cup \{0, 1\} \times \{0, 1\}$, where $J = [0, 1/4]$.

The fact that I and therefore A are not semi-algebraic is essential. Indeed, since X and Y are finite, and the transitions are polynomial, if A and B were definable in some o-minimal structure, the game would have a uniform value, see (Bolte et al., 2014, Theorem 4).

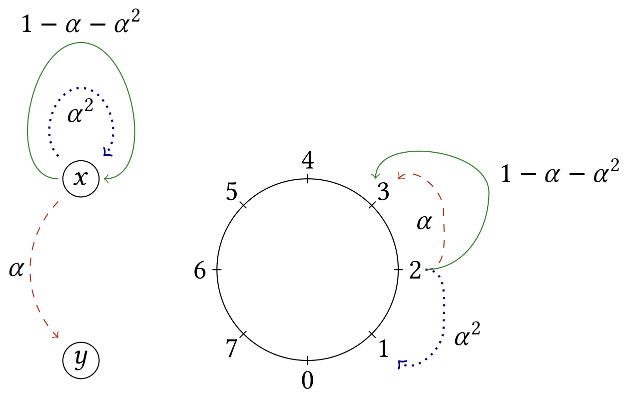

Figure 2.6: Transitions when playing $(\alpha, +1)$, $\alpha \in I$ in state $(x, 2)$

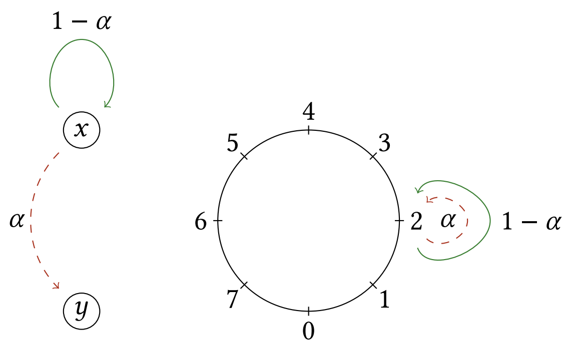

Figure 2.7: Transitions when playing $(\alpha, 0)$, $\alpha \in \{0, 1\}$ in state $(x, 2)$

Thus by playing $(0, p)$, player 1 totally controls the dynamics on C_8, and by playing $(\alpha, 0)$ with α equal to 0 or 1 he totally controls the dynamics on $\{x, y\}$. The same applies to player 2.

The payoff function is defined as follows. If the distance between the players on the circle is at least 3, the payoff is 1. If the distance between the players on the circle is at most 1, the payoff is 0. Finally, if the distance between the players on the circle is 2, then the payoff is 1 if players 1 and 2 are in the same state on $\{x, y\}$, and 0 if they are in different states.

The game has an interpretation in terms of a discounted pursuit-evasion game that is as follows. Player 1 wishes to maximize his distance to player 2, who wishes to minimize his distance to player 1. If the distance between the players on the circle is at most 1 or at least 3, then their position on $\{x, y\}$ does not matter. On the other hand, if they are at a distance of 2 on the circle, player 1 does not want to have the same position as player 2 on $\{x, y\}$.

It is noteworthy that if the distance between the players on the circle is at least 3, then player 1 can play in such a way that it remains at least 3 forever. Similarly, if the distance between the players on the circle is at most 1, then player 2 can play in such a way that it remains at most 1 forever. Thus these joint states act as absorbing states with payoffs 1 and 0 respectively.

We then show that the game built has the same Shapley equations as the one of Counterexample 2.1.

2.3 Search games

In a second step, we are interested in another form of dynamic interaction. In a *search game*, two players interact on a *search space*. The first one is called *the Searcher* or Sally, and the second is called *the Hider* or Harry. Harry chooses a place to hide in the search space. The two paradigms that Sally considers are then typically

1. either to minimize the time it takes to find Harry;
2. or to maximize the probability of finding Harry in a given time.

Harry's objective is the opposite. The dynamics present in the interaction is therefore not due to a changing state of nature, as was the case for stochastic games introduced in Section 2.2, or the games of information provision that we will present in Section 2.4, but to Sally's search trajectory. However, we will see through stochastic search games, Section 2.3.4, that these two dynamical aspects may be considered simultaneously.

2.3.1 The standard model

The standard model of search games is presented here. It has been introduced by Isaacs (1965). This model considers the paradigm of minimizing discovery time. The game is given by

- a search space Q, which is a compact subset of $\mathbb{R}^n$ endowed with a norm $\|\cdot\|$ that induces a metric d;

- a detection radius $r \geq 0$;
- an origin of Sally's trajectory $O \in Q$.

This last point could also be relaxed so that Sally can choose her starting point, which will be the case for patrolling games presented in Section 2.3.3.

Sally's set of strategies $\mathcal{S}$ is that of trajectories s which are 1-Lispchitz continuous from $\mathbb{R}_+$ to Q, and such that $s(0) = O$. Hence, Sally starts her trajectory at O. Moreover, $s(t) \in Q$ represents Sally's position at time t. Finally, Sally moves at a speed at most 1. Harry's strategy set is Q. We then define for any $(s, h) \in \mathcal{S} \times Q$ the payoff

$$g(s, h) = \inf\{t \in \mathbb{R}_+ \mid d(s(t), h) \leq r\},$$

where the infimum on the empty set is $+\infty$. Thus, the capture occurs when Sally and Harry are at a distance less than Sally's detection radius, and the payoff $g(s, h)$ represents the capture time when Sally follows the search trajectory s and Harry hides in h. The search game has a value in mixed strategies, which we denote v.

First, we give a simple example of a search game. Let us consider the game where the search space Q is the circle with radius 1, the origin O is any point in Q, and the detection radius r is 0. Consider the searcher's strategy of running through the circle at a speed of 1 with a probability of $1/2$ in trigonometric and anti-trigonometric directions respectively. Let $h \in Q$ be the hiding point. The payoff induced by these strategies is

$$\frac{1}{2}h + \frac{1}{2}(2\pi - h) = \pi,$$

and therefore does not depend on the hiding point.

Suppose that the hider plays the uniform distribution on the circle. A best response of the searcher is to walk the circle in a trigonometric direction at a speed of 1. This induces the payoff

$$\frac{1}{2\pi} \int_0^{2\pi} t dt = \pi.$$

Hence the value of the game is π, and we displayed a couple of optimal strategies.

We consider the rate at which Sally can discover new points of Q. The Lebesgue measure of a measurable subset $B \subset \mathbb{R}^n$ is denoted $\lambda(B)$. In addition, $B_r(0)$ refers to the closed ball with radius r and center 0. Finally, $\lambda(B_r)$ refers to the Lebesgue measure of any closed ball of radius r.

Definition 2.26. The *maximal discovery rate* ρ of the searcher is

$$\rho = \sup_{s \in \mathcal{S}, t > 0} \frac{\lambda(s([0, t]) + B_r(0)) - \lambda(B_r)}{t},$$

where $s([0, t]) = \{s(\tau) \mid \tau \in [0, t]\}$ is the image of $[0, t]$ by s, and $s([0, t]) + B_r(0) = \{y \in \mathbb{R}^n \mid d(s([0, t], y)) \leq r\}$.

Thus, in $\mathbb{R}^2$ and $\mathbb{R}^3$ endowed with the Euclidean norm, and under reasonable assumptions, the maximal discovery rate is $2r$ and πr^2 respectively, i.e., Sally's sweep width.

The following proposition gives a lower-bound on the value, guaranteed by the *uniform distribution* on Q of the hider. Intuitively, if the hider plays the uniform distribution, the best the researcher can do is to cover Q as quickly as possible.

Proposition 2.27. *Let Q be a positive Lebesgue measure search space. The value of the search game satisfies*

$$v \geq \frac{(\lambda(Q) - \lambda(B_r))^2}{2\lambda(Q)\rho}.$$

2.3.2 Search games on networks

A significant part of the literature on search games is devoted to the study of *networks*. While these particular search spaces have a natural interest in terms of applications, their study is also theoretically successful.

Informally, a network is defined as a continuous graph. So Harry can hide on any point of an edge, and Sally moves continuously along them. We use a vocabulary similar to that of graph theory, and we will speak for example of edge, vertex, path, cycle, tree, etc. In addition, a distance is naturally defined on networks, as well as a Lebesgue measure also denoted λ. These notions are formally defined in Chapter 4.

Two classes of networks are of particular importance with regard to search games: the *Eulerian networks* and the *trees*. These classes represent the limit cases in terms of the value of the game at a fixed network length.

We assume here that Sally's detection radius r is 0, i.e., Sally meets Harry if they are at the same point. In a network, the maximal discovery rate ρ is 1.

Definition 2.28. We call *minimal cycle* or *Chinese postman cycle* of a network $\mathcal{N}$, a cycle visiting all points of $\mathcal{N}$ of minimal length. We note $\bar{\lambda}(\mathcal{N})$ the length of such a cycle.

Definition 2.29. A network $\mathcal{N}$ is said to be Eulerian if it has a minimal cycle with length $\lambda(\mathcal{N})$. Such a cycle is called an *Eulerian cycle.*

Sally's *uniform Chinese postman strategy* is to uniformly choose a minimal cycle starting from the origin O. In an Eulerian network we speak of the *uniform Eulerian strategy.*

In a tree, Harry's equal branching density is the probability measure with support on the leaves of the tree, such that for any vertex v with outgoing edges $e_1, \ldots, e_n$, the probability that Harry hides in the subtree originating from e_i is equal to the length of this subtree divided by the length of the subtree with root

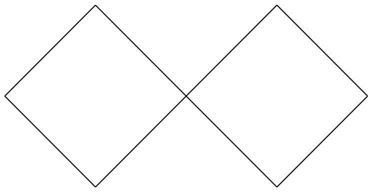

Figure 2.8: An Eulerian network

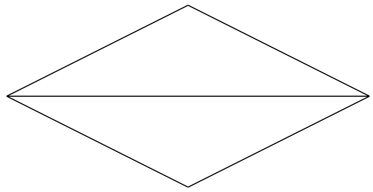

Figure 2.9: A non-Eulerian network

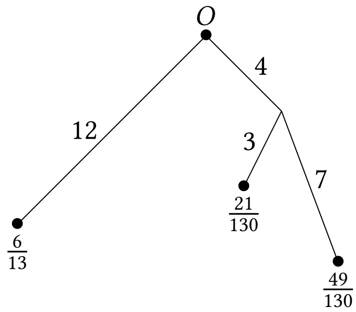

Figure 2.10: The equal branching density

v. We illustrate this strategy Fig. 2.10. The length of the edges is indicated along them, while the probability assigned to each leaf is indicated below them.

Theorem 2.30 below is one of the main theorems related to the study of search games, see (Alpern and Gal, 2003).

Theorem 2.30. *Let $\mathcal{N}$ be a network. The value of the search game on $\mathcal{N}$ satisfies*

$$\frac{\lambda(\mathcal{N})}{2} \leq v \leq \frac{\bar{\lambda}(\mathcal{N})}{2} \leq \lambda(\mathcal{N}).$$

In addition, the lower-bound is reached if and only if $\mathcal{N}$ is Eulerian, and the upper-bound $\lambda(\mathcal{N})$ is reached if and only if $\mathcal{N}$ is a tree.

If $\mathcal{N}$ is Eulerian, then Harry's uniform distribution and Sally's uniform Eulerian strategy form a couple of optimal strategies.

If $\mathcal{N}$ is a tree, then Harry's equal branching density and Sally's uniform Chinese postman strategy form a couple of optimal strategies.

A natural question is for which networks the uniform Chinese postman strategy is optimal. In other words, what are the networks for which the bound $\bar{\lambda}(\mathcal{N})/2$ in Theorem 2.30 is reached? The answer is given by a class of networks containing both Eulerian networks and trees. A *weakly Eulerian network* is a network obtained by replacing the vertices of a tree with an Eulerian network.

The following theorem due to Gal (2000) answers our question.

Theorem 2.31. *The uniform Chinese postman strategy is optimal if and only if $\mathcal{N}$ is weakly Eulerian, and we have*

- *if $\mathcal{N}$ is weakly Eulerian, then $v = \frac{\bar{\lambda}(\mathcal{N})}{2}$;*
- *if $\mathcal{N}$ is not weakly Eulerian, then $v < \frac{\bar{\lambda}(\mathcal{N})}{2}$.*

2.3.3 Contribution of the thesis: continuous patrolling and hiding games

This part is in relation with Chapter 4 entitled *Continuous patrolling and hiding games*. The main results are presented here. The model studied here, called *patrolling game* considers the paradigm 2 mentioned above, namely the maximization of the probability of capture in a given time.

In a patrolling game, an *attacker* chooses a time and place to attack in the search space Q, while a *patroller* continuously walks at a maximum speed of 1, having chosen his starting point. When the attack occurs, the patroller has a delay of $m \in \mathbb{R}_+$ to be at a maximum distance of $r \in \mathbb{R}_+$ from the point of attack, in which case the attack is defused. Otherwise, the attack is a success. The parameter m represents the time required to set up the attack. One can think, for example, of the time it takes to start a bomb.

A patrolling game is therefore given by a triplet (Q, m, r). We denote $\mathcal{W}$ the strategy set of the patroller. This is the set of 1-Lipschitz continuous functions from $\mathbb{R}_+$ to Q. The strategy set of the attacker is $\mathcal{A} = Q \times \mathbb{R}_+$.

The payoff to the patroller $g_{m,r} : \mathcal{W} \times \mathcal{A} \to \{0, 1\}$ is given by

$$g_{m,r}(w, (y, t)) = \begin{cases} 1 & \text{if } d(y, w([t, t + m])) \leq r \\ 0 & \text{otherwise,} \end{cases}$$

where $w([t, t + m]) = \{w(\tau) \mid \tau \in [t, t + m]\}$.

It should be noted that patrolling games were introduced, in the discrete case when the search space is a graph, by Alpern et al. (2011), see also (Alpern et al., 2016a, 2019). In the continuous case, let us mention (Alpern et al., 2016b).

It results from Theorem 2.5 as well as from Ascoli's theorem that the patrolling game (Q, m, r) played in mixed strategies has a value, denoted $V_Q(m, r)$. An upper-bound is given on the value which, as for search games, is based on the maximal discovery rate (Definition 2.26) and on the uniform strategy of the attacker, which consists in uniformly attacking the search space at time 0. Intuitively, a best response of the patroller is to cover as much points of Q as possible between the times 0 and m.

Proposition 2.32. *Let Q be a search space such with positive Lebesgue measure. Then the value of the patrolling game (Q, m, r) satisfies*

$$V_Q(m, r) \leq \frac{m\rho + \lambda(B_r)}{\lambda(Q)}.$$

The following decomposition result is also given. A search space Q is considered as the union of the search spaces $Q_1, \dots, Q_n$, for which the value of the corresponding patrolling game may be known. The value of the patrolling game on Q is then lower-bounded according to the values of the patrolling games involved in the decomposition.

Proposition 2.33. *Let Q and $Q_1, \dots, Q_n$ be search spaces such that $Q = \cup_{i=1}^n Q_i$. Then for all $m, r \in \mathbb{R}_+$ one has*

$$V_Q(m,r) \geq \frac{1}{\sum_{i=1}^n V_{Q_i}(m,r)^{-1}}.$$

Concerning patrolling games on networks, we show the following result giving value and optimal strategies when the network is Eulerian. The uniform strategy of the patroller on an Eulerian network consists in fixing an Eulerian cycle and uniformly choosing a starting point.

Theorem 2.34. *Let $\mathcal{N}$ be an Eulerian network, then*

$$V_{\mathcal{N}}(m,0) = \min\left(\frac{m}{\lambda(\mathcal{N})}, 1\right).$$

Moreover the uniform strategies of the patroller and the attacker are optimal.

A simple search space is a search space in $\mathbb{R}^2$ whose boundary verifies some weak regularity conditions. For these search spaces the following asymptotic result is proved.

Theorem 2.35. *Let Q be a simple search space endowed with the euclidean norm, then*

$$V_Q(m,r) \sim \frac{2rm}{\lambda(Q)},$$

as r goes to 0.

Suppose that the time required for the attack to be a success m is 0. In this case we show that the game is equivalent to a static game called a *hiding game*, introduced by Gale and Glassey (1974) in the case where the search space is a disk, see also (Ruckle, 1983). The set of strategies of both players is Q and the payoff function is

$$h_r : (x,y) \mapsto \begin{cases} 1 & \text{if } \|x-y\| \leq r \\ 0 & \text{otherwise.} \end{cases}$$

That is, both players choose a point of Q, if they are at a distance at most r, it is a success for the searcher, otherwise it is a failure.

In a hiding game, the two players therefore have the same set of strategies. In addition, the payment function is symmetrical in the sense that, if $\mu \in \Delta(Q)$ and $y \in Q$, then $h_r(\mu, y) = h_r(y, \mu) = \mu(B_r(y) \cap Q)$.

A player's strategy is said to be *equalizing*, if the payoff when this player plays this strategy does not depend on the other player's action. We then have the following proposition.

Proposition 2.36. *Let $\mu \in \Delta(Q)$. Then μ is an equalizing strategy (with constant payoff c) if and only if μ is an optimal strategy of both players (and in that case $V_Q(r) = c$).*

The following theorem is our main result regarding hiding games. It asymptotically gives the value of a hiding game on a positive Lebesgue measure search space.

Theorem 2.37. *Let Q be a search space such that $\lambda(Q) > 0$, then*

$$V_Q(r) \sim \frac{\lambda(B_r)}{\lambda(Q)},$$

as r goes to 0.

A consequence of Theorem 2.37 is that for a search space with positive Lebesgue measure, we have $V_Q(r) \sim r^n \lambda(B_1)/\lambda(Q)$ when r goes to 0. We show that for a null Lebesgue measure search space, it is not always true that V_Q admits an equivalent of the form Mr^α, with α and M positive, when r goes to 0. Our counterexample is based on a Cantor set.

2.3.4 Contribution of the thesis : stochastic search games

This part is in relation with Chapter 5 entitled *When Sally found Harry: A Stochastic search game*. The main results are presented here. A search game model is introduced on a graph that evolves over time. In this way, the dynamic aspects related to the evolution of the graph are mixed with those related to the searcher's trajectory. This model is called a *stochastic search game*.

We study a search game where Harry hides on an edge of a graph and Sally travels around the graph in search of Harry. Her goal is to find him as soon as possible.

The novelty of the model is that, due to various circumstances, at any given time, some edges may be unavailable, so the graph randomly evolves over time. At each stage, each edge e of the graph is, independently of the others, active with probability p_e and inactive with probability $1 - p_e$. Let us describe the model in more detail.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a connected undirected graph, where $\mathcal{V}$ is the set of vertices and $\mathcal{E}$ is the set of edges. There exists a special vertex $O \in \mathcal{V}$, called the *origin* of the graph $\mathcal{G}$. Let $\mathbb{G}$ be the set of subgraphs of $\mathcal{G}$. For all $v \in \mathcal{V}$, we call $\mathcal{N}(\mathcal{G}, v)$ the *immediate neighborhood* of v in $\mathcal{G}$:

$$\mathcal{N}(\mathcal{G}, v) = \{v\} \cup \{u \in \mathcal{V} \mid \{v, u\} \in \mathcal{E}\}.$$

The graph will evolve in discrete time as follows. Let $\boldsymbol{p} = (p_e)_{e\in\mathcal{E}} \in (0, 1]^{\mathcal{E}}$. At each stage $t \geq 1$, each edge $e \in \mathcal{E}$ is active with probability p_e or inactive with probability $1 - p_e$, independently of the other edges. This defines a random graph process on $\mathbb{G}$ denoted $(\mathcal{G}_t)_t = (\mathcal{V}, \mathcal{E}_t)_{t\geq 1}$, where $\mathcal{E}_t$ is the random set of active edges at time t.

The game is played as follows. At stage 0 both players know $\mathcal{G}_0 = \mathcal{G}$ and the initial position of the searcher $v_0 = O$. The hider chooses an edge $e \in \mathcal{E}$. Then the graph $\mathcal{G}_1$ is drawn and the searcher chooses $v_1 \in \mathcal{N}(\mathcal{G}_1, v_0)$. If $\{v_0, v_1\} = e$, then the game ends and the payoff to the hider is 1, otherwise the graph $\mathcal{G}_2$ is drawn and the game continues. Inductively, at each stage $t \geq 1$, knowing $h_t = (\mathcal{G}_0, v_0, \ldots, \mathcal{G}_{t-1}, v_{t-1}, \mathcal{G}_t)$, the searcher chooses $v_t \in \mathcal{N}(\mathcal{G}_t, v_{t-1})$. If $\{v_{t-1}, v_t\} = e$, then the game ends and the payoff to the hider is t, otherwise the graph $\mathcal{G}_{t+1}$ is drawn and the game continues.

One can bound the value of the stochastic search game for any $\boldsymbol{p} \in (0,1]^{\mathcal{E}}$, denoted $\operatorname{val}(\boldsymbol{p})$, with respect to the value of the deterministic search game, i.e., when the activation parameters $\boldsymbol{p}$ are all 1.

Proposition 2.38. *For all $\boldsymbol{p} \in (0,1]^{\mathcal{E}}$ the value of the stochastic search game satisfies*

$$\frac{\operatorname{val}(\mathbf{1})}{1-(1-\min_{e\in\mathcal{E}} p_e)^{\delta}} \leq \operatorname{val}(\boldsymbol{p}) \leq \frac{\operatorname{val}(\mathbf{1})}{\min_{e\in\mathcal{E}} p_e},$$

where δ is the maximum degree of $\mathcal{G}$. As a consequence

$$\operatorname{val}(\boldsymbol{p}) \longrightarrow \operatorname{val}(\mathbf{1}), \quad \text{as } \min_{e\in\mathcal{E}} p_e \longrightarrow 1.$$

First, we adapt the concepts and results defined (in the deterministic framework) in the case of continuous networks, to the case of graphs. In particular, we prove the following theorem, which is the analog of Theorem 2.30 for networks.

Theorem 2.39. *For any graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the value of the deterministic search game satisfies*

$$\frac{\operatorname{card}\mathcal{E}+1}{2} \leq \operatorname{val}(1) \leq \operatorname{card}\mathcal{E}.$$

Moreover, if $\operatorname{card}\mathcal{E} > 1$, the upper bound is reached if and only if $\mathcal{G}$ is a tree and the lower bound is reached if and only if $\mathcal{G}$ is an Eulerian graph.

If $\mathcal{G}$ is an Eulerian graph, then the uniform distribution on $\mathcal{E}$ and the uniform Eulerian strategy are optimal.

If $\mathcal{G}$ is a tree, then the equal branching density and the uniform Chinese postman tour are optimal.

We see here also that trees and Eulerian graphs represent the limit classes in terms of value of the deterministic search game with fixed number of edges. Our goal is to study these two classes in the stochastic framework. We place ourselves in the case where $\boldsymbol{p}_e = p$ for all $e \in \mathcal{E}$, i.e., the activation parameter is the same for each edge.

We define a particular class of strategies of the searcher in trees, called depth-first strategies. They have the property of never going backward at a vertex before having visited the whole subtree. They generalize the Chinese postman cycles of the deterministic case.

Definition 2.40. A depth-first strategy on a tree is a strategy of the searcher that prescribes the following, when arriving at a vertex:

- if the set of un-searched and active outgoing edges is non-empty, take one of its edges (possibly at random);
- if all the un-searched outgoing edges are inactive, wait;
- if all outgoing edges have been searched and the backward edge is active, take it;
- if all outgoing edges have been searched and the backward edge is inactive, wait.

Definition 2.41. A depth-first strategy on $\mathcal{T}$ induces an expected time to travel from the origin O back to it, covering the entire tree. This is called the *cycle time* of $\mathcal{T}$ and is denoted $\tau(O)$. For any vertex or edge z, the cycle time of $\mathcal{T}_z$ is denoted $\tau(z)$.

Notice that $\tau(O)$ depends on $\boldsymbol{p}$, but is independent of the choice of depth-first strategy. For example, in the tree represented Fig. 2.11, the cycle time is $\tau(O) = \frac{2}{1-(1-p)^2} + \frac{6}{p}$.

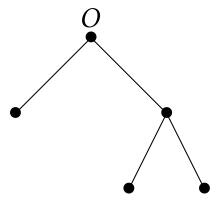

Figure 2.11: A simple tree

The *biased depth first strategy* is a depth first strategy that at any vertex v, when un-searched edges are active simultaneously, gives a certain probability of taking each of them, which depends on p as well as the geometry of the subtree with root v.

The equal branching density is also generalized to the stochastic setting. The relevant parameter is no longer the length or number of edges of the subtrees, but rather the cycle time of the subtrees. The equal branching density is the probability measure with support on the leaf edges of the tree, such that for any vertex v with outgoing edges $e_1, \ldots, e_n$, the probability that Harry hides in the subtree originating from e_i is equal to the cycle time of this subtree divided by the sum of the cycle times of the subtree originating from the edges $e_1, \ldots, e_n$.

We prove the following theorems for binary trees. The quantity Λ is defined recursively, it depends on p as well as the geometry of the tree.

Theorem 2.42. *There exists $p_0 \in (0,1)$ such that for all $p \geq p_0$, the time to reach any leaf edge using the biased depth-first strategy is*

$$\frac{1}{2}\tau(O) + \Lambda(O).$$

Hence for all $p \geq p_0$, we have

$$\mathrm{val}(p) \leq \frac{1}{2}\tau(O) + \Lambda(O).$$

Theorem 2.43. *The equal branching density of the hider yields the same payoff against any depth-first strategy of the searcher, and this payoff is*

$$\frac{1}{2}\tau(O) + \Lambda(O).$$

Theorems 2.42 and 2.43 imply that in a binary tree $\mathcal{G}$, if depth-first strategies are best responses to the equal branching density, then there exists $p_0 \in (0,1)$ such that for all $p \geq p_0$ the value of the game is $\frac{1}{2}\tau(O) + \Lambda(O)$. Moreover the biased depth-first strategy and the equal branching density are optimal.

However, we prove that there exist binary trees for which depth-first strategies are not best responses to the equal branching density. Nevertheless, we have the following corollary. The *uniform depth-first strategy* is the depth-first strategy choosing uniformly between active un-searched outgoing edges, it is independent of p.

Corollary 2.44. *If the graph $\mathcal{G}$ is a line with L edges, then depth-first strategies are best responses to the equal branching density. The value of the game is*

$$\mathrm{val}(p) = \frac{1}{2}\tau(O) + \Lambda(O) = \frac{L}{p} + \frac{1}{1-(1-p)^2} - \frac{1}{p},$$

for all $p \in (0,1]$, if the origin O is not an extreme vertex.

Moreover the equal branching density and the uniform depth-first strategy are optimal.

For Eulerian graphs, the uniform Eulerian strategy is generalized to the stochastic framework, and a strategy is called a *Eulerian strategy* if at each vertex it chooses an active edge that has not yet been searched and such that the induced path belongs to an Eulerian cycle. The Eulerian strategy that at each vertex chooses uniformly among the edges is the *uniform Eulerian strategy*. The uniform Eulerian strategy induces an expected time of travel from the origin covering the entire graph $\mathcal{G}$. It is noted $\theta(\mathcal{G})$.

A *parallel Eulerian graph* is composed of an even number of parallel paths that join two vertices, one of these vertices being the origin O, see figure Fig. 2.12. Such a graph is noted $\mathcal{P}_{2m}(\boldsymbol{\lambda})$, where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{2m})$ is the vector of parallel path lengths.

For parallel Eulerian graphs we prove the following two theorems. The quantity Φ is defined recursively. It depends on p as well as on the number of parallel paths.

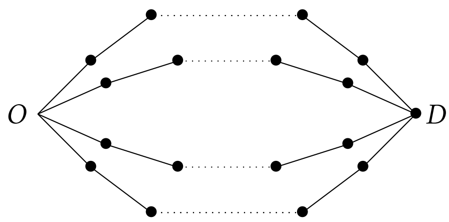

Figure 2.12: A parallel Eulerian graph

Theorem 2.45. *On a parallel Eulerian graph $\mathcal{P}_{2m}(\boldsymbol{\lambda})$, the expected time to reach any edge using the uniform Eulerian strategy is*

$$\frac{\theta(\mathcal{P}_{2m}(\boldsymbol{\lambda})) + p^{-1}}{2} + \Phi_m.$$

Hence, for all $p \in (0, 1]$, we have

$$\mathrm{val}(p) \leq \frac{\theta(\mathcal{P}_{2m}(\boldsymbol{\lambda})) + p^{-1}}{2} + \Phi_m.$$

Theorem 2.46. *On a parallel Eulerian graph $\mathcal{P}_{2m}(\boldsymbol{\lambda})$, the uniform distribution of the hider yields the same payoff*

$$\frac{\theta(\mathcal{P}_{2m}(\boldsymbol{\lambda})) + p^{-1}}{2} + \Phi_m$$

against any Eulerian strategy of the searcher.

We show that Eulerian strategies are not always better responses to uniform distribution. However, we have the following corollary.

Corollary 2.47. *The value of the game played on the circle with L edges is*

$$\mathrm{val}(p) = \frac{\theta(\mathcal{G}) + 1/p}{2} + \Phi_2 = \frac{1}{1-(1-p)^2} + \frac{L-1}{2p},$$

of all $p \in (0, 1]$. Moreover the uniform distribution and the uniform Eulerian strategy are optimal.

2.4 Repeated games with incomplete information

Until now, the game models presented had in common the players' knowledge of the state of nature as well as the actions played previously. We now place ourselves in the situation where players only receive *signals* on the state as well as on the actions, which may not completely reveal them.

2.4.1 The general model of repeated games with signals

First, let us present a general model of *repeated game with signals*, see (Mertens et al., 2015). A repeated game with signals is given by a 7-tuple $\Gamma = (\Omega, A, B, C, D, g, \rho)$ where

- Ω is the state space;
- A and B are the action sets of players 1 and 2 respectively;
- C and D are the signal sets of players 1 and 2 respectively;
- $g : \Omega \times A \times B \longrightarrow \mathbb{R}$ is the payoff function;
- $\rho : \Omega \times A \times B \longrightarrow \Delta(\Omega \times C \times D)$ is the transition probability.

We suppose that the sets Ω, A, B, C and D are nonempty and finite. Let $p \in \Delta(\Omega \times C \times D)$ be an initial probability distribution. The game $\Gamma(p)$ is played as follows.

- Before the beginning of the game, a triplet (ω_1, c_0, d_0) is drawn according to p. The initial state is ω_1, players 1 and 2 receive the signals c_0 and d_0 respectively.
- At each step $t \geq 1$, players 1 and 2 simultaneously choose actions $a_t \in A$ and $b_t \in B$ respectively. A triplet (ω_{t+1}, c_t, d_t) is drawn according to $\rho(\cdot|\omega_t, a_t, b_t)$. Players 1 and 2 receive the signals c_t and d_t respectively. The stage payoff is $g_t = g(\omega_t, a_t, b_t)$. The game goes in state ω_{t+1} and moves on to the next stage.

The notion of strategy is formally given in a repeated game with signals. We denote $H_t = \Omega \times C \times D \times (\Omega \times A \times B \times C \times D)^{t-1}$ the set of all possible histories at stage $t \geq 1$. We endow H_t with the product σ-algebra denoted $\mathcal{H}_t$. We denote $H_\infty = \Omega \times C \times D \times (\Omega \times A \times B \times C \times D)^{\mathbb{N}^*}$ the set of infinite histories, endowed with the product σ-algebra induced by $\bigcup_{t\geq 1} \mathcal{H}_t$.

The set of private histories of player 1 at stage t is $H_t^1 = C \times (A \times C)^{t-1}$. It is endowed with the product σ-algebra denoted $\mathcal{H}_t^1$. The set of private histories of player 2 which is defined in a similar way.

Definition 2.48. A *behavior strategy* of player 1 is a sequence of mappings $\sigma = (\sigma_t)_{t\geq 1}$ such that for all $t \geq 1$, σ_t is a measurable mapping from $(H_t^1, \mathcal{H}_t^1)$ to $\Delta(A)$. The behavior strategies of player 2 are defined in a similar way.

We denote $\mathcal{S}$ and $\mathcal{T}$ the behavior strategy sets of players 1 and 2 respectively. A behavior strategy is called *pure* if for each finite history $h_t^1 \in H_t^1$, we have $\sigma_t(h_t^1) \in A$. A *mixed strategy* is a probability distribution on the set of pure strategies endowed with the product σ-algebra. We have two similar definitions for player 2.

An initial probability $p \in \Delta(\Omega \times C \times D)$ as well as a pair of (behavior or mixed) strategies (σ, τ) naturally induce a probability measure on the set of finite histories $\bigcup_{t\geq 1} H_t$. Kolmogorov's extension theorem tells us that this measure extends uniquely to H_∞. We denote this probability measure $\mathbb{P}^p_{\sigma,\tau}$, and $\mathbb{E}^p_{\sigma,\tau}$ the corresponding expectation.

As with stochastic games, Kuhn's theorem, see Theorem 2.11, applies in this context.

There are also two main ways to aggregate stage payoffs, namely through Abel's and Caesaro's means.

Let $\lambda \in (0,1]$, the *λ-discounted game*, denoted $\Gamma_\lambda(p)$, is the game with strategy sets $\mathcal{S}$ and $\mathcal{T}$ and with payoff function $\gamma^p_\lambda : \mathcal{S} \times \mathcal{T} \to \mathbb{R}$ defined by

$$\gamma^p_\lambda(\sigma,\tau) = \mathbb{E}^p_{\sigma,\tau}\left(\sum_{t\geq 1} \lambda(1-\lambda)^{t-1} g_t\right).$$

Let $N \in \mathbb{N}^*$, the *N-stage repeated game*, denoted $\Gamma_N(p)$ is the game with strategy sets $\mathcal{S}$ and $\mathcal{T}$ and with payoff function $\gamma^p_N : \mathcal{S} \times \mathcal{T} \to \mathbb{R}$ defined by

$$\gamma^p_N(\sigma,\tau) = \mathbb{E}^p_{\sigma,\tau}\left(\frac{1}{N}\sum_{t=1}^{N} g_t\right).$$

Theorem 2.6 implies the existence of the value in the games $\Gamma_\lambda(p)$ and $\Gamma_N(p)$, as well as optimal strategies. We denote

$$\begin{aligned} v_\lambda(p) &= \max_{\sigma\in\mathcal{S}} \min_{\tau\in\mathcal{T}} \gamma^p_\lambda(\sigma,\tau) = \min_{\tau\in\mathcal{T}} \max_{\sigma\in\mathcal{S}} \gamma^p_\lambda(\sigma,\tau), \\ v_N(p) &= \max_{\sigma\in\mathcal{S}} \min_{\tau\in\mathcal{T}} \gamma^p_\lambda(\sigma,\tau) = \min_{\tau\in\mathcal{T}} \max_{\sigma\in\mathcal{S}} \gamma^p_N(\sigma,\tau). \end{aligned}$$

The uniform value is defined in a similar way to the one defined for stochastic games.

A plethora of literature has developed on repeated games with incomplete information. Examples include the work of Cardaliaguet et al. (2012); Forges (1982); Gensbittel (2014); Gensbittel et al. (2014); Gensbittel and Renault (2015); Gimbert et al. (2016); Laraki (2001, 2002); Mertens and Zamir (1971); Neyman (2008); Renault (2006, 2012); Renault and Venel (2017); Rosenberg et al. (2003, 2004); Sorin (1984, 1985); Ziliotto (2016b). For a literature review, we also refer to (Renault, 2018).

2.4.2 Repeated games with incomplete information on one side

Here we present the model of *repeated games with incomplete information on one side* introduced by Aumann and Maschler (1995). Before the start of the game, the initial probability $p \in \Delta(\Omega)$ selects a state $\omega \in \Omega$ which is fixed for the whole duration of the game. It is revealed to player 1 but not to player 2. Then at each stage

$t \geq 1$, players 1 and 2 simultaneously select actions $a_t \in A$ and $b_t \in B$ respectively. These are revealed to both players before the next stage. The probability p therefore represents the initial *belief* of player 2 on the state of nature.

Here is how player 1 uses his information. He chooses its first action (or more generally a message or signal k from a finite set K) according to the state of nature ω selected. Let $x \in \Delta(K)^{\Omega}$ be the transition probability played by player 1, i.e., if the state is ω, he sends the signal k with probability $x^{\omega}(k)$.

We denote λ_k the total probability that k be drawn. It is equal to

$$\sum_{\omega \in \Omega} p(\omega) x^{\omega}(k),$$

moreover if $\lambda_k > 0$, then the conditional probability on Ω knowing k, also called *posterior belief* is

$$\hat{p}(x,k) = \left(\frac{p(\omega) x^{\omega}(k)}{\lambda_k} \right)_{\omega \in \Omega}.$$

Since $\sum_{k \in K} \lambda_k \hat{p}(x,k) = p$, posterior beliefs contain the *prior belief* p in their convex envelope. The next lemma, called the *splitting lemma*, fundamental in repeated games with incomplete information, is in a way a reciprocal: player 1 can induce any posterior belief containing the initial probability p in their convex envelope.

Lemma 2.49 (Splitting lemma). *Suppose $p = \sum_{k \in K} \lambda_k p_k$, with for all $k \in K$, $\lambda_k > 0$, $p_k \in \Delta(\Omega)$ and $\sum_{k \in K} \lambda_k = 1$.*

Then there exists a transition probability $x \in \Delta(K)^{\Omega}$ such that for any $k \in K$

$$\lambda_k = \sum_{\omega \in \Omega} p(\omega) x^{\omega}(k) \text{ and } \hat{p}(x,k) = p_k.$$

For any upper semi-continuous function f defined from $\Delta(\Omega)$ to $\mathbb{R}$, the smallest concave function at any point above f is denoted cav f. The function cav f is continuous and for every $p \in \Delta(\Omega)$ we have

$$\begin{aligned} \operatorname{cav} f(p) = \max\{ \sum_{k \in K} \lambda_k f(p_k) \mid K \text{ finite}, \forall k \in K\ \lambda_k > 0,\ p_k \in \Delta(K), \\ \sum_{k \in K} \lambda_k = 1,\ \sum_{k \in K} \lambda_k p_k = p \}. \end{aligned}$$

The next lemma is a consequence of Lemma 2.49.

Lemma 2.50. *In any finitely or infinitely repeated game, if player 1 has all the information, then if he guarantees f he also guarantees cav f.*

Aumann and Maschler (1995) prove the following theorem.

Theorem 2.51. *Let $u(p)$ be the value of the matrix game with action sets A and B and payoff function $g_p : A \times B \to \mathbb{R}$ defined for all $(a, b) \in A \times B$ by*

$$g_p(a, b) = \sum_{\omega \in \Omega} p(\omega) g(\omega, a, b).$$

Then the repeated game with incomplete information on one side $\Gamma(p)$ has a uniform value which is $\operatorname{cav} u(p)$.

2.4.3 Contribution of the thesis : dynamic control of information with observed return on investment

This section is in relation with Chapter 6 entitled *Dynamic control of information with observed return on investment.* The main results are presented here.

We study a dynamic game with incomplete information in line with Bayesian persuasion models. In this respect, we mention the work of Kamenica and Gentzkow (2011) when the state of nature is fixed, and of Ely (2017). Renault et al. (2017) study a dynamic information control model close to ours, except that returns on investment are not observed, see below.

The protagonists are an advisor and an investor. At each stage n, the advisor privately observes the realization of the state of nature $\omega_n \in \Omega$ which evolves over time. The advisor decides what information to disclose to the investor, who in turn decides whether or not to invest. If there is an investment, the investor observes whether it is a success or a failure.

For each investment, the advisor receives a fixed commission normalized to 1, and discounts his earnings by the discount factor δ <1. In case of investment, the investor pays a fee of 1 and earns $M > 1$ with probability $r(\omega_n)$ and 0 with probability $1 - r(\omega_n)$, where $r : \Omega \to [0, 1]$.

We assume that $(\omega_n)_{n \geq 0}$ follows a Markov chain with transition $(\rho(\omega'|\omega))_{\omega', \omega \in \Omega}$. The investor knows the distribution of the sequence $(\omega_n)_{n \geq 0}$. The additional information received during the game comes from the advisor and the observed returns on investment. The investor invests if and only if her expected payoff is positive, i.e., if the current belief is $p \in \Delta(\Omega)$, she invests if $\langle p, r \rangle \geq c$, where $c = 1/M$. The investment region is $I = \{p \in \Delta(\Omega),\ \langle p, r \rangle \geq c\}$ and the non-investment region is $J = \Delta(\Omega) \setminus I$.

The game reduces to a Markov decision process denoted Γ, in which the advisor manipulates the investor's posterior beliefs in order to maximize the expected frequency of stages with investment.

The investor uses her knowledge of the distribution of $(\omega_n)_{n \geq 0}$ as well as the returns on investment she observes to update her posterior belief. We can thus distinguish three beliefs of the investor at the n stage. The *prior belief* $p_n \in \Delta(\Omega)$ is the belief at stage n before receiving the message from the advisor. The *intermediate belief* q_n is the belief updated after receiving the message, in particular the investor invests if and only if $q_n \in I$. Finally, the *posterior belief* s_n is the updated belief of q_n after the return on investment, if any, has been observed.

The belief p_{n+1} differs from s_n because the states are not fully persistent. For every $\omega' \in \Omega$ we have $p_{n+1}(\omega') = \sum_{\omega\in\Omega} s_n(\omega)\rho(\omega'|\omega)$. We therefore define the linear mapping $\phi : \Delta(\Omega) \to \Delta(\Omega)$ such that $p_{n+1} = \phi(s_n)$.

The belief q_n differs from p_n because of the information disclosed. Let $p \in \Delta(\Omega)$ be a belief, let $\mathcal{S}(p) \subset \Delta(\Delta(\Omega))$ be the set of probability measures on $\Delta(\Omega)$ with expectation p. The elements of $\mathcal{S}(p)$ are called *splitting* at p. A consequence of Bayesian updating is that regardless of the information disclosure strategy, the conditional distribution of q_n belongs to $\mathcal{S}(p_n)$. Conversely, the splitting lemma, Lemma 2.49, states that for a given belief $p \in \Delta(\Omega)$ and for a given splitting $\mu \in \mathcal{S}(p)$, the advisor can correlate his messages with the state of nature so that the distribution of the updated belief q is μ.

Finally, the belief s_n differs from q_n due to the observation of returns on investment. In the case where the investor does not invest, the belief is unchanged, i.e., $q_n = s_n$. In case of investment, with probability $r(\omega_n)$ the investor observes a success and we have, for every $\omega \in \Omega$, $s_n(\omega) = \frac{q_n(\omega)r(\omega)}{\langle q_n, r\rangle}$. And besides, with probability $1 - r(\omega_n)$ the investor observes a failure, and we have, for every $\omega \in \Omega$, $s_n(\omega) = \frac{q_n(\omega)(1-r(\omega))}{1-\langle q_n, r\rangle}$.

We define the mappings ψ^+ and ψ^- from $\Delta(\Omega)$ to $\Delta(\Omega)$ as follows: for any $p \in \Delta(\Omega)$, $\psi^+(p)(\omega) = \frac{p(\omega)r(\omega)}{\langle p, r\rangle}$ and $\psi^-(p)(\omega) = \frac{p(\omega)(1-r(\omega))}{1-\langle p, r\rangle}$. Thus in case of investment at stage n, we have $s_n = \psi^+(q_n)$ in case of success, and $s_n = \psi^-(q_n)$ in case of failure.

