



HAL
open science

Modèles et algorithmes pour la décision séquentielle dans l'incertain

Régis Sabbadin

► **To cite this version:**

Régis Sabbadin. Modèles et algorithmes pour la décision séquentielle dans l'incertain. Mathématiques [math]. Université Toulouse III - Paul Sabatier, 2009. tel-02818771

HAL Id: tel-02818771

<https://hal.inrae.fr/tel-02818771>

Submitted on 6 Jun 2020

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

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

Habilitation à Diriger les Recherches

DE L'UNIVERSITÉ PAUL SABATIER DE TOULOUSE
Spécialité : Intelligence Artificielle

MODÈLES ET ALGORITHMES POUR LA DÉCISION SÉQUENTIELLE DANS L'INCERTAIN

Régis SABBADIN

Soutenue le 5 février 2009

Devant le jury composé de :

| | | |
|---------------------|--|--------------|
| Jérôme LANG | Directeur de recherche, LAMSADE | Directeur |
| Hector GEFNER | Professeur, Universitat Pompeu Fabra | Rapporteur |
| Philippe LERAY | Professeur, Ecole polytechnique de l'Université de Nantes | Rapporteur |
| Patrice PERNY | Professeur, Université Pierre et Marie Curie (UPMC, Paris 6) | Rapporteur |
| Claudette CAYROL | Professeur, Université Paul Sabatier | Examinatrice |
| François CHARPILLET | Directeur de recherche, INRIA Nancy-Grand Est | Examineur |
| Hélène FARGIER | Directeur de recherche, IRIT-UPS-CNRS | Examinatrice |

sabbadin@toulouse.inra.fr
Tél : (33)-(0)5-61-28-54-76
Unité MIA
Centre INRA Toulouse

Abstract

Decision theory is a discipline at the intersection of *economics science*, *operations research* and *artificial intelligence*. The classical decision theoretic framework is grounded on the expected utility criterion, which is widely used for modelling problems of decision under uncertainty. The artificial intelligence (AI) community has adopted this framework and its extension to sequential decision under uncertainty : *Markov decision processes (MDP)*, as a useful tool for modelling and solving problems in *planning under uncertainty*. The AI community has also extended and improved this framework in order to tackle the specific challenges of planning under uncertainty :

- the need for structured representation languages and dedicated solution algorithms,
- the partial observability of the state of the world and
- the partial observability of the planning problems *models*, requiring specialized *learning* algorithms.

In addition, the expected utility criterion has also been challenged as the unique framework for decision under uncertainty and several alternative frameworks have been proposed. These alternative frameworks have been first proposed by economists but researchers in AI have followed, focusing mainly on *qualitative* alternatives.

My own research has followed, at its scale, the same path. It has first consisted in the elaboration of a qualitative decision theory for planning under uncertainty : decision criteria, sequential decision under uncertainty, factored representations... Then, I have also been working within the MDP framework, on factored representations and dedicated algorithms and on problems where observability is no more complete.

This manuscript summarizes my work of the past ten years in this field. More precisely, the three following themes are described :

1. Study of a family of qualitative decision criteria based on possibility theory and Sugeno measures.
2. Proposition of a *possibilistic* counterpart of Markov decision processes and of structured languages (inspired by logic or influence diagrams) and algorithms.
3. Extensions of the MDP framework to factored, *graph-based*, problems and to *purely epistemic* partially observable problems.

The first part of this manuscript is a short state-of-the-art of decision under uncertainty within AI. Of course, it is far from being exhaustive. Its unique goal is to put into perspective my own work in this wide domain. The remaining of the manuscript then provides a description of my work in the three above-mentioned themes.

Résumé

La théorie de la décision est une discipline qui se situe à l'intersection des domaines de l'économie, de la recherche opérationnelle et de l'intelligence artificielle. Dans cette théorie, le cadre "classique", basé sur le critère de l'utilité espérée a pris une place prépondérante pour la modélisation des problèmes de décision dans l'incertain. La communauté de l'intelligence artificielle a adopté ce cadre et son extension à la décision séquentielle, le cadre des processus décisionnels de Markov (PDM), pour la planification dans l'incertain. Elle a également développé et amélioré ce cadre, afin de répondre aux défis particuliers de la planification dans l'incertain. Cela a conduit, entre autres, à i) l'élaboration de langages de représentation structurée et d'algorithmes de planification adaptés à ces langages ii) la prise en compte de l'observabilité partielle de l'état du monde dans ces problèmes et iii) l'apprentissage automatique, basé sur la simulation, de plans ou stratégies optimales.

D'autre part, l'universalité du critère de l'utilité espérée a été remise en cause depuis longtemps et plusieurs alternatives, qualitatives ou quantitatives ont été proposées. Celles-ci ont tout d'abord été proposées dans le domaine des sciences de la décision, mais la communauté de l'intelligence artificielle s'est également lancée dans l'étude de telles alternatives, en particulier qualitatives.

A son échelle, ma recherche poursuit la même démarche double avec, en premier lieu, l'étude d'une théorie de la décision qualitative (critères, décision séquentielle, langages et algorithmes...). Dans le même temps, j'ai aussi contribué au cadre des processus décisionnels de Markov, factorisés ou partiellement observables.

Ce document résume l'ensemble de ces travaux, réalisés ces dix dernières années. Plus précisément, mes travaux sont ici regroupés en trois thèmes :

1. Étude d'une famille de critères de décision qualitative dans l'incertain basés sur la théorie des possibilités et les mesures de Sugeno.
2. Proposition d'une contrepartie possibiliste des processus décisionnels de Markov et adaptation à la décision structurée (logique, diagrammes d'influence possibilistes).
3. Travail de recherche autour des processus décisionnels de Markov "classiques", dans le cadre de la décision structurée et dans le cadre partiellement observable.

La première partie de ce document présente un bref état de l'art de la décision dans l'incertain et de ses liens avec l'intelligence artificielle. Cet état de l'art n'est pas exhaustif, loin de là, mais vise simplement à situer mes travaux à l'intérieur de ce thème extrêmement vaste. Le reste du document décrit mes travaux dans les trois thèmes mentionnés ci-dessus.

Table des matières

| | | |
|-----------|---|-----------|
| I | Introduction, motivation et situation de mes travaux | 1 |
| 1 | Introduction | 3 |
| 1.1 | Contexte : décision dans l'incertain et intelligence artificielle | 3 |
| 1.2 | Positionnement scientifique | 5 |
| 1.3 | Plan du document | 6 |
| 2 | Décision dans l'incertain et IA | 7 |
| 2.1 | Décision dans l'incertain | 7 |
| 2.1.1 | Problèmes de décision dans l'incertain : définitions générales | 7 |
| 2.1.2 | Problèmes de décision dans l'incertain non probabilisé | 8 |
| 2.1.3 | Problèmes de décision dans l'incertain probabilisé | 8 |
| 2.1.4 | Probabilités objectives ou subjectives ? | 10 |
| 2.2 | Décision séquentielle dans l'incertain | 10 |
| 2.2.1 | Politique d'actions | 10 |
| 2.2.2 | Le cadre des processus décisionnels de Markov | 11 |
| 2.3 | Planification dans l'incertain et PDM | 12 |
| 2.3.1 | Processus décisionnels de Markov partiellement observables | 12 |
| 2.3.2 | Processus décisionnels de Markov et apprentissage | 14 |
| 2.3.3 | Processus décisionnels de Markov factorisés | 14 |
| 2.4 | Représentations non-classiques de l'incertain | 16 |
| 2.4.1 | Limites et généralisations du critère de l'utilité espérée | 16 |
| 2.4.2 | Approches qualitatives de la décision dans l'incertain | 18 |
| 2.5 | Résumé, situation de mes travaux | 20 |
| II | Décision dans l'incertain et théorie des possibilités | 21 |
| 3 | Utilité qualitative monotone | 23 |
| 3.1 | Introduction | 23 |
| 3.2 | Utilité qualitative monotone et décision dans l'incertain | 24 |
| 3.2.1 | Utilité qualitative monotone | 24 |
| 3.2.2 | Utilités qualitatives possibilistes | 25 |
| 3.2.3 | Commensurabilité entre incertitudes et préférences | 26 |
| 3.2.4 | Justifications axiomatiques des critères d'utilité qualitative monotone | 26 |
| 3.3 | Raffinements des utilités qualitatives monotones | 26 |

| | | |
|------------|--|-----------|
| 3.3.1 | Utilités qualitatives possibilistes et effet de noyade | 26 |
| 3.3.2 | Raffinements des critères de décision possibilistes | 27 |
| 3.3.3 | Travaux connexes | 28 |
| 4 | Décision séquentielle possibiliste | 29 |
| 4.1 | Introduction | 29 |
| 4.2 | Programmation dynamique possibiliste | 30 |
| 4.2.1 | Horizon fini | 30 |
| 4.2.2 | Horizon infini | 31 |
| 4.2.3 | Décision en environnement partiellement observable | 31 |
| 4.2.4 | Autres travaux sur les PDM possibilistes | 32 |
| 4.3 | Décision possibiliste structurée | 33 |
| 4.3.1 | Définitions générales | 34 |
| 4.3.2 | Une approche basée sur la logique possibiliste | 35 |
| 4.3.3 | Diagrammes d'influence possibilistes | 37 |
| 4.3.4 | Diagramme d'influence possibiliste et PDM possibiliste à horizon fini | 39 |
| 4.3.5 | Complexité de la recherche de politiques optimales | 39 |
| 4.3.6 | Travaux connexes | 40 |
| 4.4 | Conclusion | 41 |
| III | Processus décisionnels de Markov factorisés ou épistémiques | 43 |
| 5 | Extensions du cadre des PDM | 45 |
| 5.1 | Généralités | 45 |
| 5.1.1 | Problèmes de décision séquentielle dans l'incertain factorisés | 45 |
| 5.1.2 | Planification épistémique | 47 |
| 5.1.3 | Plan de la troisième partie | 47 |
| 5.2 | Processus décisionnels de Markov Factorisés | 48 |
| 5.2.1 | Une méthode de décomposition de PDM de grande taille faiblement couplés, basée sur la bipartition spectrale | 48 |
| 5.2.2 | Processus décisionnels de Markov sur graphe | 51 |
| 5.2.3 | Conclusion intermédiaire | 55 |
| 5.3 | Processus décisionnels de Markov épistémiques | 56 |
| 5.3.1 | Problèmes de planification épistémique | 56 |
| 5.3.2 | Processus décisionnels de Markov épistémique | 58 |
| 5.4 | Conclusion | 59 |
| 6 | Applications en gestion agri-environnementale | 61 |
| 6.1 | Introduction | 61 |
| 6.2 | Gestion forestière et biodiversité | 61 |
| 6.2.1 | Influence de la fragmentation forestière sur des dynamiques spatialisées de populations d'oiseaux | 61 |
| 6.2.2 | Gestion spatio-temporelle de réserves naturelles | 63 |
| 6.2.3 | Gestion forestière et risque de tempête | 66 |
| 6.3 | Problèmes de gestion spatialisée en épidémiologie | 69 |

| | | |
|-----------|--|-----------|
| 6.3.1 | Gestion de la durabilité des résistances aux pathogènes : exemple du phoma du colza | 69 |
| 6.3.2 | Epidémiologie humaine : contrôle de maladies infantiles | 71 |
| 6.3.3 | Détection et éradication d'espèces nuisibles : cas des "fire ants" | 73 |
| IV | Conclusion | 75 |
| 7 | Conclusion et perspectives | 77 |
| 7.1 | Résumé de mes travaux et de mes perspectives | 77 |
| 7.1.1 | Décision qualitative possibiliste | 77 |
| 7.1.2 | Processus décisionnels de Markov | 77 |
| 7.1.3 | Résumé de mes perspectives | 78 |
| 7.2 | Perspectives détaillées autour de la décision qualitative possibiliste | 79 |
| 7.2.1 | Axiomatisation des raffinements des utilités qualitatives possibilistes | 79 |
| 7.2.2 | Raffinements de la décision séquentielle possibiliste, apprentissage par renforcement | 79 |
| 7.2.3 | Processus décisionnels de Markov possibilistes structurés | 80 |
| 7.3 | Perspectives détaillées autour des PDM structurés | 80 |
| 7.3.1 | Processus décisionnels de Markov sur graphe | 80 |
| 7.3.2 | Processus décisionnels de Markov partiellement observables sur graphe | 81 |
| A | Travaux et publications | 91 |
| A.1 | Revue à comité de lecture | 91 |
| A.1.1 | Revue internationale | 91 |
| A.1.2 | Revue nationale | 91 |
| A.1.3 | Article en cours de soumission/révision | 92 |
| A.2 | Congrès avec comité de lecture | 92 |
| A.3 | Contributions à des ouvrages collectifs | 94 |
| A.4 | Rapports diplômants | 94 |
| A.5 | Développements informatiques | 95 |
| B | Décision dans l'incertain | 97 |
| B.1 | Axiomatisation du critère de l'utilité espérée | 97 |
| B.2 | Algorithmes classiques de résolution des PDM | 100 |
| B.3 | Mesures d'incertitude monotones | 102 |
| B.4 | Axiomatisation de l'utilité qualitative monotone | 103 |
| B.5 | Axiomatisation des utilités qualitatives possibilistes | 105 |
| B.6 | Raffinements discri-monotones | 107 |
| B.7 | Raffinements des utilités qualitatives possibilistes basés sur l'utilité espérée | 108 |
| B.8 | Raffinement et ordre lexicographique | 110 |
| B.9 | Algorithmes de programmation dynamique possibiliste | 112 |
| B.10 | Décision et logique possibiliste | 115 |

| | | |
|----------|--|------------|
| C | Extensions du cadre des PDM | 119 |
| C.1 | Programmation linéaire approchée | 119 |
| C.2 | Itération de la politique approchée en “champ moyen” | 122 |
| C.3 | Processus décisionnels de Markov épistémiques | 125 |
| C.4 | Complexité du problème de décision associé à la résolution d’un PDME | 127 |
| D | Organigramme du laboratoire | 129 |
| E | Réseau de collaborations | 133 |
| F | Animation scientifique | 137 |
| F.1 | Jurys et comités de thèse, concours de recrutement | 137 |
| F.1.1 | Jury et comité de thèse | 137 |
| F.1.2 | Concours de recrutement | 137 |
| F.2 | Comités de programme de conférences, relecture d’articles | 137 |
| F.2.1 | Membre de comités de programme et d’organisation de conférences | 137 |
| F.2.2 | Relecteur pour conférences | 137 |
| F.2.3 | Reviewer pour revues | 137 |
| F.3 | Expertise | 138 |
| F.4 | Groupes de travail | 138 |
| F.5 | Participation à la vie scientifique de l’INRA | 138 |
| G | Projets de recherche | 139 |
| G.1 | Projet FOFRAMIP | 139 |
| G.2 | Projet CEDRE | 139 |
| G.3 | Projet Australian Research Council (ARC) | 139 |
| G.4 | Projet ComEvol | 139 |
| H | Activités d’encadrement | 141 |
| H.1 | Encadrement de thèses | 141 |
| H.2 | Encadrement de Master 2 (DEA) | 141 |
| H.3 | Encadrements de stages | 141 |
| H.4 | Encadrement de CDD | 142 |
| I | Activités d’enseignement | 143 |
| J | Sélection de publications | 145 |

Première partie

Introduction, motivation et situation de mes travaux

Chapitre 1

Introduction

1.1 Contexte : décision dans l'incertain et intelligence artificielle

La théorie de la décision est une discipline qui se situe à l'intersection des domaines de l'économie, de la recherche opérationnelle et de l'intelligence artificielle. Elle vise à définir des critères permettant de modéliser le comportement d'un agent interagissant avec son environnement. Ces interactions peuvent être modélisées par le schéma suivant :

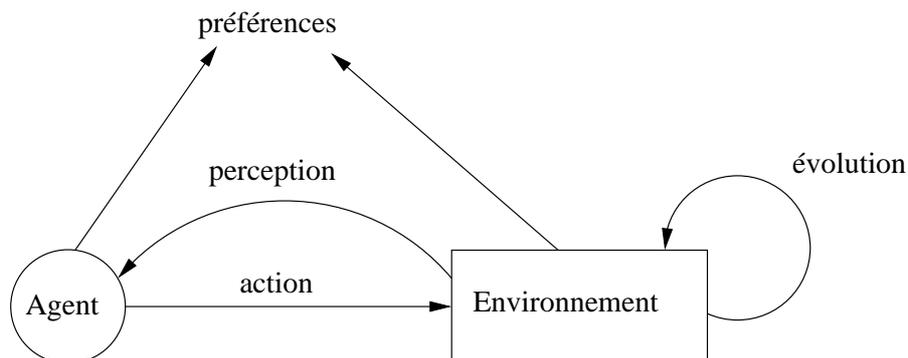


FIG. 1.1 – Interactions entre un agent et son environnement.

L'agent peut influencer sur l'évolution de son environnement par l'intermédiaire d'actions. Dans le même temps il *perçoit*, éventuellement de manière imparfaite, l'état de l'environnement. Un tel agent se trouve face à un problème de choix parmi un ensemble d'actions disponibles. Afin de faciliter ce choix, il est nécessaire de modéliser les *préférences* de l'agent sur les conséquences (perçues) de ses actions. Les préférences de l'agent peuvent être mesurées par une *fonction d'utilité* sur les conséquences ou par une relation de préférence (éventuellement partielle) construite à partir de l'agrégation des degrés de satisfaction de différents critères.

Parfois, les informations dont dispose l'agent sur l'effet de ses actions sont entachées d'incertitude. Il peut n'avoir qu'une vision incomplète ou déformée de l'environnement du problème de décision. Dans ce cas, les effets de ses actions peuvent être variables en fonction de l'état de l'environnement et l'agent se trouve alors face à un problème de *décision dans l'incertain*. Le paradigme classique de la décision dans l'incertain est celui de l'*utilité espérée*, dans lequel les préférences de l'agent sur les conséquences d'une action sont modélisées par une fonction d'utilité et ses connaissances par une

distribution de probabilité sur ces mêmes conséquences. Le critère classique de choix entre les actions est *l'espérance mathématique* de la fonction d'utilité sur les conséquences.

Dans de nombreux problèmes de décision dans l'incertain, un agent est confronté à une succession de problèmes de choix d'action. Dans ces problèmes de *décision séquentielle dans l'incertain*, les actions de l'agent ont plusieurs types d'effets :

- Elles changent tout d'abord l'état courant de l'environnement, ce changement étant gouverné par une distribution de probabilité sur l'ensemble des états successeurs potentiels de l'état courant par l'intermédiaire de l'action choisie.
- Elles peuvent aussi apporter de l'information sur l'état courant, si celui-ci est mal connu.
- Enfin, elles rapportent également une récompense immédiate à l'agent, fonction de la transition (état courant – action – état successeur) observée.

Dans ces problèmes de décision séquentielle, une séquence de transitions est appelée *trajectoire*. Un problème de décision séquentielle dans l'incertain ne consiste plus à rechercher une unique action optimale mais une *politique de choix*, définissant à tout instant l'action à appliquer en fonction de l'état courant et éventuellement de la trajectoire suivie jusqu'ici. Une telle politique de choix sera dite *optimale* si elle maximise l'espérance mathématique de la somme des récompenses obtenues par l'agent, le long de toutes les trajectoires futures possibles. Le cadre des *Processus Décisionnels de Markov* (PDM) (Puterman, 1994) permet de modéliser et de résoudre efficacement ces problèmes de décision séquentielle dans l'incertain, sous certaines hypothèses sur les probabilités de transition. C'est pourquoi il a été privilégié dans de nombreux domaines traitant de la décision séquentielle dans l'incertain (automatique, recherche opérationnelle, intelligence artificielle).

Le cadre classique, *utilité espérée / processus décisionnels de Markov*, est le cadre dominant depuis de nombreuses années pour la représentation et la résolution de problèmes de décision séquentielle dans l'incertain. Toutefois, il a été questionné depuis longtemps en tant que cadre unique pour la représentation de problèmes de décision dans l'incertain. Il a reçu de nombreuses contributions visant à l'amender afin d'élargir son champ d'application. C'est tout d'abord en économie que le critère de l'utilité espérée a été remis en cause et plus particulièrement sa capacité à modéliser le comportement de décideurs réels (paradoxes d'Allais (Allais, 1953) et d'Ellsberg (Ellsberg, 1961)). De nouveaux critères étendant l'utilité espérée ont été proposés, basés sur *l'intégrale de Choquet* par exemple (voir (Gilboa, 2004) pour un ouvrage de synthèse sur ces extensions). La longue tradition du traitement symbolique de l'information en *intelligence artificielle* a également conduit au développement de plusieurs théories qualitatives de l'incertain, puis de la décision dans l'incertain (Doyle and Thoma-son, 1999; Dubois et al., 2000). Ce qui caractérise ces théories qualitatives de la décision, c'est tout d'abord l'abandon de la représentation probabiliste de l'incertain. Les probabilités peuvent être remplacées par des notions *ensemblistes* (Brafman and Tennenholtz, 1997) ou par des représentations *ordinales* de l'incertain, comme les fonctions *kappa* de Spohn (Pearl, 1993; Giang and Shenoy, 2000) ou les *distributions de possibilité* (Dubois and Prade, 1995; Godo and Zapico, 2001). Bien entendu, l'abandon des mesures classiques d'incertitude et de préférence implique d'abandonner le critère de l'utilité espérée au profit de critères qualitatifs ou construits à partir de relations d'incertitude et de préférence *ordinales*.

L'interaction entre intelligence artificielle et théorie de la décision ne repose pas simplement sur la remise en cause du cadre probabiliste de la représentation de l'incertain et de l'utilité espérée. La communauté de la *planification* en intelligence artificielle s'est intéressée depuis les années 1970 à

la génération de *plans* (séquences d'actions) visant à atteindre un objectif fixé pour des problèmes exprimés dans un langage concis d'opérateurs de transformation d'état, souvent proche de la logique propositionnelle (voir (Gallhab et al., 2004), par exemple, pour une synthèse de cette famille d'approches). Dans un premier temps la planification ne considérait que des actions ayant des effets *déterministes*. Plus récemment, le domaine de la *planification dans l'incertain* s'est développé, proposant d'intégrer des actions à effet probabiliste puis des fonctions d'utilité additives sur les buts, conduisant à une famille d'approches pour la planification basées sur la théorie de la décision (Blythe, 1999). Le cadre des processus décisionnels de Markov est alors naturellement devenu l'approche privilégiée pour la représentation et la résolution de problèmes de planification dans l'incertain en intelligence artificielle. De nombreux travaux récents ont visé à améliorer ce cadre en le dotant du pouvoir expressif des langages de représentation traditionnellement utilisés en intelligence artificielle : logique, contraintes ou réseaux Bayésiens. L'utilisation de tels langages de représentation fait "exploser" la complexité des algorithmes de résolution classiques des PDM. La résolution de ces problèmes *structurés* est ainsi devenue un défi pour la communauté de l'intelligence artificielle et de nombreuses approches, centralisées (Boutilier et al., 2000; Jensen, 2001; Givan et al., 2003) ou distribuées (Guestrin et al., 2003) ont été développées ces dernières années.

Enfin, d'autres contributions à l'extension du cadre des PDM ont été proposées en intelligence artificielle ou en recherche opérationnelle :

- L'hypothèse d'observabilité complète de l'état courant a été remise en cause dès les années soixante (Akström, 1965; Sondik, 1971; Smallwood and Sondik, 1973; Sondik, 1978), conduisant au cadre des *PDM Partiellement Observables* (PDMPO) qui généralise le cadre des PDM. Depuis, le surcroît de complexité induit par la prise en compte de l'observabilité partielle dans les PDM a été à la source de nombreuses recherches en intelligence artificielle visant à élaborer des algorithmes efficaces de résolution exacte ou approchée des PDMPO : (Cassandra, 1998; Hauskrecht, 2000; Pineau et al., 2003; Smith and Simmons, 2005), etc.
- Plus tard, l'hypothèse d'observabilité du *modèle* d'un PDM a également été remise en question. On s'est alors intéressé au cas où le modèle de transitions et de récompenses du PDM n'est pas connu a priori et n'est accessible que via la *simulation* de transitions. *L'Apprentissage par Renforcement* (AR) (Sutton, 1988; Watkins, 1989) propose de nombreux algorithmes basés sur l'exploitation des transitions simulées pour calculer des politiques optimales ou approchées pour les PDM. L'apprentissage par renforcement est un domaine de recherche fructueux, au croisement des communautés *Machine Learning* et PDM (voir (Sutton and Barto, 2002) pour une revue des méthodes d'apprentissage par renforcement pour les PDM).
- Enfin, la prise en compte dans le cadre des PDM de décideurs multiples interagissant a également reçu beaucoup d'attention ces dernières années. Dans le cadre des *processus décisionnels de Markov multi-agents*, on recherche la *politique jointe* de plusieurs agents optimisant l'espérance d'un critère unique, alors que ces agents ont une visibilité totale (Boutilier, 1996) ou partielle (Bernstein et al., 2002) de l'état courant. Dans le cadre des *jeux stochastiques* (Filar and Vrieze, 1997; Hansen et al., 2004) les agents ont en outre des fonctions de récompense différentes et on recherche un point d'équilibre entre les politiques des différents agents plutôt qu'une politique jointe.

1.2 Positionnement scientifique

Ces dix dernières années, j'ai abordé des questions de recherche différentes, à l'intersection des domaines de la décision dans l'incertain et de l'intelligence artificielle.

J’ai tout d’abord fini d’élaborer les justifications axiomatiques, ébauchées durant ma thèse, de la *théorie des possibilités* et des *mesures qualitatives monotones* en tant que modèles de l’incertain pour la décision. Par la suite, j’ai continué à travailler sur ce modèle possibiliste de la décision dans l’incertain :

- en explorant ses liens avec la théorie de l’utilité espérée,
- en proposant une contrepartie des principaux algorithmes de programmation dynamique pour la décision séquentielle possibiliste et,
- en proposant des algorithmes et résultats de complexité pour la décision possibiliste structurée.

Depuis 2001, j’ai également commencé à étudier des problèmes de décision séquentielle dans l’incertain *spatialisés* rencontrés par exemple en étudiant le contrôle de processus agri-environnementaux : gestion forestière, épidémiologie, gestion de ressources naturelles, etc. Ce sont ces questions de recherche d’ordre *appliquées* qui m’ont conduit à travailler sur un cadre de PDM structuré dédié à ces problèmes (le cadre des *Processus Décisionnels de Markov sur Graphe*) et à élaborer plusieurs algorithmes de résolution approchée basés sur la programmation linéaire ou l’approximation en *champ moyen* de distributions de probabilité complexes [21], [28].

Enfin, mon intérêt s’est porté plus récemment (depuis 2004) sur une forme particulière de PDM *partiellement observables*, dédiée à la *planification épistémique*. Dans ce domaine, l’objectif n’est pas de modifier l’état courant du monde pour atteindre un certain but mais uniquement *d’identifier* l’état réel du monde, initialement mal connu, en appliquant des actions d’acquisition d’information. Le domaine de la planification épistémique recouvre un grand nombre d’applications : la gestion agri-environnementale¹ qui m’a intéressé tout particulièrement, le diagnostic de pannes, les jeux (Mastermind, bataille navale) ou l’élicitation de préférences.

1.3 Plan du document

Dans le chapitre suivant je vais commencer par décrire plus précisément la motivation et la situation de mes travaux. Je décrirai en particulier l’approche classique de la décision dans l’incertain et son extension à la décision séquentielle. Ensuite, je mentionnerai certains apports de l’intelligence artificielle à la décision dans l’incertain, à travers des représentations “non-classiques” de l’incertitude et des préférences, et à la planification dans l’incertain.

Dans la deuxième partie de ce rapport, j’évoquerai plus particulièrement ma contribution dans le domaine de la décision qualitative possibiliste. Tout d’abord, ma contribution à l’axiomatisation de la théorie de la décision dans l’incertain basée sur les utilités qualitatives monotones et possibilistes, ainsi que ses liens avec la théorie de l’utilité espérée (chapitre 3). Puis, l’extension de la théorie de l’utilité qualitative possibiliste à la décision séquentielle et la proposition de langages structurés pour la planification dans l’incertain (4).

Enfin, dans la troisième partie j’évoquerai mes deux contributions méthodologiques dans le domaine des processus décisionnels de Markov adaptés i) au contrôle de processus “spatialisés” et ii) au diagnostic sous observations incertaines et imprécises (chapitre 5). Dans cette même partie je présenterai également un certain nombre d’applications de ces avancées dans le domaine de la gestion agri-environnementale.

¹A travers le problème de cartographie de l’étendue d’une invasion par une espèce nuisible, par exemple.

Chapitre 2

Décision dans l'incertain et intelligence artificielle

2.1 Décision dans l'incertain

2.1.1 Problèmes de décision dans l'incertain : définitions générales

L'objet d'une théorie de la décision dans l'incertain est de proposer un ou des critères permettant de modéliser le comportement d'un agent face à un problème de choix parmi un ensemble d'actions disponibles. Dans le cas de la décision *sans incertitude*, une action a peut être vue comme une application d'un ensemble d'états du monde possibles $\mathcal{S} = \{s_1, s_2, \dots\}$ vers un ensemble de conséquences $\mathcal{X} = \{x_1, x_2, \dots\}$. L'ensemble \mathcal{S} représente l'ensemble des états du monde dans lesquels l'action peut être appliquée et pour tout état $s \in \mathcal{S}$, $a(s) \in \mathcal{X}$ est le résultat de l'action a appliquée en s .

On parle de *décision dans l'incertain* lorsque les conséquences d'une action, ou l'état du monde, ne sont pas complètement connus avant qu'on l'exécute. Ces deux cas sont résumés dans la figure (2.1.1), où σ et $\sigma_{s,a}$ représentent respectivement des mesures d'incertitude sur \mathcal{S} ou \mathcal{X} : mesures de probabilités, de possibilité (Dubois and Prade, 1988), ou simplement des sous-ensembles d'états ou de conséquences.

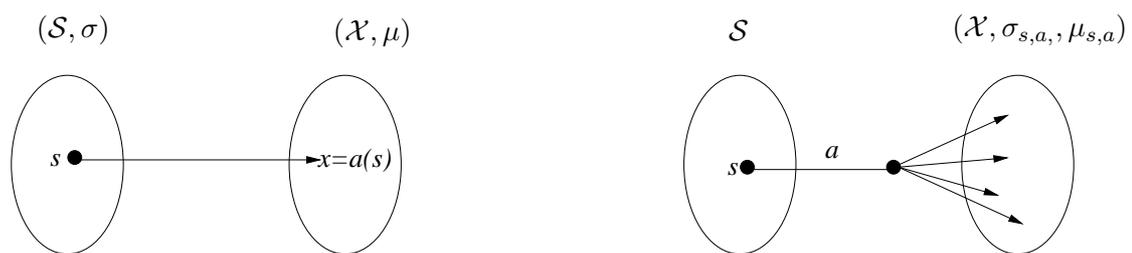


FIG. 2.1 – Décision dans l'incertain. A gauche : incertitude sur l'état initial. A droite : incertitude sur l'effet d'une action.

Dans le premier cas (à gauche), une action est encore une application de \mathcal{S} vers \mathcal{X} , mais l'état initial est incertain, la connaissance sur cet état étant représentée par la mesure σ sur \mathcal{S} . Dans le second cas (à droite), l'état initial s est supposé connu, mais la conséquence de l'action a appliquée dans l'état s est incertaine, l'incertitude étant représentée par la mesure d'incertitude $\sigma_{s,a}$ sur \mathcal{X} .

Le choix de l'agent entre les actions est supposé guidé par une *relation de préférence* sur les conséquences éventuelles des actions applicables. Dans mes travaux, je me suis placé dans le cas où les préférences de l'agent sur les conséquences sont exprimées par l'intermédiaire d'une fonction d'utilité, numérique ou ordinale, μ (ou $\mu_{s,a}$) sur \mathcal{X} . Les préférences de l'agent sur les actions sont alors modélisées via certains *critères de décision*, reflétant à la fois ses préférences sur les conséquences et ses connaissances sur l'état du monde ou l'effet des actions. La nature des problèmes de décision dans l'incertain diffère alors suivant la nature et la représentation de l'incertain choisies.

2.1.2 Problèmes de décision dans l'incertain non probabilisé

Dans certains cas, l'information disponible sur l'état du monde est très grossière, simplement représentée par un sous-ensemble *d'états possibles* du monde. Dans ce cas, (Luce and Raiffa, 1957) recensent quatre critères de décision dans l'incertain non probabilisé, à même de classer les actions disponibles d'un agent :

- Le **critère *maximin***, ou critère de Wald (Wald, 1950), évalue les actions selon l'utilité de leur pire conséquence possible. Ce critère est absolument pessimiste et ne permet pas de compensation entre les utilités des différentes conséquences possibles d'une action.
- Le **critère *minimax regret*** proposé par Savage (Savage, 1951) calcule, comme le critère maximin, l'utilité de chaque action dans chaque état possible du monde. Néanmoins, à la différence du précédent il associe à chaque action son *regret maximum* qui est le maximum des différences entre l'utilité de l'action et celle de la meilleure possible, pour chaque état du monde. Les actions sont ainsi évaluées relativement les unes aux autres, contrairement au critère *maximin* qui les évalue de manière absolue.
- Le **critère d'Hurwicz** généralise le critère *maximin*, puisqu'il évalue les actions en fonction d'une moyenne pondérée (par un indice de pessimisme) des utilités de leur pire et de leur meilleure conséquence possible.
- Enfin, le **critère de Laplace** évalue les actions en fonction de la moyenne arithmétique des utilités de toutes leurs conséquences possibles. Ici, la notion de compensation entre les utilités de toutes les conséquences possibles d'une action prend tout son sens.

Ces quatre critères requièrent des échelles d'utilité de plus en plus raffinées : une échelle ordonnée suffit pour le critère de Wald, la notion de différence entre utilités est nécessaire pour le critère de Savage, et les critères d'Hurwicz et Laplace nécessitent de calculer des moyennes.

2.1.3 Problèmes de décision dans l'incertain probabilisé

L'approche de Savage

Dans l'approche de la décision dans l'incertain proposée par Savage (Savage, 1954), mais aussi partagée par Anscombe et Aumann (Anscombe and Aumann, 1963), un problème de décision est décrit par un quadruplet $(\mathcal{S}, \mathcal{X}, \mathcal{A}, \succeq)$, où \mathcal{S} est un ensemble d'états et \mathcal{X} un ensemble de conséquences. $\mathcal{A} = \mathcal{X}^{\mathcal{S}}$ est l'ensemble des *actes*¹, où un acte $f \in \mathcal{A}$ est une application de \mathcal{S} vers \mathcal{X} . Enfin, \succeq est une relation de préférence entre les actes, supposée complète, transitive et réflexive. \sim est la relation d'équivalence associée à \succeq et \succ est la relation de préférence stricte associée à \succeq .

¹Dans une vision *à la Savage* où $\mathcal{A} = \mathcal{X}^{\mathcal{S}}$, nous utiliserons le terme *acte* pour nommer les éléments de \mathcal{A} , mais lorsque nous nous placerons dans le cadre de la décision séquentielle (à la von Neumann et Morgenstern), nous utiliserons le terme *action*. Pour distinguer les deux cadres, les actes seront notés $\{f, g, h, \dots\}$ alors que les actions seront notées $\{a, a', a_1, a_2, \dots\}$.

Dans le cadre de l'approche de Savage, une représentation numérique de la relation de préférence \succeq est adoptée. Celle-ci est basée sur le critère de l'*utilité espérée*. L'incertitude du décideur sur l'état du monde est représentée par une distribution de probabilité p sur \mathcal{S} et ses préférences sur les conséquences par une fonction d'utilité u (à valeurs réelles) sur \mathcal{X} . Le quadruplet $(\mathcal{S}, \mathcal{X}, p, u)$ est alors appelé *modèle d'utilité probabiliste* et la relation de préférence $\succeq_{EU,p,u}$ est construite à partir de la comparaison des espérances mathématiques ($EU_{p,u}$) de la fonction d'utilité des deux actes comparés :

$$\begin{aligned} f \succeq_{EU,p,u} g &\Leftrightarrow EU_{p,u}(f) \geq EU_{p,u}(g), \\ \text{où } \forall h \in \mathcal{A}, EU_{p,u}(h) &= \sum_{s \in \mathcal{S}} p(s).u(h(s)). \end{aligned} \quad (2.1)$$

On trouvera en annexe B.1 une description de l'axiomatisation du critère de l'utilité espérée proposée par Savage (Savage, 1954). Nous nous en sommes inspiré pour notre propre axiomatisation des critères d'utilité qualitative monotones [3] (voir chapitre suivant).

L'approche de von Neumann et Morgenstern

La vision de la théorie de la décision en termes de comparaison d'actes est intéressante lorsque l'on s'en sert pour *fonder* une théorie de la décision dans l'incertain ou pour éliciter les préférences ou les connaissances d'un décideur. Elle se révèle toutefois peu pratique lorsqu'il s'agit d'*optimiser* le comportement d'un décideur face à un problème de décision dans l'incertain déjà formulé, surtout si une *séquence* de décisions doit être choisie. Dans ce cas, il peut être plus commode, pour des raisons théoriques et pratiques, d'adopter une vision à la *von Neumann et Morgenstern* (von Neumann and Morgenstern, 1944) dans laquelle l'ensemble \mathcal{A} des actions est construit indépendamment de \mathcal{S} et \mathcal{X} . Une action est maintenant notée $a \in \mathcal{A}$ et son effet sur l'état courant du monde s est décrit par une probabilité conditionnelle sur les conséquences $p(x|s, a), \forall x \in \mathcal{X}$. Dans l'approche de (von Neumann and Morgenstern, 1944), la distribution de probabilité conditionnelle $p(x|s, a)$ est supposée connue et donnée a priori (probabilité *objective*) alors que dans l'approche de Savage, p peut être déduite des préférences exprimées par un décideur entre les actes. On parle, dans le cadre de (von Neumann and Morgenstern, 1944), de *décision dans le risque*, par opposition à la *décision dans l'incertain probabilisé* du cadre de Savage, dans lequel les probabilités sont *subjectives*.

Von Neumann et Morgenstern (von Neumann and Morgenstern, 1944) ont défini la notion de *loterie* associée à une action a , où une loterie $\lambda_a \in \mathcal{L}(\mathcal{X})$ est une distribution de probabilité sur l'ensemble \mathcal{X} . La relation d'incertitude porte alors sur l'effet des actions et non sur l'état du monde. $\mathcal{L}(\mathcal{X})$ est construit par induction : \mathcal{X} appartient à $\mathcal{L}(\mathcal{X})$, et toute *combinaison de loteries*, au sens de la combinaison des distributions de probabilité, est une loterie. $\mathcal{L}(\mathcal{X})$ est équipé d'une relation de préférence $\succeq_{\mathcal{L}(\mathcal{X}),u}$ et (von Neumann and Morgenstern, 1944) ont proposé un ensemble d'axiomes assurant que $\succeq_{\mathcal{L}(\mathcal{X}),u}$ soit représentable par l'espérance de la fonction d'utilité u :

$$\begin{aligned} \forall \lambda, \lambda' \in \mathcal{L}(\mathcal{X}), \lambda \succeq_{\mathcal{L}(\mathcal{X}),u} \lambda' &\Leftrightarrow u(\lambda) \geq u(\lambda'), \\ \text{où } u(\lambda) &=_{def} \sum_{x \in \mathcal{X}} p_\lambda(x).u(x). \end{aligned} \quad (2.2)$$

p_λ est la combinaison des distributions de probabilité sur lesquelles λ est construite.

Le cadre de von Neumann et Morgenstern se réduit donc à un triplet $(\mathcal{X}, \mathcal{L}(\mathcal{X}), \succeq)$, où l'axiomatique assure que \succeq peut être déduite d'une fonction d'utilité sur les conséquences, u , unique à transformation affine près. Ce cadre convient tout naturellement à l'expression de problèmes de décision

séquentielle dans l'incertain, puisqu'il permet d'écrire des loteries (représentant les effets d'actions), dont les conséquences sont elles-mêmes des loteries sur d'autres loteries, etc., et dont l'utilité espérée peut être calculée.

Il existe très clairement un lien entre les approches de Savage et de von Neumann. En fait, tout cadre à la Savage $(\mathcal{S}, \mathcal{X}, \mathcal{A}, \succeq_{EU,p,u})$ peut être transformé en un unique (à transformation affine de u près) cadre à la von Neumann et Morgenstern $(\mathcal{X}, \mathcal{L}(\mathcal{X}), \succeq_{\mathcal{L}(\mathcal{X}),u})$. Il suffit de constater que tout acte $f : \mathcal{S} \rightarrow \mathcal{X}$ peut être représenté par une loterie λ_f , définie par

$$p_{\lambda_f}(x) = \sum_{s \in \mathcal{S}, f(s)=x} p(s), \forall x \in \mathcal{X}.$$

2.1.4 Probabilités objectives ou subjectives ?

La théorie de von Neumann et Morgenstern concerne la décision sous risque. Dans ce cadre, l'incertitude est représentée par une distribution de probabilités *objective* sur l'ensemble des conséquences. Cela signifie que ces probabilités, déterminées par les conditions extérieures du problème de décision sont connues par le "décideur". Le problème se complique lorsque ces probabilités ne sont pas disponibles ou lorsqu'elles ne peuvent pas être déterminées de manière objective. Dans ce cas, l'incertitude ne peut être mesurée par la simple observation de phénomènes physiques extérieurs au décideur (statistiques, symétries du problème...). C'est au contraire l'agent (éventuellement un "expert") qui devra déterminer les degrés de vraisemblance des conséquences possibles des actions. Un des problèmes de la décision dans l'incertain est donc de faire révéler au décideur les probabilités *subjectives* qu'il attache aux états du monde ou aux conséquences des actions, avant d'appliquer éventuellement le critère de l'utilité espérée. Les méthodes proposées classiquement pour révéler ces probabilités, comme celle de (Savage, 1954), procèdent d'une analyse de la relation de préférence de l'agent entre les actions mises à sa disposition. C'est uniquement à partir de cette relation de préférence que sont déterminées les probabilités (subjectives) des états du monde. Ces probabilités ne sont pas des données du problème comme dans l'approche de von Neumann et Morgenstern.

L'approche proposée par Anscombe et Aumann (Anscombe and Aumann, 1963) généralise les théories de von Neumann et Morgenstern et de Savage puisqu'elle considère des *actes généralisés*, qui sont des applications de \mathcal{S} vers l'ensemble des loteries. Cette approche a été proposée afin d'unifier dans un même cadre l'incertain objectif (modélisée par des loteries) et l'incertain subjectif (modélisé à travers la notion d'acte).

2.2 Décision séquentielle dans l'incertain

2.2.1 Politique d'actions

Jusqu'ici, nous nous sommes placés dans des problèmes de décision où une action, unique, devait être choisie puis appliquée. En décision séquentielle, plusieurs actions doivent être appliquées successivement, la conséquence de la première action devenant l'état dans lequel la deuxième est appliquée, etc. C'est pourquoi les notions *d'état* et de *conséquence* ne peuvent plus être distinguées, et un seul ensemble, \mathcal{X} , est utilisé pour représenter l'ensemble des états du monde pouvant être atteints après qu'une séquence d'actions (éventuellement vide) ait été appliquée. Le résultat d'une séquence d'actions (a_0, \dots, a_{T-1}) appliquées n'est pas une conséquence unique mais un ensemble d'états rencon-

trés $(x_0, \dots, x_{T-1}, x_T)$. Une séquence $\tau = (x_0, a_0, \dots, x_{T-1}, a_{T-1}, x_T)$ est appelée *trajectoire* du système. En décision séquentielle, l'équivalent de la notion d'action à la von Neumann et Morgenstern qui définit une distribution de probabilité sur les conséquences, éventuellement à partir d'un état initial connu, est une *politique* $\delta = \{\delta_t\}$ qui associe à toute trajectoire partielle (x_0, \dots, x_t) une action $a_t = \delta_t(x_0, \dots, x_t) \in A$. Si l'on définit $p(\tau|x_0, \delta)$, la probabilité de suivre la trajectoire τ en appliquant la politique δ à partir de x_0 et $u(\tau)$ la récompense obtenue lorsque l'on suit la trajectoire τ , nous pouvons définir formellement l'utilité espérée d'une politique δ :

$$EU_\delta(x_0) = \sum_{\tau} p(\tau|x_0, \delta) \cdot u(\tau). \quad (2.3)$$

2.2.2 Le cadre des processus décisionnels de Markov

Résoudre un problème de décision séquentielle dans l'incertain consiste à trouver une politique δ^* d'utilité espérée maximale, pour un état initial x_0 , fixé ($EU_{\delta^*}(x_0) \geq EU_\delta(x_0), \forall \delta$), ou pour un ensemble d'états initiaux. Les *Processus Décisionnels de Markov* (PDM) forment un cadre de modélisation et proposent des algorithmes efficaces de calcul de politiques optimales.

Dans le cadre des PDM, l'hypothèse suivante est faite sur la probabilité conditionnelle $p(\tau|x_0, \delta)$:

$$p(\tau|x_0, \delta) = \prod_{t=0}^{T-1} p_t(x_{t+1}|x_t, \delta_t(x_0, \dots, x_t) = a_t). \quad (2.4)$$

Cette hypothèse revient à dire que le processus stochastique défini à partir d'une politique fixée δ est *markovien*.

De même, la fonction d'utilité sur les conséquences est construite à partir de la somme de récompenses attachées aux transitions entre états :

$$\forall \tau = (x_0, a_0, \dots, x_{T-1}, a_{T-1}, x_T), u(\tau) = \sum_{t=0}^{T-1} r_t(x_t, a_t, x_{t+1}). \quad (2.5)$$

Cette forme additive suppose que la trajectoire considérée est de longueur finie. Néanmoins, il est parfois plus commode de représenter et résoudre des problèmes de décision séquentielle dans l'incertain sur un horizon de temps infini, à condition que ces problèmes soient "stationnaires" (les probabilités de transition p_t et les récompenses r_t sont indépendantes du temps). Rien ne garantit alors que l'utilité $u(\tau)$ d'une trajectoire infinie, décrite par l'équation (2.5), soit finie. C'est pourquoi l'une des deux fonctions d'utilité suivantes est utilisée en général pour les problèmes dont l'horizon est infini, plutôt que la forme (2.5) :

$$u(\tau) = \lim_{T \rightarrow \infty} \sum_{t=0}^T \gamma^t \cdot r(x_t, a_t, x_{t+1}), 0 < \gamma < 1, \quad (2.6)$$

$$u(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \left(\sum_{t=0}^{T-1} r(x_t, a_t, x_{t+1}) \right). \quad (2.7)$$

Si la fonction de récompense r est bornée, les deux expressions (2.6) et (2.7) sont bien définies, quelle que soit la longueur de la trajectoire τ . La fonction d'utilité (2.6) est appelée *critère gamma-pondéré* alors que la fonction (2.7) est appelée *critère moyen*.

Résoudre un problème de décision séquentielle dans l'incertain revient à *optimiser* le choix d'une politique. En d'autres termes, on cherche à trouver, pour un problème de décision séquentielle dans l'incertain $(\mathcal{X}, \mathcal{A}, p, r, H)$, une politique δ^* telle que :

$$EU_{\delta^*}(x_0) \geq EU_{\delta}(x_0), \forall \delta, \forall x_0 \in \mathcal{X}. \quad (2.8)$$

A priori, rien n'indique qu'une telle politique optimale existe si nous exigeons que *la même politique* δ^* soit optimale en tout état de départ x_0 . Néanmoins, on peut montrer qu'une telle politique optimale existe, pour les trois critères évoqués (Bellman, 1957; Puterman, 1994). Qui plus est, dans le cas où l'horizon est infini, il existe une politique optimale *stationnaire* (indépendante de t), $\delta^* : \mathcal{X} \rightarrow \mathcal{A}$. Le problème d'optimisation défini par (2.8) est classiquement résolu par des méthodes de type *Programmation Dynamique Stochastique* (Bellman, 1957; Bertsekas, 1987; Puterman, 1994). L'algorithme de *Recherche Arrière* est utilisé pour résoudre des problèmes à horizon fini et les algorithmes *d'Itération de la Politique* et *d'Itération de la Valeur* sont les plus couramment utilisés pour le critère gamma-pondéré. Une description succincte de ces algorithmes peut être trouvée en annexe B.2.

2.3 Planification dans l'incertain et processus décisionnels de Markov

Le cadre des processus décisionnels de Markov s'est largement imposé comme modèle pour la planification dans l'incertain en intelligence artificielle ces dernières années. Néanmoins, un certain nombre de limitations le rendent insuffisant pour modéliser et résoudre la plupart des problèmes de planification dans l'incertain. Parmi ces limitations, citons :

- L'hypothèse d'observabilité complète de l'état du monde à chaque instant.
- L'hypothèse (différente de la précédente) de connaissance parfaite du modèle (transitions, récompenses). En effet, parfois ce modèle n'est accessible qu'indirectement, par simulation ou expérimentation. Parfois également, seules des évaluations "qualitatives" des préférences et des connaissances sont disponibles (voir section suivante).
- L'hypothèse de représentation des états et décisions *en extension*, beaucoup plus limitée que les langages de représentation habituellement utilisés en planification qui permettent de modéliser des problèmes beaucoup plus complexes.

Pour pallier ces différentes limitations, plusieurs pistes ont été suivies, que nous allons décrire succinctement dans cette section :

- Processus décisionnels de Markov partiellement observables.
- Processus décisionnels de Markov et apprentissage.
- Processus décisionnels de Markov factorisés.

2.3.1 Processus décisionnels de Markov partiellement observables

Dans de nombreux problèmes de décision séquentielle dans l'incertain, l'agent n'a pas une connaissance complète de l'état réel de son environnement, contrairement à ce qui est supposé dans le cadre des PDM. Ce sont ses actions qui, en plus de modifier l'état du système, lui apportent des informations permettant de préciser sa connaissance du monde. Ainsi, on peut distinguer deux types d'effets des actions, en environnement partiellement observable :

- Effet *physique* des actions : une action a_t , appliquée par l'agent à l'instant t , transforme l'état du monde x_t en un état x_{t+1} éventuellement de manière stochastique, suivant une probabilité de transition $p_t(x_{t+1}|x_t, a_t)$.

- Effet *épistémique* des actions : l'état résultant x_{t+1} n'est pas observé directement, mais une *observation indirecte* o , élément d'un ensemble d'observations \mathcal{O} est retournée par le système. Cette observation peut être "déterministe", dans le cas où il existe une fonction $O : \mathcal{X} \times \mathcal{A} \rightarrow \mathcal{O}$ telle que $o = O(x_{t+1}, a_t)$. Cette observation peut, plus généralement, être stochastique. Dans ce cas, $O(x_{t+1}, a_t, o) = p(o|x_{t+1}, a_t)$ est la probabilité d'observer o lorsque l'action a_t est appliquée et résulte en x_{t+1} .

La prise en compte de ces deux types d'effet conduit à étendre le cadre des processus décisionnels de Markov (Akström, 1965; Sondik, 1978). La connaissance imparfaite de l'état du monde à l'instant t est modélisée par un *état de croyance* $b_t \in \mathcal{B}$, où \mathcal{B} est l'ensemble des distributions de probabilité sur \mathcal{X} . b_0 est l'état de croyance initial et pour $t \geq 0$, $b_t(x)$ représente la probabilité associée par l'état de croyance à l'état x . Cet état de croyance évolue sous l'effet combiné des actions et observations successives.

Une première approche pour la résolution d'un PDMPO consiste à associer directement à chaque observation possible de l'état du monde une action. Mais cette méthode associe la même action à tous les états du monde conduisant à la même observation. Or, en général, de telles politiques sont sous-optimales (Littman, 1994). Afin de trouver une politique tirant au mieux parti des observations imparfaites, il est nécessaire de différencier des états conduisant à la même observation en prenant en compte les états rencontrés précédemment. Ceci peut être fait (dans le cas où l'horizon est fini) en définissant une politique δ_t déterminant l'action à effectuer à l'instant t , non pas simplement en fonction de la dernière observation o_{t-1} , mais en fonction de l'*historique* des actions et observations, $\{a_1, o_1, \dots, a_{t-1}, o_{t-1}\}$. Ainsi, une politique δ_t peut être représentée par un *arbre de décision*, de profondeur t , dont les nœuds sont des actions, et les arêtes des observations (Figure 2.3.1).

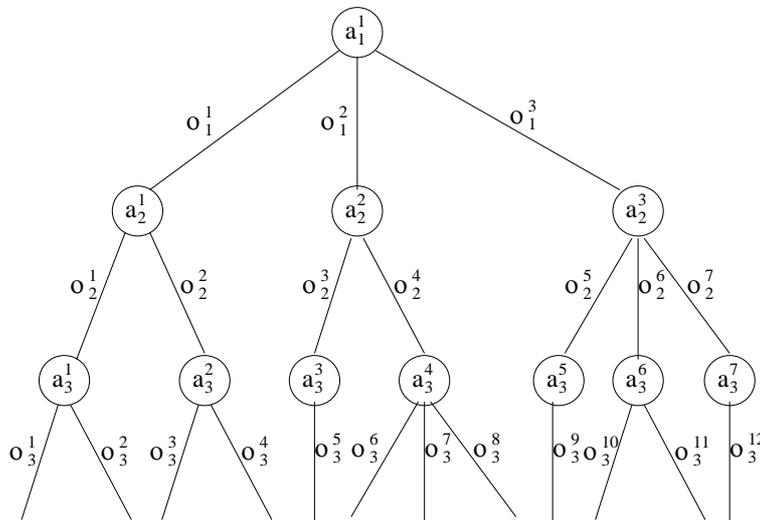


FIG. 2.2 – Politique δ_t d'un processus décisionnel de Markov partiellement observable.

Dans le cas où l'horizon est infini, bien qu'un PDMPO puisse être traduit en un PDM classique, la nature du problème obtenu (espace d'états continu) rend impossible sa résolution par les méthodes classiques de programmation dynamique. Cette complexité a été à la source de plusieurs voies de

recherche en intelligence artificielle, et plusieurs familles d'algorithmes de résolution de PDMPO ont été proposées :

- Une première famille de méthodes met à profit la structure (linéaire par morceaux et convexe) de la fonction de valeur $V : \mathcal{B} \rightarrow \mathbb{R}$ pour proposer des algorithmes de type *Itération de la valeur*. Ces algorithmes utilisent une structure d'arbre pour représenter une politique et sa fonction de valeur associée et définir une méthode itérative de calcul de la fonction de valeur (approchée) (Kaelbling et al., 1998).
- Les méthodes de résolution les plus récentes sont basées sur une discrétisation de l'espace d'états ou combinent les deux types d'approches (Pineau et al., 2003), (Smith and Simmons, 2005).

2.3.2 Processus décisionnels de Markov et apprentissage

Le cadre des processus décisionnels de Markov permet de représenter et résoudre des problèmes de planification dans l'incertain. Grâce à la Programmation Dynamique, le surcroît de complexité de la résolution de ces problèmes lié à leur aspect séquentiel est limité. Nous venons de voir qu'il est possible d'étendre le cadre des PDM à la planification en environnement "partiellement observable". On parle parfois dans le cadre des PDM d'observabilité partielle dans un sens différent alors que l'état du monde est parfaitement connu à chaque instant. Il s'agit du cas où le *modèle* du PDM est imparfaitement connu, i.e. lorsque les fonctions p et r du modèle $\langle \mathcal{X}, A, p, r \rangle$ sont inconnues a priori mais accessibles par expérimentation, soit parce qu'on peut *simuler* la dynamique du système, soit parce qu'on peut l'expérimenter en temps réel. Les méthodes de type *Apprentissage par Renforcement* visent à résoudre de tels problèmes dans lesquels le "modèle" du PDM est appris en même temps que sa solution optimale. Pour ce faire, il existe deux types de méthodes : les méthodes *indirectes* et les méthodes *directes*.

Les *méthodes indirectes* (Kumar and Varaiya, 1986; Sutton, 1991; Peng and Williams, 1993; Moore and Atkeson, 1993) supposent d'apprendre dans un premier temps (par simulation ou expérimentation) le modèle (p, r) du PDM, puis de le résoudre par un algorithme de Programmation Dynamique. De manière un peu plus évoluée, on peut focaliser l'effort lié à l'apprentissage du modèle sur des zones de l'espace d'états-actions $(\mathcal{X} \times A)$ prometteuses, tout en ne négligeant pas totalement le reste de l'espace d'états-actions, afin de garantir qu'on ne passe pas à côté d'une politique optimale. Les méthodes indirectes permettent de résoudre des PDM dont on ne connaît pas le modèle, à la condition de pouvoir expérimenter ou simuler ce modèle. Ces méthodes présentent toutefois un inconvénient : elles nécessitent de stocker au moins partiellement les fonctions \hat{p} et \hat{r} , ce qui peut nécessiter jusqu'à $O(|\mathcal{X}|^2|A|)$ d'espace de stockage.

Les *méthodes directes* (Sutton, 1988; Watkins, 1989; Watkins and Dayan, 1992) permettent de se passer de stocker le modèle (p, r) en entier et de ne garder que ce qui est nécessaire à l'évaluation de politiques ou au calcul de politiques optimales. En contrepartie, des expérimentations / simulations plus nombreuses peuvent être nécessaires. Le choix d'une méthode directe sera donc préféré lorsque les simulations ont un coût faible, et qu'un problème de taille mémoire peut se poser.

Les méthodes directes et indirectes entrelacent en général apprentissage et programmation dynamique, afin de gagner en efficacité.

2.3.3 Processus décisionnels de Markov factorisés

Jusqu'ici, nous avons décrit une représentation *en extension* des états et actions d'un PDM dans laquelle tous les états et actions sont énumérés explicitement. Les divers algorithmes de résolution de

PDM, PDMPO, avec modèle connu ou inconnu sont basés sur cette hypothèse de représentation en extension. Néanmoins, dans le cadre de l'intelligence artificielle et de la planification dans l'incertain, en particulier, la nature intrinsèquement combinatoire des domaines fait qu'on se concentre en général sur des représentations concises des états et actions. Ces représentations concises, ou *en intension*, consistent à décrire l'état du monde particulier par un ensemble de propriétés le caractérisant, plutôt que de l'identifier individuellement. Cette représentation en intension permet une représentation plus concise des probabilités de transition et des récompenses qu'une représentation explicite.

Un *Réseau Bayésien* (RB) (Pearl, 1988a) peut permettre de représenter de manière concise de telles probabilités, dans le cas où les variables sont sous forme factorisée. Un *Réseau Bayésien Homogène Stationnaire* (RBHS) est un RB particulier, adapté à la représentation des probabilités de transition dans les chaînes de Markov stationnaires (Dean and Kanazawa, 1989).

Plus précisément, un RBHS est constitué :

- d'un modèle graphique, exprimant les dépendances / indépendances entre variables, entre deux instants consécutifs quelconques, t et $t + 1$,
- d'un modèle quantitatif, représentant explicitement les probabilités conditionnelles *entre les variables dépendantes uniquement*, sous la forme de *tables de probabilités conditionnelles* ou sous forme d'*arbres de probabilités conditionnelles*, par exemple.

Il est donc possible de représenter de manière concise des probabilités de transition via l'utilisation de RBHS. Néanmoins, quoique plus compacte que la représentation en extension, cette représentation nécessite de créer un RBHS par action applicable. Ceci peut être difficile, surtout si les actions elles-mêmes sont représentées par les valeurs d'un ensemble de variables. En fait, on peut éviter de construire un RBHS par action, en intégrant celles-ci à l'intérieur d'un RBHS unique. Il suffit de rajouter au RBHS des nœuds "actions". De ces nœuds partent des arcs influençant les variables d'états à l'instant $t + 1$. Il est également possible de représenter directement des politiques stationnaires dans ce formalisme, par l'intermédiaire d'arcs "entrant" vers les variables d'action.