Let E be the set of functions from $\Delta(\Omega)$ to $[0, 1]$. The dynamic programming operator T is defined from E to E by

$$T(f) : p \mapsto \begin{cases} \delta f(\phi(p)) & \text{if } p \in J \\ 1 - \delta + \delta\langle p, r\rangle f(\phi(\psi^+(p))) + \delta(1 - \langle p, r\rangle)f(\phi(\psi^-(p))) & \text{if } p \in I. \end{cases}$$

We denote $V_\delta(p)$ the value of the dynamic optimization problem $\Gamma(p)$ as a function of the initial belief p. It is the only solution of the functional equation

$$\operatorname{cav} T(f) = f.$$

At each stage, if the current belief p is not in the investment region, the advisor must compromise. To have a positive payoff at the current stage, he must disclose some information. However, this may lower future payoffs because of the concavity of the value function. To address this problem, the greedy strategy is introduced, which minimizes the amount of information disclosed while maximizing the stage payoff.

The greedy strategy, denoted σ_*, is such that if $p \in I$, then the advisor does not disclose any information. If on the contrary $p \in J$ then we consider the following problem:

$$\begin{aligned} \text{maximize} \quad & a_I \\ \text{subject to} \quad & p = a_I q_I + a_J q_J, \\ & q_I \in I, \\ & a_I + a_J = 1, \ a_I, \ a_J \geq 0. \end{aligned} \tag{2.4.1}$$

Let (a_I, a_J, q_I, q_J) be a solution to problem 2.4.1, then $\sigma_*(p) \in \mathcal{S}(p)$ is the splitting that selects q_I and q_J with probabilities a_I and a_J respectively.

Let g denote the payoff function induced by the greedy strategy. The latter satisfies

$$g : p \mapsto \begin{cases} a_I g(q_I) + a_J g(q_J) & \text{si } p \in J \\ 1 - \delta + \delta \langle p, r \rangle g(\phi(\psi^+(p))) + \delta(1 - \langle p, r \rangle) g(\phi(\psi^-(p))) & \text{si } p \in I. \end{cases} \tag{2.4.2}$$

The proposition below, also valid in the context studied by Renault et al. (2017), gives a convenient necessary and sufficient condition for the greedy strategy to be optimal.

Proposition 2.52. *The greedy strategy is optimal if and only if*

1. *g is concave;*
2. *for all $p \in J$, one has $\delta g \circ \phi(p) \leq g(p)$.*

Thanks to Proposition 2.52, we prove the following theorem.

Theorem 2.53. *If* $\operatorname{card}(\Omega) = 2$*, then the greedy strategy is optimal.*

If there are more than three states of nature, the greedy strategy may not be optimal. This is the object of the theorem below, which is based on a counterexample of Renault et al. (2017).

Theorem 2.54. *Suppose* $\operatorname{card}(\Omega) = 3$*. The greedy strategy need not be optimal.*

Chapter 3

Communicating zero-sum product stochastic games

Journal of Mathematical Analysis and Applications 477 (2019) 60-84

We study two classes of zero-sum stochastic games with compact action sets and a finite product state space. These two classes assume a communication property on the state spaces of the players. For strongly communicating on one side games, we prove the existence of the uniform value. For weakly communicating on both sides games, we prove that the asymptotic value, and therefore the uniform value, may fail to exist.

3.1 Introduction

3.1.1 Problem and contribution

In a zero-sum stochastic game, two players interact repeatedly at discrete times, with opposite interests. At each stage, players face a zero-sum game given by the state of nature which evolves according to the current state, and the pair of actions players choose given the history. Therefore, the actions played at each stage impact both the payoff today and the law of the state of nature tomorrow. Players intend to optimize their expected overall payoff. The n-stage repeated and the λ-discounted games are the games in which the overall payoffs are respectively the Cesàro and Abel means of the stage payoffs. Under mild assumptions both games have a value denoted respectively v_n and v_λ.

A fundamental question arising in the theory of dynamic games is the asymptotic behavior of these values. We shall focus on the two following approaches of this issue. The asymptotic approach studies the convergence of the values of the n-stage repeated game and the λ-discounted game, as n goes to infinity and λ goes to 0, that is as players become more patient. If these quantities converge and are equal, the game is said to have an asymptotic value. The uniform approach is dedicated to the existence, for both players, of strategies that are ε-optimal in every

n-stage repeated game, provided that n is large enough. If such strategies exist, players are able to play optimally in every game long enough without knowing the length of the game. In that case, the game is said to have a uniform value. These strategies are also ε-optimal in every λ-discounted game, provided that λ is small enough. While the existence of the uniform value implies the existence of the asymptotic value, the converse is not true.

The aim of this paper is to study the asymptotic and uniform values in two classes of zero-sum product stochastic games. A product state space is of the form $X \times Y$. Moreover, players control the transitions on their own components of the state space, that is the next state in X only depends on the current state in X and the action of player 1, and similarly for Y. We consider the case where X and Y are finite, and action sets are compact. The two classes we are interested in assume a communication property on the state spaces of the players. These are called the strong and the weak communication properties — the strong communication property implying the weak one.

The first class is the class of strongly communicating on one side zero-sum product stochastic games. In such games, for one player, there exists a time T such that independently of his choice of policy, there is a positive probability of passing from any state to any other state in his component of the state space in exactly T stages. This assumption, of ergodic nature, implies that the current state of the player having the strong communication property has in the long run little importance.

The second class is the class of weakly communicating on both sides zero-sum product stochastic games. In such games, for each player there exists a time T and a policy such that, for any two states in their components of the state space, they can move from one to the other with positive probability in exactly T stages. Thus players totally control the dynamics on their components of the state space.

Strongly communicating on one side and weakly communicating on both sides zero-sum product stochastic games have, to our knowledge, never been studied before.

For strongly communicating on one side games, we prove the existence of the uniform value, which does not depend on the initial state of the player having the strong communication property. Furthermore we prove that this player has ε-optimal strategies that have a simple structure. We call them Markov periodic strategies. Under these strategies, the action chosen at each stage does not depend on the whole history but only on the current state and stage modulo the period. Hence they are a particular case of Markov strategies, for which the actions chosen depend only on the current state and stage, and are more general than stationary strategies, which only depend on the current state. Our proof is based on a classification of the state space of the player who is not assumed to have the strong communication property (note that if both players have it, the proof is considerably simpler and the decomposition is actually not needed). This decomposition relies on recurrent classes induced by stationary policies. It has been introduced for Markov Decision Processes (MDP) by Ross and Varadarajan

(1991), similar classifications have been used by Bather (1973); Flesch et al. (2008); Solan (2003). Building on that classification, we consider a family of auxiliary stochastic games and prove that they have a uniform value independent of the initial state. Finally we build an auxiliary MDP for the player who does not have the strong communication property, whose payoffs are the uniform values of the previous auxiliary games. We conclude by proving that the uniform value of the MDP is also the uniform value of the initial game.

Regarding weakly communicating on both sides games, we provide an example of a game which does not have an asymptotic value (and hence neither has a uniform one). Our proof is based on a reduction of this example to a simpler game of perfect information with two absorbing and two non absorbing states introduced by Renault (2019). A key ingredient of the non-existence of the asymptotic value is the non semi-algebraic aspect of the action set of one player. The example in particular shows that in weakly communicating on both sides games, even if players can go from any state to any other state of their component in finite time, they can make mistakes that are irreversible with regards to the joint state.

3.1.2 Related literature

Zero-sum stochastic games were introduced by Shapley (1953) in the finite setting (finite state and action sets), for which he proved the existence of the value in the λ-discounted game. Mertens and Neyman (1981) proved the existence of the uniform value. Their proof is based on the fact that the value of the λ-discounted game has bounded variations in λ, as shown by Bewley and Kohlberg (1976). It is a key question whether the existence of the uniform value extends to non finite zero-sum stochastic games.

This question has been answered positively for several classes of zero-sum stochastic games with a finite state space and non finite action sets, as we consider in this paper. For absorbing games (Mertens et al., 2009) and recursive games (Li and Sorin, 2016), the proofs use the operator approach of Rosenberg and Sorin (2001) that relies on the Shapley operator which entirely contains the dynamics of the game. Still for a finite state space, Bolte et al. (2014) showed that games with semi-algebraic (or more generally definable) transitions and actions set have a uniform value. Finally, Renault (2010) proved the existence of the uniform value in MDPs with a finite state space and arbitrary action set.

However, in the last few years, several counterexamples to the existence of the asymptotic value in zero-sum stochastic games with finite state space and compact action sets have been proposed, see (Sorin and Vigeral, 2015; Vigeral, 2013; Ziliotto, 2016b), ending the long standing idea that such games had an asymptotic value. Laraki and Renault (2017) provided such a counterexample with a product state space. As it is the case for the counterexample presented in this paper, all these counterexamples have in common to have non semi-algebraic transition probabilities or non semi-algebraic action sets. This is a key element to make the value oscillate. It should be put into perspective with the work of Bolte et al. (2014)

on definable zero-sum stochastic games. Another feature these counterexamples have in common is to have absorbing states. These states which cannot be left once reached are incompatible be the weak communication property. This is a major difference between previous counterexamples and the present one.

Zero-sum stochastic games on a product state space have been introduced by Altman et al. (2005), who examined the case where each player only observes his component of the state and his actions, and showed that these games can be solved by linear programming. Flesch et al. (2008, 2009) studied equilibria in N-players finite product stochastic games. They however considered an overall payoff which is the limit inferior of the n-stage repeated game payoff, while we are interested in the existence of the uniform value. Finally, Laraki and Renault (2017) showed the existence of the asymptotic value in zero-sum product stochastic games under a strong acyclicity condition. This strong acyclicity condition encompasses the irreversibility in the transitions of several classes of repeated games for which the asymptotic value is known to exist. The strong acyclicity condition is incompatible with our weak communication property. It is important to understand in which classes of stochastic games that do not assume an irreversibility condition the asymptotic value may exist.

Gillette (1957) introduced games in which independently of the choice of strategies, there is a positive probability of passing from any state to any other state in exactly T stages. He called them cyclic stochastic games. These were also investigated by Hoffman and Karp (1966), Bewley and Kohleberg (1978) and Vrieze (2003). A similar assumption has also been examined by Fudenberg and Yamamoto (2011) for games where players observe the state and a public signal related to the actions played. However in these articles the property is considered on the whole state space and not only on one component of a product state space.

3.1.3 Organization of the paper

The article is organized as follows. In Section 3.2 we describe the model of zero-sum product stochastic games and recall some elementary facts. In Section 3.3 we give formal definitions of strongly and weakly communicating, and state the two main theorems. Finally, Sections 3.4 and 3.5 are dedicated to the proofs of the two main theorems.

3.2 Preliminaries on zero-sum product stochastic games

3.2.1 Model and course of the game

Let X and Y be two nonempty finite sets. Let A and B be two nonempty compact metric sets endowed with their Borel σ-algebras. Let $p : X \times A \to \Delta(X)$ and $q : Y \times B \to \Delta(Y)$, be such that for all $x, x' \in X$ and all $y, y' \in Y$, $p(x'|x, \cdot)$ and $q(y'|y, \cdot)$ are continuous. $\Delta(X)$ denotes the set of probability measures over X, and similarly for $\Delta(Y)$. Let $u : X \times Y \times A \times B \longrightarrow [0,1]$, be such that for all $(x, y) \in X \times Y$ and all $a \in A$ and $b \in B$, $u(x, y, \cdot, b)$ and $u(x, y, a, \cdot)$ are continuous.

X is the state space of player 1, Y is the state space of player 2. A is the action set of player 1, B is the action set of player 2. It is without loss of generality that the action sets do not depend on the current state. p is the transition probability of player 1, q is the transition probability of player 2. u is the payoff to player 1.

Let $\Gamma = (X, Y, A, B, p, q, u)$. The game Γ is played in stages as follows: an initial state $(x_1, y_1) \in X \times Y$ is given and known by the players. Inductively at stage n, knowing the past history $h_n = (x_1, y_1, a_1, b_1, \ldots, x_{n-1}, y_{n-1}, a_{n-1}, b_{n-1}, x_n, y_n)$, player 1 and 2 simultaneously choose an action, respectively $a_n \in A$ and $b_n \in B$. A new state $x_{n+1} \in X$ is selected according to the distribution $p(\cdot|x_n, a_n)$ on X and a new state $y_{n+1} \in Y$ is selected according to the distribution $q(\cdot|y_n, b_n)$ on Y. The payoff to player 1 at stage n is $u_n = u(x_n, y_n, a_n, b_n)$.

3.2.2 Policies and Strategies

Since the state space of the game Γ is a product of two sets X and Y, we distinguish policies of the players, which only depend on their own history, i.e., on their actions and component of the product state space, from strategies, which depend on the joint history. We denote strategies and strategy sets depending on the joint history with bold letters.

We denote the set of positive integers by $\mathbb{N}^*$. For $n \in \mathbb{N}^*$, let $H_n^1 = X \times (A \times X)^{n-1}$ be the set of histories of player 1 at stage n and $H_\infty^1 = (A \times X)^\infty$ be the set of infinite histories. H_n^1 is endowed with the product σ-algebra $\mathcal{H}_n^1$, and H_∞^1 with the product σ-algebra $\mathcal{H}_\infty^1$ spanned by $\bigcup_{n\geq 1} \mathcal{H}_n^1$. Let $\mathcal{S}$ denote the set of behavior policies of player 1 depending only on his own history. A policy $\sigma \in \mathcal{S}$ is a sequence $(\sigma_n)_{n\in\mathbb{N}^*}$, where σ_n is a measurable map from $(H_n^1, \mathcal{H}_n^1)$ to $\Delta(A)$. A policy σ together with an initial state $x \in X$ define a unique probability distribution over H_∞^1 which we denote $\mathbb{P}_\sigma^x$. We define analogous objects for player 2 and denote $\mathcal{T}$ the set of behavior policies of player 1 depending only on his own history.

For $n \in \mathbb{N}^*$, let $H_n = X \times Y \times (A \times B \times X \times Y)^{n-1}$ be the set of joint histories at stage n and $H_\infty = (A \times B \times X \times Y)^\infty$ be the set of infinite joint histories. H_n is endowed with the product σ-algebra $\mathcal{H}_n$, and H_∞ with the product σ-algebra $\mathcal{H}_\infty$ spanned by $\bigcup_{n\geq 1} \mathcal{H}_n$. Let $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{\mathcal{T}}$ denote the sets of behavior strategies of player 1 and player 2 respectively. A strategy $\boldsymbol{\sigma} \in \boldsymbol{\mathcal{S}}$ is a sequence $(\boldsymbol{\sigma}_n)_{n\in\mathbb{N}^*}$, where $\boldsymbol{\sigma}_n$ is a measurable map from $(H_n, \mathcal{H}_n)$ to $\Delta(A)$, and likewise for $\boldsymbol{\mathcal{T}}$. A pair of strategies $(\boldsymbol{\sigma}, \boldsymbol{\tau})$ together with an initial state (x, y) define a unique probability distribution over H_∞ which we denote $\mathbb{P}_{\boldsymbol{\sigma},\boldsymbol{\tau}}^{x,y}$.

A strategy is a Markov strategy if the mixed action played at every stage depends only on the current stage and state. Markov periodic strategies are Markov strategies depending on the stage modulo the period and on the current state. Let us give a formal definition.

Definition 3.1. For $N \in \mathbb{N}^*$, a strategy $\boldsymbol{\sigma}$ of player 1 is called an N-periodic

Markov strategy if there exists $(\boldsymbol{\mu}_n)_{n\in\{1,\dots,N\}} \in \Delta(A)^{X\times Y\times N}$ such that for all $n \in \mathbb{N}^*$ and all $h_n = (x_1, y_1, a_1, b_1, \dots, x_{n-1}, y_{n-1}, a_{n-1}, b_{n-1}, x_n, y_n) \in H_n$, $\boldsymbol{\sigma}_n(h_n) = \boldsymbol{\mu}_{n'}(x_n, y_n)$, where $n' \in \{1, \dots, N\}$ is equal to n modulo N. Markov N-periodic strategies of player 2 are defined likewise.

Stationary strategies are Markov strategies depending only on the current state, hence they are Markov 1-periodic strategies. Again, stationary strategies on the product state space $X\times Y$ (elements of $\Delta(A)^{X\times Y}$ and $\Delta(B)^{X\times Y}$ for player 1 and 2 respectively), are denoted with bold letters. Stationary policies of the players on their components of the state space (elements of $\Delta(A)^X$ and $\Delta(B)^Y$ respectively) are denoted with letters that are not bold.

Finally, a pair of stationary strategies $(\boldsymbol{\mu}, \boldsymbol{\nu}) \in \Delta(A)^{X\times Y} \times \Delta(B)^{X\times Y}$ induces a Markov chain $(X_n, Y_n)_{n\geq 1}$ over $X \times Y$. A state $(x', y') \in X \times Y$ is said to be accessible from (x, y) (in t stages) under $(\boldsymbol{\mu}, \boldsymbol{\nu})$ if $\mathbb{P}^{x,y}_{\boldsymbol{\mu},\boldsymbol{\nu}}((X_t, Y_t) = (x', y')) > 0$. More generally, a property is said to hold under $(\boldsymbol{\mu}, \boldsymbol{\nu})$ if it holds for the Markov chain induced on $X \times Y$. We use similar vocabulary for stationary policies on X and on Y.

3.2.3 N-stage and λ-discounted games

For all $N \in \mathbb{N}^*$, the N-stage game Γ_N starting in $(x, y) \in X \times Y$, is the game in which the payoff is

$$\gamma_N(\boldsymbol{\sigma}, \boldsymbol{\tau})(x, y) = \frac{1}{N}\sum_{n=1}^{N} \mathbb{E}^{x,y}_{\boldsymbol{\sigma},\boldsymbol{\tau}}(u_n),$$

for all $(\boldsymbol{\sigma}, \boldsymbol{\tau}) \in \boldsymbol{\mathcal{S}} \times \boldsymbol{\mathcal{T}}$. The value of the N-stage game starting at (x, y) is denoted $v_N(x, y)$. It is characterized by the following recursive equation:

$$v_{N+1}(x, y) = \operatorname*{val}_{\mu\in\Delta(A),\, \nu\in\Delta(B)} \left[\frac{1}{N+1} u(x, y, \mu, \nu) + \frac{N}{N+1}\mathbb{E}^{x,y}_{\mu,\nu}(v_N)\right], \tag{3.2.1}$$

where

$$u(x, y, \mu, \nu) = \int_{A\times B} u(x, y, a, b) d\mu(a) d\mu(b), \text{ and}$$
$$\mathbb{E}^{x,y}_{\mu,\nu}(v_N) = \sum_{x',y'\in X\times Y} v_N(x', y') \int_{A\times B} p(x'|x, a) q(y'|y, b) d\mu(a) d\nu(b).$$

Moreover, by (Sorin, 2002, proposition 5.3), both players have optimal Markov strategies.

For all $\lambda \in (0, 1]$, the λ-discounted game Γ_λ starting in $(x, y) \in X \times Y$, is the game in which the payoff is

$$\gamma_\lambda(\boldsymbol{\sigma}, \boldsymbol{\tau})(x, y) = \lambda \sum_{n=1}^{+\infty} (1-\lambda)^{n-1} \mathbb{E}^{x,y}_{\boldsymbol{\sigma},\boldsymbol{\tau}}(u_n),$$

for all $(\boldsymbol{\sigma}, \boldsymbol{\tau}) \in \boldsymbol{\mathcal{S}} \times \boldsymbol{\mathcal{T}}$. The value of the λ-discounted game starting at (x, y) is denoted $v_\lambda(x, y)$. It is characterized by the following fixed point equation:

$$v_\lambda(x, y) = \operatorname*{val}_{\mu \in \Delta(A),\, \nu \in \Delta(B)} \left[\lambda \cdot u(x, y, \mu, \nu) + (1 - \lambda)\mathbb{E}^{x,y}_{\mu,\nu}(v_\lambda)\right].$$

If v_n and v_λ converge as n goes to infinity and λ goes to 0, and the limits are equal, then the game is said to have an asymptotic value. The Tauberian theorem of Ziliotto (2016a) applies in this setting and the asymptotic value exists if $(v_\lambda)_{\lambda \in (0,1]}$ converges as λ goes to 0.

3.2.4 Uniform value and optimal strategies

Fix an initial state $(x, y) \in X \times Y$. Player 1 is said to uniformly guarantee $v_\infty \in [0, 1]$ if he has a strategy which guarantees v_∞ (up to ε) against any strategy of player 2, in any game Γ_N, provided that N is large enough. Formally,

$$\forall \varepsilon > 0\ \exists \boldsymbol{\sigma} \in \boldsymbol{\mathcal{S}}\ \exists M \in \mathbb{N}^*\ \forall \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}\ \forall N \geq M\ \gamma_N(\boldsymbol{\sigma}, \boldsymbol{\tau})(x, y) \geq v_\infty - \varepsilon.$$

And similarly for player 2. If both players guarantee v_∞, then it is called the uniform value of the game Γ starting at (x, y).

Let $\varepsilon \geq 0$. A strategy $\boldsymbol{\sigma} \in \boldsymbol{\mathcal{S}}$ is said to be (uniformly) ε-optimal for player 1 if

$$\exists M \in \mathbb{N}^*\ \forall \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}\ \forall N \geq M\ \gamma_N(\boldsymbol{\sigma}, \boldsymbol{\tau})(x, y) \geq v_\infty - \varepsilon.$$

And similarly for player 2.

The next proposition states that if one of the players plays a stationary strategy, then the other player has an ε-optimal best response which is also stationary.

Proposition 3.2. *Let $\boldsymbol{\mu} \in \Delta(A)^{X\times Y}$ be a stationary strategy of player 1 in the game Γ. Then, for all $\varepsilon > 0$ there exist $\boldsymbol{\nu} \in \Delta(B)^{X\times Y}$ and $M \in \mathbb{N}^*$ such that for all $N \geq M$ and all $\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}$ $\gamma_N(\boldsymbol{\mu}, \boldsymbol{\nu})(x, y) \leq \gamma_N(\boldsymbol{\mu}, \boldsymbol{\tau})(x, y) + \varepsilon$.*

Proof. Let $\varepsilon > 0$, $\boldsymbol{\mu} \in \Delta(A)^{X\times Y}$ and $(x, y) \in X \times Y$. When player 1 plays the stationary strategy $\boldsymbol{\mu}$, player 2 faces a Markov decision process having a uniform value w_∞, and for which he has a uniformly ε-optimal stationary strategy $\boldsymbol{\nu}$, consult (Sorin, 2002, corollary 5.26).

Thus, there exists $M \in \mathbb{N}^*$ such that for all $N \geq M$ one has $\gamma_N(\boldsymbol{\mu}, \boldsymbol{\nu})(x, y) \leq w_\infty + \varepsilon$ and for all $\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}$ one has $\gamma_N(\boldsymbol{\mu}, \boldsymbol{\tau})(x, y) \geq w_\infty - \varepsilon$. □

3.3 Main results

In the present article, we examine two communication properties on state spaces of the players. In words, a player has the strong communication property if there exists a time T such that independently of his choice of policy (depending only on his own history), there is a positive probability of moving from any initial state of his component of the state space to any other state in exactly T stages.

Definition 3.3. Player 1 has the strong communication property if there exists $T \in \mathbb{N}^*$ such that for all policies $\sigma \in \mathcal{S}$ and all states $x, x' \in X$, one has $\mathbb{P}^x_\sigma(X_T = x') > 0$.

A similar definition stands for player 2. A game is strongly communicating on one side if at least one player has the strong communication property.

In words, a player has the weak communication property if there exists a time T and a policy (depending only on his own history) such that, for any initial state, he can reach any other state in his component of the state space in exactly T stages, with positive probability.

Definition 3.4. Player 1 has the weak communication property if there exists $T \in \mathbb{N}^*$ and a policy $\sigma \in \mathcal{S}$ such that for all states $x, x' \in X$, one has $\mathbb{P}^x_\sigma(X_T = x') > 0$.

A similar definition stands for player 2. A game is weakly communicating on both sides if both players have the weak communication property. Note that the strong communication property implies the weak one.

The next simple example illustrates the notions of strong and weak communication.

Example 3.1. In Fig. 3.1 below we represent the state space and transition probabilities of player 1. The state space is $X = \{x, y, z\}$. Transition probabilities are represented by arrows between states. In state x player 1 chooses $\alpha \in A$. With probability α the next state is x and with probability $1 - \alpha$ it is y. In state y (resp. z) the next state is y (resp. x) with probability 1.

Hence if the action set A equals $[0, 1]$, player 1 has the weak communication property but does not have the strong communication property. Whereas if A equals $[\varepsilon, 1]$ for some $\varepsilon > 0$, player 1 has the strong communication property.

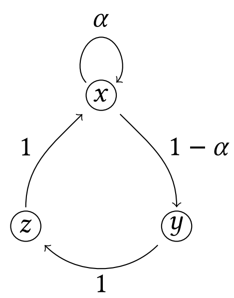

Figure 3.1: Strong and weak communication

We prove in this article the following results.

Theorem 3.5. *Any strongly communicating on one side zero-sum stochastic game has a uniform value.*

Moreover, assuming player 1 has the strong communication property, the uniform value only depends on the initial state of player 2 and for all $\varepsilon > 0$ player 1 has an ε-optimal Markov periodic strategy.

Theorem 3.6. *There exists a weakly communicating on both sides zero-sum product stochastic game which does not admit an asymptotic value.*

Remark that it is sufficient that one player has the strong communication property for the uniform value to exist, while both players having the weak communicating property does not ensure the existence of the asymptotic value.

The proof of Theorem 3.5 relies on a classification of the state space of the player who is not assumed to have the strong communication property. We then consider a family of auxiliary stochastic games and prove that they have a uniform value independent of the initial state. Finally we build an auxiliary MDP for the player who is not assumed to have the strong communication property, whose payoffs are the uniform values of the previous auxiliary games. We conclude by proving that the uniform value of the MDP is also the uniform value of the initial game.

The proof of Theorem 3.6 relies on a counterexample. We show that the Shapley equations of this game are also the Shapley equations of a simpler game with two absorbing and two non absorbing states, for which the asymptotic value does not exist.

3.4 Proof of Theorem 3.5

We assume in the proof of Theorem 3.5 that player 1 has the strong communication property.

In Section 3.4.1 we decompose the state space of player 2 in L maximal communicating sets C_i (Definition 3.7). These sets, forming a partition of Y (Proposition 3.8), allow us in Section 3.4.2 to introduce L auxiliary stochastic games Γ_i having uniform values v^i_∞ which do not depend on the initial state (Proposition 3.12).

In Section 3.4.3 we introduce a Markov Decision Process $\mathcal{G}$ for player 2 with the same transition probability q as in the initial stochastic game, but with payoff v^i_∞ when the state is in C_i. This MDP has a uniform value w_∞, which is also the uniform value of the initial stochastic game Γ.

In Section 3.4.4 we prove that player 2 uniformly guarantees w_∞ in Γ (Proposition 3.13). The idea is for player 2 to first play optimally in the MDP $\mathcal{G}$ (disregarding player 1) and then to switch to an optimal strategy in one of the games Γ_i.

Finally in Section 3.4.5 we prove that player 1 uniformly guarantees w_∞ (Proposition 3.16). This is one of the main difficulties of the proof. The natural idea is for player 1 to play optimally in each game Γ_i. However player 1 does not control the transitions from one C_i to another (which may happen infinitely often). We deal with this issue by letting player 2 play a Markov periodic best response which prevents the state from jumping infinitely often between maximal communicating sets (Lemma 3.15). We conclude showing that the payoff this yields can also be obtained by player 2 as a limit payoff in $\mathcal{G}$ (Proposition 3.16).

3.4.1 Classification of states

The state space Y is classified in a similar way to (Ross and Varadarajan, 1991), using recurrent classes induced by stationary policies on Y.

Definition 3.7. A subset C of Y is said to be a maximal communicating set if

i) There exists a stationary policy on Y such that C is a recurrent class of the induced Markov chain on Y. Such a policy is said to be a stationary policy associated to C.

ii) C is maximal, i.e., if there exists C' a subset of Y such that i) holds for C' and $C \subseteq C'$, then $C' = C$.

Let $C_1, \dots, C_L$ denote the maximal communicating sets. D denotes the set of transient states under every stationary policy (sometimes just called transient states for short).

Remark 3.1. Player 1 having the strong communication property, there is only one maximal communicating set in X, which is the whole state space X itself.

Proposition 3.8 below can be found in (Ross and Varadarajan, 1991) when the action space B is finite, since the proof is similar with a compact action space we omit it. This result, showing that the maximal communicating sets and the set of states that are transient under any stationary policy form partition of Y, is fundamental in our proof of Theorem 3.5 since it allows us (in the next section) to consider independent auxiliary games over each maximal communicating set.

Proposition 3.8. *$\{C_1, \dots, C_L, D\}$ is a partition of Y.*

The two examples given below allow to show how the state space of player 2 decomposes in maximal communicating sets, and what behavior the state process may have with regard to this decomposition. In particular, they show behaviors of the state process that may happen in the compact action setting and cannot happen in the finite one.

Example 3.2. In Fig. 3.2 the state space of player 2 is $Y = \{x, y, z\}$ and the action space is $B = [0, 1/2]$. Again, the transitions probabilities are represented on the arrows between states, e.g., playing $\beta \in B$ in state y, the next state is x with probability β, y with probability $1-\beta-\beta^2$ and z with probability β^2. The maximal communicating sets are $\{y\}$ and $\{z\}$ and the set of transient states under every stationary policy is $\{x\}$.

Remark that, if the initial state is x, playing $\beta = \frac{1}{2n}$ in state y at stage $n \geq 1$, player 2 has a positive probability of switching infinitely often between the

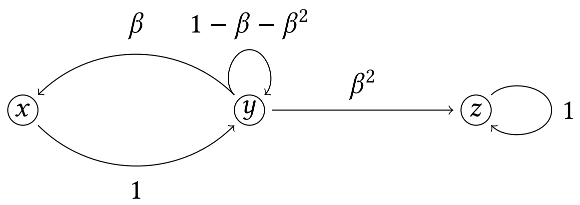

Figure 3.2: States, actions and transitions of player 2

transient state x and the maximal communicating set $\{y\}$. This cannot happen when the action space is finite. In that case, for any policy of player 2, after finitely many stages the process $(Y_n)_{n\geq 1}$ remains forever in one of the maximal communicating sets with probability 1, see (Ross and Varadarajan, 1991, lemma 2 and proposition 2).

The next example is derived from the previous one and shows another possible behavior of the state process in the compact action setting.

Example 3.3. Here, see Fig. 3.3, the state space of player 2 is $Y = \{x, y, z\}$ and the action space is $B = [0, 1/2]$. The maximal communicating sets are $\{x\}$, $\{y\}$ and $\{z\}$.

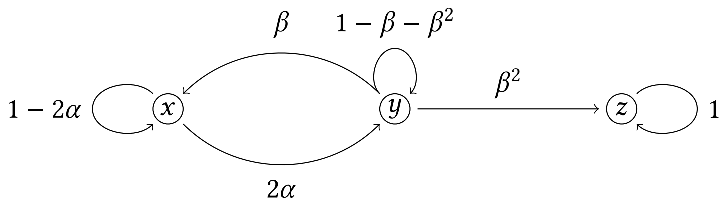

Figure 3.3: States, actions and transitions of player 2

Again, remark that if the initial state is x, playing $\alpha = 1/2$ in state x and $\beta = \frac{1}{2n}$ in state y at stage $n \geq 1$, player 2 has a positive probability of switching infinitely often between maximal communicating sets $\{x\}$ and $\{y\}$.

As we mentioned earlier, this behavior the state of player 2 may have gives rise to a difficulty since player 1 cannot control the jumps of player's 2 state between maximal communicating sets or transient states. This difficulty is defused by letting the players play Markov periodic strategies.

Another way of defining maximal communicating sets is via pairs of stationary strategies of player 1 and 2 on $X \times Y$, as in the following definition. Recall that a pair of stationary strategies $(\boldsymbol{\mu}, \boldsymbol{\nu}) \in \Delta(A)^{X\times Y} \times \Delta(B)^{X\times Y}$ induces a Markov chain over $X \times Y$, which may have recurrent classes or transient states.

Definition 3.9. A subset C of $X \times Y$ is said to be a joint maximal communicating set if

i) There exists a pair of stationary strategies in $\Delta(A)^{X\times Y} \times \Delta(B)^{X\times Y}$ such that C is a recurrent class of the induced Markov chain on $X \times Y$.

ii) C is maximal, i.e., if there exists C' a subset of $X \times Y$ such that i) holds for C' and $C \subseteq C'$, then $C' = C$.

Advantageously, maximal communicating sets and joint maximal communicating sets match in following sense.

Proposition 3.10. *The joint maximal communicating sets are $X \times C_1, \ldots, X \times C_L$, and the set of states that are transient under any pair $(\boldsymbol{\mu}, \boldsymbol{\nu})$ of stationary strategies is $X \times D$.*

Proof. Let $i \in \{1, \ldots, L\}$. $X \times C_i$ is a recurrent class in $X \times Y$ for some pair of stationary policies of player 1 on X and player 2 on Y (just take the stationary policy of player 2 on Y associated to C_i and any stationary policy of player 1 on X).

Let us show that $X \times C_i$ is maximal. Suppose that there exists a subset C of $X \times Y$ such that C is a recurrent class under a pair of stationary strategies $(\boldsymbol{\mu}, \boldsymbol{\nu}) \in \Delta(A)^{X\times Y} \times \Delta(B)^{X\times Y}$ and $X \times C_i \subseteq C$.

Let $C' = \{y \in Y \mid \exists x \in X\ (x, y) \in C\}$ be the projection of C over Y. For all $y \in C'$, define $C_y = \{x \in X \mid (x, y) \in C\}$. Finally, define $\nu \in \Delta(B)^Y$ by, for all $y \in C'$

$$\nu(y) = \frac{1}{|C_y|} \sum_{x \in C_y} \boldsymbol{\nu}(x, y),$$

and arbitrarily outside C'.

C' is closed under ν and every state in C' is accessible from any other state. Hence C' is a recurrent class. Thus $C' \subseteq C_i$, and $C = X \times C_i$.

Clearly $X \times D = X \times Y \setminus \left(\bigcup_{i=1}^{L} X \times C_i \right)$ is the set of transient states. □

3.4.2 The auxiliary games over each maximal communicating set

Consider now a family of L zero-sum product stochastic games $(\Gamma_i)_{i\in\{1,\ldots,L\}}$. For all $i \in \{1, \ldots, L\}$, if $y \in C_i$, define the set of actions of player 2 at state y under which the state has probability 1 of staying in C_i,

$$B_y = \{b \in B \mid q(C_i|y, b) = 1\}.$$

We remark that the set B_y is closed. Then the game Γ_i is given by

$$\Gamma_i = (X, C_i, A, (B_y)_{y\in C_i}, p, q, u).$$

The sets of strategies for player 1 and player 2 in Γ_i are respectively denoted $\mathcal{S}_i$ and $\mathcal{T}_i$. The value of the N-stage game starting at $(x, y) \in X \times C_i$, is denoted $v_N^i(x, y)$.

In Lemma 3.11 and Proposition 3.12 below we prove that each of these games have a uniform value that does not depend on the initial state of the game in $X \times C_i$.

Lemma 3.11. *Let $v^i : X \times C_i \to [0,1]$ be any uniform limit point of the sequence $(v^i_N)_{N\geq 1}$. Then v^i is constant over $X \times C_i$.*

Proof. Let $i \in \{1, \dots, L\}$ and $(x_1^*, y_1^*) \in \arg\max_{X\times C_i} v^i(\cdot)$. Since player 1 has the strong communication property

$$\forall t \geq T \; \forall x' \in X \; \forall \sigma \in \mathcal{S} \; \mathbb{P}_\sigma^{x_1^*}(X_T = x') > 0, \tag{3.4.1}$$

and since C_i is a maximal communicating set

$$\exists \nu \in \Delta(B)^Y \; \forall y' \in C_i \; \exists t \geq T \; \mathbb{P}_\nu^{y_1^*}(Y_t = y') > 0. \tag{3.4.2}$$

We now fix $(x', y') \in X \times C_i$ and let $\nu \in \Delta(B)^Y$ and $t \geq T$ be as in statement (3.4.2). We denote $(x_t^*, y_t^*) = (x', y')$. Then there exist $y_2^*, ..., y_{t-1}^* \in C_i$ such that

$$q(y_t^* | y_{t-1}^*, \nu(y_{t-1}^*)) \dots q(y_2^* | y_1^*, \nu(y_1^*)) > 0.$$

Passing to the limit (uniformly) in the Shapley equation (Eq. (3.2.1)) one has

$$v^i(x_1^*, y_1^*) = \max_{\mu_1 \in \Delta(A)} \min_{\nu_1 \in \Delta(B_{y_1^*})} \sum_{\substack{x_2 \in X \\ y_2 \in C_i}} p(x_2 | x_1^*, \mu_1) q(y_2 | y_1^*, \nu_1) v^i(x_2, y_2).$$

Hence

$$v^i(x_1^*, y_1^*) \leq \max_{\mu_1 \in \Delta(A)} \sum_{\substack{x_2 \in X \\ y_2 \in C_i}} p(x_2 | x_1^*, \mu_1) q(y_2 | y_1^*, \nu(y_1^*)) v^i(x_2, y_2) \leq v^i(x_1^*, y_1^*).$$

Thus there exists $\mu_1^* \in \Delta(A)$ such that

$$v^i(x_1^*, y_1^*) = \sum_{\substack{x_2 \in X \\ y_2 \in C_i}} p(x_2 | x_1^*, \mu_1^*) q(y_2 | y_1^*, \nu(y_1^*)) v^i(x_2, y_2).$$

Finally there exists $\mu_1^* \in \Delta(A)$ such that

$$v^i(x_1^*, y_1^*) = \sum_{x_2 \in X} p(x_2 | x_1^*, \mu_1^*) v^i(x_2, y_2^*).$$

Let $k \in \{2, \dots, t-1\}$ and suppose there exist $\mu_1^* \in \Delta(A)$ and $\mu_2^*, \dots, \mu_{k-1}^* \in \Delta(A)^X$ such that

$$v^i(x_1^*, y_1^*) = \sum_{x_2, \dots, x_{k-1} \in X} p(x_{k-1} | x_{k-2}, \mu_{k-2}^*) \dots p(x_2 | x_1^*, \mu_1^*) v^i(x_{k-1}, y_{k-1}^*).$$

Then there exist $\mu_1^* \in \Delta(A)$ and $\mu_2^*, \ldots, \mu_{k-1}^* \in \Delta(A)^X$ such that

$$\begin{aligned}
v^i(x_1^*, y_1^*) \leq & \sum_{x_2,\ldots,x_{k-1} \in X} p(x_{k-1}|x_{k-2}, \mu_{k-2}^*) \ldots p(x_2|x_1^*, \mu_1^*) \\
& \max_{\mu_{k-1} \in \Delta(A)} \min_{\nu_{k-1} \in \Delta(B_{y_{k-1}^*})} \sum_{\substack{x_k \in X \\ y_k \in C_i}} p(x_k|x_{k-1}, \mu_{k-1}^*) q(y_k|y_{k-1}^*, \nu_{k-1}) v^i(x_k, y_k) \\
\leq & \sum_{x_2,\ldots,x_{k-1} \in X} p(x_{k-1}|x_{k-2}, \mu_{k-2}^*) \ldots p(x_2|x_1^*, \mu_1^*) \\
& \max_{\mu_{k-1} \in \Delta(A)} \sum_{\substack{x_k \in X \\ y_k \in C_i}} p(x_k|x_{k-1}, \mu_{k-1}^*) q(y_k|y_{k-1}^*, \nu(y_{k-1}^*)) v^i(x_k, y_k) \\
\leq & \, v^i(x_1^*, y_1^*).
\end{aligned}$$

Thus there exists $\mu_k^* \in \Delta(A)^X$ such that

$$\begin{aligned}
v^i(x_1^*, y_1^*) = & \sum_{x_2,\ldots,x_{k-1} \in X} p(x_{k-1}|x_{k-2}, \mu_{k-2}^*) \ldots p(x_2|x_1^*, \mu_1^*) \\
& \sum_{\substack{x_k \in X \\ y_k \in C_i}} p(x_k|x_{k-1}, \mu_{k-1}^*) q(y_k|y_{k-1}^*, \nu(y_{k-1}^*)) v^i(x_k, y_k).
\end{aligned}$$

Finally,

$$\begin{aligned}
v^i(x_1^*, y_1^*) = & \sum_{x_2,\ldots,x_{k-1},x_k \in X} p(x_k|x_{k-1}, \mu_{k-1}^*) p(x_{k-1}|x_{k-2}, \mu_{k-2}^*) \ldots \\
& \qquad p(x_2|x_1^*, \mu_1^*) v^i(x_k, y_k^*).
\end{aligned}$$

Thus by induction, there exist $\mu_1^* \in \Delta(A)$ and $\mu_2^*, \ldots, \mu_{t-1}^* \in \Delta(A)^X$ such that for all $x_2, \ldots, x_{t-1} \in X$

$$0 = p(x_t^*|x_{t-1}, \mu_{k-1}^*(x_{t-1})) \ldots p(x_3|x_2, \mu_2^*(x_2)) p(x_2|x_1^*, \mu_1^*)(v^i(x_t^*, y_t^*) - v^i(x_1^*, y_1^*)).$$

Let us define the policy $\sigma \in \mathcal{S}$ by $\sigma_1(x_1^*) = \mu_1^*$ and for all $k \in \{2, \ldots, t-1\}$ $\sigma_k(h_k^1) = \mu_k^*(x_k)$. Finally σ is defined arbitrarily for all $k \geq t$. We have from statement (3.4.1) that there exists $x_2^*, \ldots, x_{t-1}^* \in X$ such that

$$p(x_t^*|x_{t-1}^*, \mu_{t-1}^*) \ldots p(x_2^*|x_1^*, \mu_1^*) > 0.$$

Finally, one has $v^i(x_t^*, y_t^*) = v^i(x_1^*, y_1^*)$. □

Since any uniform limit point of $(v_N^i)_{N \geq 1}$ is constant over $X \times C_i$, we show that players guarantee $\limsup_{N \to +\infty} v_N^i$ by playing an ε-optimal Markov N_0-periodic strategy for a sufficiently large N_0.

Proposition 3.12. *For all $i \in \{1, \ldots, L\}$ the game Γ_i has a uniform value v_∞^i, which is constant over $X \times C_i$.*

Moreover, for all $\varepsilon > 0$ there exists $N_0 \in \mathbb{N}^*$ such that both players have an ε-optimal Markov N_0-periodic strategy in each Γ_i.