Les fonctions récompenses peuvent également être représentées dans le cadre des RBHS par des nœuds spécifiques ne comportant que des arcs entrants, provenant de nœuds états (de l'étape $t + 1$ aussi bien que de l'étape t) ou de nœuds actions, voire d'autres nœuds récompenses. La sémantique d'un nœud récompense dans un RBHS peut être représentée par une table ou un arbre de récompenses tout comme les probabilités de transition ou les politiques. Dans le cas où la récompense immédiate peut être vue comme une somme de récompenses, un RBHS peut comporter plusieurs nœuds récompenses, chacun représentant une composante de la récompense immédiate totale. Un RBHS comprenant à la fois des nœuds "variables d'états", "actions" et "récompenses" est un *diagramme d'influence homogène stationnaire*. Lorsque le PDM sous-jacent est à horizon fini, et pas forcément stationnaire, le cadre plus classique des *Diagrammes d'influence* ((Pearl, 1988a; Jensen, 2001), par exemple) peut être utilisé pour représenter et résoudre de tels problèmes. La *fonction de valeur* associée à une politique (ou toute fonction de valeur, plus généralement) peut également s'exprimer sous forme tabulaire ou arborescente. (Boutilier et al., 2000) décrivent les algorithmes *Itération de la valeur structurée* et *Itération de la politique structurée*, générant les politiques optimales et leur fonction de valeur associée, sous forme arborescente.

(Givan et al., 2003) ont également présenté un certain nombre d'algorithmes itératifs permettant de résoudre efficacement des PDM factorisés à partir de telles partitions présentant la propriété de *bisimilarité stochastique* (ou approchant cette propriété), construites automatiquement. Ces algorithmes

permettent de construire des problèmes “réduits”, dont les états sont les éléments de la partition. D’autres auteurs (Koller and Parr, 2000; Kearns et al., 1999; Dietterich and Flann, 1995) ont également proposé des approches originales pour la résolution de PDM factorisés : une méthode approchée basée sur l’utilisation d’une combinaison linéaire de fonctions de valeurs élémentaires pour les premiers et des méthodes basées sur la simulation (Apprentissage par Renforcement, “sampling”) pour les deux autres.

2.4 Représentations non-classiques de l’incertain

Comme nous l’avons déjà mentionné, la théorie de la décision dans l’incertain basée sur l’utilité espérée est la théorie dominante. Son extension naturelle à la décision séquentielle est le cadre des processus décisionnels de Markov. Néanmoins, malgré ses fondements théoriques solides, sa pertinence a été remise en cause par des travaux expérimentaux d’économistes ou de psychologues (Allais, 1953; Ellsberg, 1961; Kahneman and Tversky, 1979). De nouveaux modèles de décision dans l’incertain, basés sur des extensions des probabilités ont ensuite été proposés (Schmeidler, 1989; Sarin and Wakker, 1992; Gilboa, 1987; Jaffray, 1989). Puis, des approches plus *qualitatives* de la décision dans l’incertain ont vu le jour en intelligence artificielle, sans doute motivées par l’approche symbolique du traitement de l’information (connaissances, préférences), classiquement utilisée dans cette discipline. Dans cette section, je présenterai brièvement ces modèles.

2.4.1 Limites et généralisations du critère de l’utilité espérée

Paradoxes d’Allais et d’Ellsberg

Un premier paradoxe, proposé par Allais (Allais, 1953), présente un problème de décision sous risque pour lequel la plupart des décideurs auxquels on le soumet font un choix incompatible avec le critère de l’utilité espérée. Nous présentons ici une version simplifiée de cet exemple, présenté dans un cadre à la von Neumann et Morgenstern :

Exemple 2.4.1 Paradoxe d’Allais

Soient les quatre loteries suivantes :

- l_1 donne accès à un gain de 15000 euros avec la probabilité 0.09 et 0 euro avec la probabilité 0.91,
- l_2 donne accès à un gain de 10000 euros avec la probabilité 0.1 et 0 euro avec la probabilité 0.9,
- l_3 donne accès à un gain de 15000 euros avec la probabilité 0.9 et 0 euro avec la probabilité 0.1,
- l_4 offre un gain certain de 10000 euros.

Il ne semble pas illogique de préférer strictement l_1 à l_2 (car l_1 permet d’obtenir un gain significativement plus grand que l_2 , avec une probabilité assez voisine) et de préférer strictement l_4 à l_3 car l_4 représente un gain certain, alors que la probabilité de ne rien gagner avec l_3 , quoique faible, est non nulle.

On peut montrer qu’un tel comportement ($l_1 \succ l_2$ et $l_4 \succ l_3$, voir figure (2.3)) est en contradiction avec le critère de l’utilité espérée. Pour cela, cherchons des utilités $u(15000)$, $u(10000)$ et $u(0)$ compatibles avec ces deux préférences. Dans la théorie de von Neumann et Morgenstern, u n’est déterminée qu’à une transformation affine près par les préférences du décideur, on peut donc librement poser $u(0) = 0$. Exprimées en termes d’utilité, les deux préférences donnent les inégalités suivantes :

- $l_1 \succ l_2$: $0,09 \times u(15000) > 0,1 \times u(10000)$, soit $0,9 \times u(15000) > u(10000)$ et
- $l_4 \succ l_3$: $u(10000) > 0,9 \times u(15000)$.

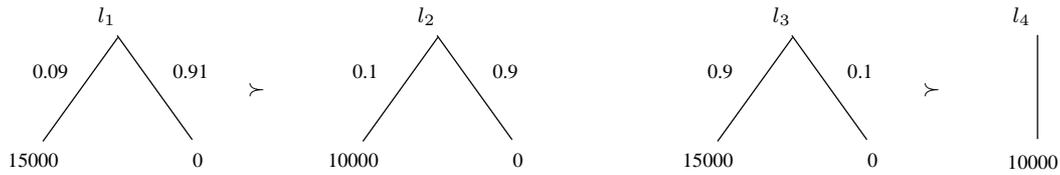


FIG. 2.3 – Violation du critère de l'utilité espérée.

Ce système d'inéquations n'admet pas de solution, ce qui signifie que pour ce problème il n'existe aucune fonction d'utilité sur les conséquences telle que les préférences de l'utilisateur soient modélisables par l'espérance mathématique de cette fonction.

Le contre-exemple suivant, proposé par Ellsberg (Ellsberg, 1961), montre comment la théorie de l'utilité espérée, et plus généralement l'usage des probabilités pour représenter l'incertitude, peuvent être mises en défaut par des problèmes mélangeant risque et incertitude. Ceci a mené certains chercheurs (Gilboa (Gilboa, 1987), Schmeidler (Schmeidler, 1989)) à proposer une généralisation de l'utilité espérée, utilisant des probabilités non-additives, ou *capacités* (Choquet (Choquet, 1954)).

Exemple 2.4.2 *Paradoxe d'Ellsberg*

Nous avons une urne remplie de 90 boules de couleurs rouge (R), blanche (B) ou jaune (J). Nous savons qu'exactly 30 boules sont rouges, et que 60 sont de couleur blanche ou jaune. Les proportions de boules blanches et jaunes sont indéterminées. Une boule est tirée au hasard et le problème de décision consiste à ordonner les deux paires de paris suivantes :

| Pari | Rouge | Blanc | Jaune |
|------|------------|------------|------------|
| f | 1000 euros | 0 euros | 0 euros |
| g | 0 euros | 1000 euros | 0 euros |
| f' | 1000 euros | 0 euros | 1000 euros |
| g' | 0 euros | 1000 euros | 1000 euros |

Le pari f rapporte 1000 euros si la boule tirée est rouge, et rien sinon, etc...

Il apparaît raisonnable de préférer f à g , car la proportion de boules rouges vaut exactement un tiers alors que la proportion de boules blanches est imprécisément connue : elle peut être de deux tiers, mais elle peut également être nulle. Préférer f à g signifie que le décideur a une *aversion pour l'ambiguïté* (ou l'incertitude).

Maintenant, si il préfère f à g , il doit pour la même raison (aversion pour l'ambiguïté) préférer g' à f' car g' offre exactement deux chances sur trois de gagner alors que f' offre entre une chance sur trois et cent pour cent de chances de gain. Ces préférences ($f \succ g$ et $g' \succ f'$) sont une nouvelle fois incompatibles avec la théorie de l'utilité espérée. On observe facilement que le "principe de la chose certaine" (**Sav2**, voir annexe B.1) est violé : f et g donnent la même conséquence pour l'état "J" (0 euros), et la modification de cette conséquence (1000 euros) conduit à un renversement de préférences. Notons que l'attitude négative du décideur envers l'ambiguïté n'est pas le point crucial : cette attitude doit simplement être cohérente pour tous les problèmes de décision qui s'offrent à lui. Un raisonnement similaire pour un décideur ayant un "attrait" pour l'ambiguïté conduirait également à une violation de **Sav2**. En effet, les préférences seraient cette fois ($f \prec g$ et $g' \prec f'$).

Détaillons les contraintes imposées aux probabilités $P(B)$, $P(J)$ et $P(R)$ par les préférences $f \succ g$ et $g' \succ f'$. Les deux contraintes obtenues sont : $P(R) > P(B)$ et $P(R \cup J) < P(B \cup J)$, donc $P(B) + P(R \cup J) < P(B \cup J) + P(R)$. Or si on suppose que les probabilités sont additives, puisque B , R et J sont disjoints, on obtient $P(R \cup B \cup J) < P(R \cup B \cup J)$, ce qui est impossible. Ainsi, aucun jeu de probabilités subjectives justifiant ces préférences ne peut exister.

Généralisations du critère de l'utilité espérée basée sur l'intégrale de Choquet

Pour résoudre les paradoxes précédents, il nous faut abandonner l'hypothèse d'additivité des probabilités (subjectives ou objectives). C'est ce qu'ont proposé un certain nombre d'auteurs qui ont défendu l'utilisation de critères de décision dans l'incertain basés sur l'intégrale de Choquet (Choquet, 1954), généralisant le critère de l'utilité espérée (voir (Chateauneuf and Cohen, 2006) pour une synthèse complète de ces travaux). L'intégrale de Choquet a été appliquée à la décision dans l'incertain et le risque par plusieurs auteurs (Gilboa (Gilboa, 1987), Schmeidler (Schmeidler, 1989), Sarin et Wakker (Sarin and Wakker, 1992)).

Sarin et Wakker (Sarin and Wakker, 1992) ont proposé une axiomatisation de ce critère, basée sur un relâchement des axiomes de Savage. Cette axiomatisation fait appel à la notion d'événements *non-ambiguous* ("probabilisables").

(Jaffray, 1989) et (Jaffray and Wakker, 1994) ont proposé et axiomatisé une forme particulière de critère de décision basé sur l'intégrale de Choquet, dans laquelle la mesure ν considérée est une *fonction de croyance* ou de *plausibilité*. Enfin, l'intégrale de Choquet a également été appliquée en décision sous risque. Dans ce cadre, le modèle *d'espérance d'utilité dépendant du rang* (UDR) a été initialement introduit par (Quiggin, 1982) (sous le terme *utilité anticipée*) et étudié et axiomatisé par (Yaari, 1987; Segal, 1987; Wakker, 1994; Chateauneuf, 1999).

2.4.2 Approches qualitatives de la décision dans l'incertain

La difficulté à quantifier numériquement l'incertitude et les préférences d'un agent en décision dans l'incertain a conduit à la recherche de *théories qualitatives de la décision dans l'incertain*, basées sur des modèles non numériques (voir (Doyle and Thomason, 1999)).

Dans une vision de l'incertain dans le style de Savage, les préférences sont exprimées par une fonction d'utilité $\mu : \mathcal{X} \rightarrow L_P$, où L_P est une échelle de degrés de préférence. Les connaissances incertaines de l'agent sont exprimées par une mesure monotone $\sigma : 2^S \rightarrow L_V$, où L_V est une échelle de degrés de vraisemblance, potentiellement différente de L_P . Dans le cadre de l'utilité espérée, la fonction d'utilité μ est à valeurs réelles et σ est une mesure de probabilité à valeurs dans l'intervalle $[0,1]$.

Dans la vision la plus *qualitative*, les deux échelles L_P et L_V n'ont que deux éléments : $L_P = \{\perp_P, \top_P\}$ et $L_V = \{\perp_V, \top_V\}$. La fonction μ ne sert alors qu'à distinguer les conséquences satisfaisantes (ou *buts*) des conséquences non satisfaisantes (*non-buts*). De son côté, la mesure σ permet de distinguer un sous ensemble $A \subseteq S$ d'états du monde *possibles*, les autres états étant impossibles. Dans ce cadre, les actes qui permettent d'obtenir une conséquence satisfaisante dans *tous* les états possibles du monde sont satisfaisants, alors que ceux qui ne permettent *jamais* d'obtenir une conséquence satisfaisante seront bien entendu les moins intéressants.

En élargissant progressivement cette notion de décision qualitative, l'échelle de préférence L_P s'enrichit, et (L_P, \geq_P) devient un ensemble fini, totalement² ordonné. Alors, le critère de Wald (Wald,

²Nous ne considérons pas dans cette brève vision de la décision qualitative, les cas où les échelles L_P et L_V sont partiellement ordonnées.

1950) devient un critère naturel pour évaluer les actes :

$$U_W(f) = \min_{s \in A} \mu(f(s)).$$

Continuons à élargir le cadre de la décision qualitative dans l'incertain : les deux échelles (L_P, \geq_P) et (L_V, \geq_V) sont maintenant toutes deux complètement ordonnées (et finies). La question qui se pose alors est celle de la *commensurabilité* des deux échelles L_P et L_V . Est-il possible de comparer deux éléments de L_P et L_V ? Différentes familles d'approches pour la décision qualitative dans l'incertain ont été proposées, supposant ou non la commensurabilité des deux échelles.

L'approche la plus simple (adoptée par (Boutillier, 1994; Lang, 1996), par exemple) ne suppose pas la commensurabilité entre les deux échelles. Elle consiste, à partir d'une mesure de vraisemblance σ sur les états de \mathcal{S} , à identifier l'ensemble A^* des états maximisant σ . Alors, le critère de Wald peut être appliqué, en ne considérant que les états de vraisemblance maximale :

$$U_{W,\sigma}(f) = \min_{s \in A^*} \mu(f(s)), A^* = \arg \max_{A \subseteq \mathcal{S}} \sigma(A).$$

Toujours dans le cas où l'on ne suppose pas la commensurabilité entre incertitudes et préférences, (Dubois et al., 2003) ont proposé et axiomatisé une méthode très générale de comparaison entre les actes, non représentable par un critère d'utilité :

$$f \geq_{\sigma,\mu} g \Leftrightarrow \sigma(\{s, \mu(f(s)) >_P \mu(g(s))\}) \geq_V \sigma(\{s, \mu(g(s)) >_P \mu(f(s))\}).$$

Dans le cas le plus général, cette relation de préférence entre actions n'est pas forcément transitive et ne le devient que dans des cas très particuliers (par exemple lorsque σ induit un ordre strict sur l'ensemble des états).

Si on considère maintenant le cas où incertitudes et préférences sont commensurables, c'est-à-dire exprimées sur une échelle unique, L , il devient alors possible d'agrèger les degrés d'incertitude $\sigma(A)$, $A \subseteq \mathcal{S}$ et les degrés de préférence $\mu(f(s))$, $s \in \mathcal{S}$. Nous avons proposé et axiomatisé [45] un critère de décision dans l'incertain basé sur l'intégrale de Sugeno (Sugeno, 1977), qui utilise une telle agrégation :

$$U_{SUG,\sigma,\mu}(f) = \max_{\lambda \in L} \min(\lambda, \sigma(F_\lambda)).$$

Ce critère de décision, très général, peut être vu comme une contrepartie qualitative (dans le sens où seuls des opérateurs *min* et *max* sont impliqués) de l'intégrale de Choquet. En choisissant comme mesure d'incertitude σ une *mesure de possibilité* Π ou une *mesure de nécessité* N (Dubois and Prade, 1988), ce critère se spécialise en l'un des deux critères de *décision dans l'incertain possibiliste* proposés par (Dubois and Prade, 1995) :

$$\begin{aligned} U_{OPT,\pi,\mu}(f) &= U_{SUG,\Pi,\mu}(f) = \max_{s \in \mathcal{S}} \min\{\pi(s), \mu(f(s))\}, \\ U_{PES,\pi,\mu}(f) &= U_{SUG,N,\mu}(f) = \min_{s \in \mathcal{S}} \max\{n(\pi(s)), \mu(f(s))\}, \end{aligned}$$

où π est la *distribution de possibilité* associée soit à la mesure Π , soit à la mesure N . Une partie importante de mes travaux de recherche ayant porté sur cette approche possibiliste (ou basée sur l'intégrale de Sugeno), j'y consacre le chapitre suivant de ce rapport.

Enfin, à la frontière entre les approches qualitatives de la décision dans l'incertain (du type “intégrales de Sugeno”) et les approches quantitatives (du type “intégrales de Choquet”), se situent les approches de type *ordre de magnitude d'utilité espérée* (Giang and Shenoy, 2000; Pearl, 1993; Wilson, 1995), basées sur une représentation de l'incertain dans le cadre des *fonctions kappa*, proposées par (Spohn, 1990), qui sont très proches des mesures possibilistes. Dans cette approche, les degrés d'incertitude et de préférences sont exprimés sur l'échelle des entiers naturels équipée de l'ordre entre les entiers, mais degrés de préférence et degrés d'incertitude sont agrégés via une *somme* dans le critère de décision proposé, ce qui nécessite un degré de raffinement supplémentaire par rapport aux approches précédentes.

2.5 Résumé, situation de mes travaux

La théorie de la décision est une discipline qui se situe à l'intersection des domaines de l'économie, de la recherche opérationnelle et de l'intelligence artificielle. Dans cette théorie, le cadre “classique”, basé sur le critère de *l'utilité espérée* a pris une place prédominante pour la modélisation des problèmes de décision dans l'incertain. La communauté de l'intelligence artificielle a adopté ce cadre et son extension à la décision séquentielle, le cadre des *processus décisionnels de Markov*, pour la planification dans l'incertain. Elle a également développé et amélioré ce cadre, afin de répondre aux défis particuliers de la planification dans l'incertain. Cela a conduit à i) l'élaboration de langages de représentation structurée et d'algorithmes de planification adaptés à ces langages ii) la prise en compte de l'observabilité partielle de l'état du monde dans ces problèmes, ou iii) l'apprentissage automatique, basé sur la simulation, de plans ou stratégies optimales.

Néanmoins, l'universalité du critère de l'utilité espérée a été remise en cause très tôt et plusieurs alternatives qualitatives ou quantitatives ont été proposées. Celles-ci ont tout d'abord été proposées dans le domaine des sciences de la décision, mais la communauté de l'intelligence artificielle s'est également lancée dans l'étude de telles alternatives, en particulier qualitatives.

A son échelle, ma recherche a suivi la même démarche, avec l'étude d'une théorie de la décision qualitative (critères, décision séquentielle, algorithmes...) et l'étude de langages et algorithmes pour la planification dans l'incertain, basés sur le cadre traditionnel des processus décisionnels de Markov.

Plus précisément, mes travaux peuvent être regroupés en trois thèmes :

1. Étude d'une famille de critères de décision qualitative dans l'incertain basés sur la *théorie des possibilités* et les *mesures de Sugeno*.
2. Proposition d'une contrepartie possibiliste des processus décisionnels de Markov et adaptation à la décision structurée (diagrammes d'influence possibilistes).
3. Travail de recherche autour des processus décisionnels de Markov “classiques”, dans le cadre de la décision structurée et dans le cadre partiellement observable.

Dans la deuxième partie de ce document je vais décrire mes travaux dans les deux premiers thèmes, puis, dans la troisième partie, je décrirai mes travaux de recherche dans le troisième thème.

Deuxième partie

Décision dans l'incertain et théorie des possibilités

Chapitre 3

Décision dans l'incertain basée sur les fonctions d'utilité qualitative monotone

Ce chapitre résume en particulier les travaux suivants, dédiés à la décision dans l'incertain basée sur les fonctions d'utilité qualitative monotone :

Justifications axiomatiques des utilités qualitatives monotones

- [45] Dubois (D.), Prade (H.) et Sabbadin (R.). – *Fuzzy Measures and Integrals : Theory and Applications*, chap. Qualitative decision theory with Sugeno integrals, pp. 314–322. – Physica-Verlag. Eds. M. Grabisch, T. Murofushi and M. Sugeno, 2000.
- [3] Dubois (D.), Prade (H.) et Sabbadin (R.). – Decision-theoretic foundations of qualitative possibility theory. *European Journal of Operations Research*, vol. 128, 2001, pp. 459–478.

Raffinements des utilités qualitatives possibilistes

- [46] Fargier (H) et Sabbadin (R.). – *Technologies for Constructing Intelligent Systems*, chap. Can qualitative Utility Criteria obey the Sure Thing Principle ?, pp. 167–178. – Heidelberg, New-York, Physica Verlag, *Studies in Fuzzyness and Soft Computing*, volume 1. Eds. B. Bouchon-Meunier, J. Gutierrez Rios, L. Magdalena et R. R. Yager, 2002.
- [4] Fargier (H.) et Sabbadin (R.). – Qualitative decision under uncertainty : back to expected utility. *Artificial Intelligence*, vol. 165, 2005, pp. 245–280.

3.1 Introduction

Ces dix dernières années, mes travaux autour de la décision qualitative se sont structurés autour de deux axes principaux :

1. Fondations et compréhension des critères de décision qualitative, possibilistes et monotone.
2. Algorithmique pour la planification dans l'incertain, dans le cadre de la théorie des possibilités.

Mes travaux sur l'axiomatisation des critères de décision qualitative possibilistes et monotone ont visé en premier lieu à expliciter les hypothèses sur lesquelles se basent ces critères de décision. Ils permettent également de proposer des procédures d'élicitation de préférences et de connaissances dans un cadre possibiliste, ou plus généralement d'utilité qualitative monotone, à partir de la comparaison

d'actes élémentaires [3], [45]. Mes travaux plus récents sur les liens entre utilité qualitative possibiliste et utilité espérée se situent également dans cet axe [19], [4]. Ces travaux sont décrits dans ce chapitre.

3.2 Utilité qualitative monotone et décision dans l'incertain

3.2.1 Utilité qualitative monotone

Nous avons défini [45] un critère d'*utilité qualitative monotone*, basé sur la représentation de l'incertitude par une mesure monotone, dans un cadre de décision à la Savage $(\mathcal{S}, \mathcal{X}, \sigma, \mu)$. Dans ce cadre, \mathcal{S} et \mathcal{X} représentent respectivement les ensembles (finis) d'états et de conséquences. $\sigma : 2^{\mathcal{S}} \rightarrow L$ est une mesure d'incertitude monotone, à valeurs dans un ensemble fini et complètement ordonné, L . La *fonction d'utilité qualitative*, $\mu : \mathcal{X} \rightarrow L$ représente les préférences d'un décideur. $\mu(x)$ exprime à quel degré la conséquence x est satisfaisante. μ représente donc les préférences du décideur de manière qualitative, en ordonnant les conséquences possibles des décisions. Un état x_{\perp} tel que $\mu(x_{\perp}) = 0_L$ sera forcément la *pire conséquence possible* (mais il peut y avoir plusieurs "pires conséquences" ex-aequo) et un état x^{\top} tel que $\mu(x^{\top}) = 1_L$ sera la *meilleure conséquence possible* (encore une fois, il peut y avoir des ex-aequos). Nous supposons que σ et μ prennent leurs valeurs dans une même échelle L . Ceci peut sembler une hypothèse forte, mais nous verrons dans la section (B.4) comment la justifier.

Considérons maintenant un acte $f \in \mathcal{A} = \mathcal{X}^{\mathcal{S}}$. L'utilité qualitative monotone de l'acte f est définie par une intégrale de Sugeno (Sugeno, 1977) :

Définition 3.2.1 *Utilité qualitative monotone d'une action [45].*

$$U_{SUG,\sigma,\mu}(f) = \max_{\lambda \in L} \min(\lambda, \sigma(F_{\lambda}))$$

où $F_{\lambda} = \{s \in \mathcal{S}, \mu(f(s)) \geq \lambda\}$, σ est une mesure d'incertitude monotone et μ une fonction d'utilité qualitative.

$U_{SUG,\sigma,\mu}$ peut être réécrite en faisant varier la conséquence $x \in \mathcal{X}$ ou l'état $s \in \mathcal{S}$, plutôt que le niveau λ dans l'échelle L :

Proposition 3.2.1

$$\begin{aligned} U_{SUG,\sigma,\mu}(f) &= \max_{x \in \mathcal{X}} \min\{\mu(x), \sigma(F_x)\} \text{ où } F_x = \{s \in \mathcal{S}, \mu(f(s)) \geq_P \mu(x)\}, \\ U_{SUG,\sigma,\mu}(f) &= \max_{s \in \mathcal{S}} \min\{\mu(f(s)), \sigma(F_s)\} \text{ où } F_s = \{s' \in \mathcal{S}, \mu(f(s')) \geq_P \mu(f(s))\}. \end{aligned}$$

Examinons comment se forme le degré d'utilité qualitative monotone de certains actes simples pour comprendre ce degré de façon intuitive. En premier lieu, il est facile de constater que l'utilité d'un acte constant x est $\mu(x)$. Plus généralement, pour les actes binaires xAy , nous avons :

Proposition 3.2.2 *Si $\mu(x) \geq \mu(y)$, alors $U_{SUG,\sigma,\mu}(xAy) = \max\{\mu(y), \min(\mu(x), \sigma(A))\}$.*

Par application de cette proposition, nous avons évidemment $U_{SUG,\sigma,\mu}(x^{\top}Ax_{\perp}) = \sigma(A)$. De plus, $U_{SUG,\sigma,\mu}(xAy)$ est la *médiane* de l'ensemble $\{\mu(y), \mu(x), \sigma(A)\}$ si $\mu(x) \geq \mu(y)$.

Exemple 3.2.1 *Considérons le cas, représenté dans la figure 3.1, d'un conducteur devant se rendre à une réunion et devant choisir un itinéraire. Ce conducteur a le choix entre deux itinéraires : centre-ville (c) et périphérique (p). Les deux états possibles du monde sont $\mathcal{S} = \{e, f\}$ où e signifie qu'il y a des embouteillages et f que le trafic est fluide. On supposera qu'il est plus vraisemblable qu'il y ait un embouteillage étant donné l'heure de départ, mais que ce n'est pas certain et qu'il n'est pas impossible que le trafic soit fluide. Cette connaissance se traduit par : $\sigma(\mathcal{S}) >_L \sigma(\{e\}) >_L \sigma(\{f\}) >_L \sigma(\emptyset)$. Le choix du conducteur influe sur son retard possible à la réunion : "à l'heure" (x), "léger retard" (y) et "fort retard" (z). Sa fonction d'utilité reflète la préférence du conducteur pour arriver à l'heure : $\mu(x) >_L \mu(y) >_L \mu(z)$. Si le conducteur choisit l'itinéraire c , il arrivera à l'heure (x) s'il n'y a pas d'embouteillage, mais avec un fort retard (z) s'il y en a. Le trajet par le périphérique (p) est plus long, mais moins risqué : quel que soit l'état du trafic, le conducteur aura un léger retard (y).*

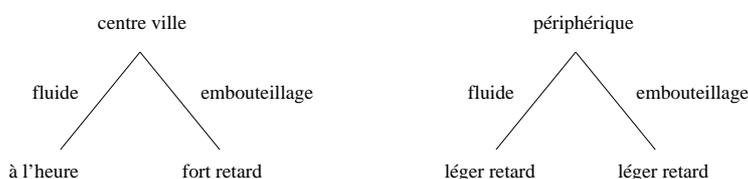


FIG. 3.1 – Conséquences du choix d'itinéraire (c ou p) en fonction du trafic (fluide ou embouteillé).

On peut calculer $U_{SUG,\sigma,\mu}(c)$ et $U_{SUG,\sigma,\mu}(p)$. p est un acte constant pour lequel $U_{SUG,\sigma,\mu}(p) = \mu(y)$. c est un acte binaire et $U_{SUG,\sigma,\mu}(c) = \max\{\mu(z), \min\{\mu(x), \sigma(\{f\})\}\}$, par application de la proposition (3.2.2). On peut alors en déduire que $U_{SUG,\sigma,\mu}(c) >_L U_{SUG,\sigma,\mu}(p)$ si et seulement si $\sigma(\{f\}) >_L \mu(y)$. En d'autres termes, le conducteur choisira de passer par le centre-ville si la plausibilité que le trafic soit fluide est plus grande que l'utilité d'avoir un léger retard.

3.2.2 Utilités qualitatives possibilistes

Dans le cadre de la décision dans l'incertain basée sur le critère d'utilité qualitative monotone, le choix de la mesure σ permet de modéliser toute une gamme d'attitudes face à l'incertitude, de la plus pessimiste à la plus optimiste. En particulier, si on choisit comme mesure σ une mesure de possibilité ou de nécessité, l'utilité qualitative monotone d'un acte prend l'une des formes suivantes :

Proposition 3.2.3 (Utilité qualitative possibiliste optimiste)

$$U_{OPT,\pi,\mu}(f) = U_{SUG,\Pi,\mu}(f) = \max_{s \in \mathcal{S}} \min\{\pi(s), \mu(f(s))\}.$$

Proposition 3.2.4 (Utilité qualitative possibiliste pessimiste)

$$U_{PES,\pi,\mu}(f) = U_{SUG,N,\mu}(f) = \min_{s \in \mathcal{S}} \max\{n(\pi(s)), \mu(f(s))\}.$$

Ces deux critères de décision qualitatifs ont été initialement proposés et axiomatisés dans un cadre à la von Neumann et Morgenstern (Dubois and Prade, 1995), puis nous les avons étudiés dans un cadre à la Savage [3]. En choisissant une mesure de nécessité ($\sigma = N$), on modélise un comportement pessimiste et en choisissant une mesure de possibilité ($\sigma = \Pi$), on modélise un comportement optimiste. $U_{OPT,\pi,\mu}$ peut être vu comme une extension du critère *maximax* qui évalue les actes suivant l'utilité de leur meilleure conséquence possible, alors que $U_{PES,\pi,\mu}$ est une extension du critère *maximin* qui les évalue suivant la pire conséquence possible.

Exemple 3.2.2 Reprenons l'exemple précédent de choix d'un itinéraire. Dans le cas de l'utilité qualitative possibiliste optimiste, $\sigma = \Pi$. Mais alors, puisque $\Pi(A \cup B) = \max\{\Pi(A), \Pi(B)\}, \forall A, B \subseteq \mathcal{S}$ et $\mathcal{S} = \{e, f\}$, nous devons avoir $\max\{\Pi(\{e\}), \Pi(\{f\})\} = \max\{\pi(e), \pi(f)\} = 1_L$.

Donc, si nous supposons qu'il est plus plausible que le trafic soit embouteillé, nous devons avoir $1_L = \Pi(\{e\}) > \Pi(\{f\}) > 0_L$. En conséquence, $U_{OPT,\pi,\mu}(c) > U_{OPT,\pi,\mu}(p)$ dès lors que $\pi(f) > \mu(y)$, soit dès qu'il est suffisamment possible que le trafic soit fluide. Pour choisir l'itinéraire "centre-ville", il nous suffit de croire qu'il est possible qu'il n'y ait pas d'embouteillage. En particulier, dans le cas où notre ignorance est totale ($\pi(f) = \pi(e) = 1_L$), nous choisirons de passer par le centre-ville. Ceci justifie le terme "optimiste" pour qualifier l'attitude d'un tel décideur !

Dans le cas de l'utilité qualitative possibiliste pessimiste où $\sigma = N$, puisque pour les mesures de nécessité, $N(A \cap B) = \min\{N(A), N(B)\}, \forall A, B \subseteq \mathcal{S}$, nous avons :

$$1_L > N(\{e\}) > N(\{f\}) = N(\emptyset) = 0_L.$$

Alors, $U_{PES,\pi,\mu}(c) = \mu(z) < U_{PES,\pi,\mu}(p) = \mu(y)$. Justifiant d'une attitude pessimiste, nous choisirons toujours l'itinéraire le plus "sûr", puisque notre connaissance ne nous garantit avec aucune certitude l'absence d'embouteillages ($\pi(e) = 1_L$).

3.2.3 Commensurabilité entre incertitudes et préférences

Les critères basés sur l'intégrale de Sugeno ordonnent les actes en les plaçant sur une échelle ordinale unique. Ceci est possible à condition de supposer une forme de commensurabilité entre échelles d'incertitude et de préférence, ce que j'ai fait dans la plupart de mes travaux. Néanmoins, cette hypothèse peut être relâchée. (Dubois et al., 2003) ont axiomatisé une relation de préférence entre actes (dans le cas non séquentiel), lorsqu'on dispose d'une relation de *confiance* entre les événements \succeq_V et d'une relation de préférence entre les conséquences \succeq_P , non commensurables (voir aussi [15] [44]).

3.2.4 Justifications axiomatiques des critères d'utilité qualitative monotone

En nous basant sur les travaux de (Savage, 1954) exposés dans le chapitre 2, nous avons proposé une justification axiomatique à la Savage des critères d'utilité qualitative monotone et possibilistes [45], [3] (voir sections B.4 et B.5 en annexe).

3.3 Raffinements des utilités qualitatives monotones

3.3.1 Utilités qualitatives possibilistes et effet de noyade

Bien que séduisantes dans le cadre d'une théorie qualitative de la décision dans l'incertain, les utilités qualitatives possibilistes¹ peuvent être trop peu *décisives*. Ceci résulte d'un *effet de noyade*, lié à l'utilisation des opérateurs *min* et *max* dans ces critères. Si deux actes f et g donnent la même conséquence, satisfaisante, dans un état très plausible, les deux actes peuvent être considérés *équivalents* par le critère optimiste même si l'un des deux actes domine l'autre dans tous les autres états du monde. Plus précisément, il se peut que $\forall s \in \mathcal{S}, \mu(f(s)) \geq \mu(g(s))$ et $\exists s^*, \pi(s^*) > 0$ et $\mu(f(s^*)) > \mu(g(s^*))$ mais $g \succeq f$. Le même type d'exemple peut être trouvé dans le cas pessimiste. Il se peut même que deux actes ne soient distingués par aucun des deux critères possibilistes, comme le montre l'exemple suivant :

¹Et les utilités qualitatives monotones, par extension.

Exemple 3.3.1 Soient $\mathcal{S} = \{s_1, s_2\}$ et $L = \{0, 1, 2, 3, 4, 5\}$. Soient f et g deux actes dont les utilités des conséquences dans les états s_1 et s_2 sont listées dans la table suivante, de même que les degrés de possibilité de s_1 et s_2 :

| | s_1 | s_2 |
|-------------|-------|-------|
| $\mu(f(s))$ | 3 | 4 |
| $\mu(g(s))$ | 3 | 1 |
| π | 5 | 2 |

On vérifie que $U_{OPT,\pi,\mu}(f) = U_{OPT,\pi,\mu}(g) = 3$ et $U_{PES,\pi,\mu}(f) = U_{PES,\pi,\mu}(g) = 3$ bien que f domine strictement g dans l'état s_2 et que les deux actes soient équivalents en s_1 .

La plupart des approches qualitatives de la décision proposées dans la littérature ((Pearl, 1993), (Du-bois and Prade, 1995), (Brafman and Tennenholtz, 1997), (Giang and Shenoy, 2000)...) sont sujettes à l'effet de noyade, mais ce n'est pas le cas en décision quantitative basée sur le critère de l'utilité espérée. En effet, le *principe de la chose certaine*, **Sav 2**, permet de s'assurer que les conséquences identiques (sinon de même utilité) de deux actes ne sont pas prises en compte dans la comparaison de ces deux actes. La question suivante est donc légitime [17] :

“Est-il possible, dans le cadre de la décision qualitative possibiliste, de restaurer la prin-cipe de la chose certaine afin de supprimer l'effet de noyade ?”

Malheureusement, on peut facilement montrer qu'une relation de préférence qualitative possibiliste $\succeq_{OPT,\pi,\mu}$ ou $\succeq_{PES,\pi,\mu}$ ne respecte le principe de la chose certaine qu'à la condition qu'il n'y ait pas d'incertitude. Ou en d'autres termes, que :

$$\exists ! s^* : \pi(s^*) = 1_L \text{ et } \forall s \neq s^*, \pi(s) = 0_L.$$

Devant ce résultat négatif, nous nous sommes tournés vers une question moins évidente et plus intéressante : celle de trouver des *raffinements* des critères de décision possibilistes qui satisfont le principe de la chose certaine.

3.3.2 Raffinements des critères de décision possibilistes

On dit qu'une relation de préférence \succeq' entre les actes raffine une relation de préférence \succeq dès lors que l'ordre strict associé \succ' range au moins aussi bien les actes que \succ .

Dans le cadre de la décision basée sur les utilités qualitatives possibilistes, π et μ représentent origi-nellement les connaissances et préférences d'un décideur. Puisqu'il n'y a aucune raison pour qu'un raffinement d'une relation de préférence modifie ces connaissances et préférences, ces degrés d'in-certitude sur les états, et de préférence sur les conséquences doivent être préservés. C'est uniquement par la manière d'agréger ces degrés que nous avons cherché à raffiner les modèles d'utilité qualitative possibiliste. De même, puisqu'il nous semblait que la commensurabilité entre degrés d'incertitude et de préférence est intéressante à conserver, les relations *plus décisives* devaient également respecter cette commensurabilité.

Nous avons exploré deux directions différentes pour la recherche de raffinements des critère de déci-sion possibiliste :

- Tout d’abord, nous avons recherché des raffinements *discri-monotones* [17], [46] des critères possibilistes, définis simplement en éliminant les états donnant des conséquences identiques de la comparaison de deux actes. Cette approche ne s’est pas révélée entièrement satisfaisante puisque, si elle permet de restaurer le principe de la chose certaine et de raffiner les relations de préférence possibilistes, la propriété de transitivité de la relation d’équivalence entre actes peut être perdue (voir annexe B.6).
- C’est pourquoi nous avons adopté ensuite une autre approche, cherchant à raffiner les critères de décision possibilistes par des critères basés sur l’utilité espérée [18], [19], [4] (voir annexe B.7). Cette approche s’est révélée plus fructueuse, puisque nous avons réussi à montrer qu’il était possible de raffiner les critères possibilistes par des critères d’utilité espérée et que les raffinements obtenus restaient qualitatifs (voir annexe B.8).

3.3.3 Travaux connexes

En suivant le même raisonnement que celui décrit dans ce chapitre, nous avons raffiné la relation d’ordre pessimiste $\succeq_{PES, \pi, \mu}$ par une relation d’ordre *leximinmax* portant sur les matrices dont les lignes sont des paires $(n(\pi(s)), \mu(f(s)))$. Nous avons montré que, là encore, le raffinement qualitatif obtenu représentait tous les raffinements probabilistes de type utilité espérée du critère pessimiste [4]. Plusieurs auteurs ont cherché des raffinements des utilités qualitatives possibilistes. (Dubois et al., 2000) ont proposé de départager les actes ex-aequo pour l’un des critères en utilisant l’autre critère. Néanmoins, cette approche ne permet pas de satisfaire à coup sûr le principe de la chose certaine, comme le montre l’exemple (3.3.1). De plus, le raffinement obtenu peut être biaisé.

(Giang and Shenoy, 2001, 2005) proposent de raffiner les critères de décision possibilistes en utilisant des degrés d’utilité “binaires”. Leur approche revient à construire l’échelle L à partir de l’ensemble des *paris sur les événements*², $x^\top Ax^\perp$. Le raffinement obtenu est non biaisé, mais ne supprime pas totalement l’effet de noyade. (Weng, 2005) a proposé ensuite d’utiliser conjointement les raffinements lexicographiques que nous avons proposés et l’approche binaire de Giang et Shenoy. Si il est clair que l’approche de (Weng, 2005) permet de raffiner les critères de (Giang and Shenoy, 2001), nous ne savons pas encore si elle permet de raffiner les critères que nous avons proposés, ni si elle admet une représentation en termes d’utilité espérée.

Dans [4], nous avons montré qu’il était également possible de proposer des raffinements de type (*utilité espérée / lexicographique*) pour le critère de décision OMEU (*Order of Magnitude Expected Utility*, (Giang and Shenoy, 2000; Pearl, 1993; Wilson, 1995)). Le critère OMEU est fondé sur une théorie de la décision basée sur les *fonctions kappa* (Spohn, 1990) et une forme d’agrégation ($\min, +$), là où l’utilité espérée est de la forme $(+, \times)$ et les utilités monotones de la forme (\max, \min) .

Plus récemment, (Dubois and Fargier, 2007) ont appliqué le même raisonnement afin de raffiner une relation d’ordre basée sur l’utilité qualitative monotone, ce qui conduit à des critères de décision basés sur l’*intégrale de Choquet*, qui ont été étudiés et axiomatisés par (Gilboa, 1987; Schmeidler, 1989; Sarin and Wakker, 1992).

²En fait, les auteurs se placent dans un cadre à la von Neumann et Morgenstern, et considèrent donc des *loteries* binaires en lieu et place d’actes binaires, mais leur approche peut être revisitée dans une optique à la Savage.

Chapitre 4

Décision séquentielle possibiliste

Ce chapitre recense mes travaux en décision séquentielle possibiliste. Mes principales références sur le sujet sont :

Algorithmes de programmation dynamique possibiliste

- [9] Sabbadin (R.), Fargier (H.) et Lang (J.). – Towards qualitative approaches to multi-stage decision making. *International Journal of Approximate Reasoning*, vol. 19, 1998, pp. 441–471.
- [7] Sabbadin (R.). – Possibilistic Markov Decision Processes. *Engineering Applications of Artificial Intelligence*, vol. 14, 2001, pp. 287–300.
- [48] Sabbadin (R.). – *Concepts et méthodes pour l'aide à la décision*, chap. Planification dans l'incertain et processus décisionnels de Markov, pp. 203–252. – Hermès Sciences, *Traité IC2, série Informatique et Systèmes d'Information*, volume 2. Eds. D. Bouyssou, D. Dubois, M. Pirlot et H. Prade, 2006.

problèmes de décision qualitative possibiliste factorisés

- [1] Dubois (D.), Le Berre (D.), Prade (H.) et Sabbadin (R.). – Using possibilistic logic for modeling qualitative decision : ATMS-based algorithms. *Fundamenta Informaticae*, vol. 37, 1999, pp. 1–30.
- [5] Garcia (L.) et Sabbadin (R.). – Complexity results and algorithms for possibilistic influence diagrams. *Artificial Intelligence*, vol. 172, 2008, pp. 1018-1044.

4.1 Introduction

Le second axe de ma recherche en décision qualitative possibiliste est celui de la décision séquentielle possibiliste. Plus précisément, mes travaux ont porté :

- Sur la programmation dynamique possibiliste et les PDM possibilistes, tout d'abord [7]. Ceux-ci constituaient un premier pas vers une approche qualitative de la planification dans l'incertain, basée sur la théorie des possibilités.
- Sur des approches structurées, basées sur la logique possibiliste [1] ou les diagrammes d'influence possibilistes [5], ensuite. Ceux-ci ont proposé des langages permettant de représenter de manière concise des problèmes de planification dans l'incertain possibiliste et des algorithmes de résolution associés.

Ces travaux, décrits dans ce chapitre, ont été illustrés sur des applications, en gestion de production agroalimentaire [8] ou en conservation de la biodiversité ([24], [25]).

4.2 Algorithmes de programmation dynamique possibiliste

Ayant défini une version qualitative de la décision dans l'incertain basée sur la théorie des possibilités, il était naturel d'essayer d'étendre cette théorie à la décision séquentielle. C'est ce que j'ai entrepris depuis la fin de ma thèse, d'abord pour la décision séquentielle en *horizon fini* avec Hélène Fargier et Jérôme Lang [9], puis en décision séquentielle en *horizon infini*, en proposant une contrepartie qualitative du cadre des *processus décisionnels de Markov*, complètement ou partiellement observables [7], [33].

Dans cette section, je vais décrire l'adaptation des principales méthodes de *programmation dynamique* au cadre possibiliste, d'abord dans le cas d'un horizon fini, puis infini.

4.2.1 Horizon fini

Dans [9], nous avons étendu la théorie de la décision qualitative possibiliste au cas séquentiel en horizon fini. Ce cadre est défini par un quintuplet $(\mathcal{X}, \mathcal{A}, \pi, \mu, T)$, où :

- \mathcal{X} représente l'ensemble des états atteignables à chaque instant,
- \mathcal{A} l'ensemble des actions (et non actes) applicables à chaque instant,
- T est l'horizon du problème, soit le nombre d'actions successives appliquées, ou de transitions observées,
- $\pi = \{\pi_t\}_{t=0\dots T-1}$, où $\pi_t(x_{t+1}|x_t, a_t)$ représente le degré de possibilité de la transition de x_t vers x_{t+1} lorsque l'action a_t est appliquée à l'instant $t \in \{0, \dots, T-1\}$,
- $\mu = \{\mu_t\}_{t=0\dots T-1}$, où $\mu_t(x_t, a_t, x_{t+1})$ représente le degré de satisfaction attaché à la transition observée de x_t vers x_{t+1} lorsque l'action a_t est appliquée à l'instant $t \in \{0, \dots, T-1\}$. Un degré de satisfaction attaché à l'état terminal, $\mu_T(x_T), x_T \in \mathcal{X}$, est également défini.

Dans ce cadre, l'utilité d'une politique $\delta : \mathcal{X} \times \{0, \dots, T-1\} \rightarrow \mathcal{A}$ dans un état initial x_0 est définie par un critère possibiliste (optimiste ou pessimiste), appliqué aux *trajectoires* possibles :

Définition 4.2.1 (Utilité possibiliste d'une politique) *Dans un cadre de décision séquentielle possibiliste dans l'incertain, $(\mathcal{X}, \mathcal{A}, \pi, \mu, T)$, l'utilité d'une politique $\delta = \{\delta_t\}_{t=0\dots T-1}, \delta_t : \mathcal{X} \rightarrow \mathcal{A}$ est définie, dans les cas optimiste et pessimiste, par :*

$$\begin{aligned} U_{OPT,\delta}(x_0) &= \max_{\tau} \min \left\{ \pi(\tau|x_0, \delta), \mu(\tau, \delta) \right\} \\ U_{PES,\delta}(x_0) &= \min_{\tau} \max \left\{ \pi(\tau|x_0, \delta), \mu(\tau, \delta) \right\} \end{aligned}$$

où $\tau = \{x_0, \dots, x_T\}$ est une trajectoire quelconque et

$$\begin{aligned} \pi(\tau|x_0, \delta) &= \min_{t \in 0\dots T-1} \left\{ \pi_t(x_{t+1}|x_t, \delta_t(x_t)) \right\}, \\ \mu(\tau) &= \min \left\{ \min_{t \in 0\dots T-1} \left\{ \mu_t(x_t, \delta_t(x_t), x_{t+1}) \right\}, \mu_T(x_T) \right\}. \end{aligned}$$

La décomposabilité du degré de possibilité conditionnel $\pi(\tau|x_0, \delta)$ résulte de propriétés classiques de la théorie des possibilités et d'une hypothèse "markovienne" similaire à celle du cas stochastique.

L'hypothèse de décomposabilité de $\mu(\tau)$ en un minimum de degrés de satisfaction de transitions élémentaires est un choix arbitraire. Ce choix sous-entend qu'une suite de transitions est aussi favorable que son degré le moins favorable. L'opérateur d'aggrégation max pourrait être utilisé à la place de l'opérateur min, ce qui donnerait lieu à des algorithmes de résolution similaires à ceux que nous avons développés.

Dans [9], nous avons montré qu'il était possible de calculer une politique δ^* optimisant l'un de ces deux critères possibilistes en appliquant un algorithme de type *recherche arrière*, similaire à celui développé pour les PDM classiques (voir annexe B.9).

4.2.2 Horizon infini

J'ai ensuite étendu le cadre de la décision séquentielle possibiliste aux problèmes à horizon infini [7]. On définit un *Processus Décisionnel de Markov possibiliste* (Π -PDM) par un quadruplet $(\mathcal{X}, \mathcal{A}, \pi, \mu)$, où π et μ sont maintenant *stationnaires* c'est-à-dire indépendantes du temps t . On définit également la notion d'utilité possibiliste d'une politique stationnaire $\delta : \mathcal{X} \rightarrow \mathcal{A}$ en horizon infini, de la manière suivante :

Définition 4.2.2 (Utilité possibiliste d'une politique stationnaire en horizon infini)

Soit $(\mathcal{X}, \mathcal{A}, \pi, \mu)$ un Π -PDM et $\delta : \mathcal{X} \rightarrow \mathcal{A}$ une politique stationnaire.

$$\begin{aligned} U_{OPT,\delta}(x_0) &= \lim_{T \rightarrow \infty} \max_{\tau_T} \min \left\{ \pi(\tau_T | x_0, \delta), \mu(\tau_T, \delta) \right\} \\ U_{PES,\delta}(x_0) &= \lim_{T \rightarrow \infty} \min_{\tau_T} \max \left\{ n(\pi(\tau_T | x_0, \delta)), \mu(\tau_T, \delta) \right\} \end{aligned}$$

où $\tau_T = \{x_0, \dots, x_T\}$ est une trajectoire quelconque de longueur $T + 1$ et

$$\begin{aligned} \mu(\tau_T, \delta) &= \min_{t \in 0 \dots T-1} \left\{ \mu(x_t, \delta_t(x_t), x_{t+1}) \right\}, \\ \pi(\tau_T | x_0, \delta) &= \min_{t \in 0 \dots T-1} \left\{ \pi(x_{t+1} | x_t, \delta_t(x_t)) \right\}. \end{aligned}$$

Pour ces problèmes de décision séquentielle en horizon infini, nous avons défini des contreparties possibilistes des algorithmes *d'itération de la politique* et *d'itération de la valeur* [7] (voir annexe B.9).

4.2.3 Décision en environnement partiellement observable

Dans [33], [34], j'ai étendu le cadre des PDM possibilistes au cas partiellement observable, sur la base de la théorie possibiliste du conditionnement (Dubois and Prade, 1994). Comme dans le cas stochastique (Cassandra et al., 1994), un PDM possibiliste *partiellement observable* (Π -PDMPO) peut être transformé en un PDM possibiliste *complètement observable*, dont les états sont des *états de croyance possibilistes*, soit des distributions de possibilité $\beta : \mathcal{X} \rightarrow L$. La principale différence avec le cas stochastique est que l'ensemble des états de croyance possibilistes est fini (mais de taille exponentielle en $|\mathcal{X}|$). Les algorithmes de résolution des PDM possibilistes décrits dans la section précédente restent donc applicables, au moins en théorie, aux PDM possibilistes partiellement observables.

Comme dans le cas des PDMPO classiques (Cassandra et al., 1994), on définit un PDMPO possibiliste en ajoutant à un modèle de PDM possibiliste un ensemble \mathcal{O} d'observations possibles, et une *fonction*

d'observation possibiliste, $\Omega : \mathcal{X} \times \mathcal{A} \times \mathcal{O} \rightarrow L$, telle que :

$$\Omega(x', a, o) = \pi(o|x', a),$$

la possibilité d'observer o lorsque l'action a a été appliquée et a débouché sur l'état x' .

Un état de croyance possibiliste, β , est une distribution de possibilité sur l'espace d'états \mathcal{X} . Contrairement au cas stochastique, l'ensemble des états de croyance possibilistes est fini dès lors que l'échelle L utilisée pour préciser les degrés de possibilité est finie. Le cardinal de B , l'ensemble des états de croyance possibilistes est majoré par $|L|^{|\mathcal{X}|}$. Supposons maintenant, comme dans le cas probabiliste, que les possibilités de transition $\pi(x'|x, a)$ sont données, de même que les possibilités des observations, $\pi(o|x', a) = \Omega(x', a, o)$. Alors, comme dans (Cassandra et al., 1994), on peut définir $\beta_a(x')$, la possibilité d'atteindre x' en partant d'une connaissance sur l'état initial définie par β , et en appliquant l'action a :

$$\beta_a(x') = \max_{x \in \mathcal{X}} \min\{\pi(x'|x, a), \beta(x)\}. \quad (4.1)$$

On calcule ensuite la possibilité d'observer $o \in \mathcal{O}$ après avoir appliqué a en β :

$$\beta_a(o) = \max_{x' \in \mathcal{X}} \min\{\pi(o|x', a), \beta_a(x')\}. \quad (4.2)$$

β_a^o est l'état de croyance possibiliste révisé, après avoir appliqué a en β et observé o :

$$\begin{aligned} \beta_a^o(x') &= 0_L \text{ si } \pi(o|x', a) = 0_L, \\ \beta_a^o(x') &= 1_L \text{ si } \pi(o|x', a) = \beta_a(o) > 0_L, \\ \beta_a^o(x') &= \beta_a(x') \text{ dans les autres cas.} \end{aligned} \quad (4.3)$$

Tous les éléments du nouveau PDM possibiliste sur l'espace des états de croyance sont définis dans les équations (4.1), (4.2) et (4.3). Intuitivement, l'évolution du système se définit par : si le système est dans l'état β , alors appliquer l'action a peut mener dans l'un des $|\mathcal{O}|$ états successeurs possibles β_a^o , la possibilité de rejoindre l'état β_a^o étant $\beta_a(o)$.

A partir de là, les équations de Bellman possibilistes peuvent être étendues au cas partiellement observable (ici dans le cas pessimiste)¹ :

$$U_{PES}^t(\beta) = \max_{a \in \mathcal{A}} \min\{\mu(\beta), \min_{o \in \mathcal{O}} \max\{n(\beta_a(o)), U_{PES}^{t+1}(\beta_a^o)\}\}. \quad (4.4)$$

où $\mu(\beta) = \min_{x \in \mathcal{X}} \max\{n(\beta(x)), \mu(x)\}$ et $U_{PES}^0(\beta)$ est initialisé à $\mu(\beta)$.

4.2.4 Autres travaux sur les PDM possibilistes

L'expression de la vraisemblance des transitions entre états en termes de degrés de possibilité est intuitive et permet de représenter avec moins de biais que des probabilités numériques la connaissance d'experts. Cependant, il n'existe pas de moyen formel universellement reconnu de *simuler* des transitions possibilistes ou de construire les degrés de possibilité de transition à partir d'un ensemble de transitions observées ou simulées.

Or il existe une communauté de recherche, sur le sujet de *l'apprentissage par renforcement*, dont l'objectif est de développer des algorithmes de résolution de Processus Décisionnels de Markov dont

¹Par souci de simplification des notations, on se limite ici à une fonction d'utilité μ sur les états, et non sur les transitions. Évidemment, l'équation (4.4) peut être étendue pour prendre en compte des préférences sur les transitions.

le modèle (p, r) n'est pas spécifié directement, mais accessible à partir de transitions $\langle x, a, x', r \rangle$ observées ou simulées. J'ai essayé, en m'inspirant de ces travaux, de développer des algorithmes de résolution de PDM possibilistes. Une première façon d'aborder ce problème a été de mesurer *empiriquement* la différence entre les solutions optimales de PDM classiques et possibilistes, *qualitativement semblables*².

Dans [35], j'ai comparé empiriquement les politiques optimales de PDM stochastiques construits à partir de PDM possibilistes pour une certaine classe de problèmes *types* ou vice-versa. Dans cette comparaison, les degrés de possibilité étaient transformés en degrés de probabilités en utilisant soit une approche de type *probabilités infinitésimales* (Darwiche and Goldszmidt, 1994; Henrion et al., 1999), soit de type *transformation cohérente* (Giang and Shenoy, 1999; Dubois et al., 1993). Le principal résultat de la comparaison empirique a été de montrer qu'il était possible *d'approcher* la solution d'un PDM classique par celle d'un PDM possibiliste dont la résolution est plus rapide, du fait des opérations impliquées.

Dans [36], j'ai utilisé les mêmes transformations afin de construire un estimateur de possibilités de transition à partir d'un simulateur de transitions probabilistes. J'ai défini une méthode de type *équivalent certain*, permettant d'apprendre un modèle $(\hat{\pi}, \hat{\mu})$ de PDM possibiliste à partir d'un ensemble $\{(x, a, x')\}$ de transitions observées ou simulées. Une politique approchée était ensuite calculée en résolvant le PDM possibiliste appris. Cette méthode naïve est très coûteuse en temps de calcul, puisqu'elle nécessite une exploration exhaustive de l'ensemble des transitions (x, a, x') . J'ai donc proposé un algorithme d'exploration guidée, de type *prioritized sweeping*, couplé à un algorithme d'itération de la politique possibiliste, afin de limiter l'effort calculatoire. Cet algorithme est efficace, à la fois en temps de calcul et en terme de qualité des politiques retournées. Cependant, du fait de l'idempotence des opérateurs *min* et *max* impliqués dans la définition des utilités qualitatives possibilistes, la convergence vers une politique optimale pour le PDM possibiliste *transformé* du PDM stochastique simulé, n'est pas assurée. Ces travaux devraient maintenant être revisités à la lumière des résultats exposés dans la section (3.3), afin de proposer des algorithmes d'apprentissage par renforcement réellement optimaux, en mettant à profit l'existence de raffinements de type "utilité espérée" des critères possibilistes. Cela devrait permettre de définir des contreparties possibilistes des algorithmes *Q-learning*, *TD*- λ ,...

4.3 Décision possibiliste structurée

En *planification dans l'incertain*, l'état du monde est généralement représenté par une combinaison de valeurs prises par un ensemble de variables d'états et n'est pas représenté *en extension*, comme nous l'avons supposé dans la section précédente. C'est pourquoi je me suis intéressé depuis ma thèse à l'élaboration de langages de représentation structurée et d'algorithmes de résolution associés, pour les problèmes de décision qualitative possibiliste ([1], [47], [42], [10]).

Dans un premier temps, je me suis appuyé sur des approches logiques pour représenter des problèmes de décision possibiliste *non-séquentiels*, puisque la logique possibiliste était déjà connue pour être un outil efficace de représentation compacte de distributions de possibilités (Lang, 1991). Plus tard, en collaboration avec Laurent Garcia ([23], [5], [11]), j'ai proposé une approche plus générale de la décision possibiliste structurée. Cette approche, basée sur une contrepartie possibiliste des *diagrammes d'influence* possède le double avantage, par rapport à l'approche logique, de permettre la représentation de problèmes de décision à partir de variables multivaluées et d'être applicable en décision

²Les ordres de vraisemblance des transitions et de préférence sur ces transitions étant identiques dans les deux représentations, stochastique et possibiliste.

séquentielle à horizon fini.

4.3.1 Définitions générales

En décision dans l'incertain, l'état du monde est généralement représenté par une combinaison $x = \{x_1, \dots, x_n\}$ de valeurs prises par un ensemble de variables d'états $\mathcal{X} = \{X_1, \dots, X_n\}$ ³. De la même manière, une décision peut être représentée par une combinaison $d = \{d_1, \dots, d_p\}$, choisie par un décideur, de valeurs des variables de décision $\mathcal{D} = \{D_1, \dots, D_p\}$, distinctes de \mathcal{X} .

Les connaissances incertaines du décideur sont représentées par des degrés de possibilité conditionnelle $\pi(x'|x, d)$ ou des distributions de possibilité a priori, $\pi(x)$. Ses préférences sont exprimées sur des couples (x, d) et représentées par une fonction d'utilité ordinaire $\mu : \mathcal{X} \times \mathcal{D} \rightarrow L$. Ainsi, $\mu(x, d)$ exprime à quel point l'instanciation d des variables de décision est satisfaisante lorsque l'état du monde est x .