Proof. Here we take $v^i = \limsup_{N\to+\infty} v_N^i$. Let $\varepsilon > 0$. Thanks to Lemma 3.11, there exists $N_0^i \geq 1$ such that for all $(x,y) \in X \times C_i$ one has $|v^i_{N_0^i}(x,y) - v^i| \leq \varepsilon$. Hence

$$\max_{\sigma\in\mathcal{S}_i}\min_{\tau\in\mathcal{T}_i} \mathbb{E}^{x,y}_{\sigma,\tau}\left[\sum_{n=1}^{N_0^i} u_n\right] \geq N_0^i(v^i - \varepsilon).$$

Recall that there exists a Markov strategy $\boldsymbol{\sigma}^i = (\sigma^i_n)_{n\in[1,N_0^i]}$, independent of the initial state (x,y), such that for all $\tau \in \mathcal{T}_i$

$$\mathbb{E}^{x,y}_{\sigma^i,\tau}\left[\sum_{n=1}^{N_0^i} u_n\right] \geq N_0^i(v^i - \varepsilon).$$

Let $N = pN_0^i + r$, $p \geq 1$ and $r \leq N_0^i - 1$ be two integers. The following Markov N_0-periodic strategy is still denoted $\boldsymbol{\sigma}^i$: at stage $n \geq 1$, in state $(x,y) \in X \times C_i$, play $\sigma^i_{n'}(x,y)$, where n' equals n modulo N_0^i. Then for all $\boldsymbol{\tau} \in \mathcal{T}_i$

$$\begin{aligned}\mathbb{E}^{x,y}_{\sigma^i,\tau}\left[\sum_{n=1}^{N} u_n\right] &= \sum_{k=0}^{p-1}\mathbb{E}^{x,y}_{\sigma^i,\tau}\left[\sum_{n=kN_0^i+1}^{(k+1)N_0^i} u_n\right] + \mathbb{E}^{x,y}_{\sigma^i,\tau}\left[\sum_{n=pN_0^i+1}^{N} u_n\right]\\ &\geq pN_0^i(v^i - \varepsilon).\end{aligned}$$

Finally, for all $N \geq \frac{N_0^i}{\varepsilon}$,

$$v_N^i(x,y) \geq \frac{pN_0^i(v^i-\varepsilon)}{N} \geq (1-\varepsilon)(v^i - \varepsilon).$$

That is, player 1 uniformly guarantees v^i and $\boldsymbol{\sigma}^i$ is an ε-optimal Markov N_0^i-periodic strategy. A similar proof shows that player 2 also uniformly guarantees v^i and has an ε-optimal Markov N_0^i-periodic strategy τ^i.

Note that N_0^i can be taken uniformly over $\{1,\ldots,L\}$ by setting $N_0 = \prod_{i=1}^L N_0^i$. □

3.4.3 The auxiliary Markov decision process $\mathcal{G}$

Consider the Markov decision process $\mathcal{G} = (Y, B, q, g)$, in which player 2 is the only decision maker and his aim is to minimize

$$\begin{aligned} g \;:\; Y &\to [0,1]\\ y &\mapsto \begin{cases} v^i_\infty \text{ if there exists } i \in \{1,\ldots,L\} \text{ such that } y \in C_i\\ 1/2 \text{ if } y \in D.\end{cases}\end{aligned}$$

Recall, see (Sorin, 2002), that, $\mathcal{G}$ has a uniform value $w_\infty \in [0,1]^Y$ and that for every $\varepsilon > 0$, player 2 has an ε-optimal stationary policy.

The interpretation is the following. The objective of player 2 is to reach the maximal communicating set C_i with corresponding auxiliary game Γ_i having the lowest uniform value v^i_∞ possible, and stay in C_i. Note that the payoff of $1/2$ in D is arbitrary and does not change the value of w_∞.

We will prove that the uniform value w_∞ of the MDP $\mathcal{G}$ is in fact also the uniform value of the initial game Γ.

3.4.4 Player 2 uniformly guarantees w_∞ in Γ

Proposition 3.13. *Player 2 uniformly guarantees w_∞ in Γ.*

To prove Proposition 3.13, we show that player 2 has an ε-optimal strategy which has a rather simple structure: play according to some ε-optimal stationary policy in $\mathcal{G}$ on Y, until reaching a recurrent class of the induced Markov chain on Y. Then switch to an ε-optimal Markov periodic strategy in some game Γ_i having the property that the state remains in the corresponding maximal communicating set C_i (indeed the ε-optimal strategies used in the proof of Proposition 3.13 below can be taken Markov periodic thanks to Proposition 3.12).

Proof. Let $\varepsilon > 0$, and let $\nu_{\mathcal{G}}$ be an ε-optimal stationary policy of player 2 in $\mathcal{G}$. $\nu_{\mathcal{G}}$ induces l recurrent classes $R_1, \dots, R_l$ over Y. Moreover, by Definition 3.7, there exists a mapping $\varphi : \{1, \dots, l\} \to \{1, \dots, l\}$ such that for all $i \in \{1, \dots, l\}$, $R_i \subseteq C_{\varphi(i)}$.

For all $i \in \{1, \dots, L\}$ define $T_i = \min\{n \in \mathbb{N}^* \mid Y_n \in R_i\}$ to be the hitting time of R_i by $(Y_n)_{n\in\mathbb{N}^*}$. The minimum over an empty set is taken equal to $+\infty$.

We now consider the game Γ with initial state $(x, y) \in X \times Y$. Let us define the strategy $\overline{\boldsymbol{\tau}} \in \boldsymbol{\mathcal{T}}$ of player 2 as follows. Until there exists $i \in \{1, \dots, l\}$ such that the state of player 2 is in R_i play $\nu_{\mathcal{G}}$. Let $n \in \mathbb{N}^*$ be the first stage at which the state of player 2 reaches one of the recurrent classes R_i, $i \in \{1, \dots, l\}$. From stage n on, play $\boldsymbol{\tau}^{\varphi(i)}$ which is an ε-optimal strategy in $\Gamma_{\varphi(i)}$.

Let $\boldsymbol{\sigma}$ be any strategy of player 1 in Γ. Remark that under $(\boldsymbol{\sigma}, \overline{\boldsymbol{\tau}})$, the laws of the T_i's are the same as under $\nu_{\mathcal{G}}$.

Let $N \geq 2$,

$$\frac{1}{N}\mathbb{E}^{x,y}_{\sigma,\overline{\tau}}\left[\sum_{n=1}^{N} u_n\right] = \frac{1}{N}\sum_{i=1}^{l}\mathbb{E}^{x,y}_{\sigma,\overline{\tau}}\left[\left(\sum_{n=1}^{T_i} u_n\right)\mathbb{1}\left(\sqrt{N} > T_i = \min_{k\in\{1,\dots,l\}} T_k\right)\right] \tag{3.4.3}$$

$$+\frac{1}{N}\sum_{i=1}^{l}\mathbb{E}^{x,y}_{\sigma,\overline{\tau}}\left[\left(\sum_{n=T_i+1}^{N} u_n\right)\mathbb{1}\left(\sqrt{N} > T_i = \min_{k\in\{1,\dots,l\}} T_k\right)\right] \tag{3.4.4}$$

$$+\frac{1}{N}\sum_{i=1}^{l}\mathbb{E}^{x,y}_{\sigma,\overline{\tau}}\left[\left(\sum_{n=1}^{N} u_n\right)\mathbb{1}\left(\sqrt{N} \leq T_i = \min_{k\in\{1,\dots,l\}} T_k\right)\right]. \tag{3.4.5}$$

Since the payoffs u_n are at most 1 and T_i is at most $\sqrt{N}$ in the indicator function, (3.4.3) is at most

$$\frac{\sqrt{N}}{N}\sum_{i=1}^{l}\mathbb{P}^{x,y}_{\sigma,\overline{\tau}}\left(\sqrt{N} > T_i = \min_{k\in\{1,\dots,l\}} T_k\right),$$

which itself is at most $l\frac{\sqrt{N}}{N}$ which is smaller than ε for N large enough.

Remark that in (3.4.4), $\frac{1}{N} \leq \frac{1}{N-T_i}$, hence this quantity is less than or equal to

$$\sum_{i=1}^{l}\mathbb{E}^{x,y}_{\sigma,\overline{\tau}}\left[\frac{1}{N-T_i}\left(\sum_{n=T_i+1}^{N} u_n\right)\mathbb{1}\left(\sqrt{N} > T_i = \min_{k\in\{1,\dots,l\}} T_k\right)\right],$$

the latter equals

$$\sum_{i=1}^{l}\mathbb{E}^{x,y}_{\sigma,\overline{\tau}}\left[\frac{1}{N-T_i}\sum_{n=T_i+1}^{N} u_n \middle| \sqrt{N} > T_i = \min_{k\in\{1,\dots,l\}} T_k\right]\mathbb{P}^{x,y}_{\sigma,\overline{\tau}}\left(\sqrt{N} > T_i = \min_{k\in\{1,\dots,l\}} T_k\right).$$

Recall that for all $i \in \{1,\dots,L\}$, $\boldsymbol{\tau}^i$ is ε-optimal in Γ_i which has uniform value v^i_∞. For N large enough the quantity above is less than or equal to

$$\sum_{i=1}^{l}\left(v^{\varphi(i)}_\infty + \varepsilon\right)\mathbb{P}^{y}_{\nu_{\mathcal{G}}}\left(T_i = \min_{k\in\{1,\dots,l\}} T_k\right),$$

because under $(\boldsymbol{\sigma},\overline{\boldsymbol{\tau}})$, the laws of the T_i's are the same as under $\nu_{\mathcal{G}}$. And since $\nu_{\mathcal{G}}$ is ε-optimal in $\mathcal{G}$, $\sum_{i=1}^{l} v^{\varphi(i)}_\infty \mathbb{P}^{y}_{\nu_{\mathcal{G}}}\left(T_i = \min_{k\in\{1,\dots,l\}} T_k\right)$ is less than or equal to $w_\infty(y) + \varepsilon$.

Finally, since the payoffs u_n are at most 1, (3.4.5) is less than or equal to

$$\sum_{i=1}^{l}\mathbb{P}^{x,y}_{\sigma,\overline{\tau}}\left(\sqrt{N} \leq T_i = \min_{k\in\{1,\dots,l\}} T_k\right),$$

which is at most

$$l\mathbb{P}^{y}_{\nu_{\mathcal{G}}}\left(\sqrt{N} \leq \min_{k\in\{1,\dots,l\}} T_k\right),$$

which itself is less than ε for N large enough. □

3.4.5 Player 1 uniformly guarantees w_∞ in Γ

The auxiliary games $(\widetilde{\Gamma}_i)_{i\in\{1,\dots,L\}}$

Let us fix $\varepsilon > 0$ and $N_0 \in \mathbb{N}^*$ accordingly as in Proposition 3.12. Beware that the objects that we now introduce also depend on ε.

We construct L auxiliary games $(\widetilde{\Gamma}_i)_{i\in\{1,\dots,L\}}$ which are copies of the games $(\Gamma_i)_{i\in\{1,\dots,L\}}$ with an additional clock keeping track of the time modulo N_0. The state of player 2 on $C_i \times [1, N_0]$, where $[1, N_0] = \{1,\dots,N_0\}$, moves on C_i as in

Γ_i and on $[1, N_0]$ by adding 1 at each stage (and starting back to 1 when $N_0 + 1$ is reached). For all $i \in \{1, \dots, L\}$,

$$\widetilde{\Gamma}_i = (X, C_i \times [1, N_0], A, (B_y)_{y \in C_i}, p, \tilde{q}, u),$$

and

$$\begin{array}{rccl} \tilde{q} & : & C_i \times [1, N_0] \times (B_y)_{y \in C_i} & \rightarrow & \Delta(C_i \times [1, N_0]) \\ & & (y, t, b) & \mapsto & q(\cdot|y, b) \otimes \delta_{t+1}, \end{array}$$

where t is taken modulo N_0 in δ_{t+1}.

The purpose of these games is that they have the same uniform value as the Γ_i's, but ε-optimal stationary strategies instead of ε-optimal Markov N_0-periodic strategies. Indeed, by Proposition 3.12, let $\boldsymbol{\sigma}^i$ be an ε-optimal Markov N_0-periodic strategy in Γ_i. Let $\tilde{\boldsymbol{\mu}}^i$ be the following stationary strategy in $\widetilde{\Gamma}_i$: for all $(x, y, t) \in X \times C_i \times [1, N_0]$, $\tilde{\boldsymbol{\mu}}^i(x, y, t) = \boldsymbol{\sigma}^i_t(x, y)$. This defines a stationary strategy of player 1 in $\widetilde{\Gamma}_i$ which is ε-optimal.

Remark 3.2. Since we are dealing with Markov periodic strategies, one could think of defining maximal communicating sets with regard to this class of strategies rather than stationary strategies as in Definition 3.7. However the construction we propose here appears to provide a simpler demonstration of player 1 uniformly guaranteeing w_∞ in Γ.

The auxiliary game $\widetilde{\Gamma}$

We now gather together the games $\widetilde{\Gamma}_i$ into one game $\widetilde{\Gamma}$. Let $\widetilde{Y} = Y \times [1, N_0]$. The game $\widetilde{\Gamma}$ is defined by $\widetilde{\Gamma} = \left(X, \widetilde{Y}, A, B, p, \tilde{q}, u\right)$. Where $\tilde{q}$ is extended on $\widetilde{Y} \times B$ as follows.

$$\begin{array}{rccl} \tilde{q} & : & \widetilde{Y} \times B & \rightarrow & \Delta(\widetilde{Y}) \\ & & (y, t, b) & \mapsto & \tilde{q}(\cdot|y, t, b), \end{array}$$

where $\tilde{q}(y', t+1|y, t, b) = q(y'|y, b)$ if there exists $i \in \{1, \dots, L\}$ such that $y, y' \in C_i$, and t is taken modulo N_0. $\tilde{q}(y', 1|y, t, b) = q(y'|y, b)$ if there exists $i \in \{1, \dots, L\}$ such that $y \in C_i$ and $y' \notin C_i$, or $y \in D$. Otherwise $\tilde{q}(y', t'|y, t, b) = 0$. In words, the clock on $[1, N_0]$ is incremented as in the games $\widetilde{\Gamma}_i$ as long as the state stays in C_i, and is reset to 1 when the state jumps from one C_i to another or to a transient state, or from a transient state to any other state.

Let $\widetilde{\boldsymbol{\mathcal{S}}}$ and $\widetilde{\boldsymbol{\mathcal{T}}}$ be the set of strategies of player 1 and 2 respectively in $\widetilde{\Gamma}$. It is important to note that any quantity guaranteed by a player in $\widetilde{\Gamma}$ is also guaranteed in Γ.

The following lemma states that the Y component of a recurrent class in $X \times \widetilde{Y}$ cannot have nonempty intersection with two different maximal communicating sets in Y.

Lemma 3.14. *Let $(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\nu}})$ be a pair of stationary strategies on $X \times \widetilde{Y}$. Let $R \subseteq X \times \widetilde{Y}$ be a recurrent class under $(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\nu}})$. Then there exists $i \in \{1, \ldots, L\}$ such that $R \subseteq X \times C_i \times [1, N_0]$.*

Proof. Let $R' = \{(x, y) \in X \times Y \mid \exists t \in [1, N_0]\ (x, y, t) \in R\}$ be the projection of R over $X \times Y$. For all $(x, y) \in R'$ let $R_{x,y} = \{t \in [1, N_0] \mid (x, y, t) \in R\}$.

Define the stationary strategies $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ on $X \times Y$ by, for all $(x, y) \in R'$,

$$\boldsymbol{\mu}(x, y) = \frac{1}{|R_{x,y}|} \sum_{t \in R_{x,y}} \tilde{\boldsymbol{\mu}}(x, y, t) \text{ and } \boldsymbol{\nu}(x, y) = \frac{1}{|R_{x,y}|} \sum_{t \in R_{x,y}} \tilde{\boldsymbol{\nu}}(x, y, t),$$

and arbitrarily outside R'. Under $(\boldsymbol{\mu}, \boldsymbol{\nu})$, R' is a recurrent class. Hence by Proposition 3.10 there exists $i \in \{1, \ldots, L\}$ such that $R' \subseteq X \times C_i$. □

For all $i \in \{1, \ldots, L\}$, let $\tilde{\boldsymbol{\mu}}^i$ be an ε-optimal stationary strategy of player 1 in $\widetilde{\Gamma}_i$. Define the stationary strategy $\widetilde{\boldsymbol{\mu}}$ of player 1 in $\widetilde{\Gamma}$ as follows. For all $(x, y, t) \in X \times \widetilde{Y}$

$$\tilde{\boldsymbol{\mu}}(x, y, t) = \begin{cases} \tilde{\boldsymbol{\mu}}^i(x, y, t) \text{ if there exists } i \in \{1, \ldots, L\} \text{ such that } y \in C_i \\ \text{arbitrary fixed action if } y \in D. \end{cases}$$

In the game $\widetilde{\Gamma}$, by Proposition 3.2, player 2 has a stationary strategy $\tilde{\boldsymbol{\nu}} \in \Delta(B)^{X \times \widetilde{Y}}$ and there exists $M \in \mathbb{N}^*$ such that for all $N \geq M$ and all $\boldsymbol{\tau} \in \widetilde{\mathcal{T}}$

$$\gamma_N(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\nu}})(x, y, t) \leq \gamma_N(\tilde{\boldsymbol{\mu}}, \boldsymbol{\tau})(x, y, t) + \varepsilon.$$

The pair $(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\nu}})$ induces a Markov chain on $X \times \widetilde{Y}$ with recurrent classes $R_1, \ldots, R_m$. By Lemma 3.14 there exists a mapping $\psi : \{1, \ldots, m\} \to \{1, \ldots, L\}$ such that for all $i \in \{1, \ldots, m\}$

$$R_i \subseteq X \times C_{\psi(i)} \times [1, N_0].$$

For all $i \in \{1, \ldots, m\}$ we define $\widetilde{T}_i = \min\{n \in \mathbb{N}^* \mid (X_n, Y_n, t_n) \in R_i\}$ to be the hitting time of R_i by $(X_n, Y_n, t_n)_{n \geq 1}$.

Lemma 3.15. *In the game $\widetilde{\Gamma}$ starting at $(x, y, 1) \in X \times \widetilde{Y}$, player 1 uniformly guarantees*

$$\sum_{i=1}^m v_\infty^{\psi(i)} \mathbb{P}_{\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\nu}}}^{x,y,1} \left(\widetilde{T}_i = \min_{k \in \{1, \ldots, m\}} \widetilde{T}_k \right) - 3\varepsilon.$$

Proof. Let $N \geq 2$ and $\varepsilon' > 0$,

$$\frac{1}{N} \mathbb{E}_{\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\nu}}}^{x,y,1} \left[\sum_{n=1}^N u_n \right] \geq \sum_{i=1}^m \frac{1}{N} \mathbb{E}_{\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\nu}}}^{x,y,1} \left[\left(\sum_{n=\widetilde{T}_i+1}^N u_n \right) \mathbb{1} \left(\sqrt{N} > \widetilde{T}_i = \min_{k \in \{1, \ldots, m\}} \widetilde{T}_k \right) \right].$$

This last quantity is equal to

$$\sum_{i=1}^{m} \mathbb{E}_{\tilde{\mu},\tilde{\nu}}^{x,y,1}\left[\frac{N-\widetilde{T}_i}{N}\frac{1}{N-\widetilde{T}_i}\left(\sum_{n=\widetilde{T}_i+1}^{N} u_n\right)\mathbb{1}\left(\sqrt{N}>\widetilde{T}_i=\min_{k\in\{1,\dots,m\}}\widetilde{T}_k\right)\right].$$

Since $\widetilde{T}_i$ is taken less than $\sqrt{N}$ in the indicator function, the latter is greater than

$$\sum_{i=1}^{m}\left(1-\frac{\sqrt{N}}{N}\right)\mathbb{E}_{\tilde{\mu},\tilde{\nu}}^{x,y,1}\left[\frac{1}{N-\widetilde{T}_i}\sum_{n=\widetilde{T}_i+1}^{N} u_n \,\middle|\, \sqrt{N}>\widetilde{T}_i=\min_{k\in\{1,\dots,m\}}\widetilde{T}_k\right]$$
$$\times\, \mathbb{P}_{\tilde{\mu},\tilde{\nu}}^{x,y,1}\left(\sqrt{N}>\widetilde{T}_i=\min_{k\in\{1,\dots,m\}}\widetilde{T}_k\right).$$

Recall that for all $i\in\{1,\dots,L\}$, $\tilde{\boldsymbol{\mu}}^i$ is ε-optimal in $\widetilde{\Gamma}_i$, which has value v_∞^i. Moreover, conditionally on $\widetilde{T}_i=\min_{k\in\{1,\dots,m\}}\widetilde{T}_k$, from $\widetilde{T}_i+1$ on, player 2 only plays actions that are in $(B_y)_{y\in R_i}$, otherwise the state process on Y would have a positive probability of leaving $C_{\psi(i)}$ and R_i would not be a recurrent class. Hence for N large enough the quantity above is greater than or equal to

$$\sum_{i=1}^{m}\left(1-\frac{\sqrt{N}}{N}\right)\left(v_\infty^{\psi(i)}-\varepsilon\right)\mathbb{P}_{\tilde{\mu},\tilde{\nu}}^{x,y,1}\left(\sqrt{N}>\widetilde{T}_i=\min_{k\in\{1,\dots,m\}}\widetilde{T}_k\right),$$

which, for N large enough is greater than

$$(1-\varepsilon')\sum_{i=1}^{m}\left(v_\infty^{\psi(i)}-\varepsilon\right)\left(\mathbb{P}_{\tilde{\mu},\tilde{\nu}}^{x,y,1}\left(\widetilde{T}_i=\min_{k\in\{1,\dots,m\}}\widetilde{T}_k\right)-\varepsilon'\right).$$

Hence there exists $\widetilde{N}\in\mathbb{N}^*$ such that for all $N\geq\widetilde{N}$ and all $\boldsymbol{\tau}\in\widetilde{\mathcal{T}}$,

$$\gamma_N(\tilde{\boldsymbol{\mu}},\boldsymbol{\tau})(x,y,1)+\varepsilon\geq(1-\varepsilon')\sum_{i=1}^{m}v_\infty^{\psi(i)}\left(\mathbb{P}_{\tilde{\mu},\tilde{\nu}}^{x,y,1}\left(\widetilde{T}_i=\min_{k\in\{1,\dots,m\}}\widetilde{T}_k\right)-\varepsilon'\right)$$
$$-\,\varepsilon(1-\varepsilon').$$

Finally, for ε' small enough, the right-hand side is greater than

$$\sum_{i=1}^{m}v_\infty^{\psi(i)}\mathbb{P}_{\tilde{\mu},\tilde{\nu}}^{x,y,1}\left(\widetilde{T}_i=\min_{k\in\{1,\dots,m\}}\widetilde{T}_k\right)-2\varepsilon,$$

which concludes the proof. □

To conclude, we prove that the payoff uniformly guaranteed (up to 3ε) by player 1 in $\widetilde{\Gamma}$ in Lemma 3.15 can be obtained by player 2 as a limit payoff in the MDP $\mathcal{G}$, and hence is greater than w_∞, the uniform value of $\mathcal{G}$.

Proposition 3.16. *Player 1 uniformly guarantees w_∞ in Γ.*

Proof. By Lemma 3.15, in the game $\widetilde{\Gamma}$ starting at $(x, y, 1) \in X \times \widetilde{Y}$ player 1 uniformly guarantees

$$\sum_{i=1}^{m} v_\infty^{\psi(i)} \mathbb{P}_{\tilde{\mu},\tilde{\nu}}^{x,y,1}\left(\widetilde{T}_i = \min_{k \in \{1,\dots,m\}} \widetilde{T}_k\right) - 3\varepsilon.$$

In the MDP $\mathcal{G}$, let $\tau_{\mathcal{G}}$ be the following policy of player 2. At stage 1, play $\tilde{\boldsymbol{\nu}}(x, y, 1)$, a new state $(y_2, t_2) \in \widetilde{Y}$ is selected according to $\tilde{q}(\cdot|y, 1, \tilde{\boldsymbol{\nu}}(x, y, 1))$. At stage 2, play $\tilde{\boldsymbol{\nu}}(x_2, y_2, t_2)$ where $x_2 \in X$ is selected according to $p(\cdot|x, \tilde{\boldsymbol{\mu}}(x, y, 1))$. A new state $(y_3, t_3) \in \widetilde{Y}$ is selected according to $\tilde{q}(\cdot|y_2, t_2, \tilde{\boldsymbol{\nu}}(x_2, y_2, t_2))$. Inductively at stage $n > 2$, play $\tilde{\boldsymbol{\nu}}(x_n, y_n, t_n)$ where $x_n \in X$ is selected according to $p(\cdot|x_{n-1}, \tilde{\boldsymbol{\mu}}(x_{n-1}, y_{n-1}, t_{n-1}))$.

$\tau_{\mathcal{G}}$ has the following interpretation. Player 2 plays in $\mathcal{G}$ according to $\tilde{\boldsymbol{\nu}} \in \Delta(B)^{X \times \widetilde{Y}}$ by simulating at each stage a fictitious state of player 1 on X that follows $\tilde{\boldsymbol{\mu}} \in \Delta(B)^{X \times \widetilde{Y}}$. This induces a fictitious hitting times $\widetilde{T}_i$'s of the recurrent classes $R_1, \dots, R_m$.

Under $\tau_{\mathcal{G}}$ the laws of the $\widetilde{T}_i$'s are the same as under $(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\nu}})$. After the process (X_n, Y_n, t_n) has reached R_i, the payoff to player 2 in $\mathcal{G}$ is v_∞^i at each stage.

Therefore, $\tau_{\mathcal{G}}$ yields in $\mathcal{G}$ a limit payoff of

$$\sum_{i=1}^{m} v_\infty^{\psi(i)} \mathbb{P}_{\tilde{\mu},\tilde{\nu}}^{x,y,1}\left(\widetilde{T}_i = \min_{k \in \{1,\dots,m\}} \widetilde{T}_k\right).$$

Hence this quantity is greater than $w_\infty(y)$.

Since in the initial game Γ starting at (x, y), player 1 also uniformly guarantees

$$\sum_{i=1}^{m} v_\infty^{\psi(i)} \mathbb{P}_{\tilde{\mu},\tilde{\nu}}^{x,y,1}\left(\widetilde{T}_i = \min_{k \in \{1,\dots,m\}} \widetilde{T}_k\right) - 3\varepsilon,$$

he uniformly guarantees $w_\infty(y) - 3\varepsilon$.

This is true for all $\varepsilon > 0$. Hence player 1 uniformly guarantees w_∞ in Γ. ◻

3.5 Proof of Theorem 3.6

In Section 3.5.1 we present the weakly communicating on both sides zero-sum product stochastic game used to prove Theorem 3.6. The example proposed shows that in weakly communicating on both sides games, even if players can go from any state to any other state of their component in finite time, they can make mistakes that are irreversible with regards to the joint state. Section 3.5.2 is dedicated to the simplification of the Shapley equations obtained for the weakly communicating on both sides game (Proposition 3.18). Finally in Section 3.5.3 we show that the Shapley equations obtained after simplification are also those of a simple game of perfect information, i.e., in each state only one player controls the transition, with two absorbing and two non absorbing states. The latter game does not have an asymptotic value (Lemmas 3.19–3.22).

3.5.1 A counterexample

State spaces and action sets

The state space of player 1 is $X = \{x, y\} \times C_8$ where $C_8 = \mathbb{Z}/8\mathbb{Z}$. The state space of player 2 is $Y = \{x', y'\} \times C'_8$ where $C'_8 = \mathbb{Z}/8\mathbb{Z}$.

Let $I = \{0\} \cup \left\{1/2^{2n} \mid n \geq 1\right\}$, and $J = [0, 1/4]$. Let $A = I \times \{-1, +1\} \cup \{0, 1\} \times \{0\}$ and $B = J \times \{-1, +1\} \cup \{0, 1\} \times \{0\}$ be the action sets of player 1 and 2 respectively.

It is essential that I, and therefore A, are not semi-algebraic. Indeed, since X and Y are finite, and the transitions we define below are polynomial, if A and B were definable in some o-minimal structure, the game would have a uniform value, see (Bolte et al., 2014, Theorem 4).

Transitions

For $i \in \{x, y\}$ we denote by $-i$ the element of $\{x, y\} \backslash \{i\}$.

In state $(i, k) \in X$ if player 1 plays $(\alpha, p) \in I \times \{-1, +1\}$ then with probability $1 - \alpha - \alpha^2$ the new state is $(i, k + p)$, with probability α the new state is $(-i, k + p)$, and with probability α^2 the new state is $(i, k - p)$ (see Fig. 3.4).

Still in state $(i, k) \in X$, if player 1 plays $(\alpha, 0) \in \{0, 1\} \times \{0\}$, then with probability $1 - \alpha$ the state remains in (i, k) and with probability α the new state is $(-i, k)$ (see Fig. 3.5).

Transitions for player 2 are analogous on Y.

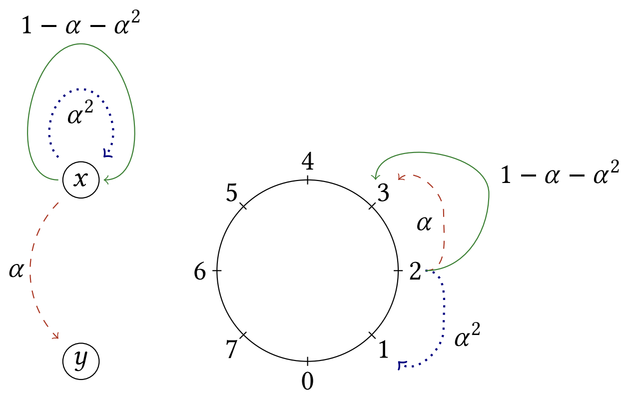

Figure 3.4: Transition of player 1 when playing $(\alpha, +1)$, $\alpha \in I$ in state $(x, 2)$

Hence by playing $(0, p)$, player 1 totally controls the dynamics on C_8, and by playing $(\alpha, 0)$ with α equal to 0 or 1 he totally controls the dynamics on $\{x, y\}$. And likewise for player 2.

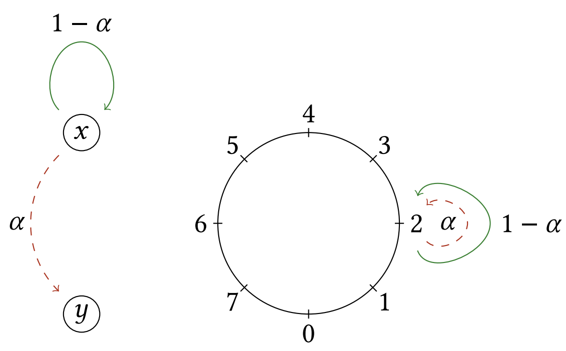

Figure 3.5: Transition of player 1 when playing $(\alpha, 0)$, $\alpha \in \{0,1\}$ in state $(x, 2)$

Payoffs

Let $(i,k) \in X$ and $(i',k') \in Y$. We denote by $d_{C_8}(k,k')$ the distance between player 1 and 2 on the circle C_8.

The payoff function u is defined as follows. If $d_{C_8}(k,k') \geq 3$ then $u((i,k),(i',k')) = 1$. If $d_{C_8}(k,k') \leq 1$ then $u((i,k),(i',k')) = 0$. Otherwise, if $d_{C_8}(k,k') = 2$ then u is defined by the following table:

$u(\cdot,\cdot)$	x'	y'
x	0	1
y	1	0

The interpretation of the game is the following. Player 1 wants to maximize his distance to player 2 who wants to minimize his distance to player 1. If the distance between them on the circle is at most 1 or at least 3, then their positions in $\{x,y\}$ and $\{x',y'\}$ do not matter. Whereas if the distance between them on the circle is equal to 2, then player 1 wants to be in x (resp. y) when player 2 is in y' (resp. x').

Note also that if the distance between the players is at least 3 (resp. at most 1), then player 1 (resp. player 2) can play such that the distance is always at least 3 (resp. at most 1), and this is optimal for him. Hence those joint states on $X \times Y$ act as absorbing states with payoff 1 and 0 respectively.

Shapley equations

For $p \in \{-1,0,+1\}$ we denote as well p by the triplet $(p_{-1},p_0,p_{+1}) \in \{0,1\}^3$ with $p_{-1} = 1$ if $p = -1$ and $p_{-1} = 0$ otherwise, and likewise for p_0 and p_{+1}. We define similarly $q = (q_{-1},q_0,q_{+1})$.

Let $(i,k) \in X$ and $(i',k') \in Y$ be the initial states of player 1 and 2. By symmetry of the game, we only consider the cases $(i,k) \in \{(x,0),(x,1),(x,2),(x,3),(x,4)\}$ and $(i',k') \in \{(x',0),(y',0)\}$. Moreover, the joint state $((i,k),(i',0))$ is denoted (i,i',k).

Let $\lambda \in (0,1)$. Let $x_\lambda = v_\lambda(x, y', 2)$ and $y_\lambda = v_\lambda(x, x', 2)$. Clearly $v_\lambda(\cdot,\cdot,3) = v_\lambda(\cdot,\cdot,4) = 1$ and $v_\lambda(\cdot,\cdot,0) = v_\lambda(\cdot,\cdot,1) = 0$. x_λ is the value of the game, played with mixed strategies, where player 1 chooses $(\alpha, p) \in A$ and player 2 chooses $(\beta, q) \in B$ with payoff $\lambda + (1-\lambda)h(x_\lambda, y_\lambda, \alpha, p, \beta, q)$ where $h(x_\lambda, y_\lambda, \alpha, p, \beta, q)$ equals

$$\begin{aligned}
&\Big[\left((1-\alpha-\alpha^2)(1-\beta-\beta^2) + \alpha\beta + \alpha^2\beta^2\right)(p_{-1}q_{-1} + p_{+1}q_{+1}) \\
&+ \left((1-\alpha)(1-\beta) + \alpha\beta\right) p_0 q_0 \\
&+ \left((1-\alpha-\alpha^2)\beta^2 + (1-\beta-\beta^2)\alpha^2\right)(p_{-1}q_{+1} + p_{+1}q_{-1}) \Big] x_\lambda \\
&+ \Big[\left((1-\alpha-\alpha^2)\beta + (1-\beta-\beta^2)\alpha\right)(p_{-1}q_{-1} + p_{+1}q_{+1}) \\
&+ \left((1-\alpha)\beta + (1-\beta)\alpha\right) p_0 q_0 + (\alpha\beta^2 + \beta\alpha^2)(p_{-1}q_{+1} + p_{+1}q_{-1}) \Big] y_\lambda \\
&+ \alpha^2(1-\beta^2)p_{-1}q_{-1} + \beta^2(1-\alpha^2)p_{+1}q_{+1} + (1-\alpha^2)p_{+1}q_0 + \alpha^2 p_{-1}q_0 \\
&+ \beta^2 p_0 q_{+1} + (1-\beta^2)p_0 q_{-1} + (1-\alpha^2)(1-\beta^2)p_{+1}q_{-1} + \alpha^2\beta^2 p_{-1}q_{+1}.
\end{aligned}$$

y_λ is the value of the game, played with mixed strategies, where player 1 chooses $(\alpha, p) \in A$ and player 2 chooses $(\beta, q) \in B$ with payoff $(1-\lambda)h(y_\lambda, x_\lambda, \alpha, p, \beta, q)$.

3.5.2 Simplification of $(x_\lambda)_{\lambda\in(0,1)}$ and $(y_\lambda)_{\lambda\in(0,1)}$

The aim of this section is to simplify the expressions of x_λ and y_λ obtained in the previous section.

Let $\lambda \in (0,1)$. It is clear that $x_\lambda > 0$, therefore $y_\lambda > 0$ because player 1 can play $(\alpha, +1)$ with $\alpha > 0$ in $(x, x', 2)$. Moreover player 2 can play $(\beta, +1)$ with $\beta > 0$ in $(x, y', 2)$ and it is easy to check that $x_\lambda < 1$.

Lemma 3.17. *For all $\lambda \in (0,1)$ the following equations hold.*

$$x_\lambda > y_\lambda \tag{3.5.1}$$

$$y_\lambda = (1-\lambda) \max_{(\alpha,p)\in A} \left(\alpha^2(p_{-1} - p_{+1})y_\lambda + \alpha p_{+1}(x_\lambda - y_\lambda) + p_{+1}y_\lambda\right) \tag{3.5.2}$$

$$x_\lambda = \lambda + (1-\lambda) \min_{(\beta,q)\in B} \Big(\beta^2(q_{-1} - q_{+1})(x_\lambda - 1) + \beta q_{+1}(y_\lambda - x_\lambda) + q_{+1}(x_\lambda - 1) + 1\Big). \tag{3.5.3}$$

The proof of Lemma 3.17 is postponed to the Appendix of Chapter 3. In the next proposition, we show that x_λ and y_λ indeed have a rather simple expression.

Proposition 3.18. *For all $\lambda \in (0,1)$,*

$$\lambda y_\lambda = (1-\lambda) \max_{\alpha\in I} \left(-\alpha^2 y_\lambda + \alpha(x_\lambda - y_\lambda)\right) \tag{3.5.4}$$

$$\lambda x_\lambda = \lambda + (1-\lambda) \min_{\beta\in J} \left(\beta^2(1 - x_\lambda) + \beta(y_\lambda - x_\lambda)\right). \tag{3.5.5}$$

The proof of Proposition 3.18 is postponed to the Appendix of Chapter 3.

3.5.3 A simple counterexample to the convergence of $(x_\lambda)_{\lambda\in(0,1)}$ and $(y_\lambda)_{\lambda\in(0,1)}$

Eqs. (3.5.4) and (3.5.5) are in fact the Shapley equations of the following game, non product, non weakly communicating on both sides, described in Fig. 3.6. This game does not have a product state space, neither has it a weakly communicating property. This example was introduced by Renault (2019). The state space is $\Omega = \{0, 1, 0^*, 1^*\}$. The action space of player 1 (resp. 2) is I (resp. J). Player 1 (resp. 2) plays in state 0 (resp. 1). The states 0^* and 1^* are absorbing. The payoff in states 0 and 0^* (resp. 1 and 1^*) is 0 (resp. 1).

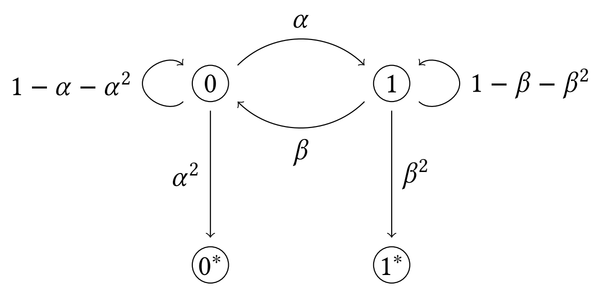

Figure 3.6: A simpler game

We conclude with the next lemmas showing that $(x_\lambda)_{\lambda\in(0,1]}$ and $(y_\lambda)_{\lambda\in(0,1]}$ do not converge. The proofs are provided for completeness.

Lemma 3.19. *For all $\lambda \le 1/17$, $\beta_\lambda = \frac{x_\lambda - y_\lambda}{2(1-x_\lambda)}$ is optimal for player 2. Moreover*

$$4\lambda(1-x_\lambda)^2 = (1-\lambda)(x_\lambda - y_\lambda)^2. \tag{3.5.6}$$

Hence $x_\lambda - y_\lambda$ goes to 0 as λ goes to 0.

Proof. Suppose that $\beta_\lambda \le 1/4$, then it is easy to verify that the lemma holds. Suppose now that $\beta_\lambda > 1/4$. Then the minimum in Eq. (3.5.5) is attained in $1/4$. It yields

$$\frac{1}{4}(1-\lambda)(x_\lambda - y_\lambda) = (1-x_\lambda)\left(\frac{1+15\lambda}{16}\right) > (1-\lambda)\frac{1-x_\lambda}{8}.$$

Hence $\lambda > 1/17$. □

Let $(\lambda_n)_{n\in\mathbb{N}^*} \in (0,1]^{\mathbb{N}^*}$ such that $\lambda_n \longrightarrow 0$ as n goes to $+\infty$.

Lemma 3.20. *If $(x_{\lambda_n})_n$ and $(y_{\lambda_n})_n$ converge to $v \in [0,1]$, then $v \le 1/2$. Moreover $x_{\lambda_n} - y_{\lambda_n} \sim 2\sqrt{\lambda_n}(1-v)$ and $\beta_{\lambda_n} \sim \sqrt{\lambda_n}$ as n goes to $+\infty$.*

Proof. Let α_λ be an optimal strategy of player 1. By Eq. (3.5.6) one gets

$$2\alpha_\lambda\sqrt{\lambda}y_\lambda \le y_\lambda(1+\alpha_\lambda^2) = \lambda y_\lambda \alpha_\lambda^2 + 2\alpha_\lambda\sqrt{\lambda}\sqrt{1-\lambda}(1-x_\lambda).$$

Dividing by $\alpha_\lambda\sqrt{\lambda}$ and passing to the limit yields $v \le 1/2$. □

In the next two lemmas, we show that the game does not have an asymptotic value. The idea is the following: player 1 would like to play in the λ-discounted game some α close to $\sqrt{\lambda}\frac{1-v}{v}$ in state 0, where v is the limit (up to some subsequence) of $(x_{\lambda_n})_n$ and $(y_{\lambda_n})_n$. If player 1 is not allowed to take any absorbing risk in $[0, 1/4]$, but player 2 is, we expect the values $(x_{\lambda_n})_n$ and $(y_{\lambda_n})_n$ to oscillate.

Lemma 3.21. *If for all $n \in \mathbb{N}^*$, $\sqrt{\lambda_n} \in I$ then*

$$\lim_{n\to+\infty} x_{\lambda_n} = 1/2.$$

Proof. Assume that, up to some subsequence, $(x_{\lambda_n})_n$ and $(y_{\lambda_n})_n$ converge to some $v \in [0, 1]$. If player 1 plays $\alpha = \sqrt{\lambda}$, Eq. (3.5.4) yields

$$\lambda y_\lambda \geq (1-\lambda)\sqrt{\lambda}(x_\lambda - y_\lambda) - (1-\lambda)\lambda y_\lambda.$$

Dividing by λ and passing to the limit, one gets $v \geq 1/2$. By Lemma 3.20, $v = 1/2$. □

Lemma 3.22. *If for all $n \in \mathbb{N}^*$, $\left(1/2\sqrt{\lambda_n}, 2\sqrt{\lambda_n}\right) \cap I = \emptyset$ then* $\limsup_{n\to+\infty} x_{\lambda_n} \leq 4/9$.

Proof. Suppose that up to some subsequence, $(x_{\lambda_n})_n$ and $(y_{\lambda_n})_n$ converge to some $v \geq 4/9$. By Lemma 3.20 $v \leq 1/2$. Let $\alpha^*_\lambda = \frac{x_\lambda - y_\lambda}{2y_\lambda} > 0$ be the argument of the maximum of the unconstrained problem associated to Eq. (3.5.4). Then $\alpha^*_\lambda \sim \sqrt{\lambda}\frac{1-v}{v}$. Hence for λ small enough in the sequence, $1/2\sqrt{\lambda} \leq \alpha^*_\lambda \leq 2\sqrt{\lambda}$.

The open interval $\left(1/2\sqrt{\lambda_n}, 2\sqrt{\lambda_n}\right)$ does not contain any point in I. Furthermore the objective function of player 1 is increasing between 0 and α^*_λ, and decreasing after.

First case, $\alpha_\lambda \leq 1/2\sqrt{\lambda}$. Then

$$\lambda y_\lambda \leq 1/2(1-\lambda)\sqrt{\lambda}(x_\lambda - y_\lambda) - 1/4(1-\lambda)\lambda y_\lambda.$$

Dividing by λ and passing to the limit yields $v \leq 4/9$.