En règle générale, l'évolution de la valeur de chaque variable d'état X_i ne dépend que de l'état d'un sous-ensemble de variables $Par(X_i) \subseteq \mathcal{X} \cup \mathcal{D}$ et nous connaissons une expression du degré de possibilité local $\pi_{X_i}(x'_i | x_{Par(X_i)}, d_{Par(X_i)})$ où $x_{Par(X_i)}$ et $d_{Par(X_i)}$ sont respectivement les projections de x et d sur $Par(X_i)$. A partir des règles classiques du conditionnement en théorie des possibilités qualitatives (Dubois and Prade, 1988), la possibilité conditionnelle $\pi(x'|x, d)$ peut être obtenue par la formule :

$$\pi(x'|x, d) = \min_{i=1 \dots n} \{\pi_{X_i}(x'_i | x_{Par(X_i)}, d_{Par(X_i)})\}. \quad (4.5)$$

De manière similaire, on peut toujours faire l'hypothèse que l'utilité qualitative globale $\mu(x, d)$ d'un couple (x, d) peut s'exprimer par combinaison de fonctions d'utilité locales, $\mu_k(x_{Par(V_k)}, d_{Par(V_k)})$, où $Par(V_k) \subseteq \mathcal{X} \cup \mathcal{D}$ représente l'ensemble des variables impliquées dans la fonction d'utilité locale μ_k . Par analogie avec l'agrégation des fonctions d'utilité dans les problèmes de satisfaction de contraintes flexibles (Dubois et al., 1996), il est également naturel d'agréger les degrés d'utilité locaux par un opérateur min :

$$\mu(x, d) = \min_{k=1 \dots q} \{\mu_k(x_{Par(V_k)}, d_{Par(V_k)})\}. \quad (4.6)$$

Considérer les degrés de possibilité et les utilités comme des agrégations de fonctions locales permet de les exprimer de manière concise en révélant des indépendances dans les possibilités de transition et dans les degrés de préférence locaux. Il est commode d'exprimer graphiquement ces dépendances locales, comme le montre l'exemple de la figure (4.1).

Exemple 4.3.1 *Le problème est de faire une omelette à six œufs à partir d'une omelette à cinq œufs, le nouvel œuf pouvant être soit frais soit pourri.*

Il y a deux variables binaires de décision :

- *CAB : casser l'œuf à part dans un bol (CAB = oui si l'œuf est cassé à part dans un bol pour être inspecté et CAB = non sinon),*
- *IDO : incorporer l'œuf dans l'omelette (IDO = oui ou non).*

Les variables d'état sont les suivantes :

³ X_i représente à la fois le nom de la variable et son domaine, la signification sera claire en fonction du contexte. De la même manière, \mathcal{X}_A , avec $A \subseteq \{1, \dots, n\}$ représente à la fois un ensemble de variables $\{X_i\}_{i \in A}$ et le produit cartésien de leurs domaines $\bigotimes_{i \in A} X_i$.

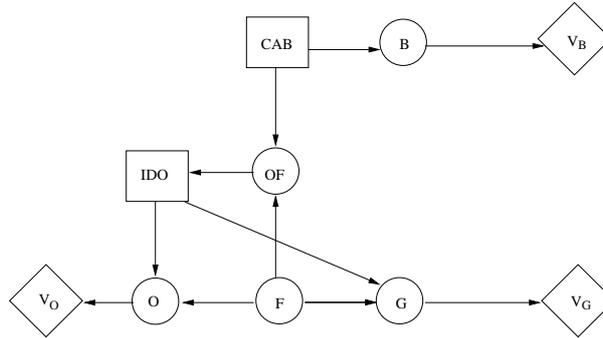


FIG. 4.1 – Exemple d'une omelette

- B : nous avons un bol à laver ($B = \text{vrai}$) ou non ($B = \text{faux}$).
- OF : nous observons que l'œuf est frais ($OF = v$) ou pourri ($OF = f$) ou nous n'observons pas la fraîcheur de l'œuf ($OF = \text{inconnu}$).
- F : l'état « réel » de l'œuf, frais ($F = v$) ou pourri ($F = f$).
- O : nous avons une omelette à six œufs ($O = 6O$), une omelette à cinq œufs ($O = 5O$) ou pas d'omelette du tout ($O = nO$).
- G : un œuf est gâché ($G = v$) ou pas ($G = f$).

On peut remarquer qu'ici, par exemple, le degré de possibilité de la variable OF n'est influencé que par l'état des variables CAB et F . Notre satisfaction globale sera fonction de degrés de satisfaction locaux associés respectivement aux variables O , B et G . Enfin, dans cet exemple, nous avons représenté un arc « informationnel » de OF vers IDO , dont nous expliquerons la signification dans la section (4.3.3).

4.3.2 Une approche basée sur la logique possibiliste

Dans un premier temps ([1], [47], [42], [10]), nous avons établi un lien entre problèmes de décision possibiliste structurée (non-séquentielle) et problèmes de raisonnement en logique possibiliste. Pour établir ce lien, nous considérons un ensemble de variables d'état booléennes, $\mathcal{X} = \{X_1, \dots, X_n\}$, un autre ensemble de variables booléennes conséquences, $\mathcal{Y} = \{Y_1, \dots, Y_m\}$ et un dernier ensemble de variables de décision booléennes, $\mathcal{D} = \{D_1, \dots, D_p\}$.

Supposons tout d'abord que les connaissances du décideur sur l'état x du monde et sur l'effet y d'une décision d en x ne sont pas graduelles. L'état x n'est pas forcément spécifié complètement, de même que la conséquence y résultant de l'application de d en x . Mais tous les états possibles x et toutes les conséquences possibles de d en x ont le même degré de plausibilité. Ce type de connaissances peut être modélisé par un ensemble de clauses $K = \{\varphi_i : A_i \rightarrow B_i\}$ où A_i est une conjonction de littéraux appartenant à $\mathcal{X} \cup \mathcal{D}$ et B_i est une disjonction de littéraux appartenant à $\mathcal{X} \cup \mathcal{Y}$. Les formules φ_i expriment donc :

- soit des connaissances sur l'état du monde (lorsque seules des variables de \mathcal{X} sont impliquées dans φ_i),
- soit des connaissances sur l'effet des actions (lorsque des variables de décision et de conséquence sont impliquées).

Les préférences du décideur sont elles aussi exprimées dans un premier temps sous la forme de buts,

via un ensemble de clauses $P = \{\psi_j\}$ portant uniquement sur les variables de \mathcal{Y} ⁴.

Exemple 4.3.2 Prenons un exemple simplifié inspiré de l'exemple (4.3.1), qui ne prend en compte que le sous-ensemble de variables propositionnelles $\{IDO, F, O, G\}$, dont les dépendances sont représentées dans la figure (4.2). La seule variable d'état est ici l'état de fraîcheur de l'œuf, F , la seule variable d'action est IDO et les variables conséquences sont O et G . Nous supposons que O est une variable binaire, prenant la valeur vrai si on obtient une omelette (à cinq ou six œufs) et faux si l'omelette est gâchée. G indique si un œuf est gâché. Soit, $\mathcal{X} = \{F\}$, $\mathcal{Y} = \{O, G\}$ et $\mathcal{D} = \{IDO\}$. Les connaissances sont les suivantes :

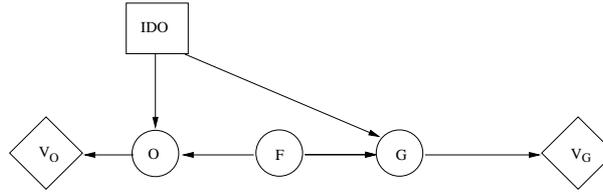


FIG. 4.2 – Exemple simplifié

- Une omelette est obtenue dès lors que l'œuf est frais ou qu'il n'est pas incorporé. Par contre, nous n'obtenons pas d'omelette si l'œuf incorporé n'est pas frais.
- Un œuf est gâché s'il est frais et non incorporé. Un œuf incorporé ou pas frais n'est pas gâché.

Ces connaissances sont modélisées par la base :

$$K = \{F \rightarrow O, \neg IDO \rightarrow O, \neg F \wedge IDO \rightarrow \neg O, IDO \rightarrow \neg G, \neg F \rightarrow \neg G, F \wedge \neg IDO \rightarrow G\}.$$

Les préférences portent sur les variables G et O et, bien évidemment, $P = \{O, \neg G\}$: on souhaite avoir une omelette et ne pas gâcher d'œuf.

Nous avons établi dans [1] le lien entre les critères de décision possibiliste pessimiste et optimiste et des problèmes de raisonnement “abductif”, ou sur la cohérence de bases propositionnelles stratifiées (voir annexe B.10).

Limites de l'approche logique de la décision qualitative possibiliste

L'approche logique de la décision qualitative possibiliste permet une première forme de représentation structurée. Les problèmes ainsi modélisés peuvent être résolus par des méthodes classiquement utilisées pour résoudre le problème **SAT**⁵. Nous avons décrit [1] des algorithmes basés sur un solveur de problème **SAT** utilisant une méthode de résolution de type *branch and bound* (algorithme de type *Davis et Putnam*) pour calculer les utilités possibilistes d'une décision.

Mais cette approche a deux limites :

- Pouvoir d'expression.

Un problème de décision structurée, tel que celui de la section (4.3.1) est exprimé via des possibilités de transition locales $\pi_{X_i}(x'_i | x_{Par(X_i)}, d_{Par(X_i)})$ et des utilités locales $\mu_k(x_{Par(V_k)}, d_{Par(V_k)})$

⁴De manière plus générale, les préférences ψ_j peuvent également porter sur les variables de décision, permettant ainsi d'exprimer des préférences sur des couples (conséquence, action). Mais dans le cadre de la décision possibiliste non-séquentielle que nous considérons ici, on peut toujours transformer ces problèmes de telle manière que la fonction d'utilité porte uniquement sur les conséquences.

⁵Satisfaction de formules booléennes

portant sur des variables qui ne sont pas forcément binaires. La représentation en logique possibiliste (par des formules exprimant connaissances et préférences) de ces possibilités de transition et fonctions d'utilité est certes possible et automatisable, mais elle peut être coûteuse en espace et en temps de calcul. La taille du modèle logique correspondant peut être exponentielle en celle du modèle initial, dans le pire des cas. De plus, la représentation obtenue est souvent moins intuitive que la représentation initiale.

– Décision séquentielle.

On recherche dans cette approche logique, une affectation *une fois pour toutes* des variables de décision. Or, dans un problème de décision séquentielle dans l'incertain, les variables de décision peuvent être instanciées séquentiellement et de nouvelles informations peuvent être intégrées avant l'instanciation d'une variable de décision. Nous verrons dans la section suivante, que cette prise en compte de la séquentialité d'un problème de décision rend celui-ci beaucoup plus difficile à résoudre. Si l'on voulait appliquer une approche logique, on se placerait dans le cadre **QBF**⁶ plutôt que dans le cadre **SAT**.

Pour ces deux raisons, nous avons décidé d'étudier une autre approche de la décision possibiliste structurée, basée sur une contrepartie possibiliste des *diagrammes d'influence* ([23], [5], [11]). Cette approche, que je vais décrire dans la section suivante, permet de dépasser les deux limites de l'approche logique que je viens de souligner. Elle permet à la fois :

- de représenter des problèmes de décision impliquant des variables multivaluées et,
- de modéliser directement des problèmes de décision séquentielle à horizon fini.

4.3.3 Diagrammes d'influence possibilistes

Nous avons défini [23] la notion de *Diagramme d'Influence Possibiliste* (DIP) afin de traiter des problèmes de décision possibiliste structurée *séquentiels*. Les dépendances entre variables se représentent graphiquement dans un DIP par un *Graphe Orienté Acyclique* (GOA), comme celui de la figure 4.1. A la différence du cas non-séquentiel décrit dans la section précédente, les nœuds représentant les décisions admettent des arcs "entrants". Les arcs entrant modélisent le fait que la valeur de la variable parent d'une variable de décision est connue au moment du choix de la décision.

Un DIP \mathcal{P} est également représenté par un quintuplet $(\mathcal{X}, \mathcal{D}, \mathcal{V}, \Phi, \Psi)$ associé à la représentation graphique par le graphe orienté acyclique :

- les **nœuds d'état** représentent les variables d'état $X_i \in \mathcal{X} = \{X_1, \dots, X_n\}$, comme dans le formalisme des *réseaux bayésiens* (RB) (Pearl, 1988b) ;
- les **nœuds de décision** représentent les variables de décision $D_j \in \mathcal{D} = \{D_1, \dots, D_p\}$;
- les **nœuds d'utilité** $\mathcal{V} = \{V_1, \dots, V_q\}$ représentent des fonctions d'utilité locales $\mu_k \in \{\mu_1, \dots, \mu_q\}$.

En complément de la partie graphique qui est identique à celle des diagrammes d'influence classiques, les DIP comprennent aussi des spécifications numériques. Dans le formalisme possibiliste, cet ordre numérique est traité de manière ordinale. Une table de possibilité conditionnelle $\pi_{X_i}(X_i | Par(X_i))$ est associée à chaque variable d'état X_i . Si X_i est une racine du GOA ($Par(X_i) = \emptyset$), on spécifie les degrés de possibilité a priori $\pi_{X_i}(x_i)$ associés à chaque valeur x_i de X_i . $\Phi = \{\pi_{X_i}\}_{X_i \in \mathcal{X}}$ est l'ensemble des tables de possibilité conditionnelles. De la même manière, pour chaque nœud d'utilité V_k , on spécifie les valeurs ordinales $\mu_k(x_{Par(V_k)}, d_{Par(V_k)})$ de chaque instanciation possible $(x_{Par(V_k)}, d_{Par(V_k)})$ des variables parents de V_k . $\Psi = \{\mu_k\}_{V_k \in \mathcal{V}}$ est l'ensemble des fonctions d'utilité

⁶Quantified Boolean Formulas.

locales.

A partir du graphe orienté acyclique, il est possible de définir un ordre partiel sur l'ensemble des variables d'état et de décision reflétant l'aspect séquentiel du problème de décision. Les valeurs des variables de décision ne sont plus forcément choisies d'emblée, mais les valeurs des variables d'état révélées auparavant peuvent influencer sur ce choix. La *politique* d'un décideur exprime alors la manière dont sont choisies les valeurs des variables de décision, en fonction des valeurs des variables d'état observées au moment du choix. Formellement, une politique est représentée par une fonction $\delta : \mathcal{X} \rightarrow \mathcal{D}$, associant une décision $d = \delta(x)$, à chaque état du monde possible. Les politiques peuvent aussi être définies par l'intermédiaire d'un ensemble de fonctions partielles $\{\delta_1, \dots, \delta_p\}$, $\delta_j : X_{j_1} \times \dots \times X_{j_{k_j}} \rightarrow D_j$ définissant la valeur prise par la variable de décision D_j en fonction des valeurs prises par un sous-ensemble des variables d'états (et pas forcément de toutes). Ces fonctions partielles permettent de rendre compte de l'aspect "séquentiel" du problème de décision : au moment du choix de d_1 , seules les valeurs de certaines variables d'état sont connues, puis de nouvelles variables d'état sont révélées et d_2 est choisie, etc. En d'autres termes, si l'ordre des variables de décision est considéré comme un ordre "temporel", et que l'on partitionne l'ensemble des variables d'état "en accord" avec cet ordre, la politique δ est "dynamique". Pour le DIP représenté graphiquement dans la figure (4.1), l'ordre partiel strict $\{CAB\} \prec \{OF\} \prec \{IDO\} \prec \{F, O, B, G\}$ est un exemple d'ordre partiel strict en accord avec le graphe, dit *ordre GOA-compatible*, reflétant par exemple un ordre temporel d'instanciation des variables.

Dans un problème exprimé sous la forme d'un DIP, les possibilités de transition $\pi(x'|x, d)$ et $\mu(x, d)$ sont décomposables et définies par les équations (4.5) et (4.6). Ainsi, le formalisme DIP, en plus de l'expression graphique des dépendances entre variables, spécifie également les tables de possibilité de transition et d'utilité qualitative locales.

Exemple 4.3.3 Dans l'exemple de la figure (4.1), on choisit $L=\{0, \dots, 5\}$. Les possibilités conditionnelles sont définies en accord avec la signification naturelle du problème. Les seules possibilités qui ne peuvent pas être déterminées directement en fonction de la signification intuitive du problème sont les possibilités a priori de F . On suppose que $\pi_F(F = v) = 5$ et $\pi_F(F = f) = 3$ (ce qui signifie que l'on s'attend à ce que l'œuf soit frais plutôt que pourri). On obtient ainsi les tables de possibilités conditionnelles et a priori suivantes :

| $\pi_B(B CAB)$ | o | n | $\pi_G(G F, IDO)$ | (v, n) | (v, o) | (f, \cdot) |
|----------------|-----|-----|-------------------|----------|----------|--------------|
| v | 5 | 0 | v | 5 | 0 | 0 |
| f | 0 | 5 | f | 0 | 5 | 5 |

| $\pi_{OF}(OF CAB, F)$ | (o, v) | (o, f) | (n, \cdot) |
|-----------------------|----------|----------|--------------|
| v | 5 | 0 | 0 |
| f | 0 | 5 | 0 |
| i | 0 | 0 | 5 |

| $\pi_O(O F, IDO)$ | (\cdot, n) | (v, o) | (f, o) | $\pi_F(F)$ | a priori |
|-------------------|--------------|----------|----------|------------|----------|
| 6O | 0 | 5 | 0 | v | 5 |
| 5O | 5 | 0 | 0 | f | 3 |
| nO | 0 | 0 | 5 | | |

Les fonctions d'utilité μ_O, μ_G et μ_B sont définies comme suit et pour chaque instanciation globale (x, d) , $\mu(x, d) = \min\{\mu_O(o), \mu_G(g), \mu_B(b)\}$.

| O | $\mu_O(O)$ |
|------|------------|
| $6O$ | 5 |
| $5O$ | 3 |
| nO | 0 |

| G | $\mu_G(G)$ |
|-----|------------|
| f | 5 |
| v | 2 |

| B | $\mu_B(B)$ |
|-----|------------|
| f | 5 |
| v | 4 |

A partir de la spécification :

- du graphe orienté acyclique,
- d'un ordre sur les variables *compatible*,
- des tables de possibilité de transition et d'utilité locales,

il est possible de définir l'utilité (optimiste ou pessimiste) d'une politique δ fixée, spécifiant la valeur de chaque variable de décision en fonction des valeurs des variables d'état qui la précèdent. On peut montrer qu'une politique δ définit, à l'aide des possibilités de transition $\pi_{X_i}(x_i|x_{Par(X_i)}, d_{Par(X_i)}^x)$, une distribution de possibilité sur \mathcal{X} :

$$\pi_\delta(x) = \pi(x|\delta(x)) = \min_{i=1..n} \pi_{X_i}(x_i|x_{Par(X_i)}, d_{Par(X_i)}^x), \quad (4.7)$$

où $d^x = \delta(x)$ est la valeur de toutes les variables de décision, définie de manière unique à partir de δ et x .

De la même manière, on peut calculer un degré de satisfaction global, $\mu_\delta(x)$, induit par une politique sur les instanciations des variables d'état :

$$\mu_\delta(x) = \mu(x, \delta(x)) = \min_{k=1..q} \mu_{V_k}(x_{Par(V_k)}, d_{Par(V_k)}^x). \quad (4.8)$$

Ainsi, $\pi_\delta(x)$ et $\mu_\delta(x)$ dépendent seulement de la valeur des variables d'état. Il est alors possible de calculer les utilités possibilistes optimiste et pessimiste d'une politique δ :

$$U_{OPT,\pi,\mu}(\delta) = \max_{x \in \mathcal{X}} \min \left\{ \pi_\delta(x), \mu_\delta(x) \right\}, \quad (4.9)$$

$$U_{PES,\pi,\mu}(\delta) = \min_{x \in \mathcal{X}} \max \left\{ n(\pi_\delta(x)), \mu_\delta(x) \right\}. \quad (4.10)$$

Résoudre un problème exprimé sous la forme d'un DIP revient donc à trouver une politique $\delta^* = \{\delta_1^*, \dots, \delta_p^*\}$ (où $\delta_j^* : \mathcal{X} \rightarrow \mathcal{D}_j$) optimisant $U_{OPT,\pi,\mu}$ ou $U_{PES,\pi,\mu}$.

4.3.4 Diagramme d'influence possibiliste et PDM possibiliste à horizon fini

Nous venons de donner un exemple de diagramme d'influence possibiliste. Remarquons que tout PDM possibiliste à horizon fini (voir section 4.2.1) peut également s'écrire sous la forme d'un DIP, ayant la structure suivante.

4.3.5 Complexité de la recherche de politiques optimales

Dans [5] et [11], nous nous sommes intéressés à la complexité des problèmes d'optimisation consistant à trouver des politiques maximisant $U_{OPT,\pi,\mu}$ ou $U_{PES,\pi,\mu}$. Nous avons en particulier montré les résultats suivants :

Proposition 4.3.1 (Optimisation de Politique Optimiste (OPO)) *Soit \mathcal{P} un DIP et $\alpha \in L$, le problème de décider si il existe une politique δ qui atteint le degré d'utilité α ($U_{OPT,\pi,\mu}(\delta) \geq \alpha$) est NP-complet.*

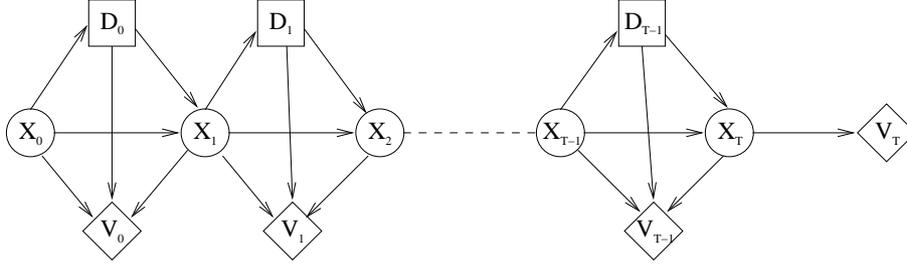


FIG. 4.3 – Représentation d'un PDM possibiliste sous forme de DIP

Proposition 4.3.2 (Optimisation de Politique Pessimiste (OPP)) Soit \mathcal{P} un DIP et $\alpha \in L$, le problème de décider si il existe une politique δ qui atteint le degré d'utilité α ($U_{PES,\pi,\mu}(\delta) \geq \alpha$) est PSPACE-complet.

La raison pour laquelle le problème OPO est moins difficile à résoudre que le problème OPP est que dans le calcul de l'utilité optimiste d'une politique, tout se passe comme si l'état du monde (inconnu) était toujours le plus favorable possible. En conséquence, la meilleure politique est "inconditionnelle" (δ_j^* n'est pas une fonction de \mathcal{X} vers \mathcal{D}_j mais simplement un élément d_j^* de \mathcal{D}_j). Alors, résoudre le problème OPO revient à "deviner" (x^*, d^*) (un couple formé de l'état du monde le plus favorable et de l'instanciation des variables de décision la plus favorable connaissant cet état) maximisant l'expression :

$$f(x, d) = \min \left\{ \pi(x|d), \mu(x, d) \right\}.$$

Dans le cas OPP, au contraire, tout se passe comme si nous alternions avec un adversaire, le choix des valeurs des variables de décision et d'état (suivant un ordre compatible avec le DIP), avec chacun un objectif opposé de maximisation ou de minimisation de l'expression :

$$g(x, d) = \max \left\{ n(\pi(x|d)), \mu(x, d) \right\}.$$

Dans [5] et [11], nous avons proposé des algorithmes de calcul de politiques (optimistes ou pessimistes) optimales. Ces algorithmes sont basés sur l'exploration d'arbres de recherche, ou sur des méthodes de type *élimination de la variable*, inspirées d'algorithmes de résolution de problèmes exprimés sous la forme de *diagrammes d'influence* ((Shachter, 1986), (Jensen, 2001)).

4.3.6 Travaux connexes

(da Costa Pereira et al., 1997a,b; da Costa, 1998) ont également proposé une approche de la planification dans l'incertain basée sur la théorie des possibilités. Celle-ci utilisait des distributions de possibilité pour représenter les effets des actions, mais se plaçait dans un cadre *non-observable* et avec des préférences non-graduelles sur des états buts. En conséquence, des plans *inconditionnels* maximisant le degré de possibilité ou de nécessité d'atteindre un état but étaient calculés. Les problèmes étaient modélisés dans un langage logique de type STRIPS (Fikes and Nilsson, 1971).

De même, (Pralet et al., 2006a) ont présenté un cadre général unifié pour représenter et résoudre de nombreux problèmes de décision structurée (depuis les CSP valués jusqu'aux diagrammes d'influence stochastiques et possibilistes). Ce cadre est plus général que celui proposé dans [5], mais n'est pas

spécifiquement dédié aux DIP. Certaines méthodes proposées dans (Pralet et al., 2006c,b) pourraient certainement aider à l'amélioration future de nos propres algorithmes d'élimination de variables.

4.4 Conclusion

Ce chapitre clôt la présentation de mes travaux autour de la décision qualitative possibiliste. Dans le dernier chapitre de ce manuscrit (chapitre 7), je présenterai l'ensemble de mes perspectives de recherche, et en particulier mes perspectives concernant la décision qualitative possibiliste. Mais avant cela, dans la partie suivante, je vais décrire mes travaux sur les PDM stochastiques, *factorisés* ou *épistémiques*.

Troisième partie

Processus décisionnels de Markov factorisés ou épistémiques

Chapitre 5

Extensions du cadre des processus décisionnels de Markov

Ce chapitre résume mes travaux dédiés à l’extension du cadre des processus décisionnels de Markov aux problèmes de décision spatialisée, puis au problème d’acquisition de connaissances, ou *planification épistémique*.

Processus décisionnels de Markov factorisés

- [38] Sabbadin (R.). – Graph partitioning techniques for Markov decision processes decomposition. *14th European Conf. on Artificial Intelligence (ECAI’02)*, pp. 670–674. – Lyon, France, 21-26 Juillet 2002.
- [21] Forsell (N.) et Sabbadin (R.). – Approximate linear-programming algorithms for graph-based Markov decision processes. *17th European Conference on Artificial Intelligence (ECAI’06)*, pp. 590–594. – Riva del Garda, Italy, Août 2006.
- [28] Peyrard (N.) et Sabbadin (R.). – Mean field approximation of the policy iteration algorithm for graph-based Markov decision processes. *17th European Conference on Artificial Intelligence (ECAI’06)*, pp. 595–599. – Riva del Garda, Italy, Août 2006.

Planification épistémique

- [41] Sabbadin (R.), Lang (J.) et Ravoanjanahry (N.). – Purely epistemic Markov decision processes. *22nd AAAI Conference on Artificial Intelligence (AAAI’07)*, pp. 1057–1062. – Vancouver, Canada, 2007.

5.1 Généralités

5.1.1 Problèmes de décision séquentielle dans l’incertain factorisés

Quel est le point commun entre des problèmes de sélection de réserves naturelles, de contrôle d’épidémies (végétales, animales ou humaines) ou de gestion forestière *durable*, prenant en compte la conservation de la biodiversité ou les risques naturels (tempêtes, incendies) auxquels elle fait face ? Ce sont tous des problèmes de décision séquentielle dans l’incertain qui partagent une même nature, *explicitement spatiale*, les rendant particulièrement complexes à représenter et à résoudre. Ce sont donc aussi forcément des problèmes intéressant un chercheur en intelligence artificielle travaillant à

l'INRA, dont l'objectif est de développer de nouvelles approches méthodologiques visant à mieux modéliser et résoudre des problèmes finalisés de gestion agri-environnementale. En ce qui me concerne, c'est en abordant de tels problèmes finalisés que je cherche à motiver et à construire, depuis plusieurs années, un cadre méthodologique basé sur les *processus décisionnels de Markov factorisés*, apte à les modéliser et les résoudre. Et ce sont ces problèmes finalisés, autant que les résultats théoriques obtenus, qui me permettent de juger de la pertinence des méthodes développées.

Contrairement aux approches de la partie précédente, les travaux présentés dans cette partie s'appuient sur un cadre probabiliste "classique" et sont de nature algorithmique et de modélisation. J'ai privilégié l'approche probabiliste pour l'étude des problèmes de décision séquentielle spatialisée en gestion environnementale. La raison est que dans ce domaine, les données nécessaires à la construction de mesures d'incertitude et de préférence quantitatives sont en général disponibles, contrairement à certains domaines de prédilection habituels de l'intelligence artificielle (élicitation de préférences ou de connaissances, aide à la décision d'experts humains...).

Néanmoins, certains problèmes de conservation d'espèces naturelles portant sur le long terme gagnent tout de même à être modélisés dans le cadre possibiliste [25], car le modèle correspondant est extrêmement incertain et les conséquences en termes de perte de biodiversité, difficilement mesurables par un critère additif.

Il est difficile de résoudre par les algorithmes classiques (itération de la valeur ou de la politique) des PDM factorisés dont les espaces d'états et d'actions sont de très grande taille. Un certain nombre d'approches de résolution ont toutefois été proposées ces dernières années. Parmi les approches proposées, on peut distinguer deux grandes catégories :

- les méthodes à base *d'agrégation d'états* et,
- les méthodes de type *décomposition des espaces d'états et de décisions : sérielle* dans le cas de PDM de grande taille faiblement couplés ou *parallèle* pour les PDM factorisés.

Le principe des méthodes de type "agrégation d'états" consiste à regrouper certains états et décisions par caractéristiques communes, et ainsi de réduire la taille des espaces d'états et de décisions du problème (Boutilier et al., 1995; Dean and Givan, 1997; Dearden and Boutilier, 1997). Souvent, ces méthodes sont couplées avec des représentations factorisées des probabilités de transition entre états ou des fonctions récompenses (Koller and Parr, 1999; Boutilier et al., 1999, 2000; Givan et al., 2003), dans des structures de type Réseaux Bayésiens (Pearl, 1988a; Dean and Kanazawa, 1989). Dans la même catégorie, on peut classer les approches de type "macro-décisions", agrégeant cette fois les décisions en politiques complexes (Koller and Parr, 2000; Hauskrecht et al., 1998).

Les PDM de grande taille peuvent parfois être décomposés en sous-problèmes relativement indépendants. Les méthodes de type "décomposition des espaces d'états et de décisions" visent à réduire la complexité du PDM étudié en le décomposant en sous-problèmes résolus séparément, les solutions élémentaires étant ensuite recombinaées. Ces méthodes peuvent elles mêmes être séparées en :

- Méthodes *sérielles*, lorsque l'espace d'états est décomposé en une union de sous-ensembles d'états distincts, faiblement couplés (pour toute politique, la probabilité de sortir d'un des sous-ensembles est faible) (Dean and Lin, 1995; Hauskrecht et al., 1998; Parr, 1998).
- Méthodes *parallèles*, lorsque l'espace d'états est décomposé en un produit cartésien de sous-espaces (Meuleau et al., 1998; Singh and Cohn, 1998).