Second case, $\alpha_\lambda \geq 2\sqrt{\lambda}$. Then

$$\lambda y_\lambda \leq 2(1-\lambda)\sqrt{\lambda}(x_\lambda - y_\lambda) - 4(1-\lambda)\lambda y_\lambda,$$

and again $v \leq 4/9$. □

Thus taking the sequences $\lambda_n = \frac{1}{2^{2n}}$ and $\lambda'_n = \frac{1}{2^{2n+1}}$, for all $n \geq 1$, one has from Lemmas 3.21 and 3.22 that $(x_\lambda)_{\lambda\in(0,1]}$ and $(y_\lambda)_{\lambda\in(0,1]}$ do not converge as λ goes to 0.

Appendix of Chapter 3: omitted proofs

Proof of Lemma 3.17. We consider the game starting in state $(x, x', 2)$. Player 2 can play $(\beta, q) = (0, +1)$, hence

$$y_\lambda \leq (1-\lambda) \max_{(\alpha,p)\in A} \left(\alpha^2(p_{-1} - p_{+1})y_\lambda + \alpha p_{+1}(x_\lambda - y_\lambda) + p_{+1}y_\lambda\right).$$

Suppose

$$\{0\} \in \arg\max_p \left(\max_{\alpha\in\{0,1\}} \left(\alpha^2(p_{-1} - p_{+1})y_\lambda + \alpha p_{+1}(x_\lambda - y_\lambda) + p_{+1}y_\lambda\right)\right).$$

Then $y_\lambda \leq (1-\lambda)\max_{\alpha\in\{0,1\}} 0$, hence $y_\lambda = 0$, contradiction.
Suppose

$$\{-1\} \in \arg\max_p \left(\max_{\alpha\in I} \left(\alpha^2(p_{-1} - p_{+1})y_\lambda + \alpha p_{+1}(x_\lambda - y_\lambda) + p_{+1}y_\lambda\right)\right).$$

Then $y_\lambda \leq (1-\lambda)\max_{\alpha\in I} \alpha^2 y_\lambda = \frac{1-\lambda}{16} y_\lambda$. Thus $y_\lambda = 0$, contradiction.
Thus

$$\arg\max_p \left(\max_{\alpha} \left(\alpha^2(p_{-1} - p_{+1})y_\lambda + \alpha p_{+1}(x_\lambda - y_\lambda) + p_{+1}y_\lambda\right)\right) = \{+1\}.$$

Hence $\lambda y_\lambda \leq (1-\lambda)\max_{\alpha\in I}\left(-\alpha^2 + \alpha(x_\lambda - y_\lambda)\right)$. Thus $x_\lambda > y_\lambda$, and Eq. (3.5.1) is proved.

To prove Eq. (3.5.2), we show that $(0, +1)$ is a dominant strategy of player 2, i.e.,

$$\forall(\alpha, p) \in A\ \forall(\beta, q) \in B\ h(y_\lambda, x_\lambda, \alpha, p, \beta, q) - h(y_\lambda, x_\lambda, \alpha, p, 0, +1) \geq 0.$$

There are 9 cases to test, corresponding to the different values of $(p, q) \in \{-1, 0, +1\}^2$.

1st case: $p_{+1} = q_{+1} = 1$

$$\begin{aligned}
&\left((1-\alpha-\alpha^2)(-\beta-\beta^2) + \alpha\beta + \alpha^2\beta^2\right) y_\lambda + \\
&\left((1-\alpha-\alpha^2)\beta + (-\beta-\beta^2)\alpha\right) x_\lambda + \beta^2(1-\alpha^2) \\
= &\left((1-2\alpha-\alpha^2-\alpha\beta)(x_\lambda - y_\lambda) + \beta(1-\alpha^2)(1-y_\lambda) + \beta\alpha^2 y_\lambda\right)\beta \geq 0.
\end{aligned}$$

2nd case: $p_{+1} = q_0 = 1 \quad 1 - \alpha^2 \geq 0$

3rd case: $p_{+1} = q_{-1} = 1$

$$\begin{aligned}
&\big((1-\alpha-\alpha^2)\beta^2 + (1-\beta-\beta^2)\alpha^2 - (1-\alpha-\alpha^2)\big)\, y_\lambda + \\
&(\alpha\beta^2 + \beta\alpha^2 - \alpha)x_\lambda + (1-\alpha^2)(1-\beta^2) \\
=& \big((\beta^2-1)(1-\alpha-2\alpha^2) - \beta\alpha^2\big)\, y_\lambda + \\
&\big(\beta\alpha^2 + \alpha(\beta^2-1)\big)\, x_\lambda + (1-\alpha^2)(1-\beta^2) \\
=& \beta\alpha^2(x_\lambda - y_\lambda) + \alpha(1-\beta^2)(1-x_\lambda) + \\
&(1-\beta^2)(1-\alpha-\alpha^2)(1-y_\lambda) + (1-\beta^2)\alpha^2 y_\lambda \geq 0
\end{aligned}$$

4th case: $p_0 = q_{+1} = 1 \quad \beta^2 \geq 0$

5th case: $p_0 = q_0 = 1$

$$((1-\alpha)(1-\beta) + \alpha\beta)\, y_\lambda + ((1-\alpha)\beta + (1-\beta)\alpha)\, x_\lambda \geq 0$$

6th case: $p_0 = q_{-1} = 1 \quad 1-\beta^2 \geq 0$

7th case: $p_{-1} = q_{+1} = 1$

$$\begin{aligned}
&\big((1-\alpha-\alpha^2)\beta^2 + (1-\beta-\beta^2)\alpha^2 - \alpha^2\big)\, y_\lambda + (\alpha^2\beta + \alpha\beta^2)x_\lambda + \alpha^2\beta^2 \\
=& \big((1-\alpha-\alpha^2)\beta^2 + (-\beta-\beta^2)\alpha^2\big)\, y_\lambda + (\alpha^2\beta + \alpha\beta^2)x_\lambda + \alpha^2\beta^2 \\
=& (1-2\alpha^2)\beta^2 y_\lambda + (\alpha^2\beta + \alpha\beta^2)(x_\lambda - y_\lambda) + \alpha^2\beta^2 \geq 0
\end{aligned}$$

8th case: $p_{-1} = q_0 = 1 \quad \alpha^2 \geq 0$

9th case: $p_{-1} = q_{-1} = 1$

$$\begin{aligned}
&\big((1-\alpha-\alpha^2)(1-\beta-\beta^2) + \alpha\beta + \alpha^2\beta^2 - \alpha^2\big)\, y_\lambda + \\
&\big((1-\alpha-\alpha^2)\beta + (1-\beta-\beta^2)\alpha\big)\, x_\lambda + \alpha^2(1-\beta^2) \\
=& \big((1-\alpha-\alpha^2)(1-\beta-\beta^2) + \alpha\beta\big)\, y_\lambda + \\
&\big((1-\alpha-\alpha^2)\beta + (1-\beta-\beta^2)\alpha\big)\, x_\lambda + \alpha^2(1-\beta^2)(1-y_\lambda) \geq 0
\end{aligned}$$

Eq. (3.5.2) is thus proved.

We now consider the game starting in state $(x, y', 2)$. To prove Eq. (3.5.3), we show that $(0, +1)$ is a dominant strategy of player 1, i.e.,

$$\forall(\alpha, p) \in A\ \forall(\beta, q) \in B\ h(x_\lambda, y_\lambda, \alpha, p, \beta, q) - h(x_\lambda, y_\lambda, 0, +1, \beta, q) \leq 0.$$

There are again 9 cases to test, corresponding to the different values of $(p, q) \in \{-1, 0, +1\}^2$.

1st case: $p_{+1} = q_{+1} = 1$

$$\begin{aligned}
&\left((1-\alpha-\alpha^2)(1-\beta-\beta^2)+\alpha\beta+\alpha^2\beta^2-(1-\beta-\beta^2)\right)x_\lambda+\\
&\left((1-\alpha-\alpha^2)\beta+(1-\beta-\beta^2)\alpha-\beta\right)y_\lambda+\beta^2(1-\alpha^2)-\beta^2\\
=&\left((-\alpha-\alpha^2)(1-\beta-\beta^2)+\alpha\beta+\alpha^2\beta^2\right)x_\lambda+\\
&\left((-\alpha-\alpha^2)\beta+(1-\beta-\beta^2)\alpha\right)y_\lambda-\alpha^2\beta^2\\
=&\alpha(1-2\beta-\beta^2-\alpha\beta)(y_\lambda-x_\lambda)-\alpha^2(1-\beta^2)x_\lambda+\alpha^2\beta^2(x_\lambda-1)\leq 0.
\end{aligned}$$

2nd case: $p_{+1} = q_0 = 1 \quad 1-1\leq 0$

3rd case: $p_{+1} = q_{-1} = 1$

$$\begin{aligned}
&\left((1-\alpha-\alpha^2)\beta^2+(1-\beta-\beta^2)\alpha^2-\beta^2\right)x_\lambda+\\
&(\alpha\beta^2+\beta\alpha^2)y_\lambda+(1-\alpha^2)(1-\beta^2)-(1-\beta^2)\\
=&\left(\beta^2(-\alpha-\alpha^2)+\alpha^2(1-\beta-\beta^2)\right)x_\lambda+\left(\beta\alpha^2+\alpha\beta^2\right)y_\lambda-\alpha^2(1-\beta^2)\\
=&\left(\beta\alpha^2+\alpha\beta^2\right)(y_\lambda-x_\lambda)+\alpha^2(1-\beta^2)(x_\lambda-1)-\alpha^2\beta^2x_\lambda\leq 0
\end{aligned}$$

4th case: $p_0 = q_{+1} = 1 \quad -(1-\beta-\beta^2)x_\lambda-\beta y_\lambda+\beta^2-\beta^2\leq 0$

5th case: $p_0 = q_0 = 1$

$$\begin{aligned}
&((1-\alpha)(1-\beta)+\alpha\beta)\,x_\lambda+((1-\alpha)\beta+(1-\beta)\alpha)\,y_\lambda-1\\
=&((1-\alpha)\beta+\alpha(1-\beta))\,(y_\lambda-x_\lambda)+x_\lambda-1\leq 0
\end{aligned}$$

6th case: $p_0 = q_{-1} = 1 \quad 1-\beta^2-(1-\beta^2)-\beta^2x_\lambda\leq 0$

7th case: $p_{-1} = q_{+1} = 1$

$$\begin{aligned}
&\left((1-\alpha-\alpha^2)\beta^2+(1-\beta-\beta^2)\alpha^2-(1-\beta-\beta^2)\right)x_\lambda+\\
&(\alpha^2\beta+\alpha\beta^2-\beta)y_\lambda+\alpha^2\beta^2-\beta^2\\
=&(\alpha^2-1)\beta y_\lambda+\alpha\beta^2(y_\lambda-x_\lambda)+\\
&(1-\alpha^2)(2\beta^2-1)x_\lambda+(\alpha^2-1)\beta(1-x_\lambda)\leq 0
\end{aligned}$$

8th case: $p_{-1} = q_0 = 1 \quad \alpha^2-1\leq 0$

9th case: $p_{-1} = q_{-1} = 1$

$$\begin{aligned}
&\left((1-\alpha-\alpha^2)(1-\beta-\beta^2)+\alpha\beta+\alpha^2\beta^2-\beta^2\right)x_\lambda+\\
&\left((1-\alpha-\alpha^2)\beta+(1-\beta-\beta^2)\alpha\right)y_\lambda+\alpha^2(1-\beta^2)-(1-\beta^2)\\
=&\left((1-\alpha-\alpha^2)\beta+(1-\beta-\beta^2)\alpha\right)(y_\lambda-x_\lambda)+\\
&\beta^2(\alpha^2-1)x_\lambda+(1-\alpha^2)(1-\beta^2)(x_\lambda-1)\leq 0
\end{aligned}$$

Thus Eq. (3.5.3) is proved. □

Proof of Proposition 3.18. It has been seen in the proof of Lemma 3.17 that

$$\arg\max_{p} \left(\max_{\alpha} \left(\alpha^2 (p_{-1} - p_{+1}) y_\lambda + \alpha p_{+1} (x_\lambda - y_\lambda) + p_{+1} y_\lambda \right) \right) = \{+1\}.$$

Thus one deduces Eq. (3.5.4). Likewise,

$$\arg\min_{q} \left(\min_{\beta} \left(\beta^2 (q_{-1} - q_{+1}) (x_\lambda - 1) + \beta q_{+1} (y_\lambda - x_\lambda) + p_{+1} y_\lambda \right) \right) = \{+1\}.$$

And Eq. (3.5.5) is proved. □

Chapter 4

Continuous patrolling and hiding games

European Journal of Operational Research 277 (2019) 42-51

We present two zero-sum games modeling situations where one player attacks (or hides in) a finite dimensional nonempty compact set, and the other tries to prevent the attack (or find him). The first game, called patrolling game, corresponds to a dynamic formulation of this situation in the sense that the attacker chooses a time and a point to attack and the patroller chooses a continuous trajectory to maximize the probability of finding the attack point in a given time. Whereas the second game, called hiding game, corresponds to a static formulation in which both the searcher and the hider choose simultaneously a point and the searcher maximizes the probability of being at distance less than a given threshold of the hider.

4.1 Introduction

To ensure the security of vulnerable facilities, a planner may deploy either dynamic or static security devices. Some examples of dynamic security devices are: soldiers or police officers patrolling the streets of a city, robots patrolling a shopping mall, drones flying above a forest to detect fires, or naval radar systems signaling the detection of an enemy ship. On the other hand, security guards positioned in the rooms of a museum, security cameras scrutinizing subway corridors, or motion detectors placed in a house are some examples of static security devices. The paradigm we adopt is the one of an adversarial threat, hence we propose a game theoretical approach to these security problems. Some game theoretic security systems are already in use, for example in the Los Angeles international airport, see (Pita et al., 2008), and in some ports of the United States, see (Shieh et al., 2012).

Motivated by the examples given above, we study two zero-sum games in which a player (the patroller or searcher) aims to detect another player (the attacker or hider). *Patrolling games* model dynamic security devices. In these games, the patroller moves continuously in a search space with bounded speed. The attacker chooses a point in the search space and a time to attack it. The attack takes a certain duration to be successful (think of a terrorist needing time to set off a bomb). The patroller wins if and only if she detects the attack before it succeeds. The case of static security devices is modeled by *hiding games*, in which both the searcher and the hider simultaneously deploy at some points in a search space, the searcher wins if and only if the hider lies within her detection radius. We provide links between patrolling and hiding games, and show how patrolling games reduce to hiding games if the attack duration is zero, that is the patroller has to detect the attack at the exact time it occurs, or if the patroller is alerted of the attack point when the attack begins.

4.1.1 Contribution

For patrolling games, we prove that the value always exists, and obtain a decomposition result to lower bound the value as well as a general upper bound. We then study patrolling games on networks. We compute the value as well as optimal strategies for the class of Eulerian networks. The special network composed of two nodes linked by three parallel arcs is also examined and bounds on the value are computed. Lastly, we study patrolling games on $\mathbb{R}^2$ and obtain an asymptotic expression for the value as the detection radius of the patroller goes to 0.

For hiding games, we first focus on a particular class of strategies for both the searcher and the hider called "equalizing". These strategies have the property that if there exists one, then it is optimal for both players. Our main result regarding hiding games is an asymptotic formula for the value when the search space has positive Lebesgue measure. A counterexample based on a Cantor-type set showing that this last result cannot be extended to compact sets with zero Lebesgue measure is also presented.

Finally, we discuss some elementary properties of monotonicity and continuity of the value function of continuous patrolling and hiding games.

4.1.2 Related literature

Continuous patrolling and hiding games belong to the literature on search and security games, consult (Hohzaki, 2016) for a survey. These games have their source in search theory, a field of operations research whose origins can be found in the works of Koopman (1956a,b, 1957). The use of game theory in a search and security context goes back to the famous book of Morse and Kimball (1951).

Patrolling games

Patrolling games were introduced by Alpern et al. (2011) in a discrete setting, that is the patroller visits nodes of a graph, where the attacker can strike, at discrete times. A companion article (Alpern et al., 2016a) is dedicated to the resolution of patrolling a discrete line. The idea of investigating a continuous version of patrolling games is suggested in (Alpern et al., 2011). In (Alpern et al., 2016b), the authors solve the continuous patrolling game played on the unit interval. Several papers in the field of search games have dealt with the transcription of discrete models to continuous ones, as in (Ruckle and Kikuta, 2000) and (Ruckle, 1981).

Other game theoretic models involving a patroller and an attacker can be found in (Basilico et al., 2015, 2012), in which the authors design algorithms to solve large instances of Stackelberg patrolling security games on graphs. Lin et al. (2013, 2014) use linear programming and heuristics to study a large class of patrolling problems on graphs, with nodes having different values. Zoroa et al. (2012) study a patrolling game with a mobile attacker on a perimeter.

Continuous patrolling games are closely related to search games with an immobile hider, introduced in the seminal book of Isaacs (1965) and developed in the monographs of Gal (1980) and Alpern and Gal (2003). In these games, a searcher intends to minimize the time necessary to find a hider. Search games have been extensively studied, let us mention (Alpern et al., 2008; Dagan and Gal, 2008; Gal, 1979) for search games on a network. See also (Bostock, 1984; Pavlovic, 1993) for the special network consisting of two nodes linked by three parallel arcs, for which the solution is surprisingly complicated.

Hiding games

The first published example of a hiding game goes back to Gale and Glassey (1974). The proposers gave a solution to the problem of hiding in the unit disc when the detection radius is $r = 1/2$. Later, Ruckle (1983) considered in his book several examples of hiding games (hiding on a sphere, hiding in a disc, among others). Computing the value of hiding games is in general a very difficult task. Danskin (1990) improved substantially the resolution of the hiding game played on a disc, he called the cookie-cutter game. However the solution is not complete for small values of r and no progress has been made since then, see also (Alpern et al., 2013) and (Washburn, 2014). Hiding games in a discrete setting, i.e., when the search space is a graph, have been studied by Bishop and Evans (2013).

Games in which the payoff is the distance between the searcher and the hider, introduced by Karlin (1959), have been extensively studied, consult (Ibragimov and Satimov, 2012) and references therein. Although these games resemble hiding games to a certain extent, the lack of continuity of the payoff function in hiding games makes their analysis much more involved.

Finally, ambush games can be seen as hiding games in which the players select not one, but several points in the unit interval, consult (Baston and Kikuta, 2004, 2009; Zoroa et al., 1999) for further details.

4.1.3 Organization of the paper

The paper is organized as follows. In Section 4.2 the models are formally presented. Section 4.3 is dedicated to patrolling games, and Section 4.4 is dedicated to hiding games. Finally in Section 4.5 we give some properties of the value function of continuous patrolling and hiding games. Proofs that are not included in the body of the paper are postponed to the Appendix of Chapter 4.

4.1.4 Notations

In all the article, $\mathbb{R}^n$ is endowed with a norm denoted $\|\cdot\|$, which induces a metric d. For all $x \in \mathbb{R}^n$ and $r > 0$, the closed ball of center x with radius r is denoted $B_r(x) = \{y \in \mathbb{R}^n \mid \|x - y\| \leq r\}$. For all Lebesgue measurable set $B \subset \mathbb{R}^n$, $\lambda(B)$ denotes its Lebesgue measure. Finally, $\lambda(B_r)$ denotes the Lebesgue measure of any ball of radius r. Let X be a topological space, the set of Borel probability measures on X is denoted $\Delta(X)$, and the set of probability measures on X with finite support is denoted $\Delta_f(X)$.

4.2 The models

4.2.1 Patrolling games

In a patrolling game two players, an attacker and a patroller, act on a set Q called the search space, which is assumed to be a nonempty compact subset of $\mathbb{R}^n$. An example of this could be a metric network, as in Section 4.3.4. The attacker chooses an attack point y in Q and a time to attack t in $\mathbb{R}_+$. The patroller walks continuously in Q with speed at most 1. When the attack occurs at time t and point y, the patroller has a time limit $m \in \mathbb{R}_+$ to be at distance at most $r \in \mathbb{R}_+$ of the attack point y. In this case she detects the attack and wins, and otherwise she does not. Thus, m represents the time needed for an attack to be successful, and r represents the detection radius of the patroller.

A patrolling game is thus a zero-sum game given by a triplet (Q, m, r). The attacker's set of pure strategies is $\mathcal{A} = Q \times \mathbb{R}_+$. An element of $\mathcal{A}$ is called an attack. The patroller's set of pure strategies is

$$\mathcal{W} = \{w : \mathbb{R}_+ \to Q \mid w \text{ is 1-Lipschitz continuous}\}.$$

An element of $\mathcal{W}$ is called a walk. $\mathcal{W}$ is endowed with the topology of compact convergence (consult the proof of Proposition 4.1 in the Appendix of Chapter 4 and (Munkres, 2000) for details).

The payoff to the patroller is given by

$$g_{m,r}(w,(y,t)) = \begin{cases} 1 & \text{if } d(y, w([t, t+m])) \leq r \\ 0 & \text{otherwise,} \end{cases}$$

where $w([t, t+m]) = \{w(\tau) \mid \tau \in [t, t+m]\}$.

4.2.2 Hiding games

In a hiding game two players, a searcher and a hider, act on a search space Q, which is again assumed to be a nonempty compact subset of $\mathbb{R}^n$. Both players choose a point in Q. The searcher has a detection radius $r \in \mathbb{R}_+$. She finds the hider if and only if the two points are at distance at most r.

Hence, a hiding game is a zero-sum game given by a couple $H = (Q, r)$. The set of pure strategies of both players, the searcher and hider, is Q. The payoff to the searcher is given by

$$h_r(x, y) = \begin{cases} 1 & \text{if } \|x - y\| \leq r \\ 0 & \text{otherwise.} \end{cases}$$

4.2.3 Links between patrolling and hiding games

Hiding games can be interpreted as two possible variants of patrolling games.

In the first variant, hiding games are considered as a particular class of patrolling games in which the attack duration m is equal to 0. Indeed, consider a hiding game $H = (Q, r)$ and a patrolling game $P = (Q, 0, r)$. In P, for all $w \in \mathcal{W}$ and $(y, t) \in \mathcal{A}$ the payoff to the patroller is

$$g_{0,r}(w, (y, t)) = \begin{cases} 1 & \text{if } \|w(t) - y\| \leq r \\ 0 & \text{otherwise.} \end{cases}$$

Let $x \in Q$ be a strategy of the searcher in the hiding game H, and consider the constant walk $w_x : t \mapsto x$ of the patroller in P. Then for any strategy (y, t) of the attacker in P, the payoff $g_{0,r}(w_x, (y, t))$ equals 1 if $\|x - y\| \leq r$ and 0 otherwise. Since the payoff is time-independent, any quantity guaranteed by the searcher in H is guaranteed by the patroller in P. Conversely, let $y \in Q$ be a strategy of the hider in H and consider the strategy $(y, 0)$ of the attacker in H. Then for any strategy w of the patroller in P, the payoff is 1 if $\|w(0) - y\| \leq r$ and 0 otherwise. Again the payoff is time-independent and any quantity guaranteed by the hider in H is guaranteed by the attacker in P. Thus, since H and P have a value (see Propositions 4.1 and 4.14), the values of these two games are the same.

The second interpretation is as follows. Alpern et al. (2011) suggest the study of patrolling games in which the patroller may be informed of the presence of the attacker. Suppose that the patroller is informed of the attack point when the attack occurs. Suppose also the search space Q is convex. The detection radius r is set to 0 for simplicity. The payoff of this game is

$$g_{m,0}(w, (y, t)) = \begin{cases} 1 & \text{if } y \in w([t, t+m]) \\ 0 & \text{otherwise.} \end{cases}$$

This patrolling game with signals is denoted P'. In P', if the patroller's strategy is to choose a point and not move until the attack, then go to the attack point in straight line when she is alerted, the attacker is time-indifferent. In particular, the attacker has a best response in the set of attacks occurring at time 0. Symmetrically, if the attack occurs at time 0, the patroller has a best response consisting in choosing a starting point in Q and going directly to the attack point when she is informed of the attack. Thus, with the same mappings of strategies in the hiding game $H' = (Q, m)$ to strategies in P' as before, any quantity guaranteed by the searcher in H' is guaranteed by the patroller in P'. Conversely, any quantity guaranteed by the hider in H' is guaranteed by the attacker in P'. Thus the values of these two games are the same.

4.3 Patrolling games

4.3.1 The value of patrolling games

Our first result is the existence of the value of patrolling games. We denote it $V_Q(m, r)$. In addition, we prove that the patroller has an optimal strategy and the attacker has an ε-optimal strategy with finite support. The fact that the patroller has an optimal strategy means that she can guarantee that the probability of detecting the attack is at least $V_Q(m, r)$, no matter what the attacker does. Similarly, the attacker can guarantee that the probability of being caught is at most $V_Q(m, r)$, up to ε, no matter what the patroller does. Hence, in patrolling games, the value represents the probability (up to ε) of the attack being intercepted when both the patroller and the attacker play (ε-)optimally.

Proposition 4.1. *The patrolling game (Q, m, r) played with mixed strategies has a value $V_Q(m, r)$.*

Moreover the patroller has an optimal strategy and the attacker has an ε-optimal strategy with finite support, i.e.,

$$\begin{aligned} V_Q(m, r) &= \max_{\mu \in \Delta(\mathcal{W})} \inf_{(y,t) \in \mathcal{A}} \int_{\mathcal{W}} g_{m,r}(w, (y, t)) d\mu(w) \\ &= \inf_{\nu \in \Delta_f(\mathcal{A})} \max_{w \in \mathcal{W}} \int_{\mathcal{A}} g_{m,r}(w, (y, t)) d\nu(y, t). \end{aligned}$$

The proof of Proposition 4.1 is provided in the Appendix of Chapter 4.

4.3.2 Decomposition

As it is the case for graphs, see (Alpern et al., 2011), it is possible to consider a search space Q as the union of simpler search spaces $Q_1, \ldots, Q_n$ for which the values of the corresponding patrolling games may be known. The value of the original patrolling game is lower bounded by a function of the values of the patrolling games involved in the decomposition.

Proposition 4.2. *Let Q and $Q_1, \dots, Q_n$ be search spaces such that $Q = \cup_{i=1}^{n} Q_i$. Then for all $m, r \in \mathbb{R}_+$*

$$V_Q(m,r) \geq \frac{1}{\sum_{i=1}^{n} V_{Q_i}(m,r)^{-1}}.$$

Proof. For all $i \in \{1, \dots, n\}$, let μ_i be an optimal strategy of the patroller in the game (Q_i, m, r). Let μ be the strategy of the patroller in the game (Q, m, r) which consists in playing strategy μ_i with probability $\frac{V_{Q_i}(m,r)^{-1}}{\sum_{k=1}^{n} V_{Q_k}(m,r)^{-1}}$.

Let $(y, t) \in \mathcal{A} = Q \times \mathbb{R}_+$ be an attack in the game (Q, m, r). Then there exists $i \in \{1, \dots, n\}$ such that $y \in Q_i$, hence $g(\sigma, (y, t)) \geq \frac{V_{Q_i}(m,r)^{-1}}{\sum_{k=1}^{n} V_{Q_k}(m,r)^{-1}} V(Q_i, m, r)$. □

4.3.3 A general upper bound

Our goal is now to obtain a general upper bound for the value of patrolling games. As in (Alpern and Gal, 2003), let us introduce the maximal rate at which the patroller can discover new points of Q.

Definition 4.3. The maximal discovery rate is given by

$$\rho = \sup_{w \in \mathcal{W}, t > 0} \frac{\lambda(w([0,t]) + B_r(0)) - \lambda(B_r)}{t},$$

where $w([0,t]) = \{w(\tau) \mid \tau \in [0,t]\}$ is the image of $[0,t]$ by w and $w([0,t]) + B_r(0) = \{y \in \mathbb{R}^n \mid d(w([0,t],y)) \leq r\}$.

Hence, in a network the maximal discovery rate ρ is 1, in $\mathbb{R}^2$ and $\mathbb{R}^3$ endowed with the Euclidean norm, under suitable assumptions, the maximal discovery rate is typically $2r$ and πr^2 respectively, that is the sweep width of the patroller.

Let us now give a general upper bound on the value of patrolling games whose search space have nonzero Lebesgue measure. It is the upper bound used to prove Theorems 4.10, 4.13 and 4.17. A similar bound is given in (Alpern et al., 2011) in the discrete case.

Proposition 4.4. *Let Q be a search space such that $\lambda(Q) > 0$. Then*

$$V_Q(m,r) \leq \frac{m\rho + \lambda(B_r)}{\lambda(Q)}.$$

To prove Proposition 4.4, we define a strategy for the attacker called the uniform strategy, under which he attacks uniformly over Q at time 0. Intuitively, a best response of the patroller is to cover as much points in Q as possible between time 0 and time m.

Definition 4.5. Let Q be a search space such that $\lambda(Q) > 0$. The attacker's uniform strategy on Q, denoted a_λ, is a random choice of the attack point at time 0 such that for all measurable sets $B \subset Q$,

$$a_\lambda(B, 0) = \frac{\lambda(B)}{\lambda(Q)}, \text{ and } a_\lambda(B,t) = 0 \text{ if } t > 0.$$

Proof of Proposition 4.4. For all $w \in \mathcal{W}$, the payoff to the patroller when the attacker plays a_λ is

$$\begin{aligned}\int_{\mathcal{A}} g_{m,r}(w,(y,t))da_\lambda(y,t) &= \frac{\lambda\big(w([0,m]) + B_r(0)\big)}{\lambda(Q)} \\ &\leq \frac{m\rho + \lambda(B_r)}{\lambda(Q)}.\end{aligned}$$

□

4.3.4 Patrolling a network

Definition of a network

We follow the construction of a network of Fournier (2018). Let (V, E, l) be a weighted undirected graph, V is the finite set of nodes and E the finite set of edges whose elements $e \in E$ have length $l(e) \in \mathbb{R}_+$. An edge $e \in E$ linking the two nodes u and v is also denoted (u, v).

We identify the elements of V with the vectors of the canonical basis of $\mathbb{R}^{|V|}$. The network generated by (V, E) is the set of points

$$\mathcal{N} = \{(u, v, \alpha) \mid \alpha \in [0, 1] \text{ and } (u, v) \in E\},$$

where $(u, v, \alpha) = \alpha u + (1 - \alpha)v$.

We denote $P(u, v)$ the set of paths between two points u and v in $\mathcal{N}$. It is the set of all sequences $(u_1, \ldots, u_n)$, $n \in \mathbb{N}^*$ such that $u_1 = u$, $u_n = v$ and such that for all $i \in \{1, \ldots, n-1\}$, u_i and u_{i+1} belong to the same edge. Let $u_1 = (u, v, \alpha_1)$ and $u_2 = (u, v, \alpha_2)$, and suppose $\alpha_1 < \alpha_2$. The set $[u_1, u_2] = \{(u, v, \alpha) \mid \alpha \in [\alpha_1, \alpha_2]\}$ is called an interval.

Finally, networks are endowed with their natural metric d, i.e., the length of the shortest path between two points, as well as their Lebesgue measure defined as a natural extension of the Lebesgue measure on real intervals.

Eulerian networks

For Eulerian networks, it is possible to compute the value and optimal strategies of the game. As stated in the next definition, an Eulerian tour is a closed path in $\mathcal{N}$ visiting all points and having length $\lambda(\mathcal{N})$.

Definition 4.6. Let $u \in \mathcal{N}$ and $\pi = (u_1, u_2, \ldots, u_{n-1}, u_n) \in P(u, u)$. If

$$\bigcup_{k=1}^{n-1} [u_k, u_{k+1}] = \mathcal{N},$$

then π is called a tour. Moreover, if $\sum_{k=1}^{n-1} \lambda([u_k, u_{k+1}]) = \lambda(\mathcal{N})$, then π is called an Eulerian tour. A network $\mathcal{N}$ is said to be Eulerian if it admits an Eulerian tour.

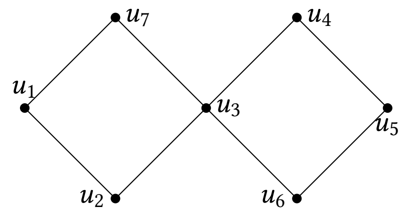

Figure 4.1: The network $\mathcal{N}_1$ is Eulerian

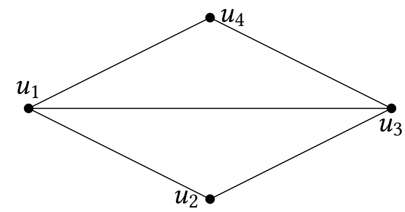

Figure 4.2: The network $\mathcal{N}_2$ is not Eulerian

Example 4.1. Figs. 4.1 and 4.2 display two examples of networks. $\mathcal{N}_1$ is an Eulerian network with Eulerian tour $\pi_1 = (u_1, u_2, u_3, u_4, u_5, u_6, u_3, u_7, u_1)$. In contrast, $\mathcal{N}_2$ is not an Eulerian network.

Our objective is now to define the uniform strategy of the patroller for Eulerian networks. This strategy is optimal for Eulerian networks. First, we need to define a parametrization of the network.

Definition 4.7. Let $\mathcal{N}$ be an Eulerian network. A continuous function w from $[0, \lambda(\mathcal{N})]$ to $\mathcal{N}$ is called a parametrization of $\mathcal{N}$ if it satisfies

i) $w(0) = w(\lambda(\mathcal{N}))$;

ii) w is surjective;

iii) $\forall t_1, t_2 \in [0, \lambda(\mathcal{N})]\ \lambda(w([t_1, t_2])) = |t_1 - t_2|$ (the speed of w is 1).

Moreover such a function w can be extended to a $\lambda(\mathcal{N})$-periodic function on $\mathbb{R}_+$ which is still denoted w.

Lemma 4.8. *Let $\mathcal{N}$ be an Eulerian network, then there exists a parametrization of $\mathcal{N}$.*

The proof of Lemma 4.8 is provided in the Appendix of Chapter 4. It is now possible to define the uniform strategy of the patroller. The idea behind this strategy is that the patroller uniformly chooses a starting point in $\mathcal{N}$, and then follows a parametrization as in Definition 4.7 above.

Definition 4.9. Suppose $\mathcal{N}$ is an Eulerian network. Let w be a parametrization of $\mathcal{N}$. Denote $(w_{t_0})_{t_0 \in [0, \lambda(\mathcal{N})]}$ the family of $\lambda(\mathcal{N})$-periodic walks such that $w_{t_0}(\cdot) = w(t_0 + \cdot)$. The patroller's uniform strategy is given by the uniform choice of $t_0 \in [0, \lambda(\mathcal{N})]$.

The next theorem is the main result on patrolling games for networks. It gives a simple expression of the value of patrolling games played on any Eulerian network. The result relies on the fact that for such networks, the patroller can achieve the upper bound of Proposition 4.4 using her uniform strategy. Note that Theorem 4.10 below is related to (Alpern et al., 2016b, Theorem 1), as well as (Alpern et al., 2011, theorem 13) for Hamiltonian graphs in the discrete case.

Theorem 4.10. *If $\mathcal{N}$ is an Eulerian network, then*

$$V_{\mathcal{N}}(m,0) = \min\left(\frac{m}{\lambda(\mathcal{N})}, 1\right).$$

Moreover the attacker's and the patroller's uniform strategies are optimal.

The proof of Theorem 4.10 is provided in the Appendix of Chapter 4. It is interesting to note that in search games with an immobile hider, the uniform strategies of the searcher and the hider are also optimal in Eulerian networks.

The three parallel arc network

In this section, we compute bounds on the value of a patrolling game played over the network with three parallel arcs, which plays an important role in the search game literature. Indeed despite its simple shape, finding the value and optimal strategies is a difficult task, as mentioned in the introduction.

We consider again the network $\mathcal{N}_2$ represented in Fig. 4.2. In this example, we take $l(u_1,u_2) = l(u_2,u_3) = l(u_1,u_4) = l(u_4,u_3) = 1/2$ and $l(u_1,u_3) = 1$. Notice that $\lambda(\mathcal{N}_2) = 3$. We compute the following bounds on the value of $(\mathcal{N}_2, m, 0)$:

$$V_{\mathcal{N}_2}(m,0) \begin{cases} = \frac{m}{3} & \text{if } m \le 2 \\ \in \left[\frac{5m-2}{3(m+2)}, 1-\frac{1}{3}\left(\frac{4-m}{2}\right)^2\right] & \text{if } m \in \left[2, \frac{10}{3}\right] \\ \in \left[\frac{14-2m}{3(6-m)}, 1-\frac{1}{3}\left(\frac{4-m}{2}\right)^2\right] & \text{if } m \in \left[\frac{10}{3}, 4\right] \\ = 1 \text{ if} & m \ge 4. \end{cases}$$

These bounds are plotted on Fig. 4.3 below.

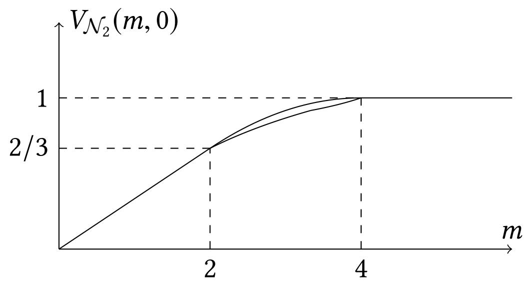

Figure 4.3: Bounds on the value of the game as a function of m

First case: $m \le 2$**.** Recall that (Proposition 4.4) $V_{\mathcal{N}_2}(m,0) \le \frac{m}{3}$ for all $m \ge 0$. Suppose the patroller uniformly chooses one of the three Eulerian sub-networks of $\mathcal{N}_2$ of length 2. Then by Theorem 4.10 she guarantees $m/2$ in these sub-networks. By symmetry, the patroller guarantees $\frac{2}{3}\frac{m}{2} = \frac{m}{3}$ in $\mathcal{N}_2$. We also give an alternative strategy of the patroller which guarantees $m/3$, and which will be useful for the next case $2 < m < 4$. Let $\pi^1 = (u_1, u_2, u_3, u_1, u_4, u_3)$ and $\pi^2 = (u_3, u_2, u_1, u_3, u_4, u_1)$ be two paths. π^1 and π^2 naturally induce two walks on $[0, 3]$ at speed 1, respectively denoted w^1 and w^2. For all $u \in \mathcal{N}_2$ and all $i \in \{1, 2\}$ there exists a unique $t_u^i \in [0, 3]$ such that $w^i(t_u^i) = u$, except for u_1 and u_3 for which the corresponding t_u^i may be taken arbitrarily. Now for all $u \in \mathcal{N}_2$ and all $t \in \mathbb{R}_+$, define

$$w_u^1(t) = \begin{cases} w^1(t + t_u^1) & \text{if } t \in \left[0, 3 - t_u^1\right] \\ w^2\left(t - (3(2k+1) - t_u^1)\right) & \text{if } t \in \left(3(2k+1) - t_u^1, 3(2k+2) - t_u^1\right] \\ w^1\left(t - (3(2k+2) - t_u^1)\right) & \text{if } t \in \left(3(2k+2) - t_u^1, 3(2k+3) - t_u^1\right] \end{cases}$$

for all $k \in \mathbb{N}$. The walk w_u^1 starts at t_u^1 and alternates between following w^1 and w^2. The walk w_u^2 is defined analogously by switching the superscripts 1 and 2 in the definition above. Denote μ^0 the uniform choice of a walk in $(w_u^i)_{u \in \mathcal{N}_2}^{i \in \{1,2\}}$. It is not difficult to check that μ^0 guarantees $m/3$ to the patroller (moreover μ^0 yields a payoff of $m/3$ for every $(y, t) \in \mathcal{A}$). Hence $V_{\mathcal{N}_2}(m, 0) = \frac{m}{3}$.

Second case: $2 < m < 4$. We detail the computation for $m = 3$. The walks w^3, w^4 and w^5 hereafter can be adapted and similar strategies can be used to derive the bounds for all $m \in (2, 4)$.

Let us define three paths π^3, π^4 and π^5 as in Figs. 4.4–4.6 respectively. That is, $\pi^3 = (u_1, u_2, u_3, u_5, u_3, u_1, u_6, u_1)$, $\pi^4 = (u_1, u_7, u_1, u_3, u_8, u_3, u_4, u_1)$ and $\pi^5 = (u_1, u_2, u_3, u_{10}, u_3, u_4, u_1, u_9, u_1)$. Where $u_5 = (u_3, u_4, 1/2)$, $u_6 = (u_1, u_4, 1/2)$, $u_7 = (u_1, u_2, 1/2)$, $u_8 = (u_2, u_3, 1/2)$, $u_9 = (u_1, u_3, 1/4)$ and $u_{10} = (u_1, u_3, 3/4)$.

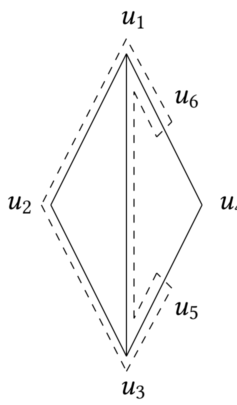

Figure 4.4: The path π^3

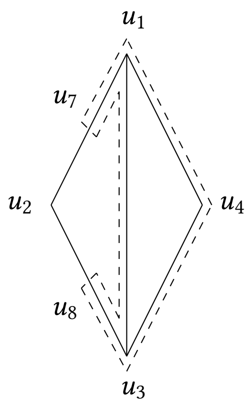

Figure 4.5: The path π^4

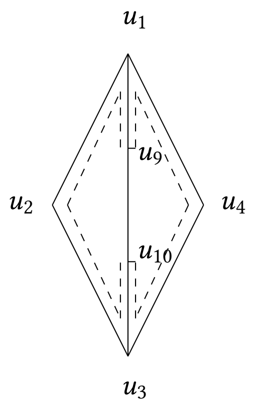

Figure 4.6: The path π^5

π^3, π^4 and π^5 naturally induce three 3-periodic walks at speed 1, denoted respectively w^3, w^4 and w^5. These are such that for $i \in \{3, 4, 5\}$, w^i intercepts any attack on $w^i([0, 3])$ with probability 1.

With a slight abuse of notation, for $y \in [0, 1/2]$, denote y the point (u_1, u_3, y) and $1 - y$ the point $(u_1, u_3, 1 - y)$. By symmetry it is enough to consider attacks occurring at y. Moreover, μ^0, w^3, w^4 and w^5 make the patroller time indifferent, hence we only consider attacks at time 0.

μ^0 intercepts the attack $(y, 0)$ with probability $1 - \frac{1-2y}{6} = \frac{5}{6} + \frac{y}{3}$. Indeed, only the walks w_u^1, such that u belongs to the open interval $\{(u_1, u_3, \alpha) \mid \alpha \in (y, 1-y)\}$ do not intercept the attack. Finally, define $\widetilde{\mu} = \frac{1}{15}(\delta_{w^3} + \delta_{w^4} + \delta_{w^5}) + \frac{4}{5}\mu^0$, where δ_w is the Dirac measure at $w \in \mathcal{W}$.

At any time, an attack at $y \leq \frac{1}{4}$ is intercepted by $\widetilde{\mu}$ with probability

$$\frac{1}{15} \cdot 3 + \frac{4}{5}\left(\frac{5}{6} + \frac{y}{3}\right) \geq \frac{3}{15} + \frac{4}{5} \cdot \frac{5}{6} = \frac{13}{15}.$$

An attack at $y > 1/4$ is intercepted by $\widetilde{\mu}$ with probability

$$\frac{1}{15} \cdot 2 + \frac{4}{5}\left(\frac{5}{6} + \frac{y}{3}\right) \geq \frac{2}{15} + \frac{4}{5}\left(\frac{5}{6} + \frac{1}{12}\right) = \frac{13}{15}.$$

Hence $V_{\mathcal{N}_2}(3, 0) \geq \frac{13}{15}$.