Les méthodes de type Apprentissage par Renforcement, enfin, surtout lorsqu'elles sont "directes", sont également aptes à traiter des problèmes de plus grande taille que les méthodes classiques de

résolution de PDM. Elles sont basées sur la simulation des probabilités de transition du PDM et ne nécessitent pas leur stockage. Elles sont souvent conjuguées avec des méthodes d'agrégation d'états, ou de décomposition des espaces d'états et de décisions (Bertsekas and Tsitsiklis, 1996; Kearns et al., 1999; Precup et al., 1998) pour résoudre des PDM factorisés.

5.1.2 Planification épistémique

Une autre caractéristique commune des applications mentionnées plus haut, en plus de leur nature "spatiale", est qu'en général l'état du monde réel n'est pas complètement observable. Ce n'est que par une approximation de la réalité que nous pouvons les modéliser comme des processus décisionnels de Markov *complètement observables*. J'ai très brièvement décrit, dans le chapitre 2, le cadre des Processus Décisionnels de Markov Partiellement Observables (PDMPO) (Smallwood and Sondik, 1973). Attardons nous à nouveau quelques instants sur les PDMPO, et plus précisément sur la notion d'*incertitude* dans les problèmes de planification.

Cette incertitude a deux sources distinctes : (i) incertitude sur les effets des actions et (ii) incertitude sur l'état courant (observé) du monde. La première source d'incertitude est liée à la dynamique du système, tandis que la seconde est purement statique. Le modèle des processus décisionnels de Markov partiellement observables traite ces deux sources à la fois. Cependant, résoudre un PDMPO est bien plus difficile que résoudre un PDM (PSPACE-complet si l'horizon est fini et court¹, EXPSPACE si il est fini mais long et indécidable si l'horizon est infini (Papadimitriou and Tsitsiklis, 1987; Mundhenk et al., 2000; Madani et al., 2003)).

Si on néglige l'incertitude sur l'état courant du monde, le cadre obtenu est celui, bien connu, des *processus décisionnels de Markov totalement observables (PDM)*. La complexité du problème obtenu se réduit alors largement, puisqu'on se retrouve dans la classe polynomiale P (pour les problèmes non factorisés, bien entendu). D'un autre côté, le cas où la seule incertitude considérée pèse sur les observations, alors que l'état du monde est supposé "fixe" au cours du temps, a été moins étudié. De nombreux problèmes appartiennent pourtant à cette classe : l'*élicitation de préférences*, applicable par exemple en commerce électronique (Boutilier, 2002), le *diagnostic* (en médecine, science de l'environnement ou dans les systèmes généraux), l'*interrogation de bases de données* ou les *jeux* (Mastermind, bataille navale, etc.). De tels problèmes peuvent être formulés dans le cadre des *belief MDP* (Kaelbling et al., 1998), qui sont une forme de PDM à espace d'états continu et généralisent le cadre des PDMPO. Mais cette approche ne permet pas de bénéficier d'éventuels avantages (efficacité calculatoire) liés au fait que l'incertitude sur les effets des actions est négligée. Dans [30], [31], [41], nous avons étudié de tels processus décisionnels de Markov dits "purement épistémiques" (PDME).

5.1.3 Plan de la troisième partie

Dans ce chapitre, je décrirai les avancées méthodologiques autour des PDM factorisés et épistémiques auxquelles j'ai contribué :

- Dans une première section, je présenterai mes travaux autour des PDM factorisés. Je décrirai tout d'abord une méthode de décomposition sérielle pour la résolution de processus décisionnels de Markov *faiblement couplés*, que j'ai proposée dans [38]. Cette approche est basée sur une méthode de *bipartition spectrale*, développée par (Hendrickson and Leland, 1995). Ensuite, je décrirai le cadre des *processus décisionnels de Markov sur Graphe*, sur lequel nous avons travaillé, avec Nathalie Peyrard et Nicklas Forsell, depuis quelques années. Je présenterai les deux méthodes de

¹On parle d'horizon court lorsque celui-ci est borné par une fonction polynomiale de la taille du problème, et long si la fonction en question est exponentielle.

décomposition parallèle que nous avons proposées pour résoudre ce type de PDM factorisé [21], [28].

- Ensuite, je décrirai les travaux sur la *planification épistémique* [30], [31], [41], qui sont au centre de la thèse de Nasolo Ravoanjanahary que j’ai coencadré avec Jérôme Lang.

Dans le chapitre 6, je décrirai brièvement un ensemble de problèmes finalisés de décision séquentielle dans l’incertain *spatialisée*, pour la gestion agri-environnementale, sur lesquels j’ai travaillé (ou je travaille encore). Ces problèmes, illustrant la plupart des approches décrites dans ce chapitre, se classent en deux catégories d’applications :

- Gestion forestière spatialisée et ses interactions avec l’écologie ou la gestion des risques naturels.
- Epidémiologie spatialisée, dans les domaines végétal, animal et humain.

5.2 Processus décisionnels de Markov Factorisés

5.2.1 Une méthode de décomposition de PDM de grande taille faiblement couplés, basée sur la bipartition spectrale

J’ai proposé dans [38] une technique de décomposition sérielle de processus décisionnels de Markov inspirée des travaux de (Dean and Lin, 1995) et (Parr, 1998). (Dean and Lin, 1995; Parr, 1998) ont proposé une méthode de décomposition de la résolution d’un processus décisionnel de Markov $\langle \mathcal{X}, \mathcal{A}, p, r \rangle$ à partir d’une partition $\mathcal{P} = \{\mathcal{X}_1, \dots, \mathcal{X}_n\}$ de \mathcal{X} , fixée a priori. Plus précisément, les auteurs proposent de définir, à partir de \mathcal{P} et de la probabilité de transition p une *décomposition en étoile* du PDM :

Définition 5.2.1 (Décomposition en étoile)

Soit $\langle \mathcal{X}, \mathcal{A}, p, r \rangle$ un processus décisionnel de Markov et $\mathcal{P} = \{\mathcal{X}_1, \dots, \mathcal{X}_n\}$, une partition de \mathcal{X} .

La *décomposition en étoile* du PDM est une partition $\mathcal{P}' = \{U, \mathcal{Y}_1, \dots, \mathcal{Y}_n\}$ de \mathcal{X} définie par :

- $Per(\mathcal{X}_i) = \{x' \in \mathcal{X} - \mathcal{X}_i, \exists x \in \mathcal{X}_i, \exists a \in \mathcal{A}, p(x'|x, a) > 0\}$,
- $U = \bigcup_{i=1..n} Per(\mathcal{X}_i)$ et
- $\mathcal{Y}_i = \mathcal{X}_i \cap \overline{U}, \forall i = 1 \dots n$.

Intuitivement, la décomposition en étoile de \mathcal{X} à partir de la partition $\mathcal{P} = \{\mathcal{X}_1, \dots, \mathcal{X}_n\}$ est obtenue en séparant les éléments des différents sous-ensembles \mathcal{X}_i que l’on ne peut atteindre directement à partir d’autres sous-ensembles \mathcal{X}_j (les “branches” de l’étoile) des éléments situés “au cœur” de l’étoile, par lesquels il est obligatoire de passer pour rejoindre une autre branche (figure 5.1).

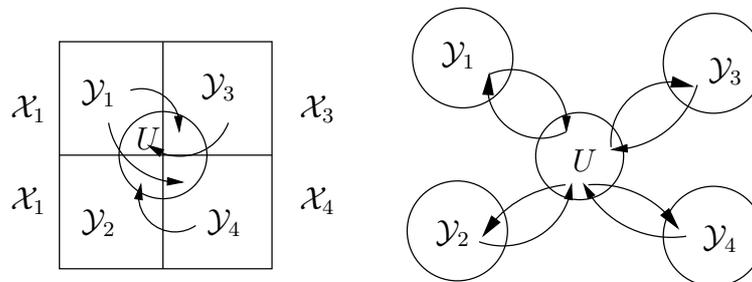


FIG. 5.1 – Décomposition en étoile. Aucune transition directe de \mathcal{Y}_i vers \mathcal{Y}_j ($i \neq j$) n’est possible.

A partir d'une décomposition en étoile $\mathcal{P}' = \{U, \mathcal{Y}_1, \dots, \mathcal{Y}_n\}$ d'un PDM $\langle \mathcal{X}, \mathcal{A}, p, r \rangle$ et d'une fonction réelle λ sur U , il est possible de définir un ensemble de PDM locaux $PDM_{\lambda}^i = \langle \mathcal{X}_i \cup Per(\mathcal{X}_i), \mathcal{A}, p_i, r_i \rangle$ de la manière suivante :

Définition 5.2.2 (PDM locaux obtenus à partir d'une décomposition en étoile)

Soit $\langle \mathcal{X}, \mathcal{A}, p, r \rangle$ un processus décisionnel de Markov et $\mathcal{P}' = \{U, \mathcal{Y}_1, \dots, \mathcal{Y}_n\}$ une décomposition en étoile de ce PDM. Soit λ , une fonction réelle sur U .

On définit pour tout $i = 1 \dots n$, un PDM local $PDM_{\lambda}^i = \langle \mathcal{X}_i \cup Per(\mathcal{X}_i), \mathcal{A}, p_i, r_i \rangle$ par :

- $\forall a \in \mathcal{A}, \forall x \in \mathcal{X}_i, \forall x' \in \mathcal{X}_i \cup Per(\mathcal{X}_i), p_i(x'|x, a) = p(x'|x, a)$,
- $\forall x \in Per(\mathcal{X}_i), p_i(x|x, a) = 1$,
- $\forall (x, x') \in \mathcal{X}_i^2, r_i(x, a, x') = r(x, a, x')$,
- $\forall (x, x') \in \mathcal{X}_i \times Per(\mathcal{X}_i), r_i(x, a, x') = r(x, a, x') + \gamma \cdot \lambda(x')$ et
- $\forall x \in Per(\mathcal{X}_i), r_i(x, a, x) = 0$.

On peut montrer (Dean and Lin, 1995; Parr, 1998) que si λ est la fonction de valeur optimale du PDM initial, restreinte à U , la politique obtenue en agrégeant les politiques solution des PDM locaux est globalement optimale.

Ce constat permet de définir un algorithme itératif de résolution exacte de PDM décomposé en étoile :

1. A partir de la partition $\mathcal{P} = \{\mathcal{X}_1, \dots, \mathcal{X}_n\}$, définir la décomposition $\mathcal{P}' = \{U, \mathcal{Y}_1, \dots, \mathcal{Y}_n\}$. Initialiser arbitrairement λ sur U .
2. Itérer les deux étapes suivantes tant que λ évolue significativement :
 - Résoudre tous les PDM locaux PDM_{λ}^i , en considérant la fonction λ courante. Les fonctions de valeur locales correspondantes sont $\{v_i^{\lambda} : \mathcal{Y}_i \rightarrow \mathbb{R}, i = 1 \dots n\}$.
 - Mettre à jour λ , en appliquant une itération de l'algorithme d'itération de la valeur restreinte au sous-ensemble d'états U .

Cet algorithme, appliquant une forme de *programmation dynamique asynchrone* (Bertsekas, 1987), converge nécessairement vers la solution optimale du PDM initial. Néanmoins, son efficacité dépend fortement de la décomposition en étoile $\mathcal{P}' = \{U, \mathcal{Y}_1, \dots, \mathcal{Y}_n\}$ résultant de la partition initiale $\mathcal{P} = \{\mathcal{X}_1, \dots, \mathcal{X}_n\}$: l'algorithme sera d'autant plus efficace que U est de petite taille et que les sous-ensembles \mathcal{Y}_i sont de tailles équilibrées.

Dans [38], j'ai proposé un algorithme permettant de trouver une partition \mathcal{P} donnant lieu à une décomposition efficace. Cet algorithme est basé sur une méthode de *partition spectrale* (Hendrickson and Leland, 1995) proposée pour la résolution du problème combinatoire (NP-complet) *min-cut graph partition*. L'analogie est la suivante, dans le cas où la partition initiale ne contient que deux éléments, \mathcal{X}_1 et \mathcal{X}_2 . Si on considère le graphe symétrique² $G = (V, E)$ construit à partir de la probabilité de transition p sur \mathcal{X} , alors la taille de U est bornée par le double du nombre d'arêtes reliant \mathcal{X}_1 et \mathcal{X}_2 , soit la taille de l'ensemble de coupure (*cut-set*) associé à \mathcal{X}_1 . Ainsi, trouver l'ensemble \mathcal{X}_1 de *cut-set* de taille minimale permet de trouver une partition en étoile intéressante, minimisant une borne supérieure sur la taille de U .

J'ai appliqué cette méthode de décomposition sérielle à des problèmes de navigation de robot en la couplant avec des méthodes d'agrégation d'états (Hendrickson and Leland, 1995) et de raffinement local (Kernighan and Lin, 1970), afin d'améliorer la partition obtenue. Lorsqu'il s'agit non pas de trouver une bipartition, mais une K -partition, la méthode de bipartition est itérée. La figure (5.2)

² $V = \mathcal{X}$ et $(x, y) \in E$ si et seulement si il existe $a \in \mathcal{A}, p(y|x, a) > 0$ ou $p(x|y, a) > 0$.

montre l'application de la méthode itérée sur un problème de taille modérée (environ 400 états), où une 4-partition équilibrée ($|\mathcal{X}_i| \geq 40$) est recherchée. Dans cet exemple, $|U| = 36$.

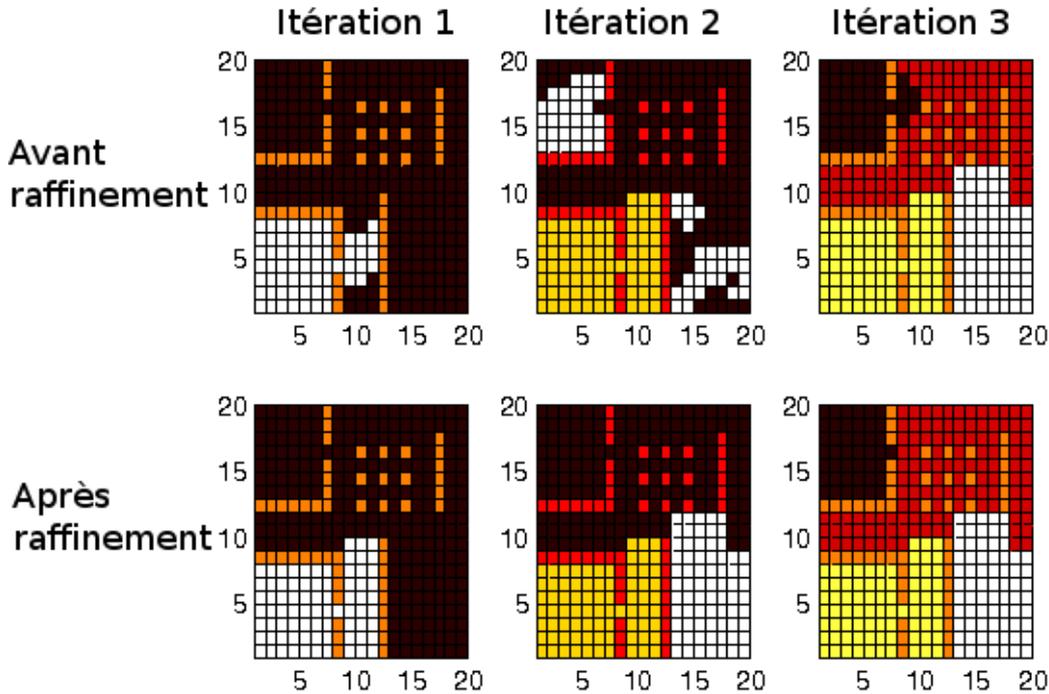


FIG. 5.2 – Bipartitions successives, avant et après raffinement local.

Décomposition sérielle ou parallèle ?

La méthode de décomposition que je viens de décrire est “sérielle” : elle est basée sur une décomposition de l'espace d'états en une *union* de sous-ensembles engendrant des sous-problèmes plus ou moins indépendants. Lorsque l'espace d'états du problème initial est factorisé, les méthodes de décomposition sérielles sont en général insuffisantes. Pour ce type de problème, on préfère en général appliquer une méthode de décomposition “parallèle”, basée sur une décomposition de l'espace d'états en un *produit cartésien* d'ensembles d'états et d'actions. Chacun des sous-problèmes engendrés par ces ensembles est alors résolu indépendamment et une politique solution du problème global est construite à partir de l'agrégation des solutions optimales des différents sous-problèmes. Dans la plupart des cas, la politique obtenue à partir d'une méthode de décomposition parallèle est sous-optimale. C'est le prix à payer pour résoudre des problèmes de très grande taille (supérieure à 2^{100}).

Dans la section suivante, je vais décrire nos travaux [20], [21], [26], [27], [28], sur des méthodes de résolution par décomposition parallèle d'un certain type de PDM factorisés : les *processus décisionnels de Markov sur graphe*.

5.2.2 Processus décisionnels de Markov sur graphe

Définitions

Un *Processus Décisionnel de Markov sur Graphe* (PDMG) est un cas particulier de *PDM factorisé* (Boutilier et al., 2000; Guestrin et al., 2003). Un PDMG est défini par un quintuplet $\langle \mathcal{X}, \mathcal{A}, p, r, G \rangle$, dans lequel l'espace d'états \mathcal{X} est un produit cartésien $\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_n$ et l'espace d'actions est, dans le cas général, un sous-ensemble \mathcal{A} d'un produit cartésien $\mathcal{A}_1 \times \dots \times \mathcal{A}_n$ de même dimension que \mathcal{X} . Nous supposons pour l'instant que $\mathcal{A} = \mathcal{A}_1 \times \dots \times \mathcal{A}_n$. $G = (V, E)$ est un graphe orienté, défini par un ensemble de sommets $V = \{1, \dots, n\}$ (n est la dimension de \mathcal{X} et de \mathcal{A}) et un ensemble d'arêtes $E \subseteq V^2$.

Une fonction de voisinage N sur G est définie par :

Définition 5.2.3 (Fonction de voisinage)

$N : V \rightarrow 2^V$ est définie par : $\forall i \in V, N(i) = \{j \in V, (j, i) \in E\}$. Le voisinage d'un sous-ensemble de sommets $I \subseteq V$ est défini par extension : $N(I) = \cup_{i \in I} N(i)$.

Dans un PDMG, les probabilités de transition et les récompenses sont supposées *locales*, au sens de la structure de voisinage définie par le graphe :

Définition 5.2.4 (Transitions locales)

Soit $\langle \mathcal{X}, \mathcal{A}, p, r, G \rangle$ un PDMG. Les transitions sont dites *locales* lorsqu'il existe une famille de probabilités de transition $\{p_i\}_{i=1..n}$, telles que :

$\forall x = (x_1, \dots, x_n) \in \mathcal{X}, x' = (x'_1, \dots, x'_n) \in \mathcal{X}, a = (a_1, \dots, a_n) \in \mathcal{A}$,

$$p(x'|x, a) = \prod_{i=1}^n p_i(x'_i | x_{N(i)}, a_i),$$

où $x_I = (x_i)_{i \in I}, \forall I \subseteq \{1, \dots, n\}$.

Définition 5.2.5 (Récompenses locales)

Soit $\langle \mathcal{X}, \mathcal{A}, p, r, G \rangle$ un PDMG. Les récompenses sont dites *locales* lorsqu'il existe une famille de fonctions $\{r_i\}_{i=1..n}$, telles que pour tout $x = (x_1, \dots, x_n) \in \mathcal{X}, a = (a_1, \dots, a_n) \in \mathcal{A}$,

$$r(x, a) = \sum_{i=1}^n r_i(x_{N(i)}, a_i).$$

Cette propriété de localité des fonctions de transition et de récompense est intéressante, puisqu'elle permet de décrire un PDMG de manière concise. La complexité spatiale de l'expression d'un PDMG est une fonction linéaire du nombre de sommets, ce qui permet de représenter de très grands problèmes, pourvu que la *largeur* du graphe ($\nu = \max_{i \in V} |N(i)|$) reste faible. Ceci est en général le cas pour des problèmes de décision "spatialisée", dans lesquels les nœuds du graphe sont attachés à des zones géographiques et les arêtes représentent des relations de voisinage entre ces zones. Néanmoins, les complexités en temps et en espace de la résolution de PDMG par des algorithmes de *programmation dynamique* sont exponentielles en n . Ceci limite grandement la taille des PDMG qui peuvent être résolus efficacement.

Une manière de réduire la complexité de la résolution est de restreindre la recherche à des politiques qui exploitent la structure du graphe G . Dans le cas général, les politiques optimales d'un PDMG sont de la forme $\delta = (\delta_1, \dots, \delta_n)$, avec $\delta_i : \mathcal{X} \rightarrow \mathcal{A}_i$. De telles politiques, dites *globales*, requièrent un espace exponentiel en n pour être exprimées. Ceci est prohibitif dans la plupart des cas. Il est donc tentant de restreindre la recherche à certaines politiques particulières, dites *politiques locales* :

Définition 5.2.6 (Politique locale)

Dans un PDMG $\langle \mathcal{X}, \mathcal{A}, p, r, G \rangle$, une politique $\delta : \mathcal{X} \rightarrow \mathcal{A}$ est dite locale ssi $\delta = (\delta_1, \dots, \delta_n)$ où $\delta_i : \mathcal{X}_{N(i)} \rightarrow \mathcal{A}_i$.

Restreindre la recherche d'une politique solution d'un PDMG à l'ensemble des politiques locales permet d'exploiter la structure du graphe. Cependant, la question qui se pose immédiatement est celle de l'existence de politiques optimales d'un PDMG, exprimables sous forme locale.

Non optimalité des politiques locales

Il est possible de montrer que si la fonction de valeur optimale $v^* : \mathcal{X} \rightarrow \mathbb{R}$ d'un PDMG est *décomposable*, alors il existe une politique locale *globalement optimale* pour ce problème (Chornei et al., 2006), [27], [28], [20], [21].

Définition 5.2.7 (Fonction de valeur optimale décomposable)

La fonction de valeur optimale v^* d'un PDMG $\langle \mathcal{X}, \mathcal{A}, p, r, G \rangle$ est dite décomposable si et seulement si il existe $(v_i^*)_{i \in V}, v_i^* : \mathcal{X}_i \rightarrow \mathbb{R}$, telles que :

$$\forall x \in \mathcal{X}, v^*(x) = \sum_{i=1}^n v_i^*(x_i).$$

Proposition 5.2.1 (Optimalité des politiques locales)

Si la fonction de valeur optimale v^* d'un PDMG $\langle \mathcal{X}, \mathcal{A}, p, r, G \rangle$ est décomposable, alors la politique locale $\delta^* = (\delta_1^*, \dots, \delta_n^*)$, définie par : $\forall i \in \{1, \dots, n\}, \forall x_{N(i)} \in \mathcal{X}_{N(i)}$,

$$\delta_i^*(x_{N(i)}) = \arg \max_{a_i \in \mathcal{A}_i} \left\{ r_i(x_{N(i)}, a_i) + \gamma \sum_{x'_i \in \mathcal{X}_i} p_i(x'_i | x_{N(i)}, a_i) \cdot v_i^*(x'_i) \right\},$$

est globalement optimale.

Malheureusement, et contrairement à ce qu'affirment à tort (Chornei et al., 2006), rien ne garantit que la fonction de valeur optimale d'un PDMG quelconque soit décomposable. Et donc rien ne garantit a priori que la politique optimale d'un PDMG quelconque puisse être locale. Au contraire, l'exemple suivant montre un problème très simple pour lequel la valeur de la meilleure politique locale est strictement inférieure à la valeur de la politique optimale, en tout état.

Exemple 5.2.1

Considérons le PDMG $\langle \mathcal{X}, \mathcal{A}, p, r, G \rangle$ dont le graphe est représenté dans la figure (5.3). Dans ce problème, $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}_3 = \{1 \dots m\}$, $\mathcal{A}_1 = \mathcal{A}_2 = \emptyset$ et $\mathcal{A}_3 = \{1 \dots m\}$.

Les probabilités de transition sont définies par :

- $p_1(x'_1 | x_1) = \frac{1}{m}, \forall (x_1, x'_1) \in \{1 \dots m\}^2$,
- $p_2(x'_2 = x_1 | x_1) = 1, \forall x_1 \in \{1 \dots m\}$,

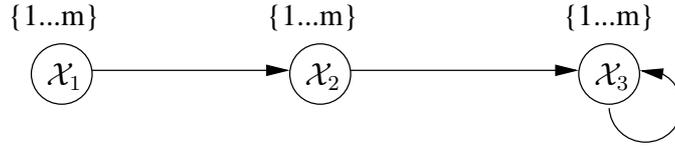


FIG. 5.3 – Graphe du PDMG contre-exemple.

$$- p_3(x'_3 = a_3 | x_2, x_3, a_3) = 1, \forall (x_2, x_3) \in \{1 \dots m\}^2, \forall a_3 \in \{1 \dots m\}.$$

Seule la fonction de récompense locale r_3 est non nulle : $r_3(x_2, x_3) = 1$ si $x_2 = x_3$ et 0 sinon.

Intuitivement, dans ce problème, on cherche à assurer l'égalité de x_2 et x_3 au cours du temps en ne décidant que de la valeur de x_3 . x_2 "suit" exactement le processus aléatoire (X_1^t) , markovien d'ordre 0, avec un pas de temps de retard.

Pour un tel problème, toute politique locale δ_3 est de la forme $a_3^t = \delta_3(x_2^t, x_3^t)$. Il est facile de montrer que pour n'importe quelle politique locale de cette forme,

$$Pr(x_2^{t+1} = x_3^{t+1} | \delta_3, x_2^t, x_3^t) = \frac{1}{m}, \forall x_2^t, x_3^t, x_2^{t+1}, x_3^{t+1}.$$

Ceci signifie que les fonctions de valeur de toutes les politiques locales sont identiques :

$$\begin{aligned} v_{\delta_3}(x) &= 1 + \sum_{t=1}^{+\infty} \frac{\gamma^t}{m} = 1 + \frac{\gamma}{m(1-\gamma)}, \text{ si } x_2 = x_3 \text{ et} \\ v_{\delta_3}(x) &= \sum_{t=1}^{+\infty} \frac{\gamma^t}{m} = \frac{\gamma}{m(1-\gamma)}, \text{ si } x_2 \neq x_3, \forall \delta_3 \text{ locale.} \end{aligned}$$

D'un autre côté, il est également facile de vérifier que la politique δ_3^* (qui n'est pas locale, puisqu'elle choisit a_3^t en fonction de x_1^t) définie par $a_3^t = \delta_3^*(x_1^t) = x_1^t$, est optimale pour le PDMG. En effet, elle garantit une récompense immédiate de 1 à chaque pas de temps, à partir de $t = 1$:

$$\begin{aligned} v_{\delta_3^*}(x) &= \frac{1}{1-\gamma}, \text{ si } x_2 = x_3 \text{ et} \\ v_{\delta_3^*}(x) &= \frac{\gamma}{1-\gamma}, \text{ si } x_2 \neq x_3. \end{aligned}$$

Donc, si $m \geq 2$, $v_{\delta_3}(x) < v_{\delta_3^*}(x), \forall x \in \mathcal{X}$, puisque :

$$\begin{aligned} \frac{v_{\delta_3}(x)}{v_{\delta_3^*}(x)} &= (1-\gamma) + \frac{\gamma}{m}, \text{ si } x_2 = x_3 \text{ et} \\ \frac{v_{\delta_3}(x)}{v_{\delta_3^*}(x)} &= \frac{1}{m}, \text{ si } x_2 \neq x_3. \end{aligned}$$

En faisant croître m , la valeur de la meilleure politique locale en certains états (pour lesquels $x_2 \neq x_3$) peut être arbitrairement moins bonne que la valeur en ces états de la politique optimale. De plus, dans cet exemple la valeur de la meilleure politique locale est strictement moins bonne, en tout état, que celle de la politique optimale.

Méthodes de résolution approchées pour les PDMG

A travers cet exemple simple, nous voyons qu'il est illusoire d'espérer trouver une politique locale qui soit globalement optimale, ou même dont la perte d'optimalité soit bornée (au pire cas ou en moyenne), pour un PDMG quelconque. Il n'en reste pas moins que les politiques locales sont particulièrement intéressantes par leur simplicité d'expression et de calcul. C'est pourquoi nous nous sommes intéressés, avec Nicklas Forsell et Nathalie Peyrard, à la conception d'algorithmes de résolution approchée de PDMG retournant des politiques locales. Nous avons également testé empiriquement ces algorithmes sur des problèmes de gestion forestière [16] ou d'épidémiologie végétale [29].

Il existe plusieurs algorithmes permettant de calculer la fonction de valeur optimale, v^* , d'un PDM, ou une valeur approchée dans le cas où la taille du problème ne permet pas une résolution exacte. Les méthodes de recherche de politiques locales pour les PDMG que nous avons proposées sont basées sur certains de ces algorithmes, de type *programmation linéaire approchée (PLA)* ou *approximation en champ moyen (CM)*.

Dans un PDM $\langle \mathcal{X}, \mathcal{A}, p, r \rangle$, la fonction de valeur optimale v^* est l'unique solution du programme linéaire suivant (de Ghellinck, 1960) :

$$\begin{aligned} \min \quad & \sum_{x \in \mathcal{X}} v(x) \\ \text{t.q.} \quad & \begin{cases} v(x) \geq r(x, a) + \gamma \sum_{x' \in \mathcal{X}} p(x'|x, a)v(x'), \forall x \in \mathcal{X}, a \in \mathcal{A}, \\ v(x) \text{ libre}, \forall x \in \mathcal{X} \end{cases} \end{aligned} \quad (5.1)$$

Ce programme linéaire comprend $|\mathcal{X}|$ variables et $|\mathcal{X}| \times |\mathcal{A}|$ contraintes. La méthode de programmation linéaire "exacte" ne peut donc être utilisée pour résoudre des PDM factorisés comportant un grand nombre de variables. Afin de résoudre de tels problèmes, il faut trouver un moyen de diminuer à la fois le nombre de variables et de contraintes du programme linéaire correspondant.