Define the following attack $\widetilde{a}$: choose uniformly a point in $\mathcal{N}_2 \times [0, 3]$. The tour $(u_1, u_2, u_3, u_1, u_4, u_3, u_2, u_1, u_3, u_4, u_1)$ induces a 6-periodic walk w^6 which is a best response for the patroller. Moreover $g_{3,0}(w^6, \widetilde{a}) = 11/12$. Hence $V_{\mathcal{N}_2}(3, 0) \leq \frac{11}{12}$.

Third case: $m \geq 4$. The tour $(u_1, u_2, u_3, u_1, u_4, u_3, u_1)$ induces a 4-periodic walk which guarantees 1 to the patroller. Hence $V_{\mathcal{N}_2}(m, 0) = 1$.

4.3.5 Patrolling a simple search space in $\mathbb{R}^2$

In this section, we are interested in patrolling games in $\mathbb{R}^2$ for a large class of search spaces called simple search spaces. To introduce this class of search spaces, we first need to recall the notion of total variation of a function f.

Definition 4.11. Let $a > 0$. Let $f : [0, a] \to \mathbb{R}^n$ be a continuous function. Then the total variation of f is the quantity:

$$TV(f) = \sup \left\{ \sum_{i=1}^{n} \|f(t_i) - f(t_{i-1})\|_2 \;\middle|\; n \in \mathbb{N}^*, 0 = t_0 < t_1 < \cdots < t_n = a \right\}.$$

If $TV(f) < +\infty$, then f is said to have bounded variation.

The next definition introduces a classical assumption on the boundary of a search space in $\mathbb{R}^2$. This is a weak assumption already made in (Gal, 1980) and (Alpern and Gal, 2003).

Definition 4.12. Let $a > 0$, let f_1 and f_2 be two continuous functions from $[0, a]$ to $\mathbb{R}$ such that $f_1 \geq f_2$, $f_1 \neq f_2$, and f_1 and f_2 have bounded variation. Then the nonempty compact set $\{(x, t) \in [0, a] \times \mathbb{R} \mid f_2(x) \leq t \leq f_1(x)\}$ is called an elementary search space.

Let Q be the finite union of elementary search spaces such that any two have disjoint interiors. If Q is path-connected, then it is called a simple search space.

The next theorem is the main result on patrolling games on simple search spaces. It gives a simple asymptotic expression of the value as the detection radius r goes to 0. The result relies on the fact that the patroller can use a uniform strategy in the spirit of what has been done in the previous section for Eulerian networks. This strategy yields a lower bound that asymptotically matches the upper bound of Proposition 4.4.

As one would expect, the value goes to 0 as r goes to 0. It is interesting to note that due to the movement of the patroller the convergence is linear in r and not quadratic. Indeed, the relevant parameter is the sweep width of the patroller and not the area of detection.

Theorem 4.13. *If Q is a simple search space endowed with the Euclidean norm, then*

$$V_Q(m, r) \sim \frac{2rm}{\lambda(Q)},$$

as r goes to 0.

The proof of Theorem 4.13 is provided in the Appendix of Chapter 4.

4.4 Hiding games

Recall that the value of a hiding game is equal to the value of a patrolling game with time limit m equal to 0. Hiding games have a value which represents the probability (up to ε) that the searcher and the hider are at distance less that r when they play (ε-)optimally.

Proposition 4.14. *The hiding game (Q, r) played in mixed strategies has a value denoted $V_Q(r)$. Moreover the searcher has an optimal strategy and the hider has an ε-optimal strategy with finite support.*

4.4.1 Equalizing strategies

We now study particular strategies called equalizing, these have been introduced by (Bishop and Evans, 2013, definition 7.3 and proposition 7.3) when the search space is a graph. We adapt those considerations to our compact setting. A similar notion also appears in (Lidbetter, 2013, lemma 2.6).

In hiding games, both players have the same strategy sets and the payoff function is symmetric in the sens that, if $\mu \in \Delta(Q)$ and $y \in Q$, then $h_r(\mu, y) = h_r(y, \mu) = \mu(B_r(y) \cap Q)$.

A strategy for one player is equalizing if the payoff, when this player uses this strategy, does not depend on the strategy of the other player. The interest of equalizing strategies lies in the fact that if such a strategy exists, then it is optimal for both players.

Definition 4.15. Let Q be a search space. A strategy $\mu \in \Delta(Q)$ is said to be equalizing if there exists $c \in [0, 1]$ such that $h_r(\mu, y) = c$ for all $y \in Q$.

Proposition 4.16. *Let $\mu \in \Delta(Q)$. Then μ is an equalizing strategy (with constant payoff c) if and only if μ is optimal for both players (and in that case $V_Q(r) = c$).*

Proof. Suppose $\mu \in \Delta(Q)$ is an equalizing strategy. If the searcher plays μ, then for all $y \in Q$ $\mu(B_r(y) \cap Q) = c$, hence $V_Q(r) \geq c$. Symmetrically, if the hider plays μ, then for all $x \in Q$ $\mu(B_r(x) \cap Q) = c$, hence $V_Q(r) \leq c$, and $V_Q(r) = c$.

Conversely, suppose $\mu \in \Delta(Q)$ is optimal for both players. Then the searcher guaranties $V_Q(r)$ that is for all $y \in Q$ $\mu(B_r(y) \cap Q) \geq V_Q(r)$, and the hider guaranties $V_Q(r)$ that is for all $x \in Q$ $\mu(B_r(x) \cap Q) \leq V_Q(r)$. Hence for all $y \in Q$

$$\mu(B_r(y) \cap Q) = V_Q(r).$$

□

The following game is an example of a hiding game with finite search space without equalizing strategies.

Example 4.2. Let $r = 1$ and $Q = \{x_1, x_2, x_3, x_4, x_5\}$ be the finite subset of $\mathbb{R}^2$ such that $x_1 = (0, 0)$, $x_2 = (0, 1)$, $x_3 = (1, 1)$, $x_4 = (1, 0)$ and $x_5 = (1/2, 0)$. Denote for $i \in \{1, \ldots, 5\}$ $Q_i = \{j \in \{1, \ldots, 5\} \mid \left\|x_i - x_j\right\|_2 \leq r\}$. That is $Q_1 = \{1, 2, 4, 5\}$, $Q_2 = \{1, 2, 3\}$, $Q_3 = \{2, 3, 4\}$, $Q_4 = \{1, 3, 4, 5\}$ and $Q_5 = \{1, 4, 5\}$.

The game (Q, r) admits an equalizing strategy if and only if the following system of equations admits a solution $p = (p_i)_{1 \leq i \leq 5}$:

$$\begin{cases} p_i \geq 0 \text{ for all } i \in \{1, \ldots, 5\} \\ \sum_{i=1}^{5} p_i = 1 \\ \sum_{i \in Q_1} p_i = \sum_{i \in Q_j} p_i \text{ for all } j \in \{2, \ldots, 5\}. \end{cases} \tag{4.4.1}$$

It is easy to verify that this system does not admit a solution, hence the game (Q, r) does not have an equalizing strategy.

4.4.2 An asymptotic result for hiding games

The next theorem is the main result on hiding games. For any search space $Q \subset \mathbb{R}^n$ with positive Lebesgue measure, it gives a simple asymptotic expression of the value when the detection radius goes to 0. In this static setting the value is equivalent, as r goes to 0, to the ratio of the volume of the ball of radius r over the volume of Q. This result relies on the fact that the searcher has a strategy that yields a lower bound which asymptotically matches the upper bound of Proposition 4.4.

Theorem 4.17. *Let Q be a compact subset of $\mathbb{R}^n$. Suppose $\lambda(Q) > 0$. Then*

$$V_Q(r) \sim \frac{\lambda(B_r)}{\lambda(Q)}$$

as r goes to 0.

The proof of Theorem 4.17 is provided in the Appendix of Chapter 4. A consequence of Theorem 4.17 is that for a compact set Q included in $\mathbb{R}^n$ such that $\lambda(Q) > 0$, $V_Q(r) \sim r^n \frac{\lambda(B_1)}{\lambda(Q)}$ as r goes to 0. When $\lambda(Q) = 0$, it is not always the case that V_Q admits an equivalent of the form Mr^α, with α and M positive, as r goes to 0, as it is shown in Example 4.3.

Example 4.3. Let $Q \subset [0,1]$ be the following Cantor-type set. Define $C_0 = [0,1]$, and for all $n \in \mathbb{N}^*$ $C_n = \frac{1}{4}C_{n-1} \cup \left(\frac{3}{4} + \frac{1}{4}C_{n-1}\right)$. Finally, let $Q = \bigcap_{n\in\mathbb{N}} C_n$. Q is compact and $\lambda(Q) = 0$.

The value of the hiding game played on Q is given by the following formula:

$$V_Q(r) = \begin{cases} \frac{1}{2^n} & \text{if } r \in \left[\frac{1}{2^{2n}}, \frac{3}{2^{2n}}\right), \\ \frac{1}{2^{n-1}} & \text{if } r \in \left[\frac{3}{2^{2n}}, \frac{1}{2^{2(n-1)}}\right), n \in \mathbb{N}^*. \end{cases}$$

Indeed, let $\Sigma_1 = \{0,1\}$ and for all $n \in \mathbb{N}^* \setminus \{1\}$ let $\Sigma_n = \frac{1}{4}\Sigma_{n-1} \cup \left(\frac{3}{4} + \frac{1}{4}\Sigma_{n-1}\right)$. For $n \in \mathbb{N}^*$, consider the following strategy σ_n: choose uniformly a point in Σ_n, that is with probability $\frac{1}{|\Sigma_n|} = \frac{1}{2^n}$. Let $n \in \mathbb{N}^*$ suppose $r \in \left[\frac{1}{2^{2n}}, \frac{3}{2^{2n}}\right)$. Then for all $q \in Q$ there is exactly one point s in Σ_n such that $|q-s| \leq r$. Hence σ_n is an equalizing strategy which guarantees $\frac{1}{2^n}$ to both players.

Let $\Sigma'_1 = \{\frac{1}{4}\}$ and for all $n \in \mathbb{N}^* \setminus \{1\}$ let $\Sigma'_n = \frac{1}{4}\Sigma'_{n-1} \cup \left(1 - \frac{1}{4}\Sigma'_{n-1}\right)$. For $n \in \mathbb{N}^*$ consider the following strategy σ'_n: choose uniformly a point in Σ'_n, that is with probability $\frac{1}{|\Sigma'_n|} = \frac{1}{2^{n-1}}$. Suppose now that $r \in \left[\frac{3}{2^{2n}}, \frac{1}{2^{2(n-1)}}\right)$. Then for all $q \in Q$ there is exactly one point s in Σ'_n such that $|q-s| \leq r$. Hence σ'_n is an equalizing strategy which guarantees $\frac{1}{2^{n-1}}$ to both players.

In particular, for all $n \in \mathbb{N}^*$

$$V_Q\left(\frac{1}{2^{2n-1}}\right) = V_Q\left(\frac{1}{2^{2n}}\right) = \frac{1}{2^n}.$$

Let $(r_n)_{n\in\mathbb{N}^*} = \left(\frac{1}{2^n}\right)_{n\in\mathbb{N}^*}$ and let $\alpha > 0$. Then for all $n \in \mathbb{N}^*$

$$\frac{V_Q(r_{2n-1})}{(r_{2n-1})^\alpha} = \frac{1}{2^\alpha}2^{(2\alpha-1)n} \text{ and } \frac{V_Q(r_{2n})}{(r_{2n})^\alpha} = 2^{(2\alpha-1)n}.$$

Thus we have

$$\lim_{n\to+\infty} \frac{V_Q(r_{2n-1})}{(r_{2n-1})^\alpha} = \begin{cases} +\infty & \text{if } \alpha > 1/2 \\ \frac{1}{\sqrt{2}} & \text{if } \alpha = 1/2 \\ 0 & \text{if } \alpha < 1/2 \end{cases} \text{ and } \lim_{n\to+\infty} \frac{V_Q(r_{2n})}{(r_{2n})^\alpha} = \begin{cases} +\infty & \text{if } \alpha > 1/2 \\ 1 & \text{if } \alpha = 1/2 \\ 0 & \text{if } \alpha < 1/2. \end{cases}$$

Hence $r \mapsto V_Q(r)$ does not admit an equivalent of the form $r \mapsto Mr^\alpha$, with α and M positive numbers, as r goes to 0.

4.5 Properties of the value function of patrolling and hiding games

In this section we give some elementary properties of the function V_Q for patrolling and hiding games.

4.5.1 The value function of patrolling games

Proposition 4.18. *Let Q be a search space. The function*

$$\begin{array}{rcccc} V_Q & : & \mathbb{R}_+ \times \mathbb{R}_+ & \to & [0,1] \\ & & (m,r) & \mapsto & V_Q(m,r) \end{array}$$

is

i) non decreasing in m and r;

ii) upper semi-continuous in r for all m;

iii) upper semi-continuous in m for all r.

Proof. Since i) is direct we only prove ii). For all $(m,r) \in \mathbb{R}_+^2$,

$$\begin{aligned} V_Q(m,r) &= \max_{\mu \in \Delta(\mathcal{W})} \inf_{(y,t) \in \mathcal{A}} \int_{\mathcal{W}} g_{m,r}(w,(y,t)) d\mu(w) \\ &= \inf_{(y,t)\in\mathcal{A}} \int_{\mathcal{W}} g_{m,r}(w,(y,t)) d\mu^*(w), \end{aligned}$$

where $\mu^* \in \Delta(\mathcal{W})$ is an optimal strategy of the patroller. Let $(y,t) \in \mathcal{A}$ and $m \geq 0$. For all $w \in \mathcal{W}$, the function $r \mapsto g_{m,r}(w,(y,t))$ is upper semi-continuous, as the indicator function of a closed set. Let $r_n \to r$, then by Fatou's lemma,

$$\begin{aligned} \limsup_n \int_W g_{m,r_n}(w,(y,t)) d\mu^*(w) &\leq \int_W \limsup_n g_{m,r_n}(w,(y,t)) d\mu^*(w) \\ &\leq \int_{\mathcal{W}} g_{m,r}(w,(y,t)) d\mu^*(w). \end{aligned}$$

Thus the function $r \mapsto \int_{\mathcal{W}} g_{m,r}(w,(y,t)) d\mu^*(w)$ is upper semi-continuous. Hence

$$V_Q(m,\cdot) : r \mapsto \inf_{(y,t)\in\mathcal{A}} \int_{\mathcal{W}} g_{m,r}(w,(y,t)) d\mu^*(w)$$

is upper semi-continuous.

Since for all $w \in \mathcal{W}$, the function $m \mapsto g_{m,r}(w,(y,t))$ is upper semi-continuous, as the indicator function of a closed set, the proof of iii) is strictly analogous. □

Example 4.4 in the next section shows that in general, for fixed m, $V_Q(\cdot, m)$ is not lower semi-continuous.

Remark 4.1. Let $m, r \geq 0$, and Q_1, Q_2 be two search spaces, it is clear that if $Q_1 \subset Q_2$ then the attacker is better off in Q_2 hence $V_{Q_1}(m,r) \geq V_{Q_2}(m,r)$.

4.5.2 The value function of hiding games

Recall that the value of a hiding game is equal to the value of a patrolling game with time limit m equal to 0. Hence, the negative results presented in this section also hold for patrolling games when $m = 0$.

The following simple example of a hiding game on the unit interval was first solved by Ruckle (1983). It shows that in general, V_Q is not lower semi-continuous.

Example 4.4. Let Q be the $[0, 1]$ interval, then

$$V_Q(r) = \sigma(s, i) = \begin{cases} \min\left(\lceil \frac{1}{2r} \rceil^{-1}, 1\right) & \text{if } r > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Indeed, it is clear when r equals 0 and $r \geq 1/2$. Let $n \in \mathbb{N}^*$ and suppose $r \in \left[\frac{1}{2(n+1)}, \frac{1}{2n}\right)$. Then the patroller guarantees $\frac{1}{n+1}$ by choosing equiprobably a point in $\left\{\frac{1+2k}{2(n+1)}\right\}_{0 \leq k \leq n}$. And the attacker, choosing equiprobably a point in $\left\{\frac{(2+\varepsilon)k}{2(n+1)}\right\}_{0 \leq k \leq n}$, with $0 < \varepsilon \leq 2/n$, also guarantees $\frac{1}{n+1}$.

The next proposition disproves the somehow intuitive belief that the value of hiding games is continuous with respect to the Haussdorff metric between nonempty compact sets.

Proposition 4.19. *Let $r \geq 0$. The function which maps any search space Q to $V_Q(r)$ is in general not continuous with respect to the Hausdorff metric between nonempty compact sets.*

Proof. Let $D_s = \{x \in \mathbb{R}^2 \mid \|x\|_2 \leq s\}$ be the Euclidean disc of radius $s > 0$ centered at 0. From Danskin (1990), it is known that

$$V_{D_s}(1) = \begin{cases} 1 & \text{if } s \in [0, 1] \\ \frac{1}{\pi} \arcsin\left(\frac{1}{s}\right) & \text{if } s \in \left(1, \sqrt{2}\right]. \end{cases}$$

Hence $\lim_{s \to 1, s > 1} V_{D_s}(1) = \frac{1}{2} < 1$.

The intuition is the following: it is clear that when s equals 1 the searcher guarantees 1 by playing $x = (0, 0)$. Suppose now that s equals $1 + \varepsilon$. Then the searcher covers almost all the area of the disc but less than half of its circumference. Hence the hider guarantees 1/2 by choosing uniformly a point on the boundary of D_s. □

Appendix of Chapter 4: omitted proofs

Omitted proofs of Section 4.3.1

Let us first define a metric D on the set $\mathcal{W}$ inducing the topology of compact convergence. For $n \in \mathbb{N}$, define $K_n = [0, n]$. Then

$$\begin{array}{rcccc} D & : & \mathcal{W} \times \mathcal{W} & \to & \mathbb{R}_+ \\ & & (f, g) & \mapsto & \sum_{n=1}^{\infty} \frac{1}{2^n} \sup_{x \in K_n} \|f(x) - g(x)\|, \end{array}$$

is a metric on $\mathcal{W}$, which induces the topology of compact convergence.

We recall the following fact about the topology of compact convergence.

Proposition 4.20 (Application of Theorem 46.2 in (Munkres, 2000)**).** *Let Q be a search space. A sequence $f_n : \mathbb{R}_+ \to Q$ of functions converges to the function f in the topology of compact convergence if and only if for each compact subspace K of $\mathbb{R}_+$, the sequence $f_n|_K$ converges uniformly to $f|_K$.*

The following corollary follows from Sion's theorem, Sion (1958).

Corollary 4.21 (Proposition A.10 in (Sorin, 2002)**).** *Let (X, Y, g) be a zero-sum game, i.e., X and Y are the action sets of player 1 and 2 respectively and g is the payoff to player 1, such that: X is a compact metric space, for all $y \in Y$, the function $g(\cdot, y)$ is upper semi-continuous. Then the game $(\Delta(X), \Delta_f(Y), g)$ has a value and player 1 has an optimal strategy.*

We are now able to complete the proof of Proposition 4.1.

Proof of Proposition 4.1. By Ascoli's theorem (application of (Munkres, 2000, Theorem 47.1)), $\mathcal{W}$ is compact for the topology of compact convergence. Moreover, for all $(y, t) \in \mathcal{A}$ the function $g_{m,r}(\cdot, (y, t))$ is upper semi-continuous. The conclusion follows from Corollary 4.21. □

Omitted proofs of Section 4.3.4

Proof of Lemma 4.8. Let $\pi = (u_1, u_2, \ldots, u_{n-1}, u_n)$, $u_1 = u_n = u \in V$ be an Eulerian tour. Without loss of generality, suppose $l(u_i, u_{i+1}) \neq 0$ for all $i \in \{1, \ldots, n-1\}$. The parametrization is constructed in the following way.

If $t \in [0, l(u_1, u_2)]$ then

$$w(t) = \left(u_1, u_2, \frac{t}{l(u_1, u_2)}\right).$$

Else, suppose $n \geq 3$. For all $k \in \{2, \ldots, n-1\}$ if

$$t \in \left(\sum_{i=1}^{k-1} l(u_i, u_{i+1}), \sum_{i=1}^{k} l(u_i, u_{i+1})\right]$$

then

$$w(t) = \left(u_k, u_{k+1}, \frac{t - \sum_{i=1}^{k-1} l(u_i, u_{i+1})}{l(u_k, u_{k+1})}\right).$$

It is not difficult to verify that such w is appropriate. □

Proof of Theorem 4.10. If $m \geq \lambda(\mathcal{N})$, the patroller guarantees 1 by playing a parametrization of $\mathcal{N}$. Suppose that $m < \lambda(\mathcal{N})$. Let $(y, t) \in \mathcal{N} \times \mathbb{R}_+$ be a pure strategy of the attacker and let w be as in Definition 4.7. There exists $t_y \in [0, \lambda(\mathcal{N})]$ such that $w(t_y) = y$. Now let $t_0 \in [t_y - t - m, t_y - t]$. Then $w_{t_0}(t_y - t_0) = w(t_y) = y$. And $t_y - t_0 \in [t, t + m]$. Thus $y \in w_{t_0}([t, t + m])$. Hence under the patroller's uniform strategy

$$\mathbb{P}(y \in w_{t_0}([t, t + m])) \geq \mathbb{P}(t_0 \in [t_y - t - m, t_y - t]) = \frac{m}{\lambda(\mathcal{N})}.$$

The other inequality follows from Proposition 4.4 since in this case, ρ equals 1. □

Omitted proofs of Section 4.3.5

To prove Theorem 4.13 we first need some preliminary definition and lemmas.

Definition 4.22. Let Q be a search space. A continuous function $L : [0, 1] \longrightarrow Q$ such that $L(0) = L(1)$ is called an r-tour if for any $x \in Q$ there exists $l \in L([0, 1])$ such that $d(x, l) \leq r$.

The next lemma shows that when the radius of detection r is small, one can find in Q an r-tour with length not exceeding $\lambda(Q)/2r$, up to some ε.

Lemma 4.23 (Lemma 3.39 in (Alpern and Gal, 2003)). *Let $Q \subset \mathbb{R}^2$ be a simple search space. Endow Q with the Euclidean norm. Then for any $\varepsilon > 0$ there exits $r_\varepsilon > 0$ such that for any $r < r_\varepsilon$ there exists an r-tour $L : [0, 1] \longrightarrow Q$ such that*

$$TV(L) \leq (1 + \varepsilon)\frac{\lambda(Q)}{2r}.$$

The next lemma gives a parametrization of $L([0, 1])$ in terms of walks.

Lemma 4.24. *Let L be an r-tour as in Lemma 4.23. Then for all $\varepsilon' > 0$ there exists $w : [0, TV(L) + \varepsilon'] \longrightarrow L([0, 1])$ continuous such that:*

i) $w(0) = w(TV(L) + \varepsilon')$;

ii) w is surjective;

iii) w is 1-Lipschitz continuous;

iv) $TV(w) = TV(L)$.

w is extended to a $(TV(L) + \varepsilon')$-periodic function on $\mathbb{R}$, which is still denoted w.

Proof of Lemma 4.24. Let $\varepsilon' > 0$ and let

$$\begin{array}{rcrcl} f & : & [0,1] & \longrightarrow & [0, TV(L)+\varepsilon'] \\ & & s & \longmapsto & TV(L|_{[0,s]}) + \varepsilon' s. \end{array}$$

The function f is increasing and continuous on $[0,1]$, hence f is an homeomorphism. Define w as $L \circ f^{-1}$ on $[0, TV(L)+\varepsilon']$. It is not difficult to prove that such w verifies the condition of the lemma. □

We are now able to prove Theorem 4.13.

Proof of Theorem 4.13. Let L and w be as in Lemmas 4.23 and 4.24 respectively. For all $t_0 \in [0, TV(L)+\varepsilon']$ define $w_{t_0}(\cdot)$ as $w(t_0 + \cdot)$.

Let $(l,t) \in L([0,1]) \times \mathbb{R}_+$. By Lemma 4.24 ii), there exists $t_l \in [0, TV(L)+\varepsilon']$ such that $w(t_l) = l$. Now let $t_0 \in [t_l - t - m, t_l - t]$. Then $w_{t_0}(t_l - t_0) = w(t_l) = l$. And $t_l - t_0 \in [t, t+m]$. Hence $l \in w_{t_0}([t, t+m])$.

Suppose t_0 is chosen uniformly in $[0, TV(L)+\varepsilon']$. By Lemma 4.24 iii) this is an admissible strategy for the patroller. Let $(y,t) \in \mathcal{A}$ be a pure strategy of the attacker. Then if $l \in L([0,1])$ is such that $d(y,l) \le r$,

$$\begin{aligned} \mathbb{P}(d(y, w_{t_0}([t,t+m])) \le r) &\ge \mathbb{P}(l \in w_{t_0}([t,t+m])), \\ &\ge \mathbb{P}(t_0 \in [t_l - t - m, t_l - t]) \\ &= \frac{m}{TV(L)+\varepsilon'}. \end{aligned}$$

By Lemma 4.23, this last quantity is greater than or equal to $\frac{m}{\frac{(1+\varepsilon)\lambda(Q)}{2r}+\varepsilon'}$. Hence the patroller guarantees $\frac{m}{\frac{(1+\varepsilon)\lambda(Q)}{2r}+\varepsilon'}$ for all $\varepsilon' > 0$, that is

$$V_Q(m,r) \ge \frac{2rm}{(1+\varepsilon)\lambda(Q)} \sim \frac{2rm}{\lambda(Q)}$$

as r goes to 0.

In this context, Proposition 4.4 yields $V_Q(m,r) \le \frac{2rm+\pi r^2}{\lambda(Q)} \sim \frac{2rm}{\lambda(Q)}$ as r goes to 0. □

Omitted proofs of Section 4.4.2

To prove Theorem 4.17 we first need to introduce a technical lemma.

Denote $B^2_r(x) = \{y \in \mathbb{R}^n \mid \|x-y\|_2 \le r\}$ the closed ball of center x with radius r for the Euclidean norm, and $\partial B^2_r(x) = \{y \in \mathbb{R}^n \mid \|x-y\|_2 = r\}$ the sphere of center x with radius r for the Eucliean norm.

The intuition behind Lemma 4.25 below is the following. We consider the balls $B^2_\varepsilon(0)$ and $B^2_r(x)$ with x on the boundary of $B^2_\varepsilon(0)$. When r goes to zero, the ratio between the volume of the ball $B^2_r(x)$ and the ball $B^2_r(x)$ intersected with the ball $B^2_\varepsilon(0)$ goes to 2. Lemma 4.25 gives an upper bound to this ratio, as r goes to 0, for a non necessary Euclidean ball $B_r(x)$.

Lemma 4.25. *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ and $c_1, c_2 > 0$ be such that $c_1 \|\cdot\| \leq \|\cdot\|_2 \leq c_2 \|\cdot\|$. Then for all $x \in \partial B^2_\varepsilon(0)$*

$$\limsup_{r\to 0} \frac{\lambda(B_r)}{\lambda(B^2_\varepsilon(0) \cap B_r(x))} \leq 2\left(\frac{c_2}{c_1}\right)^n .$$

Proof of Lemma 4.25. Let $x \in \partial B^2_\varepsilon(0)$, let $\varepsilon > 0$. Denote I the regularized incomplete Beta function: for $a, b > 0$ and $0 < z < 1$, $I_z(a,b) = \frac{B(z;a,b)}{B(a,b)}$. Where $B(z;a,b) = \int_0^z t^{a-1}(1-t)^{b-1}dt$ and $B(a,b) = B(1;a,b)$ is the Beta function. Then we have, see Li (2011),

$$\lambda\left(B^2_\varepsilon(0) \cap B^2_r(x)\right) = \frac{\pi^{n/2}}{2\Gamma(\frac{n}{2}+1)}\left(r^n I_{1-(\frac{r}{2\varepsilon})^2}\left(\frac{n+1}{2}, \frac{1}{2}\right) \right.$$
$$\left. + \varepsilon^n I_{(\frac{r}{\varepsilon})^2\left(1-(\frac{r}{2\varepsilon})^2\right)}\left(\frac{n+1}{2}, \frac{1}{2}\right)\right).$$

Since $t \mapsto t^{\frac{n-1}{2}}(1-t)^{-1/2}$ is integrable over $[0,1)$,

$$I_{1-(\frac{r}{2\varepsilon})^2}\left(\frac{n+1}{2}, \frac{1}{2}\right) = \frac{\int_0^{1-(\frac{r}{2\varepsilon})^2} t^{\frac{n-1}{2}}(1-t)^{-1/2}dt}{B(\frac{n-1}{2}, \frac{1}{2})} \longrightarrow 1,$$

as r goes to 0. And,

$$I_{(\frac{r}{\varepsilon})^2\left(1-(\frac{r}{2\varepsilon})^2\right)}\left(\frac{n+1}{2}, \frac{1}{2}\right) = \frac{\int_0^{(\frac{r}{\varepsilon})^2\left(1-(\frac{r}{2\varepsilon})^2\right)} t^{\frac{n-1}{2}}(1-t)^{-1/2}dt}{B(\frac{n-1}{2}, \frac{1}{2})}$$

which, since $1 \leq (1-t)^{-1/2}$ when $t \in [0,1)$, is greater than

$$\frac{\frac{2}{n+1}\left(\frac{r}{\varepsilon}\right)^{n+1}\left(1-\left(\frac{r}{2\varepsilon}\right)^2\right)^{\frac{n+1}{2}}}{B(\frac{n-1}{2}, \frac{1}{2})} = \frac{2r^{n+1}}{(n+1)\varepsilon^{n+1}B(\frac{n-1}{2}, \frac{1}{2})} + o(r^{2n+2})$$

when r goes to 0. Hence we have

$$\lambda\left(B^2_\varepsilon(0) \cap B^2_r(x)\right) \geq \frac{\pi^{n/2}}{2\Gamma(\frac{n}{2}+1)}(r^n + o(r^n))$$

as r goes to 0. Moreover since

$$\begin{aligned} B^2_\varepsilon(0) \cap B_r(x) &= \{y \in \mathbb{R}^n \mid \|y\|_2 \leq \varepsilon \text{ and } \|x-y\| \leq r\} \\ &\supset \{y \in \mathbb{R}^n \mid \|y\|_2 \leq \varepsilon \text{ and } \|x-y\|_2 \leq c_1 r\}, \end{aligned}$$

and $B_r(0) \subset B^2_{c_2 r}(0)$, we have $\lambda\left(B^2_\varepsilon(0) \cap B_r(x)\right) \geq \lambda\left(B^2_\varepsilon(0) \cap B^2_{c_1 r}(x)\right)$, and $c_2^n \lambda(B^2_r) \geq \lambda(B_r)$. Finally, dividing by $\lambda\left(B^2_\varepsilon(0) \cap B_r(x)\right)$ and taking the lim sup, since $\lambda(B^2_r) = \frac{\pi^{n/2} r^n}{\Gamma(\frac{n}{2}+1)}$ we have

$$\limsup_{r\to 0} \frac{\lambda(B_r)}{\lambda\left(B^2_\varepsilon(0) \cap B_r(x)\right)} \leq \limsup_{r\to 0} \frac{c_2^n \lambda(B^2_r)}{\lambda\left(B^2_\varepsilon(0) \cap B^2_{c_1 r}(x)\right)} \leq 2\left(\frac{c_2}{c_1}\right)^n .$$

□

We are now able to prove Theorem 4.17.

Proof of Theorem 4.17. Let $\varepsilon > 0$ and $r \in (0, \varepsilon)$. We regularize the boundary of Q by defining $Q_\varepsilon = Q + B^2_\varepsilon(0)$, and $I^\varepsilon(r) = \{y \in Q_\varepsilon \mid B_r(y) \subset Q_\varepsilon\}$. Define as well $\lambda^\varepsilon_{\min}(r) = \min_{y \in Q_\varepsilon} \lambda(B_r(y) \cap Q_\varepsilon)$. Finally define $\mu \in \Delta(Q_\varepsilon)$ such that for all $B \subset Q_\varepsilon$ measurable

$$\mu(B) = \frac{\lambda\left(B \cap I^\varepsilon(r)\right)\lambda^\varepsilon_{\min}(r) + \lambda(B \cap (Q_\varepsilon \setminus I^\varepsilon(r))\lambda(B_r)}{\lambda(I^\varepsilon(r))\lambda^\varepsilon_{\min}(r) + \lambda(B_r)\lambda(Q_\varepsilon \setminus I^\varepsilon(r))}.$$

Since by definition $\lambda(B_r) \geq \lambda^\varepsilon_{\min}(r)$, for all $x \in Q_\varepsilon$

$$\mu(B_r(x) \cap Q_\varepsilon) \geq \frac{\lambda^\varepsilon_{\min}(r)\lambda(B_r)}{\lambda(I^\varepsilon(r))\lambda^\varepsilon_{\min}(r) + \lambda(B_r)\lambda(Q_\varepsilon \setminus I^\varepsilon(r))}.$$

Because the hider can play in (Q_ε, r) as he would play in (Q, r), $V_{Q_\varepsilon}(r) \leq V_Q(r)$. By Proposition 4.4,

$$\frac{\lambda^\varepsilon_{\min}(r)\lambda(B_r)}{\lambda(I^\varepsilon(r))\lambda^\varepsilon_{\min}(r) + \lambda(B_r)\lambda(Q_\varepsilon \setminus I^\varepsilon(r))} \leq V_{Q_\varepsilon}(r) \leq V_Q(r) \leq \frac{\lambda(B_r)}{\lambda(Q)}.$$

Dividing by $\lambda(B_r)/\lambda(Q)$,

$$\frac{\lambda^\varepsilon_{\min}(r)\lambda(Q)}{\lambda(I^\varepsilon(r))\lambda^\varepsilon_{\min}(r) + \lambda(B_r)\lambda(Q_\varepsilon \setminus I^\varepsilon(r))} \leq \frac{V_Q(r)\lambda(Q)}{\lambda(B_r)} \leq 1. \tag{4.5.1}$$

Let us show that for all $\varepsilon > 0$ $\bigcup_{r>0} I^\varepsilon(r) = \mathring{Q}_\varepsilon$. Indeed, let $y \in \bigcup_{r>0} I^\varepsilon(r)$. There exists $r > 0$ such that $y \in I^\varepsilon(r)$. Thus there exists $r > 0$ such that $B_r(y) \subset Q_\varepsilon$. Conversely, let $y \in \mathring{Q}_\varepsilon$. There exists $r' > 0$ such that $B'_{r'}(y) \subset \mathring{Q}_\varepsilon$, where $B'_{r'}(y) = \{x \in \mathbb{R}^n \mid \|x - y\| < r'\}$. Take $0 < r < r'$, then $B_r(y) \subset \mathring{Q}_\varepsilon$ hence $y \in I^\varepsilon(r)$.

For all $r_1, r_2 > 0$ such that $r_1 > r_2$ one has $I^\varepsilon(r_1) \subset I^\varepsilon(r_2)$. Hence $\lim_{r \to 0} \lambda(I^\varepsilon(r)) = \lambda(\mathring{Q}_\varepsilon)$. Dividing by $\lambda^\varepsilon_{\min}(r)$ and letting r go to 0 in Eq. (4.5.1), by Lemma 4.25 one has, since the minimum in $\lambda^\varepsilon_{\min}(r)$ is reached on the boundary of a Euclidean ball,

$$\frac{\lambda(Q)}{\lambda(\mathring{Q}_\varepsilon) + 2\left(\frac{c_2}{c_1}\right)^n \lambda(\partial Q_\varepsilon)} \leq \liminf_{r \to 0} \frac{V_Q(r)\lambda(Q)}{\lambda(B_r)} \leq \limsup_{r \to 0} \frac{V_Q(r)\lambda(Q)}{\lambda(B_r)} \leq 1. \tag{4.5.2}$$

Let us show that $\bigcap_{\varepsilon>0} \mathring{Q}_\varepsilon = \bigcap_{\varepsilon>0} Q_\varepsilon = Q$. Indeed, let $y \in \bigcap_{\varepsilon>0} Q_\varepsilon$. For all $\varepsilon > 0$ $\min_{z \in Q} \|y - z\|_2 \leq \varepsilon$, hence $y \in Q$. Conversely, for all $\varepsilon > 0$ $Q \subset \mathring{Q}_\varepsilon$ hence $Q \subset \bigcap_{\varepsilon>0} \mathring{Q}_\varepsilon$. Moreover for all $\varepsilon_1, \varepsilon_2 > 0$ such that $\varepsilon_1 < \varepsilon_2$ one has $Q_{\varepsilon_1} \subset Q_{\varepsilon_2}$. Hence $\lim_{\varepsilon \to 0} \lambda(\mathring{Q}_\varepsilon) = \lambda(Q)$, $\lim_{\varepsilon \to 0} \lambda(Q_\varepsilon) = \lambda(Q)$ and $\lambda(\partial Q_\varepsilon) = \lambda(Q_\varepsilon) - \lambda(\mathring{Q}_\varepsilon)$ so $\lim_{\varepsilon \to 0} \lambda(\partial Q_\varepsilon) = 0$.

Letting $\varepsilon \to 0$ in Eq. (4.5.2), $1 = \frac{\lambda(Q)}{\lambda(Q)} \leq \lim_{r \to 0} \frac{V_Q(r)\lambda(Q)}{\lambda(B_r)} \leq 1$. ☐

Chapter 5

When Sally found Harry: A Stochastic search game

Article written in collaboration with Pr. Marco Scarsini, submitted, arXiv:1904.12852

Harry hides on an edge of a graph and does not move from there. Sally, starting from a known origin, tries to find him as soon as she can. Harry's goal is to be found as late as possible. At any given time, each edge of the graph is either active or inactive, independently of the other edges, with a known probability of being active. This situation can be modeled as a zero-sum two-person stochastic game. We show that the game has a value and we provide upper and lower bounds for this value. Finally, by generalizing optimal strategies of the deterministic case, we provide more refined results for trees and Eulerian graphs.

5.1 Introduction

5.1.1 The problem

In a typical search game a hider hides on a space and a searcher, starting from a specified point, searches for the hider, trying to find him as fast as possible. Often the space where the hider hides is assumed to be a network. In almost all existing versions of the game the network is fixed and all the edges are always available to the searcher. In real life it is often the case that some edges of the network are momentarily unavailable, for various reasons. For instance when the police are looking for a suspect in a city, it is possible that the presence of traffic, or civilians, or other unexpected obstacles, forces them to deviate from the planned path. Most often the obstacles on the network are not permanent, but vary with time. For instance, traffic may be intense in an area of the city at some time and in a different area at a different time. The vehicles involved in an accidents at some point get removed from the road and traffic goes back to normal.

Similar scenarios appear for instance when a rescue team is searching for miners in a mine. Explosions or landslides may force the rescuers to change the course of actions. Although in this case we do not have an adversarial hider, we can frame the situation as a zero-sum game, by considering the worst-case scenario.

It is clear that the stochastic elements that affect the shape of the network must be taken into account by both the hider and the searcher. Consider the set of edges available to a searcher at a specific time. If the edge that she would have chosen is unavailable, she has two options: she can either wait until the edge becomes available, or she can take a different edge. Her choice clearly depends on the probability that each edge is available, on the structure of the network, and on her position in the game.

5.1.2 Our contribution

We study a hide-search model where a hider (Harry) hides on an edge of a graph and a searcher (Sally) travels around the graph in search of Harry. Her goal is to find him as soon as possible.

The novelty of the model is that, due to various circumstances, at any given time, some edges may be unavailable, so the graph randomly evolves over time. At each stage, each edge e of the graph is, independently of the others, active with probability p_e and inactive with probability $1 - p_e$.

At the beginning of the game, Harry hides on one edge of his choice and is immobile for the rest of the game. Starting from an initial vertex, called the root of the graph, Sally chooses at each stage a vertex among those reachable through active edges in the neighborhood of her current vertex. The game ends when Sally traverses the edge where Harry is hidden, and his payoff is the number of stages needed for the game to end. So, Sally tries to minimize this time needed to find Harry and Harry aims at maximizing this time. This can be modeled as a zero-sum two-person game.

We first examine the deterministic version of the game when $p_e = 1$ for each edge e. This game has a value and optimal strategies. Analogously to well-known models in continuous time, we provide an upper and lower bound for this value, which correspond, for a fixed number of edges, to the value of games played on trees and on Eulerian graphs, respectively. We also characterize optimal strategies when the graph is either a tree or an Eulerian graph. We then turn to the stochastic framework and show that, even in this case, the game has a value for all positive p_e. We provide an upper and lower bound for this value and show that it converges to the value of the deterministic game when $p_e \to 1$ for each edge e. We consider some particular instances when all p_e are equal. We generalize optimal strategies of the deterministic setting to the stochastic one and obtain upper bounds on the value of the games played on binary trees and on parallel Eulerian graphs. The upper bounds are tight when Sally is restricted to some search trajectories. Finally we solve the stochastic search games played on the line and on the circle.

5.1.3 Related literature

Several types of hide-search games (HSGs) have been studied by various authors under different assumptions. von Neumann (1953) studied a discrete version of the model where a hider hides in a cell (i, j) of a matrix and a searcher chooses a row or column of the matrix; she finds the hider if the row or column contains the cell (i, j). The problem was framed as a two-person zero-sum game. Several variations of this discrete game were studied by various authors, among them Baston et al. (1990); Berry and Mensch (1986); Efron (1964); Gittins and Roberts (1979); Neuts (1963); Roberts and Gittins (1978); Sakaguchi (1973); Subelman (1981).

The search game with an immobile hider was introduced by Isaacs (1965). Beck and Newman (1970) considered a continuous HSG with a hider hiding on a line according to some distribution and a searcher, starting from an origin and moving at fixed speed, tries to find the hider as soon as possible. The continuous model was then generalized by Gal (1972, 1974); Gal and Chazan (1976), who, among other things extended the state space from a line to a plane.

More relevantly to our paper, some authors dealt with HSGs on a network. Among them, Bostock (1984) studied a discrete version of a continuous HSG proposed by Gal (1980). This game is played on a parallel multi-graph with three edges that join two vertices A and B and the searcher, starting from B has to find an immobile hider. The fact that the network has an odd number of parallel edges and, therefore, is not Eulerian makes the problem difficult to solve. Kikuta (1990, 1991) considered a HSG where the hider hides in one of n cells on a straight line and the searcher incurs some traveling cost. Anderson and Aramendia (1990) considered a HSG on a network and framed the problem as an infinite-dimensional linear program. Alpern (2008); Cao (1995); Dagan and Gal (2008); Gal (1979); Reijnierse and Potters (1993) examined HSGs on trees, Eulerian networks, and some more general classes. Alpern et al. (2008, 2009); Gal (2000); Kikuta (2004); Pavlović (1995) extended the analysis to more general networks. Alpern (2011) considered a find-and-fetch game on a tree where the searcher has to find a hider on a network and can travel at speed 1 to find him, and then has to return to the origin at a different speed. Alpern and Lidbetter (2013, 2019) replaced the usual pathwise search with what they call expanded search, where the searched area of a rooted network expands over different paths from the origin at different speeds chosen by the searcher, in such a way that the sum of the speeds is fixed. Alpern and Lidbetter (2015) dealt with a situation where the searcher can choose one of two speeds to travel and can detect the hider, when passing in front of him, only if she travels at the lower speed. Alpern (2017) considered a model where the hider can hide anywhere in a network and the searcher has to entirely traverse an edge before being able to turn around. This constraints gives the problem a more combinatorial flavor. Related to our stochastic model, Boczkowski et al. (2018) dealt with a search model on a graph, where randomness is induced by potentially unreliable advice, that is, with some fixed probability each node is faulty and points to the wrong neighbor. von Stengel and Werchner (1997) studied the complexity of a HSG on a graph when the hider hides on one of the nodes of the graph. Jotshi and Batta (2008) proposed a heuristic algorithm to find a hider hidden uniformly at random on a network.

In the HSG studied by Alpern (2010); Alpern and Lidbetter (2014) the searcher moves on a network at a speed that depend on her location and direction. An intuitive link can be established between the speed variations considered in these two articles, and the expected time to cross some edges considered in the present article.

In a forthcoming paper Glazebrook et al. (2019) considered a search game where an object is hidden in one of many discrete locations and the searcher can use one of two search modes: a fast but inaccurate mode or a slow but accurate one. The reader is referred to the classical book by Alpern and Gal (2003) for an extended treatment of search games and to Hohzaki (2016) for a recent survey of the relevant literature.

To the best of our knowledge, the model where edges of a network are present only with some probability has not been studied before in the framework of search games, but is standard in other fields. For instance, it is at the foundations of the classical model of random graphs proposed by Erdős and Rényi (1959, 1960, 1961), where, given a set of vertices, a random graph is generated by creating an edge between any two pairs of vertices independently with probability p. A similar model is studied in percolation theory, where edges of a graph are independently active with probability p and one relevant problem is the number of clusters in the random graph and, as a consequence, the possibility of reaching one vertex starting from another one. The reader is referred, for instance, to Bollobás (2001); Bollobás and Riordan (2006); Grimmett (1999); van der Hofstad (2017) for a general treatment of random graphs and percolation. Bollobás et al. (2013) considered a cop and robbers games played on a random graph. Some intriguing interactions between percolation and game theory have been recently studied by Day and Falgas-Ravry (2018); Holroyd et al. (2019), who considered two-person zero-sum games on a graph with alternating moves.

5.1.4 Organization of the paper

The paper is organized as follows. Section 5.2 describes the model. Section 5.3 deals with the deterministic case, where all edges are active with probability 1. Section 6.3 shows existence of the value for the stochastic case and provides upper and lower bounds for this value. Section 5.5 uses dynamic programming to find best responses of the searcher against a known hiding distribution of the hider. Sections 5.6 and 5.7 are devoted to the analysis of search games on trees and Eulerian graphs, respectively. Omitted proofs can be found in the Appendix of Chapter 5.

5.2 The model

5.2.1 Notation

Given a finite set A, we call $\operatorname{card} A$ its cardinality and $\Delta(A)$ the set of probability measures on A.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a connected undirected graph, where $\mathcal{V}$ is the nonempty finite set of vertices and $\mathcal{E}$ is the nonempty finite set of edges. All edges have length 1. There exists a special vertex $O \in \mathcal{V}$, called the *root* of the graph $\mathcal{G}$. Let $\mathbb{G}$ be the set of subgraphs of $\mathcal{G}$. For all $v \in \mathcal{V}$, we call $\mathcal{N}(\mathcal{G}, v)$ the *immediate neighborhood* of v in $\mathcal{G}$:

$$\mathcal{N}(\mathcal{G}, v) = \{v\} \cup \{u \in \mathcal{V} | \{v, u\} \in \mathcal{E}\}. \tag{5.2.1}$$

The graph will evolve in discrete time as follows. Let $\boldsymbol{p} = (p_e)_{e \in \mathcal{E}} \in (0, 1]^{\mathcal{E}}$. At each stage $t \geq 1$, each edge $e \in \mathcal{E}$ is active with probability p_e or inactive with probability $1 - p_e$, independently of the other edges. This defines a random graph process on $\mathbb{G}$ denoted $(\mathcal{G}_t)_t = (\mathcal{V}, \mathcal{E}_t)_{t \geq 1}$, where $\mathcal{E}_t$ is the random set of active edges at time t.

5.2.2 The game

We consider a stochastic zero-sum game $\Gamma = \langle \mathcal{G}, O, \boldsymbol{p} \rangle$ with two players: a maximizer, called the *hider* (Harry), and a minimizer, called the *searcher* (Sally). We call this game a *stochastic search game* (SSG).

The game is played as follows. At stage 0 both players know $\mathcal{G}_0 = \mathcal{G}$ and the initial position of the searcher $v_0 = O$. The hider chooses an edge $e \in \mathcal{E}$. Then the graph $\mathcal{G}_1$ is drawn and the searcher chooses $v_1 \in \mathcal{N}(\mathcal{G}_1, v_0)$. If $\{v_0, v_1\} = e$, then the game ends and the payoff to the hider is 1, otherwise the graph $\mathcal{G}_2$ is drawn and the game continues. Inductively, at each stage $t \geq 1$, knowing $h_t = (\mathcal{G}_0, v_0, \ldots, \mathcal{G}_{t-1}, v_{t-1}, \mathcal{G}_t)$, the searcher chooses $v_t \in \mathcal{N}(\mathcal{G}_t, v_{t-1})$. If $\{v_{t-1}, v_t\} = e$, then the game ends and the payoff to the hider is t, otherwise the graph $\mathcal{G}_{t+1}$ is drawn and the game continues.

Hence in this SSG, the state space is $\mathbb{G} \times \mathcal{V}$, the action set of the hider is $\mathcal{E}$, and the action set of the searcher in state $(\mathcal{G}', v) \in \mathbb{G} \times \mathcal{V}$ is $\mathcal{N}(\mathcal{G}', v)$. We now describe the sets of strategies of the players. For $t \geq 0$, let $H_t = \mathbb{G} \times (\mathbb{G} \times \mathcal{V})^t$ be the set of histories at stage t and let $H = \bigcup_{t \geq 0} H_t$ be the set of all histories. Call $\mathcal{S}$ the set of (behavior) strategies of the searcher, that is the strategies $\sigma \colon H \to \Delta(\mathcal{V})$ such that $\sigma(h_t) \in \Delta(\mathcal{N}(\mathcal{G}_t, v_{t-1}))$.

We call pure the strategies s such that, for all $t \geq 0$ and all $h_t \in H_t$,

$$s(h_t) = v_t \in \mathcal{N}(\mathcal{G}_t, v_{t-1}).$$

A behavior strategy σ naturally induces a probability measure on each H_t, for every $t \geq 1$, which can be uniquely extended to H_∞ by Kolmogorov's extension theorem. This probability is denoted $\mathbb{P}_\sigma$ and the corresponding expectation is denoted $\mathbb{E}_\sigma$.

A mixed strategy of the searcher is a probability distribution over pure strategies, endowed with the product σ-algebra. By Kuhn's theorem, behavior and mixed strategies are equivalent (see, e.g., Aumann, 1964; Sorin, 2002). The sets of pure and mixed strategies of the hider are $\mathcal{E}$ and $\Delta(\mathcal{E})$, respectively. Pure strategies of the hider and the searcher will usually be denoted with the letters e and s respectively, while mixed and behavior strategies will usually be denoted with the letters ε and σ, respectively. We denote ε^{U} the uniform distribution (UD) on $\mathcal{E}$.

Finally, the payoff function of the hider is the function $g\colon \mathcal{E}\times\mathcal{S} \to \mathbb{R}_+ \cup \{+\infty\}$, defined as

$$g(e, \sigma) = \mathbb{E}_\sigma[\inf\{t \geq 1 | \{v_{t-1}, v_t\} = e\}], \tag{5.2.2}$$

where the infimum over the empty set is $+\infty$. The function g is linearly extended to $\Delta(\mathcal{E})$. The goal of the hider is thus to maximize the expected time by which he is found by the searcher, while the goal of the searcher is to minimize the expected time by which she finds the hider.

5.3 Deterministic search games

Proposition 5.9 below will show that the search game $\langle \mathcal{G}, O, \boldsymbol{p}\rangle$ has a value, which we denote $\mathrm{val}(\boldsymbol{p})$. If p_e is equal to 1 for all $e \in \mathcal{E}$, we then recover a search game with an immobile hider played on a graph. We call this game a *deterministic search game* (DSG). DSGs have a value $\mathrm{val}(\mathbf{1})$.

We recall some important definitions and results for DSGs. Versions of these results are well known when the game is played in continuous time over a continuous network (see, e.g., Alpern and Gal, 2003).

Definition 5.1. (i) A cycle in an graph is called *Eulerian* if it uses each edge exactly once. If such a cycle exists, the graph is called Eulerian.

(ii) A *Chinese postman cycle* is a cycle of minimal length that visits each edge. In Eulerian graphs, the Chinese postman cycles are the Eulerian cycles.

Definition 5.2. (i) The *uniform Eulerian strategy* (UES) is a mixed strategy that mixes over all Eulerian cycles with equal probability.

(ii) The *uniform Chinese postman strategy* (UCPS) is a mixed strategy that mixes over all Chinese postman cycles with equal probability.

When considering trees, we will endow them with an orientation outgoing from the root. This orientation does not affect the behavior of the searcher, who can travel any edge in any direction, but is just needed to state and prove some of our results.

Let $\mathcal{G} = \mathcal{T}$ be a tree. If v is a vertex of $\mathcal{T}$, then $\mathcal{T}_v$ is the subtree that has v as a root and contains all edges below v in the original tree $\mathcal{T}$. Hence $\mathcal{T} = \mathcal{T}_O$.

If e is an edge of $\mathcal{G}$, then $\mathcal{T}_e := \{e\} \cup \mathcal{T}_v$ where v is the head of e, i.e., $\mathcal{T}_e$ includes e and the maximal subtree below the head of e. We denote $\mathcal{E}_v$ (resp. $\mathcal{E}_e$) the set of edges of $\mathcal{T}_v$ (resp. $\mathcal{T}_e$).

The following definition is an adaptation to our framework of what Alpern and Gal (2003, Section 3.3) have in the continuous setting.

Definition 5.3. The *equal branching density* (EBD) ε^* of the hider is the unique distribution on $\mathcal{E}$ that is supported on the leaf edges and, for every branching vertex v with outgoing edges $e_1, \dots, e_n$, satisfies

$$\frac{\varepsilon^*(\mathcal{E}_{e_i})}{\operatorname{card} \mathcal{E}_{e_i}} = \frac{\varepsilon^*(\mathcal{E}_{e_1})}{\operatorname{card} \mathcal{E}_{e_1}}, \quad \text{for all } i \in \{1, \dots, n\}. \tag{5.3.1}$$

Proposition 5.4. *Let* $\Gamma = (\mathcal{V}, \mathcal{E})$. *In a DSG* $\Gamma = \langle \mathcal{G}, O, \mathbf{1} \rangle$ *we have*

$$\operatorname{val}(1) \leq \operatorname{card} \mathcal{E}. \tag{5.3.2}$$

Moreover, $\operatorname{val}(1) = \operatorname{card} \mathcal{E}$ *if and only if* $\mathcal{G}$ *is a tree. In this case, the EBD and the UCPS are optimal strategies.*

We first prove the following lemma.

Lemma 5.5. *Let* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ *be a connected graph. Any Chinese postman cycle has length*

(i) $2 \operatorname{card} \mathcal{E}$ *if* $\mathcal{G}$ *is a tree,*

(ii) *at most* $2 \operatorname{card} \mathcal{E} - 2$ *if* $\mathcal{G}$ *is not tree.*

Proof. If $\mathcal{G}$ is a tree the result follows by induction on $\operatorname{card} \mathcal{E}$.

Suppose now that $\mathcal{G}$ is not a tree. We again proceed by induction on $\operatorname{card} \mathcal{E}$. There exists an edge $e = \{u, v\} \in \mathcal{E}$ such that $\mathcal{G}' = (\mathcal{V}, \mathcal{E} \setminus \{e\})$ is connected.

If $\mathcal{G}'$ is a tree, we consider a Chinese postman cycle $\gamma \in \mathcal{G}'$ starting at u, such that the subtree with root v is the last visited. Once the vertex v is visited for the last time on γ, we replace the end of the cycle—which has already been visited—with e, going straight from v to u. This new cycle in $\mathcal{G}$ has length at most $2(\operatorname{card} \mathcal{E} - 1) + 1 - 1 = 2 \operatorname{card} \mathcal{E} - 2$, since the length of the cycle in $\mathcal{G}'$ is $2(\operatorname{card} \mathcal{E} - 1)$, the length of e is 1, and the number of the edges not visited a second time is at least 1.

If $\mathcal{G}'$ is not a tree, then it admits a Chinese postman cycle γ with length at most $2(\operatorname{card} \mathcal{E} - 1) - 2$. We now consider the cycle $\gamma' \in \mathcal{G}$ which starts at u, goes back and forth on e and then follows the cycle γ on $\mathcal{G}'$. This cycle has length $2(\operatorname{card} \mathcal{E} - 1) - 2 + 2 = 2 \operatorname{card} \mathcal{E} - 2$. ◻

The proof of Proposition 5.4 will make use of the following lemma, which refers to a model for continuous networks in continuous time. Let Q be a continuous tree network, and suppose that the edges of Q have integer length. Then Q is mapped to a tree graph $\mathcal{T}$ in the natural way. The UCPS and the EBD are defined in a similar way in $\mathcal{T}$ and in Q, and are naturally mapped from the graph setting to the continuous network setting, and vice versa.

Lemma 5.6. *[(Alpern and Gal, 2003, Theorem 3.21)] Let Q be a continuous tree network with total length μ. Then*

(i) *The UCPS is an optimal search strategy.*

(ii) *The EBD is an optimal hiding strategy.*

(iii) $\mathrm{val}(\mathbf{1}) = \mu$.

If the continuous network Q with total length μ is not a tree, then $\mathrm{val}(\mathbf{1}) < \mu$.

Proof of Proposition 5.4. If $\mathcal{G}$ is a tree, the result follows from Lemma 5.6. Indeed, in the discrete setting, hiding on edges that are not leafs is strictly dominated. Similarly in the continuous setting, hiding at a point of the tree which is not terminal is strictly dominated. Hence the UCPS guarantees the value of the continuous game in the discrete one—with the natural mapping. Moreover, since the set of hiding strategies in the discrete setting is a subset of the set of hiding strategies on the continuous setting—again with the natural mapping—the EBD guarantees in the discrete game the value of the continuous one.

If $\mathcal{G}$ is not a tree, suppose that the searcher uniformly chooses between any Chinese postman cycle, and let the hider choose an edge e. For any fixed Chinese postman cycle of length n, e has position k in the cycle and position $n - k + 1$ in the reverse cycle. By Lemma 5.5, $n \leq 2\,\mathrm{card}\,\mathcal{E} - 2$, hence, the payoff is at most

$$\frac{k + 2\,\mathrm{card}\,\mathcal{E} - 2 - k + 1}{2} = \mathrm{card}\,\mathcal{E} - \frac{1}{2} < \mathrm{card}\,\mathcal{E}. \qquad \square$$

Proposition 5.7. *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. In a DSG $\Gamma = \langle \mathcal{G}, O, \mathbf{1} \rangle$ we have*

$$\mathrm{val}(1) \geq \frac{\mathrm{card}\,\mathcal{E} + 1}{2}. \tag{5.3.3}$$

Moreover, if $\mathrm{card}\,\mathcal{E} > 1$, *then*

$$\mathrm{val}(1) = \frac{\mathrm{card}\,\mathcal{E} + 1}{2}. \tag{5.3.4}$$

if and only if $\mathcal{G}$ is Eulerian. In this case, the UD on $\mathcal{E}$ and the UES are optimal strategies.

Proof. Suppose the hider hides uniformly over $\mathcal{E}$. Now let the searcher choose any sequence of edges (without necessarily following a path in $\mathcal{G}$). Then if the searcher does not search the same edge twice during his $\mathrm{card}\,\mathcal{E}$ first picks, the payoff is $(\mathrm{card}\,\mathcal{E} + 1)/2$, hence the lower bound. Suppose $\mathrm{card}\,\mathcal{E} > 1$, it is clear that this bound is reached only in Eulerian graphs, following an Eulerian cycle, because, if the graph is not Eulerian, then an edge is visited twice. Finally, using an argument similar to the one used in Proposition 5.4, we can show that the uniform Eulerian strategy yields the payoff $(\mathrm{card}\,\mathcal{E} + 1)/2$ against any strategy of the hider. $\square$

Together, Propositions 5.4 and 5.7 yield the next theorem, whose continuous version is a cornerstone of the search game literature. It gives bounds on the value of deterministic search games played on any graphs. Moreover, it shows that Eulerian graphs and trees are the two extreme classes of graphs in term of value of the game.

Theorem 5.8. *For any graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the value of the DSG $\Gamma = \langle \mathcal{G}, O, \mathbf{1} \rangle$ satisfies*

$$\frac{\operatorname{card} \mathcal{E} + 1}{2} \leq \operatorname{val}(1) \leq \operatorname{card} \mathcal{E}. \tag{5.3.5}$$

Moreover, if $\operatorname{card} \mathcal{E} > 1$, the upper bound is reached if and only if $\mathcal{G}$ is a tree and the lower bound is reached if and only if $\mathcal{G}$ is an Eulerian graph.

If $\mathcal{G}$ is an Eulerian graph, then the UD on $\mathcal{E}$ and the UES are optimal strategies.

If $\mathcal{G}$ is a tree, then the EBD and the UCPS are optimal strategies.

In Sections 5.6 and 5.7 we focus on subclasses of these two extreme classes that are Eurelian graphs and trees. Both subclasses have a recursive structure. We generalize the strategies of interest to our stochastic setting and derive bounds on the value. We also prove that these strategies are optimal in the cases of circles and lines.

5.4 Value of the game

Proposition 5.9. *For any $\boldsymbol{p} \in (0,1]^{\mathcal{E}}$ the SSG $\langle \mathcal{G}, O, \boldsymbol{p} \rangle$ has a value $\operatorname{val}(\boldsymbol{p})$. Moreover both players have an optimal strategy.*

The proof of Proposition 5.9 is postponed to the Appendix of Chapter 5.

Proposition 5.10. *For all $\boldsymbol{p} \in (0,1]^{\mathcal{E}}$ the value of the SSG $\langle \mathcal{G}, O, \boldsymbol{p} \rangle$ satisfies*

$$\frac{\operatorname{val}(\mathbf{1})}{1 - (1 - \min_{e \in \mathcal{E}} p_e)^{\delta}} \leq \operatorname{val}(\boldsymbol{p}) \leq \frac{\operatorname{val}(\mathbf{1})}{\min_{e \in \mathcal{E}} p_e}, \tag{5.4.1}$$

where δ is the maximum degree of $\mathcal{G}$.

As a consequence

$$\operatorname{val}(\boldsymbol{p}) \longrightarrow \operatorname{val}(\mathbf{1}), \text{ as } \min_{e \in \mathcal{E}} p_e \longrightarrow 1. \tag{5.4.2}$$

Proof. The hider guarantees the lower bound by playing as in the DSG. In expectation the searcher waits at least $(1 - (1 - \min_{e \in \mathcal{E}} p_e)^{\delta(\mathcal{G})})^{-1}$ for a neighbor edge to be active.

We map a strategy of the searcher in the DSG to the strategy in the SSG following the same path, even if it means waiting for an edge to be active. The searcher guarantees the upper bound since it takes in expectation at most $1/\min_{e \in \mathcal{E}} p_e$ stages to cross a single edge. □

5.5 Dynamic programming

The next proposition is a dynamic programming formula which allows to find best responses of the searcher against a known hiding distribution of the hider. The activation parameters $\boldsymbol{p} \in (0,1]^{\mathcal{E}}$ are fixed and we omit them.

For all $\mathcal{G}_1 \in \mathbb{G}$, $v_0 \in \mathcal{V}$, $I \subset \mathcal{E}$ and $\varepsilon \in \Delta(I)$, we define

$$\mathrm{Val}(\mathcal{G}_1, v_0, I, \varepsilon) = \min_{s \in \mathcal{S}} \mathbb{E}_s \left[\sum_{e \in I} \varepsilon(e) \inf\{t \geq 1 | \{v_{t-1}, v_t\} = e\} \right]. \tag{5.5.1}$$

This quantity represents the value of the (one player) game in which the searcher knows the graph $\mathcal{G}_1$ and the distribution ε of the hider on $I \subset \mathcal{E}$, starts from v_0 and chooses immediately $v_1 \in \mathcal{N}(\mathcal{G}_1, v_0)$ at the first stage, before $\mathcal{G}_2$ is drawn (and then the game continues). In other words, in the true game, a graph $\mathcal{G}_1$ is drawn before Sally starts playing. Here the graph $\mathcal{G}_1$ is already fixed and Sally starts playing immediately.

Proposition 5.11. *If $I = \varnothing$, then $\mathrm{Val}(\mathcal{G}_1, v_0, I, \varepsilon) = 0$. Otherwise*

$$\mathrm{Val}(\mathcal{G}_1, v_0, I, \varepsilon) = 1 + \min_{v_1 \in \mathcal{N}(\mathcal{G}_1, v_0)} \varepsilon(I \setminus \{v_0, v_1\}) \, \mathbb{E} \left[\mathrm{Val}\left(\mathcal{G}_2, v_1, I \setminus \{v_0, v_1\}, \varepsilon^{\{v_0, v_1\}} \right) \right], \tag{5.5.2}$$

where $\varepsilon^{\{v_0,v_1\}}(\cdot) = \frac{1}{\varepsilon(I \setminus \{v_0, v_1\})} \varepsilon(\cdot)$, and the randomness in Eq. (5.5.2) is over $\mathcal{G}_2$.

Proof. If the searcher finds the hider in the first stage, which happens with probability $\varepsilon(\{v_0, v_1\})$, then the game ends and the continuation payoff is 0. On the other hand, if the searcher does not find the hider in the first stage, which happens with probability $1 - \varepsilon(\{v_0, v_1\})$, then the game continues with continuation payoff

$$\mathbb{E} \left[\mathrm{Val}\left(\mathcal{G}_2, v_1, I \setminus \{v_0, v_1\}, \varepsilon^{\{v_0, v_1\}} \right) \right], \tag{5.5.3}$$

since the edge $\{v_0, v_1\}$ has been visited and the next graph $\mathcal{G}_2$ is yet to be drawn. □

5.6 Stochastic search games on trees

In this section and in the following one we assume

$$p_e = p \in (0,1], \quad \text{for all } e \in \mathcal{E}. \tag{5.6.1}$$

Moreover in this section we assume that $\mathcal{G}$ is a tree $\mathcal{T}$ with origin O. Remark that in a tree, any strategy of the hider that consists in hiding in edges other than leaf edges is strictly dominated.

5.6.1 Depth-first strategies and the equal branching density

We define a particular class of strategies of the searcher in trees, called depth-first strategies. They have the property of never going backward at a vertex before having visited the whole subtree. They generalize the Chinese postman cycles of the deterministic setting.

Definition 5.12. A *depth-first strategy* (DFS) on a tree is a strategy of the searcher that prescribes the following, when arriving at a vertex:

- if the set of un-searched and active outgoing edges is non-empty, take one of its edges (possibly at random);
- if all the un-searched outgoing edges are inactive, wait;
- if all outgoing edges have been searched and the backward edge is active, take it;
- if all outgoing edges have been searched and the backward edge is inactive, wait.

The *uniform depth-first strategy* (UDFS) is the DFS that, at every vertex, randomizes uniformly between all active and un-searched outgoing edges.

Definition 5.13. A DFS on $\mathcal{T}$ induces an expected time to travel from the origin O back to it, covering the entire tree. This is called the *cycle time* of $\mathcal{T}$ and is denoted $\tau(O)$. For any vertex or edge z, the cycle time of $\mathcal{T}_z$ is denoted $\tau(z)$.

Notice that $\tau(O)$ depends on $\boldsymbol{p}$, but is independent of the choice of DFS.

We now generalize Definition 5.3 to the stochastic setting, where the relevant quantity is not the number of edges of the subtrees, but rather their cycle times.

Definition 5.14. The *equal branching density* (EBD) ε^* of the hider is the unique distribution on the leaf edges such that, for every branching vertex v with outgoing edges $e_1, \dots, e_n$, we have

$$\frac{\varepsilon^*(\mathcal{E}_{e_i})}{\tau(e_i)} = \frac{\varepsilon^*(\mathcal{E}_{e_1})}{\tau(e_1)}, \quad \text{for all } i \in \{1, \dots, n\}. \tag{5.6.2}$$

Notice that Definitions 5.3 and 5.14 coincide when $p_e = 1$ for all $e \in \mathcal{E}$.

5.6.2 Binary trees

Generalities

In this sections we consider games played on binary trees, i.e., trees with at most two outgoing edges at any vertex. We call $\mathbb{T}$ the set of binary trees. DFSs allow us to obtain an upper bound for the value, when p is large enough. We also prove that this upper bound is the value of the game in which Sally is restricted to play DFSs. As a by-product we will show that, for every $p \in (0, 1]$, the UDFS and EBD are a pair of optimal strategies when the game is played on a line.

Definition 5.15. Given a tree $\mathcal{T} = (\mathcal{V}, \mathcal{E})$, we define the function $\Lambda\colon \mathbb{T} \to \mathbb{R}$ recursively as follows, where, for the sake of simplicity we use the notations $\Lambda(e) = \Lambda(\mathcal{T}_e)$ and $\Lambda(v) = \Lambda(\mathcal{T}_v)$:

If $\mathcal{T}$ has a single edge $e = (O, v)$, as in Fig. 5.1, then

$$\Lambda(O) = \Lambda(e) = \Lambda(v) = 0. \tag{5.6.3}$$

If $\mathrm{degr}(O) = 1$ and $e = (O, v)$, as in Fig. 5.2, then $\Lambda(O) = \Lambda(e) = \Lambda(v)$.

Figure 5.1: One edge

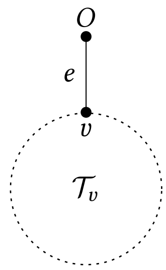

Figure 5.2: O has degree 1

If $\mathcal{T}$ has two edges and $\mathrm{degr}(O) = 2$, as in Fig. 5.3, then

$$\Lambda(O) = \frac{1}{2}\left(\frac{1}{1-(1-p)^2} - \frac{1}{p}\right). \tag{5.6.4}$$

If $\mathrm{degr}(O) = 2$, $e_1 = (O, v_1)$, and $e_2 = (O, v_2)$, as in Fig. 5.4, then

$$\Lambda(O) = \frac{\tau(v_1)}{\tau(v_1)+\tau(v_2)}\Lambda(v_1) + \frac{\tau(v_2)}{\tau(v_1)+\tau(v_2)}\Lambda(v_2) + \frac{1}{2}\left(\frac{1}{1-(1-p)^2} - \frac{1}{p}\right). \tag{5.6.5}$$

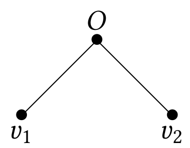

Figure 5.3: Two edges

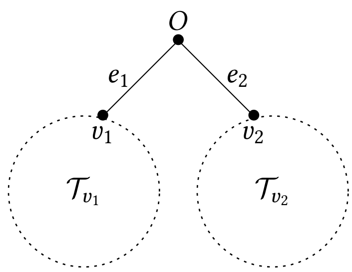

Figure 5.4: O has degree 2

The function Λ depends on p, but we do not make the dependence explicit.

Lemma 5.16. *Let v be a branching vertex with outgoing edges e_1 and e_2. Then for all $p \in (0, 1]$,*

$$\frac{|\Lambda(e_1)| + |\Lambda(e_2)|}{\tau(e_1) + \tau(e_2)} < \frac{1}{2}.$$

Proof. We proceed by induction on the number of edges. The base case is immediate since $\Lambda(e_1) = \Lambda(e_2) = 0$. For the induction step the situation is represented in Fig. 5.5. The vertex v_1 is the first vertex encountered in $\mathcal{T}_{e_1}$ with two outgoing edges, and similarly for v_2 and $\mathcal{T}_{e_2}$.

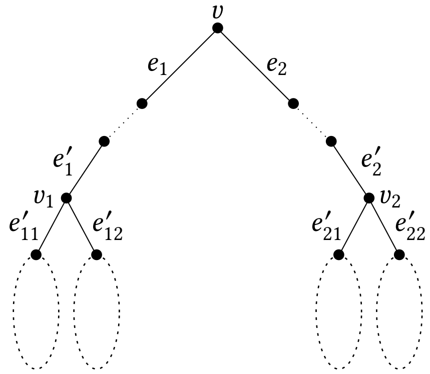

Figure 5.5: The induction step

We have

$$|\Lambda(e_1)| = |\Lambda(e_1')| = \left|\frac{\tau(e_{11}')}{\tau(e_{11}') + \tau(e_{12}')}\Lambda(e_{11}') + \frac{\tau(e_{12}')}{\tau(e_{11}') + \tau(e_{12}')}\Lambda(e_{12}') + \frac{1}{2}\left(\frac{1}{1-(1-p)^2} - \frac{1}{p}\right)\right|$$
$$< \frac{1}{2}\max(\tau(e_{11}'), \tau(e_{12}')) + \frac{1}{2}\left|\frac{1}{1-(1-p)^2} - \frac{1}{p}\right|$$

by induction, and similarly for $\Lambda(e_2)$. Moreover we have

$$\tau(e_1) > \tau(e_1') = \tau(e_{11}') + \tau(e_{12}') + \frac{1}{p} + \frac{1}{1-(1-p)^2},$$

and similarly for $\tau(e_2)$. Finally,

$$\frac{|\Lambda(e_1)| + |\Lambda(e_2)|}{\tau(e_1) + \tau(e_2)} < \frac{\frac{1}{2}\left(\max(\tau(e_{11}'), \tau(e_{12}')) + \max(\tau(e_{21}'), \tau(e_{22}')) + \frac{2}{p} - \frac{2}{1-(1-p^2)}\right)}{\tau(e_{11}') + \tau(e_{12}') + \tau(e_{21}') + \tau(e_{22}') + \frac{2}{p} + \frac{2}{1-(1-p^2)}}$$
$$< \frac{1}{2}.$$

□

We now define the biased depth-first (behavior) strategy of the hider.

Definition 5.17. Assume that vertex v has outgoing edges e_1 and e_2 and they are both active and un-searched. A DFS strategy σ_α is called the *biased depth-first strategy* (BDFS) if it takes e_1 with probability $\alpha(e_1)$ and e_2 with probability $\alpha(e_2)$, where

$$\alpha(e_1) = \mathrm{proj}_{[0,1]}\left(\frac{1}{2} + \frac{\Lambda(e_1) - \Lambda(e_2)}{\tau(e_1) + \tau(e_2)}\frac{1-(1-p)^2}{p^2}\right) \tag{5.6.6}$$

$$\alpha(e_2) = 1 - \alpha(e_1), \tag{5.6.7}$$

where $\text{proj}_{[0,1]}$ indicates the projection on $[0, 1]$.

Theorem 5.18. *There exists $p_0 \in (0, 1)$ such that for all $p \geq p_0$, the time to reach any leaf edge using the BDFS is $\frac{1}{2}\tau(O) + \Lambda(O)$. Hence for all $p \geq p_0$, we have*

$$\text{val}(p) \leq \frac{1}{2}\tau(O) + \Lambda(O). \tag{5.6.8}$$

The proof of Theorem 5.18 is postponed to the Appendix of Chapter 5.

Theorem 5.19. *The EBD of the hider yields the same payoff against any DFS of the searcher, and this payoff is $\frac{1}{2}\tau(O) + \Lambda(O)$.*

The proof of Theorem 5.19 is postponed to the Appendix of Chapter 5. Theorems 5.18 and 5.19 imply that in a binary tree $\mathcal{G}$, if DFSs are best responses to the EBD, then there exists $p_0 \in (0, 1)$ such that for all $p \geq p_0$ the value of the game is $\frac{1}{2}\tau(O) + \Lambda(O)$. Moreover the BDFS and the EBD are optimal.

However, there exist binary trees for which DFSs are not best responses to the EBD as the following example shows.

Example 5.1. We study the game played on the tree represented in Fig. 5.6.

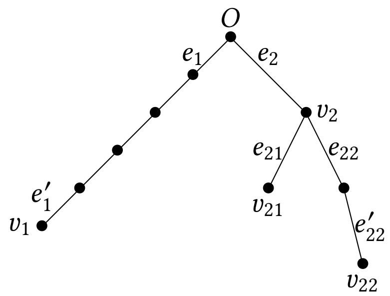

Figure 5.6: A counter-example

Consider the case where Sally visits v_{22} before any other leaf vertex. When she plays a DFS, this event has positive probability. Assume also that, when she has returned to v_2, after visiting v_{22}, the edge e_2 is active but e_{21} is not. At this point she can either take edge e_2 and visit v_1 before v_{21} or wait until e_{21} becomes active and visit v_{21} before v_1. The first choice yields a lower payoff to Sally.

Indeed, visiting v_1 first yields the continuation payoff

$$g_1 = \varepsilon^*(e'_1)\left(1 + \frac{5}{p}\right) + \varepsilon^*(e_{21})\left(1 + \frac{12}{p}\right),$$

whereas visiting v_{21} first yields the continuation payoff

$$g_2 = \varepsilon^*(e_{21})\left(1 + \frac{1}{p}\right) + \varepsilon^*(e'_1)\left(1 + \frac{8}{p}\right).$$

The sign $g_1 - g_2$ is the same as the sign of $11\varepsilon^*(e_{21}) - 3\varepsilon^*(e_1')$, which is the same as

$$\frac{11}{3}\left(\frac{7}{p} + \frac{1}{1-(1-p)^2}\right) - \frac{30}{p},$$

which is negative for all $p \in (0, 1)$.

A simple binary tree

We now present a game played on a tree (Fig. 5.7) for which we give the value and a pair of optimal strategies for any value of $p \in (0, 1]$.

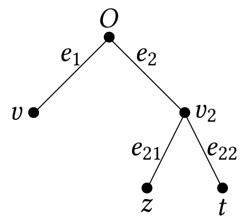

Figure 5.7: A simple binary tree

Let

$$p_0 = \frac{9 - \sqrt{65}}{8} \approx 0.12. \tag{5.6.9}$$

First case $p \geq p_0$: In this case, Sally's BDFS and Harry's EBD are a pair of optimal strategies. The value of the game is thus

$$\operatorname{val}(p) = \frac{1}{2}\tau(O) + \Lambda(O) = \frac{92 - 75p + 15p^2}{p(15 - 7p)(2 - p)}.$$

Second case $p \leq p_0$: Harry's strategy $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ is optimal. We now describe an optimal strategy of Sally.

- If no leaf edges have been visited:
  - At O: if e_1 is active, take it. Otherwise, if e_2 is active but e_1 is not, take e_2.
  - At v_2: take the first active edge between e_{21} and e_{22}, drawing uniformly, if they both are.
- If only e_1 has been visited, play the UDFS in the continuation game.
- If only e_{21} (resp. e_{22}) has been visited, at v_2:
  - If e_{22} (resp. e_{21}) is active, take it.

  - If e_2 is active but e_{22} (resp. e_{21}) is not, randomize, waiting at v_2 with probability $\zeta(p)$ and taking e_{22} (resp. e_{21}) with probability $1 - \zeta(p)$.

- If two leaf edges have been visited, go to the third leaf edge as quickly as possible.

The waiting probability $\zeta(p)$ is given by

$$\zeta(p) = \frac{8(2-p) - (1-p)(1+p)(2-p)}{8(2-p)(1-p) - p(1-p)^2}.$$

The value of the game is

$$\mathrm{val}(p) = \frac{1}{3}\frac{37 - 33p + 7p^2}{p(2-p)^2}.$$

The line

We consider a SSG played on a line. If the origin O is an extreme vertex, then the value of the game is $\mathrm{card}(\mathcal{E})/p$. We now suppose that the origin O is not an extreme vertex, and that the line has $L = \lambda_1 + \lambda_2$ edges (λ_1 on the left side of O and λ_2 on the right side) as shown in Fig. 5.8.

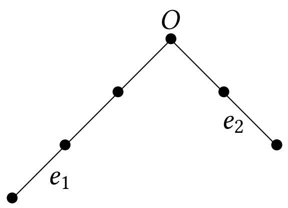

Figure 5.8: The line with $\lambda_1 = 3$ and $\lambda_2 = 2$

In this case, for all $p \in (0, 1]$ the BDFS is the UDFS σ^*, and the EBD of the hider is

$$\varepsilon^* = \left(\frac{\lambda_1}{L}, \frac{\lambda_2}{L}\right).$$

Proposition 5.20. *If the graph $\mathcal{G}$ is a line, then DFS are best responses to the EBD. Hence, $(\varepsilon^*, \sigma^*)$ is a pair of optimal strategies.*

Proof. Harry plays ε^*. At O, whatever active edge Sally takes, the continuation payoff is $(\lambda_1 + \lambda_2 - 1)/p$. Hence she does not profit from waiting for one specific edge to be active. □

Together with Theorems 5.18 and 5.19, Proposition 5.20 yields the following corollary.

Corollary 5.21. *The value of the game played on the line with L edges is*

$$\mathrm{val}(p) = \frac{1}{2}\tau(O) + \Lambda(O) = \frac{L}{p} + \frac{1}{1-(1-p)^2} - \frac{1}{p},$$

for all $p \in (0, 1]$, if the root is not an extreme vertex. Moreover the EBD and the UDFS are optimal strategies.

5.7 Stochastic search games on Eulerian graphs

5.7.1 Eulerian strategies and the uniform density

For Eulerian graphs we define a strategy of the searcher, called *Eulerian strategy* (ES), which generalizes an Eulerian cycle of the deterministic setting. At any vertex an ES chooses an active outgoing edge that had not previously been visited in such a way that the induced path is an Eulerian cycle. The ES that at any vertex randomizes uniformly over the outgoing edges is called the *uniform Eulerian strategy* (UES) and is denoted σ^*.

Definition 5.22. The UES on an Eulerian graph $\mathcal{G}$ induces an expected time to travel from the origin O covering the entire Eulerian graph. This is called the *cycle time* of $\mathcal{G}$ and is denoted $\theta(\mathcal{G})$.

5.7.2 Parallel Eulerian graphs

Generalities

We call *parallel graph* a graph where parallel paths link two vertices, one of these two vertices being the root O, as in Fig. 5.9. Such a graph is denoted $\mathcal{P}_m(\boldsymbol{\lambda})$, where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)$ is the vector of the lengths of the parallel paths. The parallel uniform strategy of Sally consists in choosing at O uniformly between active and unsearched edges and then going straight to D on the current parallel path (and similarly at D).

Remark that if the number of parallel paths $m = 2m$ is even, then the parallel graph is Eulerian and we call it a *parallel Eulerian graph*. In this case, the parallel uniform strategy is the UES. For a parallel Eulerian graph $\mathcal{P}_{2m}(\boldsymbol{\lambda})$ with $2m$ parallel lines, the cycle time of $\mathcal{P}_{2m}(\boldsymbol{\lambda})$ is

$$\theta(\mathcal{P}_{2m}(\boldsymbol{\lambda})) = \sum_{k=1}^{2m} \left(\frac{1}{1-(1-p)^k} + \frac{\lambda_k - 1}{p} \right).$$

The UES allows us to obtain an upper bound for the value. We also prove that this upper bound is the value of the game in which Sally is restricted to play ESs. As a by-product we will show that, for every $p \in (0, 1]$, the UES and UD are a pair of optimal strategies when the game is played on a circle.

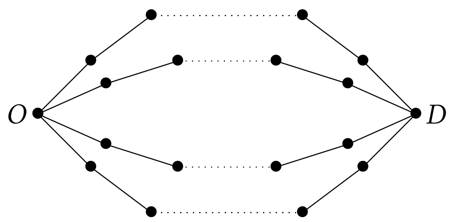

Figure 5.9: A parallel Eulerian graph

Definition 5.23. Given a parallel Eulerian graph $\mathcal{P}_{2m}(\boldsymbol{\lambda})$ with $2m$ parallel lines, let Φ_m be the following quantity defined recursively:

$$\Phi_1 = \frac{1}{2}\left(\frac{1}{1-(1-p)^2} - \frac{1}{p}\right), \tag{5.7.1}$$

and for eack $m > 1$,

$$\begin{aligned}\Phi_m = &\frac{1}{2}\frac{1}{1-(1-p)^{2m}} + \left(\frac{1}{2} - \frac{1}{2m}\right)\frac{1}{1-(1-p)^{2m-1}} \\ &- \frac{1}{2m}\left(\sum_{k=1}^{2(m-1)} \frac{1}{1-(1-p)^k} + \frac{1}{p}\right) + \frac{m-1}{m}\Phi_{m-1}.\end{aligned}$$

Remark that Φ_m only depends on the number parallel paths and not on their length.

Theorem 5.24. *On a parallel Eulerian graph $\mathcal{P}_{2m}(\boldsymbol{\lambda})$, the expected time to reach any edge using the UES is*

$$\frac{\theta(\mathcal{P}_{2m}(\boldsymbol{\lambda})) + p^{-1}}{2} + \Phi_m. \tag{5.7.2}$$

Hence, for all $p \in (0, 1]$, we have

$$\mathrm{val}(p) \leq \frac{\theta(\mathcal{P}_{2m}(\boldsymbol{\lambda})) + p^{-1}}{2} + \Phi_m. \tag{5.7.3}$$

The proof of Theorem 5.24 is postponed to the Appendix of Chapter 5.

Theorem 5.25. *On a parallel Eulerian graph $\mathcal{P}_{2m}(\boldsymbol{\lambda})$, the uniform density of the hider yields the same payoff*

$$\frac{\theta(\mathcal{P}_{2m}(\boldsymbol{\lambda})) + p^{-1}}{2} + \Phi_m$$

against any Eulerian strategy of the searcher.

The proof of Theorem 5.25 is postponed to the Appendix of Chapter 5. Theorems 5.24 and 5.25 imply that in a parallel Eulerian graph $\mathcal{P}_{2m}(\boldsymbol{\lambda})$, if Eulerian strategies are best responses to the uniform density, for all $p \in (0,1]$ the value of the game is

$$\frac{\theta(\mathcal{P}_{2m}(\boldsymbol{\lambda})) + p^{-1}}{2} + \Phi_m.$$

Moreover the UES and the UD are optimal.

However, Eulerian strategies are not always best responses to the UD, as we now argue.

Example 5.2. We study the game played on a parallel Eulerian graph with four parallel paths. Each path i has two edges $e_{i1} = \{O, v_i\}$ and $e_{i2} = \{v_i, D\}$, where v_i is the middle vertex of the i-th path.

Consider the case where Sally visits e_{41}, e_{42} and e_{12} before any other edge. When she plays an ES, this event has positive probability. Assume also that, when at v_1, the edge e_{12} is active but e_{11} is not. At this point she can either wait at v_2 until e_{11} becomes active in order to follow an ES, or she can take e_{12}, then the first active edge between e_{22} and e_{32} and continue with e_{21} or e_{31} respectively. Finally, she takes the first active edge between e_{11} and the other edge at O that has not been visited yet, and then visits the two remaining edges as quickly as possible.

Following an ES yields the continuation payoff

$$g_1 = \frac{1}{5}\left(5 + \frac{11}{p} + \frac{4}{1-(1-p)^2}\right).$$

Following the second strategy yields the continuation payoff

$$g_2 = \frac{1}{5}\left(5 + \frac{17}{2p} + \frac{8}{1-(1-p)^2}\right).$$

Hence if $p < 2/5$, the second strategy yields a lower payoff to Sally than an ES.