La méthode de *Programmation Linéaire Approchée (PLA)* (de Farias and Van Roy, 2003, 2004) consiste à approcher la fonction de valeur optimale v^* , solution de (5.1), par une fonction de valeur paramétrée v_w , élément d'une famille de fonctions de valeurs paramétrées $(v_w)_{w \in \mathbb{R}^k}$, combinaisons linéaires d'un ensemble de fonctions réelles $H = \{h_1, \dots, h_k\}$, choisies arbitrairement :

$$v_w(x) = \sum_{i=1}^k w_i h_i(x), \forall x \in \mathcal{X}. \quad (5.2)$$

Si on impose à la fonction de valeur v du programme linéaire (5.1), la forme paramétrée (5.2), on peut définir un nouveau programme linéaire, dont les variables sont maintenant les paramètres $w \in \mathbb{R}^k$:

$$\begin{aligned} \min \quad & \sum_{x \in \mathcal{X}} \sum_{i=1}^k w_i h_i(x) \\ \text{t.q.} \quad & \begin{cases} \sum_{i=1}^k w_i h_i(x) \geq r(x, a) + \gamma \sum_{x' \in \mathcal{X}} p(x'|x, a) \cdot \left(\sum_{i=1}^k w_i h_i(x') \right), \\ \forall x \in \mathcal{X}, \forall a \in \mathcal{A}, \\ w_i \text{ libre}, \forall i = 1, \dots, k. \end{cases} \end{aligned} \quad (5.3)$$

La solution optimale $w^* = (w_1^*, \dots, w_k^*)$ de ce nouveau programme linéaire ne permet pas de calculer la fonction de valeur solution du PDM initial, mais v_{w^*} en est la meilleure approximation en norme $\|\cdot\|_1$. Elle minimise l'expression :

$$\|v^* - v_w\|_1 = \sum_{x \in \mathcal{X}} |v^*(x) - v_w(x)|.$$

Une politique approchée peut ensuite être calculée de manière gloutonne à partir de v_{w^*} :

$$\delta_{w^*}(x) = \arg \max_{a \in \mathcal{A}} \left\{ r(x, a) + \gamma \sum_{x' \in \mathcal{X}} p(x'|x, a) \cdot v_{w^*}(x') \right\}.$$

(de Farias and Van Roy, 2003) fournissent des résultats sur la perte d’optimalité engendrée par le choix de δ_{w^*} à la place de δ^* . Malheureusement, il n’existe aucune garantie que la politique δ_{w^*} soit la meilleure politique paramétrée au sens de la norme infinie $\|\cdot\|_\infty$ et donc que la politique approchée soit la meilleure possible “dans le pire des cas”.

L’intérêt de rechercher une politique paramétrée est de diminuer le nombre de variables du programme linéaire à résoudre (en général, on choisit $k \ll |\mathcal{X}|$). Cependant, cette approximation ne permet pas de diminuer le nombre de contraintes du problème, qui reste $|\mathcal{X}| \times |\mathcal{A}|$. (de Farias and Van Roy, 2004) proposent une approximation supplémentaire, consistant à ne considérer dans (5.3) qu’un sous-ensemble de contraintes de taille raisonnable, choisi aléatoirement. Cette méthode est basée sur le constat que la plupart des contraintes sont “redondantes” et que la suppression aléatoire de certaines d’entre elles n’a qu’une faible probabilité de modifier la solution optimale de (5.3).

Nous avons montré [20], [21] (voir aussi annexe C.1), que la structure d’un PDMG pouvait être mise à profit pour concevoir un algorithme de résolution basé sur la programmation linéaire approchée, efficace (en termes de temps de calcul) et fournissant une borne en norme infinie sur la qualité des solutions retournées.

L’algorithme *Itération de la Politique* (IP) (Howard, 1960) permet également de calculer une politique optimale dans un PDM (voir chapitre 2). Cet algorithme itératif alterne deux étapes, une étape d’évaluation de la politique courante et une étape d’amélioration de cette politique. Lorsque la politique courante ne peut plus être améliorée, son optimalité est garantie. Lorsque les espaces d’états et d’actions sont trop grands, ces deux phases d’évaluation et d’amélioration peuvent être approchées. L’approximation des deux phases correspond à des algorithmes dits d’*Itération de la Politique Approchée* (IPA) (Bertsekas and Tsitsiklis, 1996), pour lesquels l’optimalité de la politique calculée est sacrifiée à l’efficacité calculatoire.

Nous avons proposé dans [27], [28], une méthode de résolution de PDMG de type “itération de la politique approchée”. Cette méthode est basée sur l’utilisation de *mesures d’occupation* (Altman, 1999), que nous approchons par une méthode de type *champ moyen* (Chandler, 1987) (voir annexe C.2).

5.2.3 Conclusion intermédiaire

Jusqu’ici, j’ai décrit dans ce chapitre des travaux sur les PDM factorisés, réalisés principalement avec N. Forsell et N. Peyrard. Le modèle des PDM sur graphe, sur lequel nous avons focalisé notre étude, est inspiré d’un cadre proposé par (Chorney et al., 2006). Ce cadre était d’autant plus séduisant, a priori, que les auteurs affirmaient la décomposabilité de la fonction de valeur optimale, qui entraîne la localité des politiques optimales. Ceci laissait envisager la possibilité de l’existence d’algorithmes de résolution de complexité “raisonnable” (polynomiale en espace).

Malheureusement, les résultats de (Chorney et al., 2006) étaient erronés (voir section 5.2.2). Néanmoins, les deux algorithmes que nous avons proposés permettent de calculer efficacement des politiques locales approchées, mais de qualité non garantie. Dans le chapitre 6, je décrirai deux applications du cadre des PDMG en gestion forestière et en épidémiologie, qui ont servis de validation empirique de ces algorithmes. Mais avant cela, je vais décrire ma deuxième contribution dans le

domaine des processus décisionnels de Markov, consacrée à l'étude des *processus décisionnels de Markov épistémiques*.

5.3 Processus décisionnels de Markov épistémiques

5.3.1 Problèmes de planification épistémique

Un problème de *planification épistémique* est un exemple de problème de *planification dans l'incertain*, dans lequel on ne cherche pas à atteindre un état du monde *but*, mais plutôt à préciser notre connaissance de l'état du monde. Il s'agit ici d'un problème de *méta-raisonnement* (Conitzer and Sandholm, 2003), dans lesquels on raisonne sur la connaissance d'un agent sur l'état du monde, et non sur l'état du monde lui-même.

L'exemple suivant nous permet d'illustrer cette classe de problèmes.

Exemple 5.3.1 *Supposons le jeu suivant entre deux joueurs. Le joueur A jette un dé à six faces et observe le résultat. Le joueur B n'observe pas le résultat du lancer mais peut poser à A, contre rémunération, les questions suivantes :*

Q1 Si le résultat du lancer est 1 ou 6, donne-moi le résultat exact du lancer.

Q2 Dis-moi si le résultat du lancer est inférieur ou égal à 3.

Q3 Dis-moi si le résultat du lancer est pair ou impair.

Le joueur B doit décider d'une "politique épistémique", soit de questions successives à poser à A, afin de déterminer le plus précisément possible le résultat du lancer, pour un coût minimum (en espérance).

Dans cet exemple, la connaissance du joueur B sur l'état du monde est modélisable à tout instant par un *état de croyance*, une distribution de probabilité sur les résultats possibles du lancer. Les réponses aux questions posées par B lui permettent de remettre à jour cette connaissance. Si nous supposons par exemple, que le joueur A répond forcément la vérité, alors il est possible de construire une politique permettant de connaître à *coup sûr* le résultat du lancer, en trois questions au plus et en $(2/3).3 + (1/3) = 7/3$ questions en moyenne, si les états sont a priori équiprobables (figure 5.4). Notons que dans cet exemple, si la réponse à Q1 est "1" ou "6", il est inutile de poser une nouvelle question. Une action épistémique supplémentaire *terminante* (gratuite), nommée *stop*, permet de terminer le processus à tout instant.

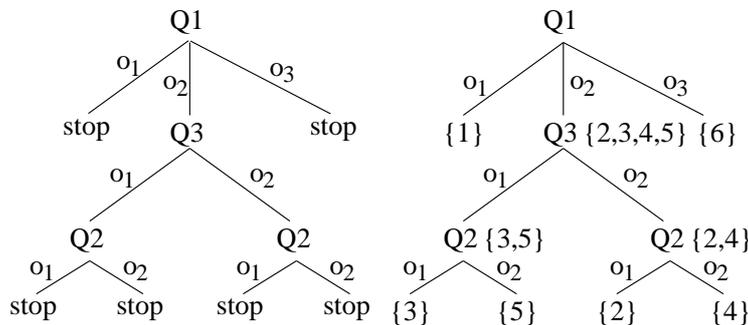


FIG. 5.4 – Gauche : politique. Droite : états de croyance accessibles.

Si le joueur A a la possibilité de mentir, il devient impossible de "désambiguïser" totalement l'état du monde en un nombre fini de questions. Dans ce cas, on recherche une politique (à horizon borné)

maximisant en espérance un critère tenant compte de l'utilité de l'état de croyance final auquel la politique permet d'accéder, et du coût des actions appliquées pour l'atteindre. L'utilité d'un état de croyance b est mesurée par une fonction d'utilité ρ . La définition d'une telle fonction dépend bien entendu du problème de méta-raisonnement considéré. Voici quelques exemples possibles :

- *Désambiguisation totale.* $\rho(b) = r(x)$ si $b(\{x\}) = 1$ et 0 sinon, où r est une fonction réelle. Une récompense n'est reçue que si l'état est identifié parfaitement (et cette récompense dépend éventuellement de l'état).
- *Désambiguisation partielle.* $\{\langle X_1, r_1 \rangle, \dots, \langle X_q, r_q \rangle\}$ où $\{X_i\}_{i=1..q}$ forme un ensemble couvrant de \mathcal{X} ($\cup_i X_i = \mathcal{X}$) et chaque r_i est un nombre réel positif. La récompense $\rho(b) = \max\{r_i | b(X_i) = 1\}$ est reçue dans l'état de croyance b si il existe i tel que $b(X_i) = 1$ (et $\rho(b) = 0$ sinon). Dans ce cas, on se contente de circonscrire l'état réel du monde à un des sous-ensembles prédéfinis³.
- *Minimisation de l'entropie.* $\rho(b) = -H(b)$, où $H(b) = \sum_{x \in \mathcal{X}} -b(x) \cdot \log(b(x))$. $H(b)$ représente l'entropie de la distribution b . En théorie de l'information, celle-ci est utilisée pour mesurer le degré d'information délivré par une distribution de probabilité : plus l'entropie est faible plus la distribution est informative.

Pourquoi le choix évident $\rho(b) = \sum_{x \in \mathcal{X}} b(x) \cdot r(x)$ ne figure-t-il pas dans cette liste ?

Nous souhaitons que les récompenses attachées au contenu épistémique des états de croyance augmentent avec la valeur de l'information représentée par b . Or, nous pouvons voir sur un exemple simple que l'espérance mathématique ne satisfait pas ce souhait : si nous considérons $\mathcal{X} = \{x_1, x_2\}$ et $u(x_1) < u(x_2)$, l'état de croyance b_* défini par $b_*(x_1) = 1$, pour lequel nous sommes sûr que l'état réel est x_1 aura une récompense minimale parmi tous les états de croyance. Ceci viole le principe souhaitable que plus les états de croyance sont informatifs, plus ils sont désirables.

De nombreux problèmes rencontrés dans la réalité peuvent être exprimés sous la forme de problèmes de planification épistémique :

- *Jeux.* Il existe bon nombre de jeux classiques se ramenant à un problème de planification épistémique. C'est le cas de la *bataille navale*, ou du *Mastermind*, dans lesquels les réponses à des suggestions de position occupée ou de couleurs de pions permettent de préciser la connaissance du joueur. Dans ces deux jeux, l'objectif de celui qui pose les questions est d'identifier parfaitement l'état du monde, grâce aux réponses (toujours fiables) de l'adversaire.
- *Elicitation de préférences.* Le problème de l'élicitation de préférences revient à construire des outils informatiques permettant de révéler le plus efficacement possible (minimum de questions, par exemple) les préférences d'un client sur une gamme de produits aux caractéristiques nombreuses. De tels systèmes permettent par exemple de proposer une location de vacances à un client, en fonction de critères tels que le prix, le lieu, la nature du logement, etc. De la même manière, le choix d'une configuration d'options pour une voiture neuve devient tellement complexe que de tels outils sont souvent utilisés. L'élicitation de préférences peut être vue comme un problème de planification épistémique, dans lequel la connaissance à révéler est la relation de préférence du client (ou simplement l'article maximisant cette préférence).
- *Diagnostic de systèmes.* La problématique du diagnostic en général est une problématique de planification épistémique. Le diagnostic utilise des *tests* ou des *mesures* (de quantités électriques dans un circuit, par exemple), afin de déterminer si un composant du système fonctionne ou non.
- *Diagnostic en sciences naturelles et environnement.* De nombreux facteurs biologiques ayant trait

³On se ramène au problème précédent si l'ensemble des X_i est l'ensemble des singletons de \mathcal{X} ($X_i = \{x_i\}$) et $r_i = r(x_i)$.

à l'environnement (indicateurs de biodiversité, de pollution, de présence de bioagresseurs dans des espaces naturels...), utiles pour la gestion environnementale, nécessitent des campagnes d'observation lourdes avant d'être révélés. La conception de ces campagnes d'observation peut être vue comme un problème de planification épistémique complexe. Nous décrivons dans le chapitre 6 un tel exemple.

5.3.2 Processus décisionnels de Markov épistémique

Nous avons défini dans [41] (voir aussi l'annexe C.3) le cadre des processus décisionnels de Markov épistémiques (PDME), permettant de modéliser les problèmes de planification épistémiques décrits précédemment. Un processus décisionnel de Markov épistémique (PDME) diffère d'un processus décisionnel de Markov partiellement observable (PDMPO), décrit dans le chapitre 2, par les deux propriétés suivantes :

- L'état *physique* du système, x , ne change pas au cours du temps. Seule la *connaissance* que l'on a de cet état évolue au cours du temps, sous l'effet d'*actions épistémiques*.
- La *fonction de récompense*, modélisant la satisfaction immédiate des préférences de l'agent, ne porte pas sur l'état physique du système mais sur la *connaissance* qu'en a l'agent. En effet, l'état physique du système n'est pas modifiables par les actions épistémiques de l'agent.

Formellement, un PDME peut être défini par un tuple $\langle \mathcal{X}, \mathcal{A}, \mathcal{O}, O, \rho, b_0, T \rangle$ où :

- \mathcal{X} est l'ensemble des *états physiques* possibles du système. La connaissance courante sur l'état du système est modélisée par un *état de croyance* b (une distribution de probabilité sur \mathcal{X}).
- \mathcal{A} est l'ensemble des *actions épistémiques* disponibles. Ces actions permettent de modifier l'état de croyance courant, mais pas l'état physique du système. C'est pourquoi, contrairement au cadre des PDMPO, le cadre des PDME n'inclut pas de fonction de transition p sur les états physiques (ou, si on préfère, $p(x|x, a) = 1, \forall x \in \mathcal{X}, \forall a \in \mathcal{A}$). Pour des raisons techniques, nous ajoutons une action supplémentaire, *terminante*, nommée *stop*.
- \mathcal{O} est un ensemble fini d'*observations* possibles résultant de l'exécution des actions épistémiques. L'effet d'une action épistémique a , appliquée dans l'état physique x , est de retourner une observation indirecte $o \in \mathcal{O}$ de l'état.
- O est une *fonction d'observation* : $O(x, a, o) = p(o|x, a)$ est la probabilité d'observer $o \in \mathcal{O}$ après avoir appliqué a en x . La fonction O peut éventuellement être *déterministe*. Dans ce cas, O est une application de $\mathcal{X} \times \mathcal{A}$ vers \mathcal{O} : la même observation $o = O(x, a)$ est toujours obtenue lorsqu'on applique a en x .
- ρ est la *fonction de récompense épistémique* immédiate. Elle est construite par extension d'une fonction d'utilité épistémique sur les états de croyance. Un coût (positif) $C(a)$ est associé à toute action a non terminante. Ce coût est indépendant de l'état de croyance courant b . Lorsque l'action *stop* est exécutée, une récompense $\rho(b)$ est reçue, mesurant l'utilité pour l'agent de l'état de croyance final b . La fonction de récompense épistémique immédiate est donc définie par :

$$\rho(b, a) = -C(a) \text{ si } a \neq \text{stop} \text{ et } \rho(b, \text{stop}) = \rho(b).$$

- b_0 est une distribution de probabilité sur \mathcal{X} , décrivant l'*état de croyance initial*.
- $T \in \mathbb{N}$ est l'*horizon* du PDME.

Dans [41], [31], nous nous sommes intéressés plus particulièrement à la complexité théorique des problèmes de planification épistémiques posés dans ce cadre. Nous avons montré que suivant la nature

de la fonction d'observation du problème, déterministe ou stochastique, de tels problèmes étaient NP-complet ou PSPACE-complet (voir annexe C.3).

5.4 Conclusion

Ces dernières années, j'ai travaillé à la définition de cadres formels pour la représentation et la résolution de problèmes de décision séquentielle dans l'incertain, factorisés ou épistémiques.

Les processus décisionnels de Markov sur graphe (PDMG) sont un cadre de représentation et de résolution approchée de PDM factorisés, dédié à la décision séquentielle dans l'incertain spatialisé. Nous avons défini la notion de politique locale dans ce cadre, et proposé des algorithmes de résolution approchée de PDMG exploitant ces politiques locales.

Nous avons également décrit le cadre des processus décisionnels de Markov épistémiques (PDME), permettant de représenter et de résoudre des problèmes de *planification épistémique*, visant à optimiser la prise d'information dans des problèmes de raisonnement (ou de décision) sous incertitude. Le cadre des PDME est un cas particulier de *belief MDP* (PDM sur les états de croyance), différant du cadre des PDMPO à la fois par son caractère *purement épistémique* (les actions ne modifient pas l'état physique du monde), et par la forme très générale que peut prendre la fonction de récompense sur les états de croyance.

Dans le chapitre suivant, je vais décrire un certain nombre de problèmes de gestion agri-environnementale spatialisée sur lesquels j'ai mis en oeuvre les approches décrites dans ce chapitre.

Chapitre 6

Applications en gestion agri-environnementale

6.1 Introduction

Au cours de ces dernières années, j'ai participé à un certain nombre de projets appliqués dans les domaines :

- de la gestion forestière et de ses effets sur la biodiversité et
- de l'épidémiologie.

Les points communs de tous les problèmes que j'ai abordés sont leur nature spatiale et dynamique et l'existence d'une problématique de gestion, explicite ou implicite. Ces projets ont également toujours impliqué des partenaires modélisateurs : écologues, forestiers ou agronomes.

Mon rôle, dans chacun de ces projets, a été de proposer (parfois en collaboration avec d'autres chercheurs en mathématique ou informatique) des approches mathématiques ou informatiques innovantes pour modéliser ces problèmes. J'ai également participé activement à la construction des modèles, à leur implémentation informatique et à l'exploitation des résultats obtenus, souvent à travers l'encadrement de stagiaires. Cette partie de mon travail de recherche, résumée dans ce chapitre, a pris une place de plus en plus importante dans mon activité ces dernières années.

6.2 Gestion forestière et biodiversité

6.2.1 Influence de la fragmentation forestière sur des dynamiques spatialisées de populations d'oiseaux

Contexte :

Projet CCRDT, Région Midi-Pyrénées : *Les forêts fragmentées dans les territoires et les exploitations agricoles de Midi-Pyrénées*, action 3 : “du diagnostic à l'aide à la décision ou comment concilier les fonctions productives et environnementales des îlots boisés des côteaux de Midi-Pyrénées”.

Période : 2002-2004.

Partenaires INRA :

- UMR DYNAFOR (DYNAmiques FORestières dans l'espace rural).

– UR CEFS (Comportement, Ecologie de la Faune Sauvage).

Description des activités concernées

Dans le cadre de cette collaboration, principalement avec l'UMR DYNAFOR, nous avons commencé à aborder l'étude des processus décisionnels de Markov "spatialisés" avec Frédéric Garcia et Roger Martin-Clouaire. Dans ces travaux "préliminaires" du point de vue méthodologique, nous nous sommes intéressés principalement à l'approximation (par simulation) de processus stochastiques spatialisés. La partie "optimisation de la décision" était absente.

Dans la méthode d'approximation retenue, une distribution de probabilité sur un ensemble de cartes de présence/absence (sur une grille régulière) est approchée par un produit de distributions indépendantes, estimées par simulation d'un processus de diffusion. Nous avons tout d'abord élaboré puis testé ce modèle sur un problème artificiel de propagation d'incendie¹ [39].

Ce problème artificiel a été choisi pour sa simplicité, en vue de développer le modèle. Celui-ci a ensuite été testé pour modéliser la dynamique d'une population de rouge-gorges dans les forêts du canton d'Aurignac². La connaissance sur la population de rouge-gorges à un instant donné résulte d'un ensemble d'observations (points d'écoute), réparties spatialement. Cette connaissance est décrite par un tableau de probabilités d'occupation de cellules d'une grille "plaquée" sur la carte du canton d'Aurignac, distinguant les zones boisées et non boisées. En "approchant" la loi jointe de la probabilité de présence des oiseaux sur le canton par le produit des probabilités du tableau, nous avons étudié sa dynamique en utilisant des connaissances expertes sur les traits de vie du rouge-gorge. Nous avons enfin comparé empiriquement [13] les distributions "limite" du processus approché avec des probabilités de présences "statiques", déduites d'un modèle de réponse des oiseaux à la fragmentation forestière (Ladet, 2000). La figure (6.2.1) montre les deux cartes de probabilités, qualitativement semblables, obtenues.

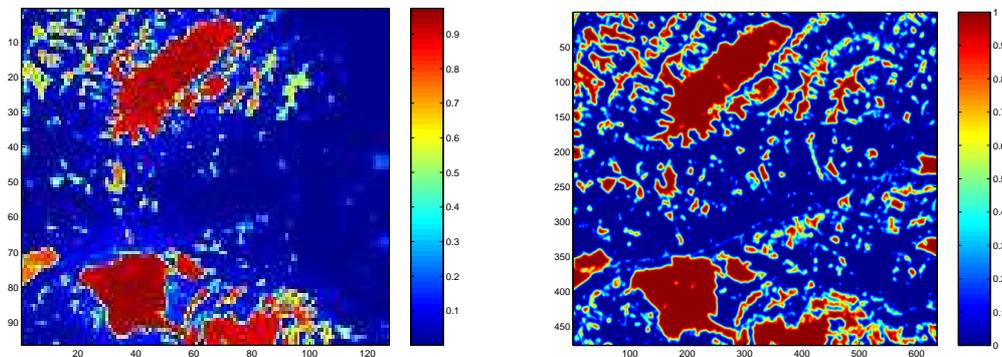


FIG. 6.1 – Probabilités de présence. Gauche : limite du processus de Markov approché par simulation. Droite : Réponse à l'indice de fragmentation forestière.

Le modèle basé sur la simulation semble surestimer (par rapport au modèle basé sur l'indice de fragmentation) les probabilités de présence élevées, et sous-estimer les probabilités de présence faibles.

¹E. Huard. Approximation d'un processus stochastique multidimensionnel : application à la modélisation de la propagation d'un incendie. Mémoire de stage INSA de 4ème année, Toulouse (septembre 2002).

²C. Paccard. Modélisation de dynamiques (stochastiques) de populations de rouge-gorges dans le canton d'Aurignac. Mémoire de stage INSA de 4ème année, Toulouse (septembre 2003).

Pour aller plus loin, il aurait été nécessaire de recalibrer les deux modèles par rapport aux données recueillies sur le terrain. Dans un second temps, notre objectif était de modéliser la dynamique forestière, et son impact sur les populations de rouges-gorges (et de 43 autres espèces d'oiseaux présentes dans la vallée de la Nère), puis d'évaluer, voire d'optimiser des stratégies d'exploitation forestière durables (du point de vue de la conservation de la biodiversité aviaire dans la région considérée). Ces objectifs supplémentaires n'ont finalement pas été mis en œuvre, la problématique d'optimisation de stratégies se révélant trop difficile à formaliser pour ce problème.

6.2.2 Gestion spatio-temporelle de réserves naturelles

Contexte : Collaboration scientifique, non formalisée par un contrat.

Période : 2004-2007.

Partenaire : Daniel Spring, écologue, Monash University, Clayton, Australie.

Description des activités concernées

La création de réserves naturelles est le moyen de protection de la biodiversité le plus efficace, car l'existence de zones protégées d'aire suffisante permet d'assurer la survie d'espèces menacées. Le processus de création de réserves est en général graduel, des sites naturels étant progressivement acquis au fur et à mesure qu'ils deviennent disponibles (Mair et al., 2004) ou que les fonds nécessaires à leur acquisition deviennent mobilisables (Costello and Polaski, 2004). Certains sites naturels sont détruits par un processus concurrent de *développement* et deviennent définitivement inaccessibles pour la création de réserves naturelles. Dans les régions tropicales, telles que l'Indonésie, l'Afrique subsaharienne ou la forêt Amazonienne, le processus de développement peut être modélisé par un *processus de contact* stochastique (Laurance et al., 2004) : de nouvelles zones risquent de se développer au fur et à mesure que des zones voisines se développent et les rendent accessibles, grâce à la création de nouvelles routes par exemple.

La conjonction de la dynamique stochastique de développement et de la disponibilité limitée des ressources allouées à la création de réserves naturelles plaident pour l'élaboration de modèles d'optimisation séquentielle dans l'incertain spatialisés pour la gestion de réserves naturelles. (Mair et al., 2004) et (Costello and Polaski, 2004) ont proposé des modèles de type *processus décisionnel de Markov factorisé* pour ce problème. Dans ces modèles, l'état courant du système est représenté par l'état de chaque site forestier : $X_t^i = (D)$ éveloppé, (R) éservé ou (N) on-développé. A chaque instant, les actions disponibles consistent à réserver un seul site non-développé, faisant ainsi passer son état de N à R . Dans le même temps, tous les sites non-développés que l'on ne réserve pas peuvent se développer, avec une probabilité dépendant de plusieurs facteurs, propres au site lui-même (taille, type de région) ou à l'état de développement des sites voisins.

Chaque site non-développé peut donc passer à chaque instant dans l'état développé ou réservé ou rester dans le même état. (Mair et al., 2004) et (Costello and Polaski, 2004) supposent en outre que le passage dans les états R et D est irréversible. Le processus stochastique prend fin, soit à un horizon T prédéfini, soit lorsque plus aucun site non-développé ne persiste (un *état absorbant* est alors atteint). A ce moment, il est possible de vérifier si les conditions de survie de chacune des espèces menacées initialement présentes sont réunies. Ces conditions consistent en général en l'existence d'un ensemble

de sites connectés de taille excédant un seuil minimal, présentant les caractéristiques nécessaires à la survie de l'espèce (altitude, climat, ...).

Le processus décisionnel de Markov factorisé représentant ce problème de gestion de réserve est en général de taille bien trop grande pour pouvoir être résolu de manière exacte (résolution exacte de complexité temporelle exponentielle en le nombre de sites). C'est pourquoi (Mair et al., 2004; Costello and Polaski, 2004) proposent des méthodes de résolution heuristiques pour ce problème. Ces politiques approchées "gloutonnes", choisissent par exemple le site à réserver maximisant le nombre d'espèces immédiatement sauvegardées (i.e. permettant d'atteindre la surface minimale nécessaire à la survie du maximum d'espèces).

Dans [32], [6], nous³ avons proposé de résoudre ce problème par une méthode de type *apprentissage par renforcement*, utilisant une représentation paramétrée de la fonction de valeur (Bertsekas and Tsitsiklis, 1996; Tsitsiklis and van Roy, 1996). Plus précisément, ce problème est un cas particulier de PDMG dans lequel l'espace des actions est un sous-ensemble strict d'un produit cartésien d'espaces d'actions binaires. Nous approchons, comme dans le chapitre précédent, la fonction de valeur optimale du problème par une fonction linéaire paramétrée de la forme :

$$v_\varepsilon(x) = \varepsilon_1(x_1)\psi_1(x_1) + \dots + \varepsilon_n(x_n)\psi_n(x_n).$$

Les fonctions $\psi_i(x_i) : \mathcal{X}_i \rightarrow \mathbb{R}$ ont été définies dans ce problème de la manière suivante :

$$\psi_i(D) = 0 \text{ et } \psi_i(U) = \psi_i(R) = 1.$$

(Tsitsiklis and van Roy, 1996) ont proposé un algorithme de type *descente de gradient* exploitant ce type de fonctions ψ_i et basé sur la simulation de transitions $\langle x^t, a^t, x^{t+1}, r^t(x^t, a^t) \rangle$, afin de calculer une approximation v_{ε^*} de la fonction de valeur optimale.

Plus précisément, $\varepsilon^* = \lim_{t \rightarrow \infty} \varepsilon^t$, où la suite (ε^t) est définie par :

$$\varepsilon_i^{t+1} = \varepsilon_i^t + \alpha_t \left(r^t(x^t, a^t) + v_{\varepsilon^t}(x_i^{t+1}) - v_{\varepsilon^t}(x_i^t) \right) \cdot \psi_i(x_i^t), \forall i = 1 \dots n,$$

où α_t est un paramètre décroissant vers 0 avec t .

La convergence de cet algorithme vers une fonction de valeur satisfaisante dépend en partie du choix de l'action courante a^t à appliquer dans l'état x^t . En règle général, ce choix est un résultat du compromis entre *exploration aléatoire* de l'espace des actions (a^t choisi aléatoirement) et *exploitation* de la connaissance courante de la fonction de valeur, v_{ε^t} :

$$a^t = \arg \max_a \left(r(x^t, a) + E \left[v_{\varepsilon^t}(x^{t+1}) | x^t, a \right] \right).$$

En pratique, ce compromis est opéré en choisissant l'exploration aléatoire avec une probabilité β_t qui décroît avec le temps (et l'exploitation avec probabilité $1 - \beta_t$).