The circle

We now examine the game played on a circle.

Lemma 5.26. *If the graph $\mathcal{G}$ is a circle, then Eulerian strategies are best responses to the uniform density.*

The proof of Lemma 5.26 is rather straightforward and we omit it. Together with Theorems 5.24 and 5.25, Lemma 5.26 yields the following corollary.

Corollary 5.27. *The value of the game played on the circle with L edges is*

$$\operatorname{val}(p) = \frac{\theta(\mathcal{G}) + p^{-1}}{2} + \Phi_2 = \frac{1}{1-(1-p)^2} + \frac{L-1}{2p},$$

for all $p \in (0,1]$. Moreover the uniform density and the uniform Eulerian strategy are optimal strategies.

Appendix of Chapter 5: omitted proofs

Omitted proofs of Section 5.4

The following lemma is a corollary of Flesch et al. (2019, Theorem 12).

Lemma 5.28. *Positive zero-sum stochastic games with finite state space and action spaces have a value. Moreover the minimizer has an optimal (stationary) strategy.*

Proof of Proposition 5.9. We restate the stochastic search game as a positive zero-sum stochastic game with finite state and action spaces and apply Lemma 5.28. The idea is that the stage payoff of the searcher is 1 at each stage until he finds the hider, transitioning then to an absorbing state in which the payoff is 0 forever. The total payoff is then the sum of the stage payoffs.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be the underlying graph. In order to cast our problem in the framework of Flesch et al. (2019), we will use a larger state space than the $\mathbb{G} \times \mathcal{V}$, the state space used in Section 5.2. Let $(\mathcal{V} \times \mathbb{G} \times (\mathcal{E} \cup \{\dagger\})) \cup \{*\}$ be the finite state space.

The state $(v_0, (\mathcal{V}, \emptyset), \dagger)$ is the initial state at stage 0, in which the hider has not chosen an edge where to hide. In this state, the finite action space of the searcher is $\mathcal{N}((\mathcal{V}, \emptyset), v_0) = \{v_0\}$, the finite action space of the hider is $\mathcal{E}$ and the payoff is 0. The state $*$ is an absorbing state in which the payoff is 0 forever. In any other state the payoff is 1.

The state moves from the initial state to $(v_0, \mathcal{G}_1, e)$ where $\mathcal{G}_1$ is the graph drawn at stage 1 and e is the edge chosen by the hider (which is fixed for the rest of the game). In any state $(v, \mathcal{G}', e) \in \mathcal{V} \times \mathbb{G} \times \mathcal{E}$ the searcher selects $v' \in \mathcal{N}(\mathcal{G}', v)$ and the hider selects $e \in \{e\}$. If $\{v, v'\} = e$ then the state next moves to the absorbing state $*$. If $\{v, v'\} \neq e$, the state moves to $(v', \mathcal{G}'', e)$ where $\mathcal{G}''$ is drawn according to the activation parameter.

Finally, since $\mathcal{E}$ is finite, the hider has an optimal strategy. □

Omitted proofs of Section 5.6

Proof of Theorem 5.18. We proceed by induction on the number of edges in the tree $\mathcal{T}$. If $\mathcal{T}$ has only one edge e, then

$$g(e, \sigma_\alpha) = \frac{1}{p} = \frac{1}{2}\left(\frac{2}{p} + 0\right). \tag{5.7.4}$$

Suppose that for any tree that has less edges than $\mathcal{T}$, the time to reach any leaf edge using the BDFS is $\frac{1}{2}\tau(O) + \Lambda(O)$.

If the origin O has degree 1 (as in Fig. 5.2), then, for any leaf edge e, we have

$$g(e, \sigma_\alpha) = \frac{1}{p} + \frac{1}{2}(\tau(v) + \Lambda(v)) = \frac{1}{2}\left(\tau(v) + \frac{2}{p} + \Lambda(v)\right) = \frac{1}{2}(\tau(O) + \Lambda(O)). \tag{5.7.5}$$

Consider the case where O has degree 2 (as in Fig. 5.4) and let e_1 be a leaf edge in $\mathcal{T}_{v_1}$. Then

$$\begin{aligned} g(e_1,\sigma_\alpha) &= (1-p)^2(1+g(e_1,\sigma_\alpha)) \\ &\quad + p(1-p)\left(1+\frac{1}{2}\tau(v_1)+\Lambda(v_1)+1+\tau(v_2)+\frac{2}{p}+\frac{1}{2}\tau(v_1)+\Lambda(v_1)\right) \\ &\quad + p^2\left(\alpha(e_1)\left(1+\frac{1}{2}\tau(v_1)+\Lambda(v_1)\right)+\alpha(e_2)\left(1+\tau(v_2)+\frac{2}{p}+\frac{1}{2}\tau(v_1)+\Lambda(v_1)\right)\right). \end{aligned}$$

and

$$\begin{aligned} g(e_1,\sigma_\alpha)(1-(1-p)^2) = 1 + p(1-p)\left(\tau(v_1)+\tau(v_2)+\frac{2}{p}+2\Lambda(v_1)\right) \\ + p^2\left(\frac{1}{2}\tau(v_1)+\Lambda(v_1)+\alpha(e_2)\left(\tau(v_2)+\frac{2}{p}\right)\right). \end{aligned}$$

Furthermore,

$$\tau(O) = \tau(v_1)+\tau(v_2)+\frac{3}{p}+\frac{1}{1-(1-p)^2}$$

and

$$\tau(e_1)+\tau(e_2) = \tau(v_1)+\tau(v_2)+\frac{4}{p} = \tau(O)+\frac{1}{p}-\frac{1}{1-(1-p)^2}.$$

Hence, by Lemma 5.16, for p large enough we do not need the projection in Eq. (5.6.6), so we have

$$\begin{aligned} g(e_1,\sigma_\alpha)(1-(1-p)^2) = 1 + p(1-p)\left(\tau(O)-\frac{1}{p}-\frac{1}{1-(1-p)^2}+2\Lambda(e_1)\right) \\ + p^2\left(\frac{1}{2}\left(\tau(e_1)-\frac{2}{p}\right)+\Lambda(e_1)+\left(\frac{1}{2}+\frac{\Lambda(e_2)-\Lambda(e_1)}{\tau(e_1)+\tau(e_2)}\frac{1-(1-p)^2}{p^2}\right)\tau(e_2)\right). \end{aligned}$$

Thus,

$$\begin{aligned} g(e_1,\sigma_\alpha) &= \frac{1}{1-(1-p)^2}+\frac{1}{2}\left(\tau(O)-\frac{1}{p}-\frac{1}{1-(1-p)^2}+2\Lambda(e_1)\right)+\frac{\Lambda(e_2)-\Lambda(e_1)}{\tau(e_1)+\tau(e_2)}\tau(e_2) \\ &= \frac{1}{2}\tau(O)+\frac{1}{2}\left(\frac{1}{1-(1-p)^2}-\frac{1}{p}\right)+\Lambda(e_1)\frac{\tau(e_1)}{\tau(e_1)+\tau(e_2)}+\Lambda(e_2)\frac{\tau(e_2)}{\tau(e_1)+\tau(e_2)} \\ &= \frac{1}{2}\tau(O)+\Lambda(O). \end{aligned}$$ ◻

Proof of Theorem 5.19. The proof is by induction on the number of edges of the tree $\mathcal{T}$. If $\mathcal{T}$ has only one edge, the result is immediate. Suppose now that the results holds for any tree with less edges than $\mathcal{T}$.

If the degree of the origin O is 1, the result follows immediately from the induction hypothesis. Assume now that the degree of O is 2 (as in Fig. 5.4). Let

$s(v_1)$ and $s(v_2)$ be two DFSs on $\mathcal{T}_{v_1}$ and $\mathcal{T}_{v_2}$, respectively. Let $s(e_1)$ be the pure DFS on $\mathcal{T}$ that, when both e_1 and e_2 are active, takes edge e_1 concatenated with $s(v_1)$ and then $s(v_2)$, in case Harry is not found in $\mathcal{T}_{v_1}$.

The pure strategy $s(e_2)$ is defined analogously.

Given a vertex v, call ε_v^* the conditional probability measure on $\mathcal{E}_v$ induced by ε^*. Then

$$\begin{aligned} g(\varepsilon^*, s(e_1)) = {} & (1-p)^2(1 + g(\varepsilon^*, s(e_1))) \\ & + p\left(\varepsilon^*(\mathcal{E}_{e_1})(1 + g(\varepsilon^*_{v_1}, s(v_1)) + \varepsilon^*(\mathcal{E}_{e_2})\left(1 + \frac{2}{p} + \tau(v_1) + g(\varepsilon^*_{v_2}, s(v_2))\right)\right) \\ & + p(1-p)\left(\varepsilon^*(\mathcal{E}_{e_2})(1 + g(\varepsilon^*_{v_2}, s(v_2)) + \varepsilon^*(\mathcal{E}_{e_1})\left(1 + \frac{2}{p} + \tau(v_2) + g(\varepsilon^*_{v_1}, s(v_1))\right)\right). \end{aligned}$$

Hence,

$$\begin{aligned} g(\varepsilon^*, s(e_1)) = g(\varepsilon^*, s(e_2)) & \iff \varepsilon^*(\mathcal{E}_{e_1})\left(\frac{2}{p} + \tau(v_2)\right) = \varepsilon^*(\mathcal{E}_{e_2})\left(\frac{2}{p} + \tau(v_1)\right) \\ & \iff \varepsilon^*(\mathcal{E}_{e_1}) = \frac{\tau(e_1)}{\tau(e_1) + \tau(e_2)}. \end{aligned}$$

□

Omitted proofs of Section 5.7

Proof of Theorem 5.24. We denote $e(i,j)$ the j-th edge of path i, starting from the root O. We proceed by induction on m.

Consider that, with probability $(1-p)^{2m}$ all edges starting from O are inactive; if this happens, Sally has to wait one turn and her payoff is $(1+g(e(i,j),\sigma^*))$. With probability $1-(1-p)^{2m}$ at least one edge is active and each of the available edges is chosen with equal probability. Given that Harry hides in $e(i,j)$, if the chosen path is i, then the game ends in $(j-1)/p$ units of time. If the chosen path is $k \neq i$, then Sally goes to D and the continuation payoff is $g_k(e(i,\lambda_i - j + 1), \sigma^*)$, where g_k is the payoff of the game played on $\mathcal{P}_{2m-1}(\boldsymbol{\lambda} \setminus \lambda_k)$, in which path k has been visited, $\boldsymbol{\lambda} \setminus \lambda_k$ is the vector $(\lambda_1, \dots, \lambda_{k-1}, \lambda_{k+1}, \dots, \lambda_{2m})$ of size $2m-1$, and the game starts in D.

In formula:

$$\begin{aligned} g(e(i,j),\sigma^*) = {} & (1-p)^{2m}(1 + g(e(i,j),\sigma^*)) \\ & + \frac{1-(1-p)^{2m}}{2m}\left(1 + \frac{j-1}{p} + \sum_{k \neq i}\left(1 + \frac{\lambda_k - 1}{p} + g_k(e(i,\lambda_i - j + 1),\sigma^*)\right)\right). \end{aligned}$$

The above expression yields

$$g(e(i,j),\sigma^*) = \frac{1}{1-(1-p)^{2m}} + \frac{1}{2m}\left(\frac{j-1}{p} + \sum_{k \neq i}\left(\frac{\lambda_k - 1}{p} + g_k(e(i,\lambda_i - j + 1),\sigma^*)\right)\right) \tag{5.7.6}$$

A similar expression holds for $g_k(e(i, \lambda_i - j + 1), \sigma^*)$. Plugging it in Eq. (5.7.6), we obtain

$$g(e(i,j),\sigma^*) = \frac{1}{1-(1-p)^{2m}} + \frac{1}{2m}\left(\frac{j-1}{p} + \sum_{k\neq i}\left(\frac{\lambda_k - 1}{p} + \frac{1}{1-(1-p)^{2m-1}} + \frac{1}{2m-1}\left(\frac{\lambda_i - j}{p} + \sum_{k'\neq k,i}\left(\frac{\lambda_{k'}-1}{p} + g_{k,k'}(e(i,j),\sigma^*)\right)\right)\right)\right), \tag{5.7.7}$$

where $g_{k,k'}$ is the payoff of the game played on $\mathcal{P}_{2(m-1)}(\boldsymbol{\lambda} \setminus \lambda_k, \lambda_{k'})$, in which both path k and path k' have been visited. The induction hypothesis is

$$g_{k,k'}(e(i,j),\sigma^*) = \frac{\theta(\mathcal{P}_{2(m-1)}(\boldsymbol{\lambda} \setminus \lambda_k, \lambda_{k'})) + p^{-1}}{2} + \Phi_{m-1}. \tag{5.7.8}$$

Therefore, plugging Eq. (5.7.8) into Eq. (5.7.7), we get

$$\begin{aligned} g(e(i,j),\sigma^*) = {} & \frac{1}{1-(1-p)^{2m}} + \frac{2m-1}{2m}\frac{1}{1-(1-p)^{2m-1}} \\ & + \frac{1}{2m}\left(1 + \frac{2m-2}{2m-1}\right)\sum_{k\neq i}\frac{\lambda_k - 1}{p} + \frac{1}{2m}\frac{\lambda_i - 1}{p} \\ & + \frac{1}{2m(2m-1)}\sum_{k\neq i}\sum_{k'\neq i,k}\left(\frac{\theta(\mathcal{P}_{2(m-1)}(\boldsymbol{\lambda} \setminus \lambda_k, \lambda_{k'})) + p^{-1}}{2} + \Phi_{m-1}\right). \end{aligned}$$

Furthermore,

$$\begin{aligned} \sum_{k\neq i}\sum_{k'\neq i,k}\left(\theta(\mathcal{P}_{2(m-1)}(\boldsymbol{\lambda} \setminus \lambda_k, \lambda_{k'})) + p^{-1}\right) = {} & \frac{(2m-1)(2m-2)}{p} \\ & + (2m-1)(2m-2)\sum_{k=1}^{2(m-1)}\frac{1}{1-(1-p)^k} \\ & + (2m-1)(2m-2)\frac{\lambda_i - 1}{p} \\ & + (2m-2)(2m-3)\sum_{k\neq i}\frac{\lambda_k - 1}{p}. \end{aligned}$$

And finally, one obtains the following simplifications

$$
\begin{aligned}
g(e(i,j),\sigma^*) &= \frac{1}{1-(1-p)^{2m}} + \frac{2m-1}{2m}\frac{1}{1-(1-p)^{2m-1}} + \frac{2m-2}{4m}\left(\frac{1}{p} + \sum_{k=1}^{2(m-1)} \frac{1}{1-(1-p)^k}\right) \\
&\quad + \frac{1}{2m}\left(1+\frac{2m-2}{2m-1}\right)\sum_{k\neq i}\frac{\lambda_k-1}{p} + \frac{1}{2m}\frac{\lambda_i-1}{p} + \frac{2m-2}{4m}\frac{\lambda_i-1}{p} \\
&\quad + \frac{(2m-2)(2m-3)}{4m(2m-1)}\sum_{k\neq i}\frac{\lambda_k-1}{p} + \frac{2m-2}{2m}\Phi_{m-1} \\
&= \frac{1}{1-(1-p)^{2m}} + \frac{2m-1}{2m}\frac{1}{1-(1-p)^{2m-1}} \\
&\quad + \frac{m-1}{2m}\left(\frac{1}{p} + \sum_{k=1}^{2(m-1)}\frac{1}{1-(1-p)^k}\right) + \frac{1}{2}\sum_{k=1}^{2m}\frac{\lambda_k-1}{p} + \frac{m-1}{m}\Phi_{m-1} \\
&= \frac{\theta(\mathcal{P}_{2m}(\boldsymbol{\lambda}))+p^{-1}}{2} + \Phi_m. \qquad \square
\end{aligned}
$$

Proof of Theorem 5.25. The proof is by induction on the number of parallel paths. Let s be a ES of Sally, and denote

$$L = \sum_{k=1}^{2m} \lambda_k$$

the number of edges of $\mathcal{P}_{2m}(\boldsymbol{\lambda})$. First,

$$g(\varepsilon^{\mathrm{U}}, s) = \frac{1}{1-(1-p)^{2m}} + \frac{1}{2m}\left(1-\frac{1}{L}\right)\sum_{k=1}^{2m} g^{\lambda_k-1}(\varepsilon^{\mathrm{U}}, s),$$

where $g^{\lambda_k-1}(\varepsilon^{\mathrm{U}}, s)$ is the payoff of the continuation game after one edge of path k has been visited. It is not difficult to prove that

$$(L-1)g^{\lambda_k-1}(\varepsilon^{\mathrm{U}}, s) = (L-\lambda_i)g_k(\varepsilon^{\mathrm{U}}, s) + \frac{L(\lambda_k-1)}{p} - \frac{\lambda_k(\lambda_k-1)}{2p},$$

where g_k is the payoff of the game played on $\mathcal{P}_{2(m-1)}(\boldsymbol{\lambda}\setminus\lambda_k)$, in which path k has been visited. Therefore

$$g(\varepsilon^{\mathrm{U}}, s) = \frac{1}{1-(1-p)^{2m}} + \frac{L}{2mp} - \frac{1}{p} + \frac{1}{2mL}\sum_{k=1}^{2m}\left(\frac{-\lambda_k(\lambda_k-1)}{2p} + (L-\lambda_k)g_k(\varepsilon^{\mathrm{U}}, s)\right).$$

Computing a similar expression for $g_k(\varepsilon^{\mathrm{U}}, s)$ and plugging it in the above equation one has

$$g(\varepsilon^{\mathrm{U}}, s) = \frac{1}{1-(1-p)^{2m}} + \frac{2m-1}{2m}\frac{1}{1-(1-p)^{2m-1}} + \frac{L}{2mp} - \frac{1}{p} - \frac{2m-1}{2mp}$$
$$+ \frac{1}{2mL}\sum_{k=1}^{2m}\Bigg(\frac{-\lambda_k(\lambda_k-1)}{2p} + \frac{(L-\lambda_k)^2}{(2m-1)p}$$
$$+ \frac{1}{(2m-1)}\sum_{k'\neq k}\left(\frac{-\lambda_{k'}(\lambda_{k'}-1)}{2p} + (L-\lambda_k-\lambda_{k'})g_{k,k'}(\varepsilon^{\mathrm{U}}, s)\right)\Bigg),$$

where $g_{k,k'}$ is the payoff of the game played on $\mathcal{P}_{2(m-1)}(\boldsymbol{\lambda}\setminus\lambda_k,\lambda_{k'})$, in which both path k and path k' have been visited. From the induction hypothesis, one has

$$g_{k,k'}(\varepsilon^{\mathrm{U}}, s) = \frac{1}{2}\left(\frac{1}{p} + \frac{L-\lambda_k-\lambda_{k'}-2(m-1)}{p} + \sum_{l=1}^{2(m-1)}\frac{1}{1-(1-p)^l}\right) + \Phi_{m-1}.$$

Plugging this expression in the previous equation, one has

$$g(\varepsilon^{\mathrm{U}}, s) = \frac{1}{1-(1-p)^{2m}} + \frac{2m-1}{2m}\frac{1}{1-(1-p)^{2m-1}} + \frac{L}{2mp} - \frac{1}{p} - \frac{2m-1}{2mp} + \frac{m-1}{m}\Phi_{m-1}$$
$$+ \frac{m-1}{2m}\left(\frac{1}{p} - \frac{2(m-1)}{p} + \sum_{l=1}^{2(m-1)}\frac{1}{1-(1-p)^l}\right)$$
$$+ \frac{1}{2mL}\sum_{k=1}^{2m}\left(\frac{-\lambda_k(\lambda_k-1)}{2p} + \frac{(L-\lambda_k)^2}{(2m-1)p} + \frac{1}{2m-1}\sum_{k'\neq k}\left(\frac{-\lambda_{k'}(\lambda_{k'}-1)}{2p} + \frac{(L-\lambda_k-\lambda_{k'})^2}{2p}\right)\right).$$

Furthermore

$$\frac{1}{2mL}\sum_{k=1}^{2m}\left(\frac{-\lambda_k(\lambda_k-1)}{2p} + \frac{(L-\lambda_k)^2}{(2m-1)p} + \frac{1}{2m-1}\sum_{k'\neq k}\left(\frac{-\lambda_{k'}(\lambda_{k'}-1)}{2p} + \frac{(L-\lambda_k-\lambda_{k'})^2}{2p}\right)\right)$$
$$= \frac{1}{2mp} + \frac{(m-1)L}{2mp}.$$

And finally one has

$$g(\varepsilon^{\mathrm{U}}, s) = \frac{1}{1-(1-p)^{2m}} + \frac{2m-1}{2m}\frac{1}{1-(1-p)^{2m-1}} + \frac{m-1}{2m}\left(\frac{1}{p} + \sum_{k=1}^{2(m-1)}\frac{1}{1-(1-p)^k}\right)$$
$$+ \frac{1}{2}\sum_{k=1}^{2m}\frac{\lambda_k-1}{p} + \frac{m-1}{m}\Phi_{m-1}$$
$$= \frac{\theta(\mathcal{P}_{2m}(\boldsymbol{\lambda})) + p^{-1}}{2} + \Phi_m. \qquad \square$$

Chapter 6

Dynamic control of information with observed return on investment

Article written in collaboration with Pr. Jérôme Renault, in preparation

We study a model of dynamic control of information between an advisor and an investor. Each day, the advisor, who privately knows the changing state of nature discloses some information to the investor, in order to manipulate her decision. The investor may decide to invest or not so as to maximize her current payoff. In case she invests, she pays a fee to the advisor and observes whether her investment was successful or not, obtaining more information on the state of nature. The advisor aims at maximizing the frequency of days in which the investor invests. Our focus is on the greedy information disclosure strategy of the advisor, which minimizes the information disclosed while maximizing his current payoff. The greedy strategy proves to be optimal when there are two states of nature, but may fail to be when there are more.

6.1 Introduction

We study a model of dynamic control of information between an advisor and an investor, or a sequence of short-lived investors. The advisor has a private knowledge of the state of nature, which randomly evolves with time. Every day the advisor chooses the amount of information he discloses to the investor through messages. In turn, the investor myopically decides whether to invest or not so as to maximize her current payoff. The investor observes previous investment outcomes. In case of investment, the advisor receives a fixed fee from the investor, hence his goal is to maximize the discounted frequency of days in which investment occurs.

Because the investor follows a myopic behavior, the advisor faces a Markov decision process (MDP). As in Bayesian persuasion models, the state space of the MDP is the set of posterior beliefs of the investor, and the action set is the set of information disclosure possibilities. In this MDP the advisor acts honestly in the sense that the realization of sent messages cannot be manipulated, however he is also strategic in the way he correlates the messages with the state of nature.

Our focus is on a special information disclosure strategy of the advisor called the greedy strategy. It is a myopic (stationary) strategy which has the property of minimizing the information disclosed while maximizing the current payoff. We prove that when there are two states of nature the greedy strategy is optimal. However, when there are more than two states of nature, the greedy strategy may fail to be optimal. This is proved via a counterexample of Renault et al. (2017).

This paper lies within the scope of the literature on repeated games with incomplete information, the roots of which can be found in (Aumann and Maschler, 1995). More precisely, it belongs to the effervescent literature on Bayesian persuasion which began with the work of Kamenica and Gentzkow (2011). Renault et al. (2017) dealt with a similar model as ours but in which the investor does not observe the outcomes of previous investments. Ely (2017) considers a model similar to Renault et al. (2017) in continuous time for which he fully solves an example and studies several variants.

In Section 6.2 we define formally the model. Then in Section 6.3 we study the value function as well as the dynamic programming operator. In Section 6.4 we define the greedy strategy and give necessary and sufficient conditions for it to be optimal. Section 6.5 is dedicated to the two-state case, for which we prove that the greedy strategy is optimal for any initial belief. We also provide examples for which we compute the value function. Finally in Section 6.6 we remark that the greedy strategy may not be optimal for any initial belief if there are more than two states of nature. Here also we provide examples for which we compute the value function.

6.2 The model

We study a two-player dynamic game between an *advisor* and short-lived *investors*, simply called the investor. At each stage n in time, the advisor privately observes the realization ω_n of a changing state of nature which has value in a finite set Ω. The advisor decides which information to disclose to the current investor, who in turn decides whether to invest or not. If the current investor chooses to invest, she observes whether the investment is successful or not. The game then moves to the next stage.

Whenever investment takes place, the advisor receives a fee which is normalized to 1, and discounts future payoffs according to the discount factor $\delta < 1$. If she chooses to invest, the investor pays the investment fee of 1, and gets a reward of $M > 1$ with probability $r(\omega_n)$ and 0 with probability $1 - r(\omega_n)$, with

$r : \Omega \to [0, 1]$.

We assume that the sequence $(\omega_n)_{n\geq 0}$ follows a Markov chain with transition $(\rho(\omega'|\omega))_{\omega',\omega\in\Omega}$. The investor knows the distribution of the sequence $(\omega_n)_{n\geq 0}$. The additional information received along the play comes from the advisor and from the observed returns on investments. The investor chooses to invest if and only if the expected payoff from investing is nonnegative. That is, given the current belief is $p \in \Delta(\Omega)$, denoting $c = 1/M$, if $\langle p, r\rangle \geq c$. Accordingly, the investment region is $I = \{p \in \Delta(\Omega),\ \langle p, r\rangle \geq c\}$. Conversely the noninvestment region is $J = \Delta(\Omega) \setminus I$.

The game reduces to a stochastic one-player game, or Markov decision process, that we denote Γ, in which the advisor manipulates the posterior beliefs of the investor, so as to maximize the expected discounted frequency of stages in which investment takes place.

The investor uses her knowledge of the distribution of $(\omega_n)_{n\geq 0}$ as well as the return on investment she observes to update her posterior belief. We distinguish three beliefs of the investor at a stage n. The *prior belief* $p_n \in \Delta(\Omega)$ is the belief in stage n before receiving the message of the advisor. The *intermediate belief* q_n is the updated belief right after the message has been received. In particular the investor invests in stage n if and only if $q_n \in I$. Finally, the *posterior belief* s_n is the belief updated from q_n after the return on investment (if any) has been observed.

The belief p_{n+1} may differ from s_n because the states are not fully persistent. For each $\omega' \in \Omega$ one has $p_{n+1}(\omega') = \sum_{\omega\in\Omega} s_n(\omega)\rho(\omega'|\omega)$. We accordingly define the linear map $\phi : \Delta(\Omega) \to \Delta(\Omega)$ such that $p_{n+1} = \phi(s_n)$.

The belief q_n may differ from p_n because of information disclosure. For a given belief $p \in \Delta(\Omega)$, let $\mathcal{S}(p) \subset \Delta(\Delta(\Omega))$ be the set of probability measures over $\Delta(\Omega)$ with mean p. Elements of $\mathcal{S}(p)$ will be called splittings at p. It is a consequence of Bayesian updating that for every information disclosure strategy, the conditional distribution of q_n belongs to $\mathcal{S}(p_n)$. Reciprocally, the splitting lemma, see (Aumann and Maschler, 1995), states that given a belief $p \in \Delta(\Omega)$ and a splitting $\mu \in \mathcal{S}(p)$, the advisor can correlate his message with the state of nature so that the distribution of the updated belief q of the investor be μ.

Finally, the belief s_n may differ from q_n due to the observed return on investment. In case the investor does not invest, the belief is unchanged, that is $q_n = s_n$. In case of investment, with probability $r(\omega_n)$ the investor observes a success and one has, for all $\omega \in \Omega$, that $s_n(\omega) = \frac{q_n(\omega)r(\omega)}{\langle q_n, r\rangle}$. On the other hand, with probability $1 - r(\omega_n)$ the investor observes a failure, and one has, for all $\omega \in \Omega$, that $s_n(\omega) = \frac{q_n(\omega)(1-r(\omega))}{1-\langle q_n, r\rangle}$.

We define the maps ψ^+ and ψ^- from $\Delta(\Omega)$ to $\Delta(\Omega)$ as follows: for any $p \in \Delta(\Omega)$ and $\omega \in \Omega$, one lets $\psi^+(p)(\omega) = \frac{p(\omega)r(\omega)}{\langle p, r\rangle}$ and $\psi^-(p)(\omega) = \frac{p(\omega)(1-r(\omega))}{1-\langle p, r\rangle}$. Hence in case of investment at stage n, one has $s_n = \psi^+(q_n)$ in case of success, and $s_n = \psi^-(q_n)$ in case of failure.

6.3 The value of the game

In this section we introduce the value of the game. It is convenient to define this value as the fixed point of a dynamic programming operator.

Let E be the set of functions from $\Delta(\Omega)$ to $[0,1]$, endowed with the distance $\|f_1 - f_2\| = \sup_{p\in\Delta(\Omega)} |f_1(p) - f_2(p)|$. The dynamic programming operator T is defined from E to E by

$$T(f) : p \mapsto \begin{cases} \delta f(\phi(p)) & \text{if } p \in J \\ 1-\delta+\delta\langle p,r\rangle f(\phi(\psi^+(p))) + \delta(1-\langle p,r\rangle) f(\phi(\psi^-(p))) & \text{if } p \in I. \end{cases}$$

Notice that T is a δ-contraction, i.e., $\|T(f_1) - T(f_2)\| \leq \delta\|f_1 - f_2\|$ for all $f_1, f_2 \in E$. We denote cav the concavification operator, so that for each f in E, the function $\operatorname{cav} T(f)$ is the smallest concave function not lower than $T(f)$.

We denote $v_\delta(p)$ the *value* of the dynamic optimization problem Γ as a function of the initial distribution p. Since the operator $\operatorname{cav} T : E \to E$ is also a δ-contraction and the set E is complete, the value function v_δ is characterized as the unique solution of the functional equation

$$\operatorname{cav} T(f) = f. \tag{6.3.1}$$

An immediate consequence of the definition is that the value function v_δ is concave. Note that Eq. (6.3.1) can also be rewritten

$$\max_{\mu\in\mathcal{S}(p)} \mathbb{E}_\mu [T(f)(q)] = f(p), \text{ for all } p \in \Delta(\Omega).$$

6.4 The greedy strategy

At each stage, if the current belief p is not in the investment region, the advisor has to compromise. To get a positive payoff in the current stage, some information has to be disclosed. However, due to the concavity of the value function, this may lower future payoffs. To deal with this matter, we introduce in this section the *greedy strategy*, which minimizes the information being disclosed while maximizing the current payoff.

We now define the greedy strategy of the advisor, which is denoted σ_*. It is such that if $p \in I$, then the advisor does not disclose any information. However if $p \in J$ then we consider the following problem

$$\begin{aligned} \text{maximize} \quad & a_I \\ \text{subject to} \quad & p = a_I q_I + a_J q_J, \\ & q_I \in I, \\ & a_I + a_J = 1,\ a_I, a_J \geq 0. \end{aligned} \tag{6.4.1}$$

Given a solution (a_I, a_J, q_I, q_J) of problem 6.4.1, $\sigma_*(p) \in \mathcal{S}(p)$ is the splitting which selects q_I and q_J with probabilities a_I and a_J respectively.

Consider for a moment the case card $\Omega = 2$. If $p \in J$, then one verifies that the greedy splitting at p selects 0 and p^* with probabilities $1 - \frac{p}{p^*}$ and $\frac{p}{p^*}$ respectively.

For card $\Omega = 3$ we give more details on Example 6.1 below.

Example 6.1. We represent the situation Fig. 6.1. The state space is $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Moreover $r(\omega_1) > c$ and $r(\omega_2), r(\omega_3) < c$. The investment frontier $\{p \in \Delta(\Omega),\ \langle p, r\rangle = c\}$ is the segment $[p_2, p_3]$. Since p_2 is closer to ω_2 than p_3 is to ω_3, one has that $r(\omega_2) > r(\omega_3)$. Since the segment $[p_2, p_3]$ has positive slope,

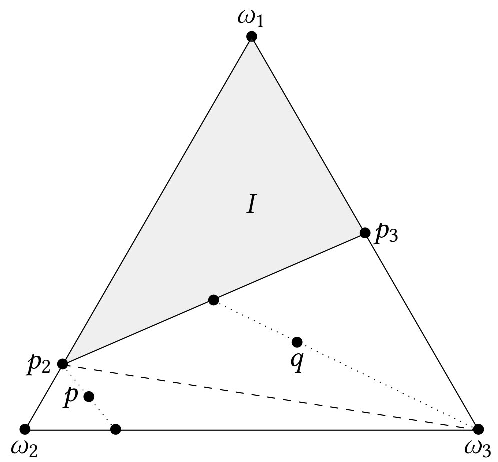

Figure 6.1: The greedy strategy for card $\Omega = 3$

a belief p in the triangle $(p_2, \omega_2, \omega_3)$ is split between p_2 and a point on the segment $[\omega_2, \omega_3]$. Similarly, a belief q in the triangle (p_2, ω_3, p_3) is split between ω_3 and a point on the investment frontier $[p_2, p_3]$.

Moreover the greedy splitting is uniquely defined, unless the segment $[p_2, p_3]$ be parallel to the segment $[\omega_2, \omega_3]$, in which case $r(\omega_2) = r(\omega_3)$. One may generalize this fact, see Lemma 6.1 below.

The following lemma about the uniqueness of the greedy strategy is proved by Renault et al. (2017).

Lemma 6.1. *Suppose that for every $\omega, \omega' \in \Omega$ one has $r(\omega) \neq r(\omega')$. Then for all $p \in J$ the greedy splitting $\sigma_*(p)$ is uniquely defined.*

Let g be the payoff function induced by the greedy strategy. It satisfies

$$g : p \mapsto \begin{cases} a_I g(q_I) + a_J g(q_J) & \text{if } p \in J \\ 1 - \delta + \delta\langle p, r\rangle g(\phi(\psi^+(p))) + \delta(1 - \langle p, r\rangle) g(\phi(\psi^-(p))) & \text{if } p \in I. \end{cases} \tag{6.4.2}$$

Let $U : E \longrightarrow E$ be the *greedy operator* defined by

$$U(f) : p \mapsto \begin{cases} a_I\left(1-\delta+\delta\langle q_I, r\rangle f(\phi(\psi^+(q_I)))+\delta(1-\langle q_I, r\rangle) f(\phi(\psi^-(q_I)))\right) & \\ +a_J \delta f(\phi(q_J)) & \text{if } p \in J \\ 1-\delta+\delta\langle p, r\rangle f(\phi(\psi^+(p)))+\delta(1-\langle p, r\rangle) f(\phi(\psi^-(p))) & \text{if } p \in I. \end{cases} \tag{6.4.3}$$

Note that the greedy operator U is a δ-contraction. The greedy payoff function g is thus the unique fixed point of the greedy operator U.

The next proposition gives a necessary and sufficient condition for the greedy strategy to be optimal, which appears to be very convenient.

Proposition 6.2. *The greedy strategy is optimal if and only if*

1. *g is concave, and;*
2. *for all $p \in J$, one has $\delta g \circ \phi(p) \leq g(p)$.*

Proof. Suppose σ_* is optimal, then $g = v_\delta$ and g is concave. Moreover the dynamic programming principle implies that for all $p \in J$, one has $\delta g \circ \phi(p) \leq g(p)$, the first term being the overall payoff if the advisor do not disclose any information in the first stage and then switches to σ_*.

Conversely, suppose g is concave and for all $q \in J$, one has $\delta g \circ \phi(q) \leq g(q)$. Let $\mu \in \mathcal{S}(p)$ and let $q \in \Delta(\Omega)$. If $q \in I$, then

$$g(q) = 1-\delta+\delta\left(\langle q, r\rangle g(\phi \circ \psi^+(q))+(1-\langle q, r\rangle) g(\phi \circ \psi^-(q))\right).$$

Hence

$$\begin{aligned} &\mathbb{1}_{q \in J}\, \delta g(\phi(q))+\mathbb{1}_{q \in I}\left(1-\delta+\delta\left(\langle q, r\rangle g(\phi \circ \psi^+(q))+(1-\langle q, r\rangle) g(\phi \circ \psi^-(q))\right)\right) \\ &\leq g(q). \end{aligned}$$

Taking the expectation with respect to μ, one has

$$\begin{aligned} &\mathbb{E}_\mu\big(\mathbb{1}_{q \in J}\, \delta g(\phi(q)) \\ &\qquad +\mathbb{1}_{q \in I}\left(1-\delta+\delta\left(\langle q, r\rangle g(\phi \circ \psi^+(q))+(1-\langle q, r\rangle) g(\phi \circ \psi^-(q))\right)\right)\big) \\ &\leq \mathbb{E}_\mu g(q) \\ &\leq g(p), \end{aligned}$$

where the last inequality comes from the concavity of g. Finally, taking the maximum with respect to μ over $\mathcal{S}(p)$, one has that $g \geq v_\delta$, and hence σ_* is optimal. □

6.5 The two-state case

We now consider the case card$(\Omega) = 2$, that is $\Omega = \{\omega_+, \omega_-\}$. We let $s = r(\omega_+)$ and $t = r(\omega_-)$, with $s > c > t$. We consider beliefs in $[0, 1]$, that is we identify a belief $p \in \Delta(\Omega)$, with the probability $p \in [0,1]$ assigned to ω_+. The belief $p^* \in \Delta(\Omega)$ is such that the investment region I is $[p^*, 1]$, that is $p^* \in (0,1)$ solves $p^*s + (1-p^*)t = c$. The Markov transition mapping ϕ may be written

$$\phi : p \mapsto m + \lambda(p - m),$$

where m is the invariant measure, which assigns probability $\frac{\rho(\omega_+|\omega_-)}{\rho(\omega_+|\omega_-)+\rho(\omega_-|\omega_+)}$ to ω_+, and λ is the ratio $1 - \rho(\omega_+|\omega_-) - \rho(\omega_-|\omega_+) \in (-1,1)$.

We introduce the operator $L : E \longrightarrow E$ defined by

$$L(f) : p \mapsto \langle p, r\rangle f \circ \phi \circ \psi^+(p) + (1 - \langle p, r\rangle) f \circ \phi \circ \psi^-(p).$$

The greedy operator U becomes

$$U(f)(p) = \begin{cases} \frac{p}{p^*}(1 - \delta + \delta L(f)(p^*)) + \left(1 - \frac{p}{p^*}\right)\delta f(\phi(0)) & \text{if } p \leq p^* \\ 1 - \delta + \delta L(f)(p) & \text{if } p \geq p^*. \end{cases} \tag{6.5.1}$$

It is straightforward that the U maps continuous functions to continuous function, a direct corollary is the following lemma.

Lemma 6.3. *The greedy payoff function g is continuous.*

In Examples 6.2 and 6.3 below we study two particular cases for which we compute the payoff of the greedy strategy, and prove it is optimal.

Example 6.2. In this example we consider the case where the transition map ϕ is the identity mapping, that is, the state of nature is fully persistent from one stage to the next one. Note that in this case, $v_\delta(0) = 0$ and $v_\delta(1) = 1$. Note also that property 2 of Proposition 6.2: $\delta g \circ \phi(p) \leq g(p)$ for all $p < p^*$ holds, and that one only needs to verify that g is concave to assert that σ_* is optimal.

We moreover suppose that $t = 0$. The greedy operator U becomes

$$U(f)(p) = \begin{cases} \frac{p}{p^*} f(p^*) & \text{if } p \leq p^* \\ 1 - \delta + \delta(ps + (1-p)t) f(\psi^+(p)) & \\ +\delta(1 - ps - (1-p)t) f(\psi^-(p)) & \text{if } p \geq p^*. \end{cases} \tag{6.5.2}$$

And g satisfies for all $p \in [0,1]$

$$g(p) = \begin{cases} \frac{p}{p^*} f(p^*) & \text{if } p \leq p^* \\ 1 - \delta(1 - ps)\left(1 - g\left(\frac{p(1-s)}{1-sp}\right)\right) & \text{if } p \geq p^*. \end{cases}$$

Let $p_0 = 0, p_1 = p^*$ and for all $n \geq 1$ $p_{n+1} = (\psi^-)^{-1}(p_n)$. g is linear on $[p_n, p_{n+1}]$ for all $n \geq 0$. Let $g(p) = \alpha_n + \beta_n p$ for all $p \in [p_n, p_{n+1}]$. One has $\alpha_0 = 0$ and $\beta_0 = \frac{1-\delta+\delta p^* s}{p^*(1-\delta(1-s))}$. Moreover for all $n \geq 1$ and all $p \in [0,1]$

$$\alpha_n + \beta_n p = 1 - \delta(1 - ps) + \delta(1 - ps)\alpha_{n-1} + \delta\beta_{n-1}(p(1-s)).$$

This yields for all $n \geq 1$

$$\alpha_n = 1 - \delta(1 - \alpha_{n-1}),$$

hence for all $n \geq 0$

$$\alpha_n = 1 - \delta^n.$$

And for all for all $n \geq 1$

$$\beta_n = \delta(1 - \alpha_{n-1}) + \delta(1 - s)\beta_{n-1},$$

and therefore

$$\beta_n = s\delta^n + \delta(1 - s)\beta_{n-1}.$$

One has $\beta_1 - \beta_0 = s\delta + (\delta(1-s)\beta_0) = s\delta - \frac{1-\delta+\delta p^* s}{p^*} < 0$. Hence $\beta_0 > \beta_1$. Assume that $\beta_n < \beta_{n+1}$ for some $n \geq 0$. Since $\delta^{n+1}s > \delta^{n+2}s$, it follows that $\beta_{n+1} < \beta_{n+2}$. Finally, one has that g is concave, hence σ_* is optimal.

Remark that the greedy strategy is not the unique optimal strategy, since for any n and any $p \in [p_n, p_{n+1}]$, splitting in $[p_n, p_{n+1}]$ would also be optimal.

Example 6.3. In this example we deal with the case $s = 1$ and $t = 0$. In this case the greedy operator becomes

$$U(f)(p) = \begin{cases} \delta f \circ \phi(0) + p\left(\frac{1-\delta}{p^*} + \delta f \circ \phi(1) - \delta f \circ \phi(0)\right) & \text{if } p \leq p^* \\ 1 - \delta + \delta f \circ \phi(0) + p(\delta f \circ \phi(1) - \delta f \circ \phi(0)) & \text{if } p \geq p^*. \end{cases} \tag{6.5.3}$$

Therefore in this case, g is linear on $[0, p^*]$ and on $[p^*, 1]$. It is moreover straightforward to verify that g is concave, and thus that the greedy strategy is optimal. It is also not difficult to compute g.

Once again, the greedy strategy is not the unique optimal strategy, since for any $p \in I$, splitting in I would also be optimal.

We now go back to the general case. The next two Lemmas 6.5 and 6.6 below, together with Proposition 6.2 allow us to state Theorem 6.4.

Theorem 6.4. *If* $\operatorname{card}(\Omega) = 2$, *then the greedy strategy is optimal.*

Lemma 6.5. *Let $f \in E$ be a concave function, then $U(f)$ is concave.*

Proof. Let $f \in E$ be a concave function. Since $f \circ \phi$ is also a concave function we may as well assume that ϕ is the identity mapping, and hence we furthermore assume that $f(0) = 0$.

We first prove that $L(f)$ is concave. One has

$$L(f)(p) = \langle p, r\rangle f(\psi^+(p)) + (1 - \langle p, r\rangle) f(\psi^-(p)).$$

Let $\lambda \in [0, 1]$, one has

$$\begin{aligned}\psi^+(\lambda p + (1-\lambda)q) &= \left(\frac{\lambda p(\omega)r(\omega) + (1-\lambda)q(\omega)r(\omega)}{\lambda\langle p, r\rangle + (1-\lambda)\langle q, r\rangle}\right)_{\omega\in\Omega} \\ &= \frac{\lambda\langle p, r\rangle}{\lambda\langle p, r\rangle + (1-\lambda)\langle q, r\rangle}\psi^+(p) + \frac{(1-\lambda)(\langle q, r\rangle}{\lambda\langle p, r\rangle + (1-\lambda)\langle q, r\rangle}\psi^+(q).\end{aligned}$$

Since f is concave, one has

$$\begin{aligned}f(\psi^+(\lambda p+(1-\lambda)q)) \geq &\frac{\lambda\langle p,r\rangle}{\lambda\langle p,r\rangle+(1-\lambda)\langle q,r\rangle}f(\psi^+(p))\\ &+\frac{(1-\lambda)(\langle q,r\rangle}{\lambda\langle p,r\rangle+(1-\lambda)\langle q,r\rangle}f(\psi^+(q)),\end{aligned}$$

and similarly for ψ^-. Finally, one has

$$\begin{aligned}L(f)(\lambda p+(1-\lambda)q) &\geq \lambda\langle p,r\rangle f\circ\psi^+(p)+(1-\lambda)\langle q,r\rangle f\circ\psi^+(q)\\ &+\lambda\langle p,r\rangle f\circ\psi^-(p)+(1-\lambda)\langle q,r\rangle f\circ\psi^-(q)\\ &=\lambda L(f)(p)+(1-\lambda)L(f)(q).\end{aligned}$$

The function $U(f)$ is hence linear on $[0,p^*]$ and concave on $[p^*,1]$. Hence one only needs to verify that the left slope of $U(f)$ at p^* is at least the right slope. This is true if and only if

$$\begin{aligned}&\frac{1}{p^*}\left(1-\delta+\delta(\langle p^*,r\rangle f(\psi^+(p^*))+(1-\langle p^*,r\rangle)f(\psi^-(p^*)))\right)\\ \geq&\delta\Big((s-t)f(\psi^+(p^*))+\frac{st}{t+p^*(s-t)}f'_r(\psi^+(p^*))\\ &+(t-s)f(\psi^-(p^*))+\frac{(1-s)(1-t)}{1-t+p^*(t-s)}f'_r(\psi^-(p^*))\Big),\end{aligned}$$

where f'_r denotes the right derivative of f. Since the inequality is linear in δ, it is enough to verify that it holds for δ equal to 0 and to 1. If $\delta=0$, it boils down to $(1-\delta)/p^*\geq 0$, which is true. If $\delta=1$, it boils down to

$$\begin{aligned}\frac{t}{p^*}f(\psi^+(p^*))+\frac{1-t}{p^*}f(\psi^-(p^*))\geq&\frac{st}{t+p^*(s-t)}f'_r(\psi^+(p^*))\\ &+\frac{(1-s)(1-t)}{1-t+p^*(t-s)}f'_r(\psi^-(p^*)).\end{aligned}$$

The latter is equivalent to

$$\frac{t}{p^*}f(\psi^+(p^*))+\frac{1-t}{p^*}f(\psi^-(p^*))\geq\frac{t}{p^*}\psi^+(p^*)f'_r(\psi^+(p^*))+\frac{(1-t)}{p^*}\psi^-(p^*)f'_r(\psi^-(p^*)),$$

which holds true because, since f is concave, one has for all $p\in(0,1)$,

$$f(p)-pf'_r(p)\geq f(0).$$

□

A direct corollary of Lemma 6.5 is that the greedy payoff function g is concave on $[0,1]$.

Lemma 6.6. *For all $p<p^*$, one has $\delta g\circ\phi(p)\leq g(p)$.*

Proof. We define the function h on $[0, p^*)$ such that $h : p \mapsto g(p) - \delta g(\phi(p))$. One has $h(0) = 0$. Moreover since g is linear on $[0, p^*]$ and concave on $[p^*, 1]$, then h is convex, hence it is enough to verify that $h'(0) = g(0) - \delta\phi'(0)g'(\phi(0))$ is nonnegative.

Furthermore, since $g(1) = 1 - \delta + \delta g(\phi(1)) \geq g(\phi(1))$ and g is concave, one has that $g'(0) \geq 0$. Thus, $g'(0) \geq |g'(\phi(0))| \geq \delta\phi'(0)g'(\phi(0))$, since $\phi'(0) = \lambda \in (-1, 1)$. □

6.6 More than two states

In the next two Examples 6.4 and 6.5, we consider the case where the Markov chain is constant, that is $\phi = \mathrm{id}$. In this case, the second condition of Proposition 6.2 is trivially satisfied, so the greedy strategy is optimal if and only if the greedy payoff function g is concave. The state space is $\Omega = \{A, B, C\}$. Elements of $\Delta(\Omega)$ are denoted $p = (p_A, p_B, p_C)$.

Example 6.4. In this example, we assume that $r(C) \leq r(B) < c < r(A) = 1$. The situation is represented Fig. 6.2, with $D = \left(\frac{c - r(B)}{r(A) - r(B)}, \frac{r(A) - c}{r(A) - r(B)}, 0\right)$ and $E = \left(\frac{c - r(C)}{r(A) - r(C)}, 0, \frac{r(A) - c}{r(A) - r(C)}\right)$.

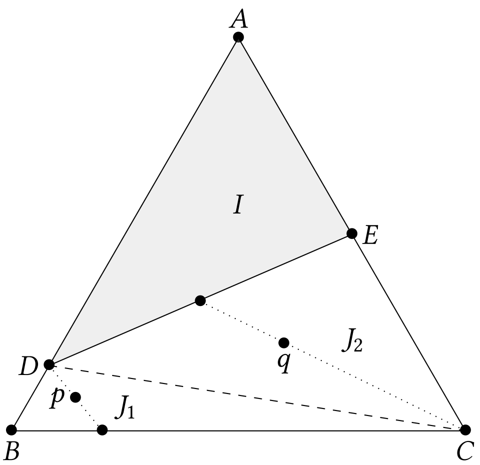

Figure 6.2: The situation of Example 6.4

Let J_1, J_2 and $J_3 = I$, be the convex hull of $\{D, C, B\}$, $\{D, C, E\}$ and $\{A, D, E\}$ respectively. For each p we write $r(p) = p_A r(A) + p_B r(B) + p_C r(C)$. Hence, $\psi^+(p) = \frac{1}{r(p)}(p_A r(A), p_B r(B), p_C r(C))$. Since $r(A) = 1$, one has that $\psi^-(p)$ always belongs to the segment $[B, C]$. Moreover, since ϕ is the identity mapping and B and C belong to J one has $v_\delta(\psi^-(p)) = 0$ for all p. The greedy payoff function is thus characterized by:

1. $g(A) = 1$ and $g(B) = g(C) = 0$;

2. g is linear on J_1;

3. if $q = a_I q_I + (1 - a_I)C \in J_2$, with $q_I \in [D, E]$ and $a_I \in [0, 1]$, then $g(p) = a_I g(q_I)$;

4. if $p \in I$, then $g(p) = 1 - \delta + \delta r(p) g\left(\frac{1}{r(p)}(p_A r(A), p_B r(B), p_C r(C))\right)$.

One easily checks that the equation of Item 4 above has a linear solution given by:

$$\begin{aligned} g_3(p) &= (1-\delta)\left(\frac{p_A}{1-\delta r(A)} + \frac{p_B}{1-\delta r(B)} + \frac{p_C}{1-\delta r(C)}\right) \\ &= p_A + \frac{1-\delta}{1-\delta r(B)} p_B + \frac{1-\delta}{1-\delta r(C)} p_C. \end{aligned}$$

In particular, one has:

$$\begin{aligned} &g_3(A) = 1,\ g_3(D) = \frac{(c - r(B))(1 - \delta r(B)) + (1-\delta)(1-c)}{(1 - r(B))(1 - \delta r(B))} \text{ and} \\ &g_3(E) = \frac{(c - r(C))(1 - \delta r(C)) + (1-\delta)(1-c)}{(1 - r(C))(1 - \delta r(C))}. \end{aligned}$$

Let g_1 be the linear map defined on $\Delta(\Omega)$ by $g_1(B) = g_1(C) = 0$ and $g_1(D) = g_3(D)$. Similarly let g_2 be the linear map such that $g_2(C) = 0$, $g_2(D) = g_3(D)$ and $g_2(E) = g_3(E)$. Using the above characterization, we have shown that the greedy payoff function σ_* is g_1 on J_1, g_2 on J_2 and g_3 on J_3.

Let us now verify that g is concave. It is the case if for all $i \in \{1, 2, 3\}$ and all $p \in J_i$

$$g(p) = \min\{g_1(p), g_2(p), g_3(p)\}. \tag{6.6.1}$$

It is enough to check Eq. (6.6.1) above for p equal to A, B, C, D and E. It is clearly satisfied for A, B, C and D. For all p, one has $g_1(p) = p_A \frac{(1-r(B))}{(c-r(B))} g_1(D)$, hence

$$\begin{aligned} g_1(E) &= \frac{(c - r(B))(1 - \delta r(B)) + (1-\delta)(1-c)}{(1 - r(B))(1 - \delta r(B))} \frac{(c - r(C))}{(1 - r(C))} \frac{(1 - r(B))}{(c - r(B))} \\ &= \frac{c - r(C)}{1 - r(C)} + \frac{(1-\delta)(1-c)(c - r(C))}{(1 - \delta r(B))(1 - r(C))(c - r(B))}. \end{aligned}$$

This expression is increasing in $r(B)$, so we get: $g_1(E) \geq \frac{c-r(C)}{1-r(C)} + \frac{(1-\delta)(1-c)(c-r(C))}{(1-\delta r(C))(1-r(C))(c-r(C))} = g_3(E)$. Eq. (6.6.1) is thus satisfied, therefore g is concave and the greedy strategy is optimal.

Example 6.5. In this second example we let $0 = r(C) < c < r(B) \leq r(A)$. Hence A and B belong to $I = \{p \in \Delta(\Omega),\ p_A r(A) + p_B r(B) \geq c\}$. Let $D_1 = (0, c/r(B), 1 - c/r(B))$ and $E_1 = (c/r(A), 0, 1 - c/r(A))$. Then I is the convex hull of $\{A, B, D_1, E_1\}$ and J is the convex hull of $\{C, D_1, E_1\}$.

Recall that $\phi = \mathrm{id}$. For each p, $\psi^+(p)$ lies in the segment $[A, B]$, so $v_\delta(\psi^+(p)) = 1$. We write $s(p) = 1 - r(p) = p_A(1 - r(A)) + p_B(1 - r(B)) + p_C$ for the probability of failure if there is investment at p. Hence for all $p \in \Delta(\Omega)$ one has

$$\psi^-(p) = \frac{1}{s(p)} \left(p_A(1 - r(A)), p_B(1 - r(B)), p_C \right).$$

The greedy payoff function is characterized by:

1. $g(A) = g(B) = 1$ and $g(C) = 0$;
2. if $q = a_I q_I + (1 - a_I)C \in J$, with $q_I \in [D_1, E_1]$ and $a_I \in [0, 1]$, then $g(p) = a_I g(q_I)$;
3. if $p \in I$, then

$$g(p) = 1 - \delta + \delta r(p) + \delta s(p) g\left(\frac{1}{s(p)} (p_A(1 - r(A)), p_B(1 - r(B)), p_C) \right).$$

Let us now define for each positive integer n, $D_{n+1} = (\psi^-)^{-1}(D_n)$, $E_{n+1} = (\psi^-)^{-1}(E_n)$ and Z_n as the convex hull of $\{D_n, E_n, D_{n+1}, E_{n+1}\}$. One has $\lim_{n\to\infty} E_n = A$ and $\lim_{n\to\infty} D_n = B$. The situation is represented Fig. 6.3.

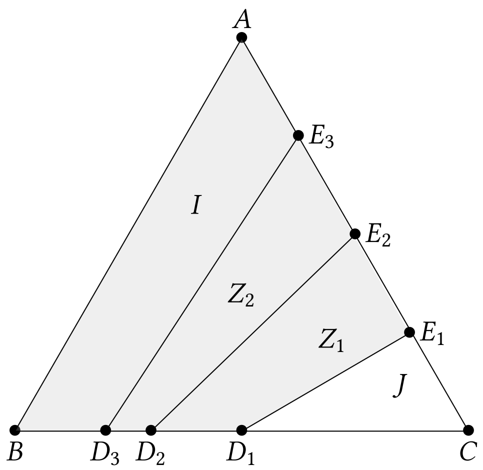

Figure 6.3: The situation of Example 6.5

One has

$$Z_1 = \{p \in \Delta(I),\ \psi^-(p) \in J\}.$$

And for every positive integer n one has $\psi^-([E_{n+1}, D_{n+1}]) = (E_n, D_n)$, so that

$$Z_{n+1} = \{p \in \Delta(I),\ \psi^-(p) \in Z_n\}.$$

If we restrict attention to states A and C, we are exactly in the situation of Example 6.2: the greedy strategy is optimal, and the value is linear on $[E_n, E_{n+1}]$ for each $n \geq 0$ (writing $C = E_0$). The situation is similar if we restrict attention to states B and C.

Let $g_0 : \mathbb{R}^3 \to \mathbb{R}$ be the linear map such that: $g_0(C) = 0$, $g_0(D_1) = v_\delta(D_1)$ and $g_0(E_1) = v_\delta(E_1)$. For each $n \geq 1$, let $g_n : \mathbb{R}^3 \to \mathbb{R}$ satisfying for all $p \in \mathbb{R}^3$:

$$g_n(p) = 1 - \delta + \delta r(p) + \delta s(p) g_{n-1}\left(\frac{1}{s(p)}(p_A(1-r(A)), p_B(1-r(B)), p_C)\right).$$

Notice that g_n is also linear. We finally define $g : \Delta(\Omega) \to [0,1]$ by:

$$g(p) = \begin{cases} g_0(p) & \text{if } p \in J \\ g_n(p) & \text{if } p \in Z_n,\ n \geq 1 \\ 1 & \text{if } p \in [A,B]. \end{cases}$$

The function g satisfies the characterization of the greedy payoff function, therefor the two are equal.

It remains to check that g is concave. This is the case if for all $p \in \Delta(\Omega)$,

$$g(p) = \min\{g_n(p),\ n \geq 0\}.$$

Since g is linear on J, on $[A,B]$, and on each Z_n for $n \geq 1$, it is enough to restrict attention to the cases where p belongs to $\{A,B,C\} \cup \{D_n, n \geq 1\} \cup \{E_n, n \geq 1\}$. For each such p, we know that $g(p) = \min\{g_n(p), n \geq 0\}$ by the analysis of the 2-state case (see Example 6.2) where we know that the greedy payoff function is concave. Thefore, g is concave on $\Delta(\Omega)$ and the greedy strategy is optimal.

A natural question is whether the greedy strategy is always optimal. It turns out that when there are more than two states of nature, the greedy strategy may fail to be optimal. This is the object of Theorem 6.7 below.

Theorem 6.7. *Suppose* $\mathrm{card}(\Omega) = 3$. *Then the greedy strategy is not necessarily optimal.*

Theorem 6.7 is proved via a counterexample of Renault et al. (2017). We reproduce it here for completeness, see Example 6.6 below.

Example 6.6. The situation is represented on Fig. 6.4 below. The state space Ω is $\{\omega_1, \omega_2, \omega_3\}$. Let $\varepsilon > 0$, the investment region is the triangle $(\omega_1, \frac{1}{2}\omega_1 + \frac{1}{2}\omega_3, \varepsilon\omega_1 + (1-\varepsilon)\omega_2)$. Suppose that the invariant measure is $m = \omega_2$ and that $\lambda = \frac{1}{2}$. Finally, let the initial belief p_1 be $2\varepsilon\omega_1 + (1-2\varepsilon)\omega_3$.

At the first stage, playing the greedy strategy σ_*, the belief p_1 is split between ω_3 with probability $1 - 4\varepsilon$ and $\frac{1}{2}\omega_1 + \frac{1}{2}\omega_2$ with probability 4ε. Since the segment $[\omega_2, \omega_3]$ is included in J and the invariant measure m is ω_2, once ω_3 is reached the payoff is 0 for ever. Hence $g(p_1) = 4\varepsilon g\left(\frac{1}{2}\omega_1 + \frac{1}{2}\omega_3\right) \leq 4\varepsilon$.

Consider now the following strategy, which does not disclose any information in the first stage. One has

$$p_2 = \phi(p_1) = \varepsilon\omega_1 + \frac{1}{2}\omega_2 + \left(\frac{1}{2} - \varepsilon\right)\omega_3.$$

Then at the second stage, the strategy is to split p_2 between $q_2 = \varepsilon\omega_1 + (1-\varepsilon)\omega_2$ with probability $\frac{1}{2(1-\varepsilon)}$ and $q_3 = \varepsilon\omega_1 + (1-\varepsilon)\omega_3$ with probability $1 - \frac{1}{2(1-\varepsilon)}$. Hence the expected payoff in stage 2 is $\frac{1}{2(1-\varepsilon)}$. Thus as soon as $4\varepsilon < \frac{\delta(1-\delta)}{2(1-\varepsilon)}$, this strategy is better than the greedy strategy. In particular this is true for ε small enough.

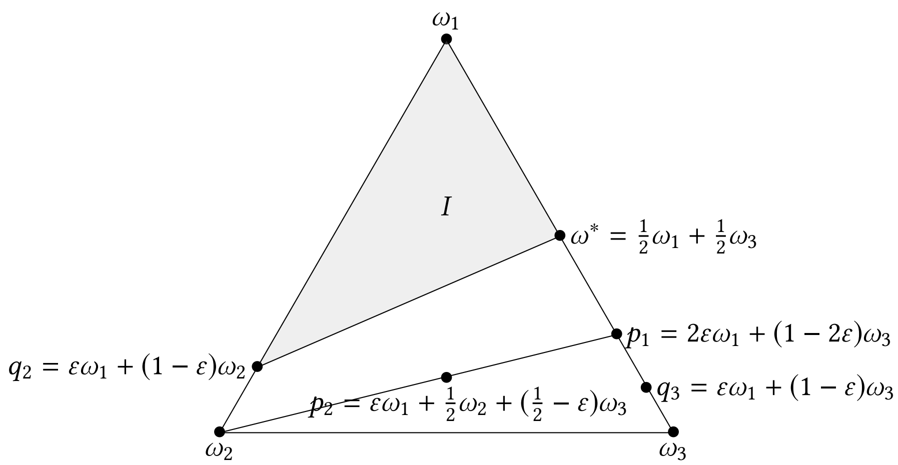

Figure 6.4: A counterexample with card $\Omega = 3$

Notice that this counterexample deeply relies on the action of the Markov chain.

Bibliography

Alpern, S. (2008). Hide-and-seek games on a tree to which Eulerian networks are attached. *Networks*, 52(3):162–166.

Alpern, S. (2010). Search games on trees with asymmetric travel times. *SIAM J. Control Optim.*, 48(8):5547–5563.

Alpern, S. (2011). Find-and-fetch search on a tree. *Oper. Res.*, 59(5):1258–1268.

Alpern, S. (2017). Hide-and-seek games on a network, using combinatorial search paths. *Oper. Res.*, 65(5):1207–1214.

Alpern, S., Baston, V., and Gal, S. (2008). Network search games with immobile hider, without a designated searcher starting point. *Internat. J. Game Theory*, 37(2):281–302.

Alpern, S., Baston, V., and Gal, S. (2009). Searching symmetric networks with utilitarian-postman paths. *Networks*, 53(4):392–402.

Alpern, S., Fokkink, R., Gasieniec, L., Lindelauf, R., and Subrahmanian, V., editors (2013). *Search Theory, A Game Theoretic Perspective*. Springer-Verlag New York.

Alpern, S. and Gal, S. (2003). *The Theory of Search Games and Rendezvous*. Kluwer Academic Publishers, Boston, MA.

Alpern, S. and Lidbetter, T. (2013). Mining coal or finding terrorists: the expanding search paradigm. *Oper. Res.*, 61(2):265–279.

Alpern, S. and Lidbetter, T. (2014). Searching a variable speed network. *Math. Oper. Res.*, 39(3):697–711.

Alpern, S. and Lidbetter, T. (2015). Optimal trade-off between speed and acuity when searching for a small object. *Oper. Res.*, 63(1):122–133.

Alpern, S. and Lidbetter, T. (2019). Approximate solutions for expanding search games on general networks. *Ann. Oper. Res.*, 275(2):259–279.

Alpern, S., Lidbetter, T., Morton, A., and Papadaki, K. (2016a). Patrolling a border. *Operations Research*, 64(6):1256–1269.

Alpern, S., Lidbetter, T., Morton, A., and Papadaki, K. (2016b). Patrolling a pipeline. In *Decision and Game Theory for Security. GameSec 2016.*, volume 9996. Springer.

Alpern, S., Lidbetter, T., and Papadaki, K. (2019). Optimizing periodic patrols against short attacks on the line and other networks. *European Journal of Operational Research*, 273(3):1065–1073.

Alpern, S., Morton, A., and Papadaki, K. (2011). Patrolling games. *Operations Research*, 59(5):1246–1257.

Altman, E., Avrachenkov, K., Marquez, R., and Miller, G. (2005). Zero-sum constrained stochastic games with independent state processes. *Mathematical Methods of Operations Research*, 6:375–386.

Anderson, E. J. and Aramendia, M. A. (1990). The search game on a network with immobile hider. *Networks*, 20(7):817–844.

Aumann, R. J. (1964). Mixed and behavior strategies in infinite extensive games. In *Advances in Game Theory*, pages 627–650. Princeton Univ. Press, Princeton, N.J.

Aumann, R. J. and Maschler, M. (1995). *Repeated games with incomplete information.* MIT press.

Basilico, N., De Nittis, G., and Gatti, N. (2015). Adversarial patrolling with spatially uncertain alarm signals. arXiv:1506.02850.

Basilico, N., Gatti, N., and Amigoni, F. (2012). Patrolling security games: Definition and algorithms for solving large instances with single patroller and single intruder. *Artificial Intelligence*, 184:78–123.

Baston, V. and Kikuta, K. (2004). An ambush game with an unknown number of infiltrators. *Operations Research*, 52(4):597–605.

Baston, V. and Kikuta, K. (2009). An ambush game with a fat infiltrator. *Operations Research*, 57(2):514–519. Technical note.

Baston, V. J., Bostock, F. A., and Ruckle, W. H. (1990). The gold-mine game. *J. Optim. Theory Appl.*, 64(3):641–650.

Bather, J. (1973). Optimal decision procedures for finite markov chains. Part III: General convex systems. *Advances in Applied Probability*, 5:541–553.

Beck, A. and Newman, D. J. (1970). Yet more on the linear search problem. *Israel J. Math.*, 8:419–429.

Berry, D. A. and Mensch, R. F. (1986). Discrete search with directional information. *Oper. Res.*, 34(3):470–477.

Bewley, T. and Kohlberg, E. (1976). The asymptotic theory of stochastic games. *Mathematics of Operations Research*, 1:197–208.

Bewley, T. and Kohleberg, E. (1978). On stochastic games with stationary optimal strategies. *Mathematics of Operations Research*, 3(2):104–125.

Bishop, S. and Evans, T. O. (2013). Static search games played over graphs and general metric spaces. *European Journal of Operational Research*, 231(3):667–689.

Boczkowski, L., Korman, A., and Rodeh, Y. (2018). Searching a tree with permanently noisy advice. In *26th European Symposium on Algorithms*, volume 112 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 54, 13. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern.

Bollobás, B. (2001). *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, second edition.

Bollobás, B. and Riordan, O. (2006). *Percolation.* Cambridge University Press, New York.

Bollobás, B., Kun, G., and Leader, I. (2013). Cops and robbers in a random graph. *Journal of Combinatorial Theory, Series B*, 103:226–236.

Bolte, J., Gaubert, S., and Vigeral, G. (2014). Definable zero-sum stochastic games. *Mathematics of Operations Research.*

Bostock, F. A. (1984). On a discrete search problem on three arcs. *SIAM J. Algebraic Discrete Methods*, 5(1):94–100.

Cao, B. (1995). Search-hide games on trees. *European J. Oper. Res.*, 80(1):175–183.

Cardaliaguet, P., Laraki, R., and Sorin, S. (2012). A continuous time approach for the asymptotic value in two-person zero-sum repeated games. *SIAM Journal on Control and Optimization*, 50:1573–1596.

Dagan, A. and Gal, S. (2008). Network search games, with arbitrary searcher starting point. *Networks*, 52(3):156–161.

Danskin, J. (1990). On the cookie-cutter game: Search and evasion on a disc. *Mathematics of Operations Research*, 15(4):573–596.

Day, A. N. and Falgas-Ravry, V. (2018). Maker-breaker percolation games i: crossing grids.

Efron, B. (1964). Optimum evasion versus systematic search. *J. Soc. Indust. Appl. Math.*, 12:450–457.

Ely, J. C. (2017). Beeps. *American Economic Review*, 107(1):31–53.

Erdős, P. and Rényi, A. (1959). On random graphs. I. *Publ. Math. Debrecen*, 6:290–297.

Erdős, P. and Rényi, A. (1960). On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61.

Erdős, P. and Rényi, A. (1961). On the evolution of random graphs. *Bull. Inst. Internat. Statist.*, 38:343–347.

Flesch, J., Predtetchinski, A., and Sudderth, W. (2019). Positive zero-sum stochastic games with countable state and action spaces. *Appl. Math. Optim.*, forthcoming.

Flesch, J., Schoenmakers, G., and Vrieze, K. (2008). Stochastic games on a product state space. *Mathematics of Operations Research*, 33:403–420.

Flesch, J., Schoenmakers, G., and Vrieze, K. (2009). Stochastic games on a product state space: The periodic case. *International Journal of Game Theory*, 38:263–289.

Forges, F. (1982). Infinitely repeated games of incomplete information: Symmetric case with random signals. *International Journal of Game Theory*, 11(3-4):203–213.

Fournier, G. (2018). General distribution of consumers in pure hotelling games. *International Journal of Game Theory*.

Fudenberg, D. and Yamamoto, Y. (2011). The folk theorem for irreducible stochastic games with imperfect public monitoring. *Journal of Economic Theory*, 146:1664–1683.

Gal, S. (1972). A general search game. *Israel J. Math.*, 12:32–45.

Gal, S. (1974). A discrete search game. *SIAM J. Appl. Math.*, 27:641–648.

Gal, S. (1979). Search games with mobile and immobile hider. *SIAM J. Control Optim.*, 17(1):99–122.

Gal, S. (1980). *Search games*, volume 149 of *Mathematics in Science and Engineering*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London.

Gal, S. (2000). On the optimality of a simple strategy for searching graphs. *Internat. J. Game Theory*, 29(4):533–542 (2001).

Gal, S. and Chazan, D. (1976). On the optimality of the exponential functions for some minimax problems. *SIAM J. Appl. Math.*, 30(2):324–348.

Gale, D. and Glassey, G. (1974). Elementary problem no. 2468. *American Mathematical Monthly*, 82(5):521–522.

Garrec, T. (2019a). Communicating zero-sum product stochastic games. *Journal of Mathematical Analysis and Applications*, 477:60–84.

Garrec, T. (2019b). Continuous patrolling and hiding games. *European Journal of Operational Research*, 277(1):42–51.

Garrec, T. and Renault, J. (2019). Dynamic control of information with observed return on investment. In preparation.

Garrec, T. and Scarsini, M. (2019). When Sally found Harry: A stochastic search game. arXiv:1904.12852.

Gensbittel, F. (2014). Extensions of the cav (u) theorem for repeated games with incomplete information on one side. *Mathematics of Operations Research*, 40(1):80–104.

Gensbittel, F., Oliu-Barton, M., and Venel, X. (2014). Existence of the uniform value in repeated games with a more informed controller. *Journal of Dynamics and Games*, 1(3):411–445.

Gensbittel, F. and Renault, J. (2015). The value of markov chain games with incomplete information on both sides. *Mathematics of Operations Research*, 40(4):820–841.

Gillette, D. (1957). *Contributions to the Theory of Games*, volume 3, chapter Stochastic Games with Zero Stop Probabilities, pages 179–187. Princeton University Press.

Gimbert, H., Renault, J., Sorin, S., Venel, X., Zielonka, W., et al. (2016). On values of repeated games with signals. *The Annals of Applied Probability*, 26(1):402–424.

Gittins, J. C. and Roberts, D. M. (1979). The search for an intelligent evader concealed in one of an arbitrary number of regions. *Naval Res. Logist. Quart.*, 26(4):651–666.

Glazebrook, K., Clarkson, J., and Lin, K. (2019). Fast or slow: search in discrete locations with two search modes. *Oper. Res.*, forthcoming.

Grimmett, G. (1999). *Percolation*, volume 321 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition.

Hoffman, A. J. and Karp, R. M. (1966). On nonterminating stochastic games. *Management Science*, 12(5):359–370.

Hohzaki, R. (2016). Search games: literature and survey. *J. Oper. Res. Soc. Japan*, 59(1):1–34.

Holroyd, A. E., Marcovici, I., and Martin, J. B. (2019). Percolation games, probabilistic cellular automata, and the hard-core model. *Probab. Theory Related Fields*, forthcoming.

Ibragimov, G. and Satimov, N. (2012). One class of simultaneous pursuit games. *Russian Mathematics (Iz. VUZ)*, 56(5):39–46.

Isaacs, R. (1965). *Differential games. A mathematical theory with applications to warfare and pursuit, control and optimization.* John Wiley & Sons, Inc., New York-London-Sydney.

Jotshi, A. and Batta, R. (2008). Search for an immobile entity on a network. *European J. Oper. Res.*, 191(2):347–359.

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101(6):2590–2615.

Karlin, S. (1959). *Mathematical Methods and Theory in Games, Programming and Economics*, volume II: The Theory of Infinite Games. Pergamon Press, Addison-Wesley Publishung Company, Inc.

Kikuta, K. (1990). A hide and seek game with traveling cost. *J. Oper. Res. Soc. Japan*, 33(2):168–187.

Kikuta, K. (1991). A search game with traveling cost. *J. Oper. Res. Soc. Japan*, 34(4):365–382.

Kikuta, K. (2004). A search game on a cyclic graph. *Naval Res. Logist.*, 51(7):977–993.

Koopman, B. (1956a). The theory of search I. Kinetic bases. *Operations Research*, 4:324–346.

Koopman, B. (1956b). The theory of search II. Target detection. *Operations Research*, 4:503–531.

Koopman, B. (1957). The theory of search III. The optimum distribution of searching effort. *Operations Research*, 5:613–626.

Laraki, R. (2001). Variational inequialites, system of functional equations, and incomplete information repeated games. *SIAM Journal on Control and Optimization*, 40:516–524.

Laraki, R. (2002). Repeated games with lack of information on one side: the dual differential approach. *Mathematics of Operations Research*, 27(2):419–440.

Laraki, R. and Renault, J. (2017). Acyclic gambling games. arXiv:1702.06866.

Li, S. (2011). Concise formulas for the area and volume of a hyperspherical cap. *Asian Journal of Mathematics and Statistics*, 4(1):66–70.

Li, X. and Sorin, S. (2016). Uniform value for recursive games with compact action sets. *Operations Research Letters*, 44:575–577.

Lidbetter, T. (2013). Search games with multiple hidden objects. *SIAM Journal on Control and Optimization*, 51(4):3056–3074.

Lin, K., Atkinson, M., and Glazebrook, K. (2013). A graph patrol problem with random attack times. *Operations Research*, 61(3):694–710.

Lin, K., Atkinson, M., and Glazebrook, K. (2014). Optimal patrol to uncover threats in time when detection is imperfect. *Naval Research Logistic*, 61(8):557–576.

Mertens, J.-F. and Neyman, A. (1981). Stochastic games. *International Journal of Game Theory*, 10:53–66.

Mertens, J.-F., Neyman, A., and Rosenberg, D. (2009). Absorbing games with compact action spaces. *Mathematics of Operations Research*, 34(2):257–262.

Mertens, J.-F., Sorin, S., and Zamir, S. (2015). *Repeated Games.* Cambridge University Press.

Mertens, J.-F. and Zamir, S. (1971). The value of two-person zero-sum repeated games with lack of information on both sides. *International Journal of Game Theory*, pages 39–64.

Morse, P. and Kimball, G. (1951). *Methods of Operations Research.* MIT Press, Cambridge.

Munkres, J. (2000). *Topology.* Pearson, 2nd edition.

Neuts, M. F. (1963). A multistage search game. *J. Soc. Indust. Appl. Math.*, 11:502–507.

Neyman, A. (2008). Existence of optimal strategies in markov games with incomplete information. *International Journal of Game Theory*, 37(4):581–596.

Pavlovic, L. (1993). Search game on an odd number of arcs with immobile hider. *Yugoslav Journal of Operations Research*, 3(1):11–19.

Pavlović, L. (1995). A search game on the union of graphs with immobile hider. *Naval Res. Logist.*, 42(8):1177–1189.

Pita, J., Jain, M., Marecki, J., Ordóñez, F., Portway, C., Tambe, M., Western, C., Paruchuri, P., and Kraus, S. (2008). Deployed ARMOR protection: The application of a game theoretic model for security at the Los Angeles international airport. *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems*, pages 125–132.

Reijnierse, J. H. and Potters, J. A. M. (1993). Search games with immobile hider. *Internat. J. Game Theory*, 21(4):385–394.

Renault, J. (2006). The value of markov chain games with lack of information on one side. *Mathematics of Operations Research*, 31(3):490–512.

Renault, J. (2010). Uniform value in dynamic programming. *Journal of the European Mathematical Society*, 13:309–330.

Renault, J. (2012). The value of repeated games with an informed controller. *Mathematics of operations Research*, 37(1):154–179.

Renault, J. (2018). Repeated games with incomplete information. *Encyclopedia of Complexity and Systems Science*.

Renault, J. (2019). A tutorial on Zero-sum Stochastic Games. hal-02130451.

Renault, J., Solan, E., and Vieille, N. (2017). Optimal dynamic information provision. *Games and Economic Behavior*, 104:329–349.

Renault, J. and Venel, X. (2017). A distance for probability spaces, and long-term values in markov decision processes and repeated games. *Mathematics of Operations Research*, 42(2).

Roberts, D. M. and Gittins, J. C. (1978). The search for an intelligent evader; strategies for searcher and evader in the two-region problem. *Naval Res. Logist. Quart.*, 25(1):95–106.

Rosenberg, D., Solan, E., and Vieille, N. (2003). The maxmin value of stochastic games with imperfect monitoring. *International journal of game theory*, 32(1):133–150.

Rosenberg, D., Solan, E., and Vieille, N. (2004). Stochastic games with a single controller and incomplete information. *SIAM journal on control and optimization*, 43(1):86–110.

Rosenberg, D. and Sorin, S. (2001). An operator approach to zero-sum repeated games. *Israel Journal of Mathematics*, 121:221–246.

Ross, K. and Varadarajan, R. (1991). Multichain markov decision processes with a sample path constraint: a decomposition approach. *Mathematics of Operations Research*, 16(1):195–207.

Ruckle, W. (1981). Ambushing random walks II: Continuous models. *Operations Research*, 29(1):108–120.

Ruckle, W. (1983). *Geometric games and their applications*. Pitman Advanced Publishing Program.

Ruckle, W. and Kikuta, K. (2000). Continuous accumulation games in continuous regions. *Journal of Optimization Theory and Applications*, 106(3):581–601.

Sakaguchi, M. (1973). Two-sided search games. *J. Operations Res. Soc. Japan*, 16:207–225.

Shapley, L. (1953). Stochastic games. *Proceeding of the National Academy of Sciences of the United States of America*, 39(10):1095–1100.

Shieh, E., An, B., Yang, R., Tambe, M., Baldwin, C., DiRenzo, J., Maule, B., and Meyer, G. (2012). PROTECT: A deployed game theoretic system to protect the ports of the United States. *Proceedings of The 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

Sion, M. (1958). On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176.

Solan, E. (2003). *Stochastic Games and Applications*, chapter Perturbations of Markov chains with applications to stochastic games, pages 265–280. 570. Springer Netherlands.

Sorin, S. (1984). "big match" with lack of information on one side (part i). *International Journal of Game Theory*, 13(4):201–255.

Sorin, S. (1985). "big match" with lack of information on one side (part ii). *International Journal of Game Theory*, 14(3):173–204.

Sorin, S. (2002). *A First Course on Zero-Sum Repeated Games*. Springer-Verlag, Berlin.

Sorin, S. and Vigeral, G. (2015). Reversibility and oscillations in zero-sum discounted stochastic games. *Journal of Dynamics and Games*, 2(1):103–115.

Subelman, E. J. (1981). A hide-search game. *J. Appl. Probab.*, 18(3):628–640.

van der Hofstad, R. (2017). *Random graphs and complex networks. Vol. 1.* Cambridge Series in Statistical and Probabilistic Mathematics, [43]. Cambridge University Press, Cambridge.

Vigeral, G. (2013). A zero-sum stochastic game with compact action sets and no asymptotic value. *Dynamic Games and Applications*, 3(2):172–186.

von Neumann, J. (1928). Zur theorie der gesellschaftspiele. *Mathematische Annalen*, 100:295–320.

von Neumann, J. (1953). A certain zero-sum two-person game equivalent to the optimal assignment problem. In *Contributions to the Theory of Games, vol. 2*, Annals of Mathematics Studies, no. 28, pages 5–12. Princeton University Press, Princeton, N. J.

von Stengel, B. and Werchner, R. (1997). Complexity of searching an immobile hider in a graph. *Discrete Appl. Math.*, 78(1-3):235–249.

Vrieze, O. (2003). *Stochastic Games and Applications*, chapter Stochastic Games and Stationary Strategies, pages 37–50. 570. Springer Netherlands.

Washburn, A. (2014). *Two-Person Zero-Sum Games*, volume 201. Springer US, 4th edition.

Ziliotto, B. (2016a). A Tauberian theorem for nonexpansive operators and applications to zero-sum stochastic games. *Mathematics of Operations Research*, 41(4):1522 – 1534.

Ziliotto, B. (2016b). Zero-sum repeated games: counterexamples to the existence of the asymptotic value and the conjecture maxmin=lim v(n). *Annals of Probability*, 44(2):1107–1133.

Zoroa, N., Fernández-Sáez, M., and Zoroa, P. (2012). Patrolling a perimeter. *European Journal of Operational Research*, 222(3):571–582.

Zoroa, N., Zoroa, P., and Fernández-Sáez, M. (1999). A generalization of Ruckle's results for an ambush game. *European Journal of Operational Research*, 119(2):353–364.

Résumé

Dans cette thèse, nous étudions divers modèles de jeux dynamiques. Ceux-ci modélisent des processus de décisions prises par des agents rationnels en interactions stratégiques et dont la situation évolue au cours du temps.

Le premier chapitre est consacré aux jeux stochastiques. Dans ces derniers, le jeu courant dépend d'un état de la nature, qui évolue d'une étape à la suivante de manière aléatoire en fonction de l'état courant ainsi que des actions des joueurs, qui observent ces éléments. On étudie des propriétés de communication entre les états, lorsque l'espace d'états Ω est sous la forme d'un produit $X \times Y$, et que les joueurs contrôlent la dynamique sur leur composante de l'espace d'états. On montre l'existence de stratégies optimales dans tout jeu répété un nombre suffisant d'étapes, c'est-à-dire l'existence de la valeur uniforme, sous hypothèse de communication forte d'un côté. On montre en revanche la non converge de la valeur du jeu escompté, qui implique la non existence de la valeur asymptotique, sous hypothèse de communication faible des deux côtés.

Les deux chapitres suivants sont consacrés à des modèles de jeux de recherche-dissimulation. Un chercheur et un dissimulateur agissent sur un espace de recherche. L'objectif du chercheur est typiquement de retrouver le dissimulateur le plus rapidement possible, ou alors de maximiser la probabilité de le trouver en un temps imparti. L'enjeu est alors de calculer la valeur et les stratégies optimales des joueurs en fonction de la géométrie de l'espace de recherche. Dans un jeu de patrouille, un attaquant choisit un temps et un lieu à attaquer, tandis qu'un patrouilleur marche continûment. Lorsque l'attaque survient, le patrouilleur a un certain délai pour repérer l'attaquant. Dans un jeu de recherche-dissimulation stochastique, les joueurs se trouvent sur un graphe. La nouveauté du modèle est qu'en raison de divers évènements, à chaque étape, certaines arêtes peuvent ne pas être disponibles, de sorte que le graphe évolue de façon aléatoire dans le temps.

Enfin, le dernier chapitre est consacré à un modèle de jeux répétés à information incomplète dit de contrôle dynamique de l'information. Un conseiller a une connaissance privée de l'état de la nature, qui évolue aléatoirement avec le temps. Chaque jour le conseiller choisit la quantité d'information qu'il dévoile à un investisseur au travers de messages. À son tour, l'investisseur choisit d'investir ou non afin de maximiser son paiement quotidien espéré. En cas d'investissement, le conseiller reçoit une commission fixe de la part de l'investisseur. Son objectif est alors de maximiser la fréquence escomptée de jours où a lieu l'investissement. On s'intéresse à une stratégie de dévoilement d'information particulière du conseiller dite stratégie gloutonne. C'est une stratégie stationnaire ayant la propriété de minimiser la quantité d'information dévoilée sous contrainte de maximiser le paiement courant du conseiller.

Abstract

In this thesis, we study various models of dynamic games. These model decision-making processes taken by rational agents in strategic interactions and whose situation changes over time.

The first chapter is devoted to stochastic games. In these, the current game depends on a state of nature, which evolves randomly from one stage to the next depending on the current state as well as the actions of the players, who observe these elements. We study communication properties between states, when the state space Ω is in the form of a product $X \times Y$, and players control the dynamics on their components of the state space. The existence of optimal strategies in any long enough repeated game, i.e., the existence of the uniform value, is proved under the assumption of strong communication on one side. We prove the non-convergence of the value of the discounted game, which implies the non-existence of the asymptotic value, under the assumption of weak communication on both sides.

The next two chapters are devoted to models of search games. A searcher and a hider act on a search space. The searcher's objective is typically to find the hider as quickly as possible, or to maximize the probability of finding him in a given time. The challenge is then to calculate the value and optimal strategies of the players according to the geometry of the search space. In a patrolling game, an attacker chooses a time and place to attack, while a patroller walks continuously. When the attack occurs, the patroller has a fixed amount of time to locate the attacker. In a stochastic search game, players act on a graph. The novelty of the model is that due to various events, at each stage, some edges may not be available, so the graph evolves randomly over time.

Finally, the last chapter is devoted to a model of repeated games with incomplete information called dynamic control of information. An advisor has a private knowledge of the state of nature, which changes randomly over time. Every day, the advisor chooses the amount of information he discloses to an investor through messages. In turn, the investor chooses whether or not to invest in order to maximize her daily expected payoff. In the event of an investment, the advisor receives a fixed commission from the investor. His objective is then to maximize the discounted frequency of days on which investment takes place. We are interested in a specific information disclosure strategy of the advisor called the greedy strategy. It is a stationary strategy with the property of minimizing the amount of information disclosed under the constraint of maximizing the advisor's current payoff.