Le calcul de $E \left[v_{\varepsilon^t}(x^{t+1}) | x^t, a \right]$ est difficile car le nombre d'états successeurs possibles de l'état courant x^t est de l'ordre de 2^{N^t} , où N^t est le nombre de sites non-développés à l'instant t : Chaque site non-développé a une probabilité non nulle de devenir développé à l'instant suivant. Il en résulte que

³C.-E. Rabier. Processus Décisionnels de Markov : application à la gestion de réserves naturelles. Mémoire de stage INSA de fin d'étude, Toulouse (Juin 2005).

le calcul exact de $E[v_{\varepsilon t}(x^{t+1})|x^t, a]$ est impossible en pratique. Aussi, dans [32], [6], nous avons suggéré d’approcher cette espérance par une moyenne empirique obtenue par la simulation des transitions.

Dans [40], nous avons appliqué ce modèle d’optimisation de gestion de réserves naturelles au cas de la forêt primitive du Costa-Rica⁴, pour lequel nous avons des données (SIG) de déforestation (voir figure (6.2)). Nous nous sommes servi de ces données pour estimer les paramètres d’un modèle

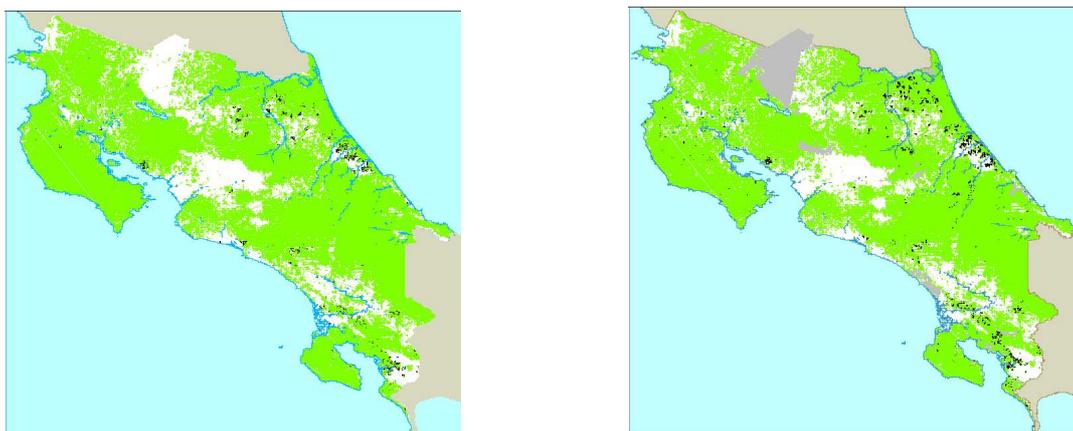


FIG. 6.2 – Gauche : déforestation observée entre 1997 et 2000. Droite : déforestation simulée entre 2000 et 2030.

logistique de probabilités de déforestation par contagion, tenant compte de l’aire de chaque site, de l’influence de l’état des sites voisins, du type de la région (altitude, climat), etc.

Ensuite, nous avons exploité des données disponibles sur les surfaces minimales requises pour la survie d’une centaine d’espèces d’oiseaux ou de mammifères présentes dans le pays, afin de construire une fonction de récompense immédiate. Une récompense d’une unité est obtenue dès lors qu’une nouvelle espèce rencontre les conditions nécessaires à sa survie. Finalement, nous avons choisi une zone d’étude particulière, représentée dans la figure (6.3). Cette zone comprend 309 sites, de différents types, permettant de protéger jusqu’à 54 espèces, dans le cas où tous les sites sont réservés.

Après détermination du vecteur de paramètres ε^* par apprentissage par renforcement, nous avons calculé le nombre d’espèces protégées par la politique correspondante au cours du temps à partir de simulations. Les résultats (moyenne et écart type) sont représentés dans la figure (6.4).

La forme de la courbe de l’espérance du nombre d’espèces protégées au cours du temps est très spécifique : dans un premier temps, un petit nombre de sites permettent de protéger un maximum d’espèces (deux sites, pour près de 30 espèces protégées), avant de subir un risque de développement (variance nulle sur les cent simulations). Puis, une phase d’accumulation de sites réservés ne permettant la protection d’aucune nouvelle espèce suit (douze pas de temps, soit 36 ans), avant que de nouvelles réservations permettent progressivement la protection d’une quinzaine d’espèces supplémentaires (avec une variance de l’ordre de deux espèces). La présence du long plateau entre les pas de temps 4 et 16 révèle l’inefficacité des méthodes d’optimisation “gloutonnes” sur ce problème. Entre ces pas de temps, toutes les actions sont considérées comme équivalentes par les méthodes gloutonnes car elles ne permettent pas d’obtenir de récompense immédiate. Les actions sont donc choisies au hasard et

⁴E. Bergonnier. Processus Décisionnels de Markov : application à la gestion de réserves naturelles. Mémoire de stage INSA de 4ème année, Toulouse (Juin 2006).

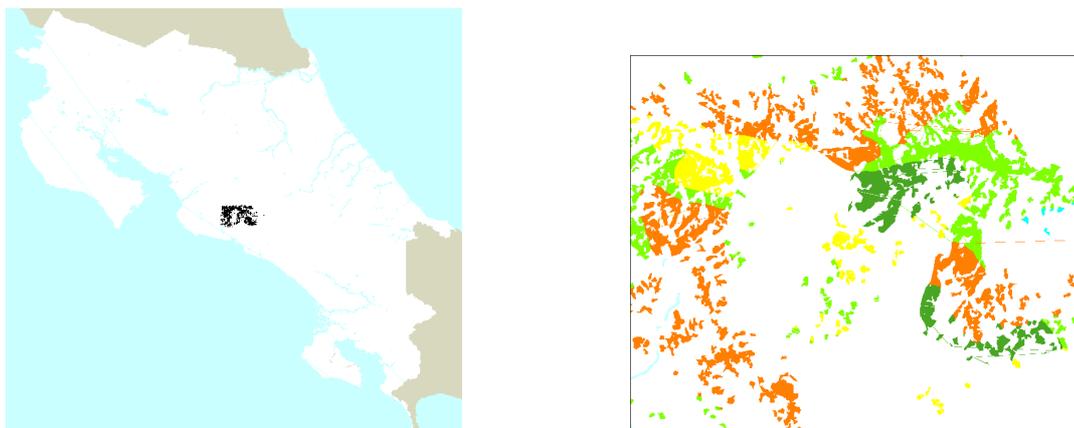


FIG. 6.3 – Gauche : zone d'étude. Droite : détails et types des différent sites.

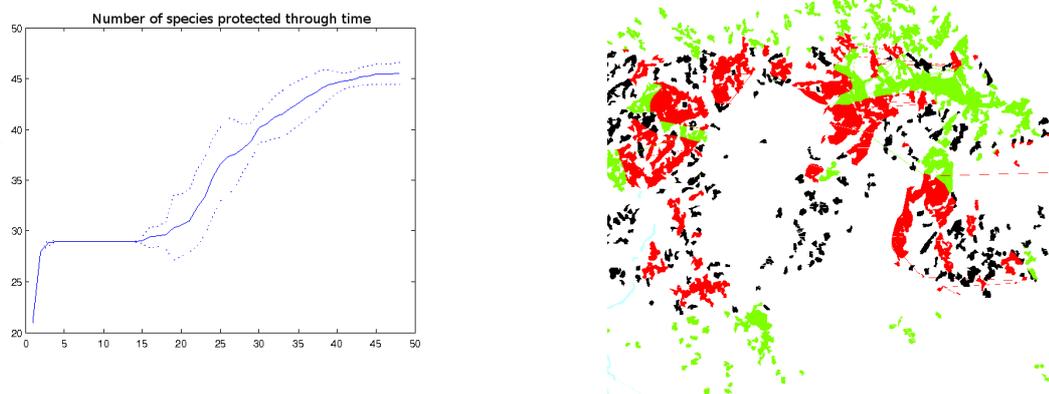


FIG. 6.4 – Gauche : espérance du nombre d'espèces protégées au cours du temps. Droite : une simulation après cinquante pas de temps (sites développés en noir, sites réservés en rouge, non-réservés en vert).

conduisent à des nombres d'espèces finalement sauvegardées de moyenne plus faible et de variance plus grande.

6.2.3 Gestion forestière et risque de tempête

Contexte : Coencadrement de la thèse de Nicklas Forsell (co-tutelle INRA-MIA Toulouse, SLU Umeå, Suède).

Période : 2005-2009 (les thèses durent quatre ans en Suède).

Partenaires : Coencadrement de la thèse : O. Eriksson, P. Wikstrom (SLU, Umeå), F. Garcia, R. Sabbadin (INRA-MIA Toulouse).

Description des activités concernées

Après de premiers travaux consistant en l'application de méthodes d'apprentissage par renforcement distribué à la gestion forestière multiparcelles sur un problème jouet [22], [14], nous avons entamé avec F. Garcia un travail plus poussé autour de la gestion forestière. Ce travail a pris la forme d'un coencadrement de thèse en cotutelle avec des chercheurs "modélisateurs" du SLU, institut Suédois spécialisé en agroforesterie. La thèse de Nicklas Forsell a débuté par des travaux méthodologiques sur la résolution de processus décisionnels de Markov sur graphes par programmation linéaire approchée (PLA) [21], [20] (voir aussi le chapitre précédent), puis s'est poursuivie par la modélisation dans le cadre des PDMG d'un problème réel de gestion forestière sous risque de tempête et sa résolution par la méthode de PLA [16].



FIG. 6.5 – Zone d'étude de Björnstorp.

Le cas d'étude considéré se situe dans la région de Björnstorp, dans le sud de la Suède (figure (6.5)). La zone en question couvre 2800 ha dont 1200 boisés et séparés en 623 parcelles (unités élémentaires de gestion). L'état courant est représenté par la classe d'âge des arbres de chaque parcelle ($\mathcal{X}_i = \{1, \dots, 5\}$ où $x_i = k$ si les arbres de la parcelle ont entre $20(k - 1)$ et $20k$ ans). La figure (6.5) montre l'état courant du système en 2007 : les parcelles grisées représentent les parcelles forestières et les niveaux de gris indiquent la classe d'âge de la parcelle.

Dans ce problème de gestion forestière à long terme, deux actions peuvent être envisagées pour chaque parcelle durant chaque période, de durée vingt ans ($|\mathcal{A}_i| = 2$) : entretien normal de la parcelle ou coupe claire durant la période. Le principal facteur d'aléa provient du risque de forte tempête aux effets dévastateurs. En 2005, par exemple, une tempête exceptionnelle a déposé 75 millions de m^3 de bois dans le sud de la Suède, mais habituellement quatre millions de m^3 sont touchés en moyenne chaque année (18 millions de m^3 en Europe) (Valinger and Fridman, 1997). L'effet potentiel d'une tempête sur chaque parcelle est double :

- Ces tempêtes affectent la dynamique de la parcelle en l'endommageant suffisamment pour "forcer" une coupe claire, même si celle-ci n'était pas décidée à l'avance.
- Elles affectent également le revenu immédiat lié aux actions car une coupe forcée, résultant d'une tempête endommageant une parcelle, engendre un revenu moindre que celui d'une coupe claire.

Considérons tout d'abord le modèle de transition $p(x^{t+1}|x^t, a^t)$. Le risque de dommage à une parcelle par tempête est lié aux caractéristiques propres de cette parcelle (âge, densité. . .) ainsi que des parcelles voisines (différence de hauteur des arbres à la frontière entre les parcelles, angle formé par la frontière avec la direction du vent dominant, etc.). (Olofsson and Blennow, 2005) ont proposé un modèle permettant d'identifier les frontières "à risque" à partir des caractéristiques des deux parcelles la partageant. Dans [16], nous avons étendu ce modèle, permettant ainsi de construire pour chaque parcelle i , une table de probabilité conditionnelle, $p_i^d(\text{dommage}_i|x_{N(i)})$. Cette table définit la probabilité de dommage par tempête à la parcelle i sur une période de 20 ans, étant donné l'âge de cette parcelle et de ses voisines. A partir de là, il est facile de définir la relation de voisinage du PDMG (à partir de la donnée des "frontières"), puis de calculer les tables de probabilités de transition locales. Nous avons également construit un modèle de récompenses locales, $\{r_i(x_{N(i)}, a_i)\}$. Les récompenses locales sont liées au revenu généré sur chaque parcelle sur une période de 20 ans. Or ce revenu est une fonction de l'âge de la parcelle, du mode de gestion choisi pour la période, ainsi que de l'occurrence de dommages liés aux tempêtes sur la période : $\hat{r}_i(x_i, a_i, d_i)$, où d_i est une variable binaire représentant l'occurrence éventuelle de dommages liés aux tempêtes. Connaissant la probabilité conditionnelle de dommages liés aux tempêtes sur une période, $p_i^d(d_i|x_{N(i)})$, nous pouvons déterminer une fonction de récompense locale "en espérance" :

$$r_i(x_{N(i)}, a_i) = \sum_{d_i=1,2} p_i^d(d_i|x_{N(i)}) \cdot \hat{r}_i(x_i, a_i, d_i). \quad (6.1)$$

Plusieurs modèles de PDMG ont été générés, sous différentes hypothèses quant à la probabilité de dommage [12]. Les politiques obtenues ont été évaluées par simulation et comparées à la politique obtenue en négligeant le risque de tempête. Les résultats sont variables suivant les hypothèses (voir [12]), mais les conclusions générales suivantes ont été tirées :

1. Les politiques obtenues en considérant ou non le risque de tempête diffèrent pour les trois quarts des sites environ (entre 60% et 80%). Dans tous les cas, cette différence se traduit par un raccourcissement de la période de rotation (intervalle de temps entre deux coupes claires) lorsque les tempêtes sont considérées.
2. La différence de valeur actuelle nette (VAN : revenu total espéré) entre les deux types de politiques est relativement faible sur la forêt entière (moins de 2%). Néanmoins, une différence de 2% est loin d'être négligeable, étant données les sommes en jeu.
3. De plus, cette différence s'accroît fortement si on ne considère que les parcelles pour lesquelles les politiques diffèrent. Restreinte à ces parcelles, l'accroissement de VAN est de l'ordre de 4% à 8% en moyenne, selon les hypothèses retenues.

6.3 Problèmes de gestion spatialisée en épidémiologie

6.3.1 Gestion de la durabilité des résistances aux pathogènes : exemple du phoma du colza

Contexte :

Participation au projet *Exploitation durable de résistances aux maladies chez les végétaux* du programme *Agriculture et Développement Durable* de l'ANR. *Work Package 2* : Optimisation de l'utilisation de résistances variétales pour une agriculture durable.

Période : 2005-2008.

Partenaires : Collaboration avec N. Peyrard (MIA-Toulouse) et J.-N. Aubertot (EA-Grignon puis EA-Toulouse), entre autres.

Description des activités concernées

Le développement d'une agriculture durable nécessite de réduire les dommages aux cultures occasionnés par des organismes pathogènes, en réduisant la taille de leurs populations et en limitant leur évolution, tout en réduisant l'usage de pesticides. Dans le domaine des productions végétales, il est possible d'agir au niveau de la plante hôte, de la parcelle ou du bassin de production. Jusqu'à présent, c'est majoritairement l'utilisation de gènes majeurs dans les variétés cultivées qui a été préconisée. Or, plusieurs de ces gènes ont été contournés par les populations d'agents pathogènes et peu de stratégies de gestion durable des résistances ont été mises en œuvre en pratique, du fait du manque d'outils permettant de proposer des stratégies de gestion performantes.

En choisissant un ensemble de pathosystèmes représentatifs, l'objectif du projet est de proposer des stratégies de gestion performantes des résistances variétales, prenant en compte la diversité des systèmes de culture et les questions d'acceptabilité et de réglementation pouvant accompagner leur mise en œuvre. Ce projet repose sur une approche pluridisciplinaire en sciences socio-économiques et biotechniques entre des équipes de recherche et des partenaires (ICTA, services régionaux, groupements de producteurs...).

Dans ce projet de grande ampleur (16 équipes de recherche et 8 instituts techniques agricoles impliqués, budget global de 600000 euros) notre rôle, en tant qu'équipe spécialiste en modélisation mathématique et informatique, est de proposer un modèle mathématique pour le problème générique d'optimisation de la durabilité des résistances variétales et d'illustrer son intérêt sur un exemple d'interaction culture hôte / pathogène. Nous avons choisi, en collaboration avec J.-N. Aubertot (chercheur en agronomie, INRA) de mettre en œuvre le cadre des PDMG sur un cas d'étude simple. Il s'agit d'optimiser la marge brute sur une région où chaque parcelle est emblavée selon une rotation classique colza-blé-orge, dans le cas de l'adaptation des populations de pathogène du colza *Leptosphaeria maculans* à une résistance spécifique. La population pathogène ne comporte que deux pathotypes : un virulent et un avirulent vis-à-vis d'une seule résistance spécifique. Deux variétés de colza sont considérées, l'une possédant la résistance et l'autre pas. La variété sans résistance spécifique a un rendement potentiel plus élevé que celle avec la résistance. Les deux seuls modes d'action considérés pour la protection des cultures sont le contrôle génétique (choix variétal) et un contrôle cultural au niveau du système de culture (enfouissement des résidus de culture par labour). La sévérité de l'attaque au niveau d'une parcelle est résumée par un indicateur, la note G2, obtenue par l'observation de la sévérité de la nécrose sur des sections de tige réalisées au niveau du collet. L'objectif est de déterminer une

stratégie optimale, maximisant les marges brutes cumulées dans le temps : L'hypothèse est faite que si une résistance est contournée, alors le rendement "s'écroule". Les deux objectifs d'optimisation des marges brutes sur le long terme et d'optimisation de la durabilité sont donc liés.

Durant la première année du projet (2005-2006), nous avons proposé une modélisation générique du problème de gestion des résistances du colza dans le cadre PDMG⁵. Le modèle proposé est résumé dans la figure (6.6). Un bassin de production est constitué d'un ensemble de parcelles cultivées suivant une rotation (annuelle) colza-blé-orge. Le pathogène *Leptosphaeria maculans* est un champignon qui se disperse au printemps sous la forme de spores et atteint les feuilles des jeunes plants de colza. Ensuite, il se développe pendant la croissance du plant, en colonisant puis nécrosant la tige, affectant sa croissance et allant éventuellement jusqu'à un phénomène de verse (rupture de la tige). Après la récolte, des résidus de tiges de colza subsistent sur les parcelles qui seront ensuite cultivées en blé l'année suivante. Le pathogène présent sur ces résidus continue son cycle de développement et émettra au printemps suivant des spores qui pourront aller coloniser les parcelles qui seront alors cultivées en colza.

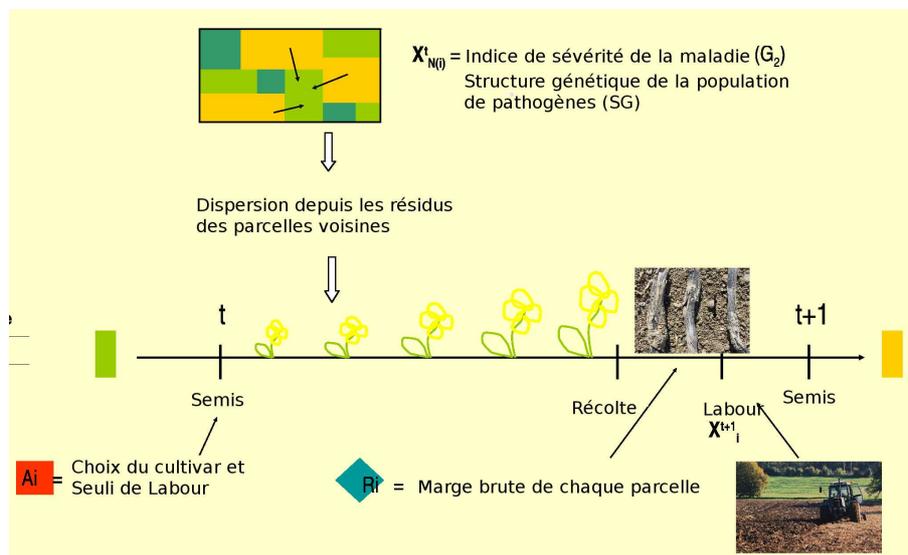


FIG. 6.6 – Cycle de développement et de dispersion de *Leptosphaeria maculans*.

Dans le modèle de PDMG construit, les nœuds du graphe correspondent aux différentes parcelles et les arêtes relient les parcelles voisines (hypothèse de dispersion des spores à distance réduite). Les variables d'état $\{X_i^t\}$ représentent l'état de chaque parcelle. Les caractéristiques déterminantes dans l'évolution de la maladie au niveau du bassin de production sont :

- la quantité de spores présentes sur les résidus de colza (dans les parcelles cultivées en blé), mesurée indirectement par l'indicateur G_2 , et
- la structure génétique de ces spores, SG (proportion de spores virulentes).

Ces deux variables d'état ont été discrétisées (3 classes de note G_2 , 5 classes pour SG) et le domaine de chaque variable X_i^t est de taille $17 : 5 \times 3$ (pour les parcelles en blé) + 2 (les parcelles en colza ou orge ne comprennent plus de résidus).

⁵E.-A. Guersch. Mise en œuvre des processus décisionnels de Markov sur graphe (PDMG) pour la gestion durable des résistances. Stage de Master I d'Ingénierie Mathématiques (Université Paul Sabatier, Toulouse), 2006.

Les actions disponibles pour chaque parcelle cultivée en colza sont le choix d'une variété résistante ou non (CV) et le choix d'un seuil (S) de note G2 à partir duquel un labour, enfouissant partiellement les résidus de colza, sera effectué après la récolte. Les domaines des variables d'action sont donc de taille 2×4 (pas de travail du sol, ou travail à partir d'un des 3 niveaux de note G2).

Les fonctions de transition locales $p_i(x_i^{t+1} | x_{N(i)}^t, a_i^t)$ du PDMG ont ensuite été déterminées à partir d'un modèle composite, comportant trois modules, *production de spores*, *dispersion* et *infection*, développés en se basant sur des modèles biologiques⁶ [29] (voir figure (6.7)). Des fonctions de récompense locales, de la forme $r_i(x_i^t, a_i^t, x_i^{t+1})$, tenant compte des rendements après observation des notes G2 et des coûts des actions appliquées, ont également été calculées.

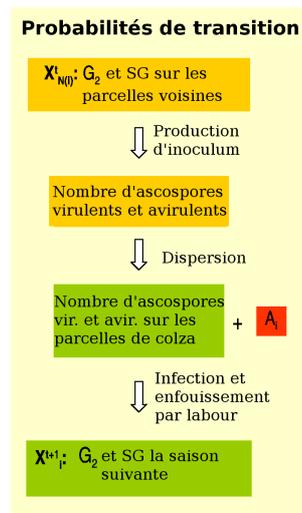


FIG. 6.7 – Les trois étapes du calcul des probabilités de transition locales.

Actuellement, nous sommes dans la phase de validation du modèle par rapport aux données biologiques et d'évaluation des politiques retournées par notre modèle, par comparaison (par simulation) avec des politiques expertes⁷.

6.3.2 Epidémiologie humaine : contrôle de maladies infantiles

Contexte : Collaboration initiée dans le cadre d'un projet financé par le GDR COMEVOL (interactions biotiques dans les communautés : théorie et modèles) : *Dynamique, Evolution et Contrôle des Maladies Infectieuses sur Réseaux d'Interactions*.

Période : 2007-2008.

Partenaires : N. Peyrard (MIA-Toulouse, INRA), M. Choisy (laboratoire de Génétique et Évolution des Maladies Infectieuses, Montpellier, IRD), A. Franc (EFPA-Bordeaux, UMR BioGeCo, INRA).

⁶M. Laviron. Mise en œuvre des Processus Décisionnels de Markov sur Graphes (PDMG) pour la gestion durable des résistances. Stage de seconde année de SUP'AERO, 2007.

⁷B. Borgy. Gestion durable de résistances variétales au phoma du colza sur un bassin de production à l'aide d'un modèle de PDMG. Stage de Master 2 SEVAB, Université Paul Sabatier, 2008.

Description des activités concernées

Au cours des trois dernières décennies, le développement des modèles mathématiques a permis de mieux comprendre et contrôler la dynamique et l'évolution des maladies infectieuses. Ainsi, l'exploitation des modèles déterministes de type SIR a été particulièrement riche dans l'étude des maladies infantiles bactériennes et virales telles que la rougeole, la coqueluche, la varicelle ou les oreillons. Les modèles de type SIR compartimentent la population hôte selon son état clinique en individus susceptibles (S), infectieux (i.e. contaminés et contagieux, I) et guéris (R). Les flux d'individus d'un compartiment à l'autre sont gouvernés par des taux fixes et la dynamique du système global est décrite par un système d'équations différentielles.

Aujourd'hui, les progrès réalisés dans les techniques mathématiques et dans la puissance de calcul informatique permettent d'aborder des questions de plus en plus complexes. Les modèles stochastiques, plus appropriés que les modèles déterministes pour modéliser un début d'épidémie ou une épidémie dans une population hôte de petite taille, sont maintenant communément utilisés. L'hétérogénéité de la population hôte, en termes de sexe, de classes d'âge, de profil génétique, de localité géographique, etc. est un autre facteur limitant l'efficacité des modèles déterministes. Dans ce contexte, nous avons commencé à développer un modèle stochastique sur réseaux d'interactions permettant d'évaluer l'influence de l'hétérogénéité des contacts entre individus hôtes sur la dynamique des maladies infectieuses (voir figure (6.8) pour un exemple de réseau d'interaction).

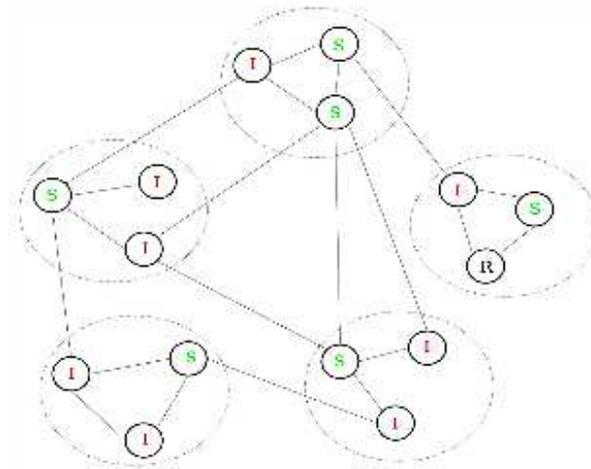


FIG. 6.8 – Graphe d'interactions représentant les contacts entre individus de classes d'âge différentes.

Dans un premier temps, nous avons élaboré un modèle de type PDMG permettant de modéliser un problème de contrôle d'épidémie par isolement d'individus⁸. Les actions d'isolement étaient modélisées par des retraits (temporaires) d'arêtes du graphe de contact, non intégrables directement dans le modèle PDMG puisque celui-ci ne permet de prendre en compte des actions ne modifiant que l'état des nœuds du graphe. Néanmoins, nous avons montré que ce problème pouvait encore être modélisé par un PDMG de taille plus grande que le problème initial. Nous sommes actuellement dans une phase de comparaison expérimentale de politiques obtenues par résolution de PDMG et de politiques expertes. A cet effet, un simulateur de dynamiques épidémiques, permettant d'évaluer l'effet de poli-

⁸L. Condet. Extension du cadre des PDMG pour l'optimisation de la structure de réseaux d'interactions dans les processus spatio-temporels. Stage de Master 2 Probabilités et Statistiques, Université Paul Sabatier, 2007.

tiques diverses (pas seulement locales), est en cours de réalisation⁹.

6.3.3 Détection et éradication d'espèces nuisibles : cas des “fire ants”

Contexte : Participation à un projet ARC (Australian Research Council) : *Applying search theory for eradicating invasive species*.

Période : 2007-2009.

Partenaires : D. Spring et R. Mac Nally (Monash University, Clayton, Australie), B. Brook (Adelaïde University, Australie), B. Leung (Mac Gill University, Canada).

Description des activités concernées

Les espèces invasives ont, partout dans le monde, un impact environnemental et économique majeur. C'est encore plus vrai en période de changement climatique brutal, qui bouleverse l'équilibre des écosystèmes locaux. L'objectif général de ce projet est de proposer un ensemble de méthodes pour le diagnostic et l'éradication d'espèces invasives. Plus précisément, les objectifs du projet sont de développer des outils mathématiques et informatiques pour :

1. la délimitation de l'étendue spatiale d'une infestation.
2. l'éradication d'une espèce invasive dans une région donnée.
3. la détermination de régions garanties sans présence de l'espèce invasive.

Les méthodes en question doivent être testées sur un certain nombre de cas réels d'invasion recensés en Australie : invasion de fourmis venimeuses *fire ants* (dont les contacts avec l'homme sont assez douloureux, voir figure (6.9), partie gauche), ou d'une espèce de crapeau venimeux *cane toad* introduite dans le Queensland en 1935 et qui s'est révélée un désastre pour de nombreuses espèces locales (concurrentes pour les ressources, ou prédatrices éventuelles empoisonnées par son venin), voir figure (6.9), partie droite. . .

Ma contribution dans ce projet en cours est de proposer une méthode de diagnostic optimal d'invasion de *fire ants*, basée sur les *processus décisionnels de Markov épistémiques* (PDME) [41]. Mais la nature “spatiale” du problème de diagnostic en question, ainsi que des données disponibles (voir figure 6.10) nécessite d'adapter le cadre des PDME à des états de croyance représentés par des *champs de Markov* (Guyon, 1995).

Les deux étapes du projet (en cours de réalisation) me concernant sont les suivantes :

- Définition du cadre des PDME avec états de croyance représentés sous la forme d'un *champ de Markov par paire* et définition de méthodes d'optimisation de stratégies épistémiques adaptées à ce cadre.
- Utilisation de méthodes *d'analyse d'image* (Forbes et al., 2003) pour estimer un état de croyance initial du problème de planification épistémique, sous la forme d'un *modèle de Potts* (Wu, 1982), qui est un cas particulier de champ de Markov par paire.

⁹M.-A. Rafik. Développement d'un simulateur de dynamiques épidémiques sur réseau. Application au contrôle de maladies infantiles dans une population structurée en classes d'âge. Stage de Master 1 Informatique et Gestion. Université Paul Sabatier, 2008.



FIG. 6.9 – Gauche : morsures de fourmis *fire ants*. Droite : un crapaud *cane toad*, de bonne taille.

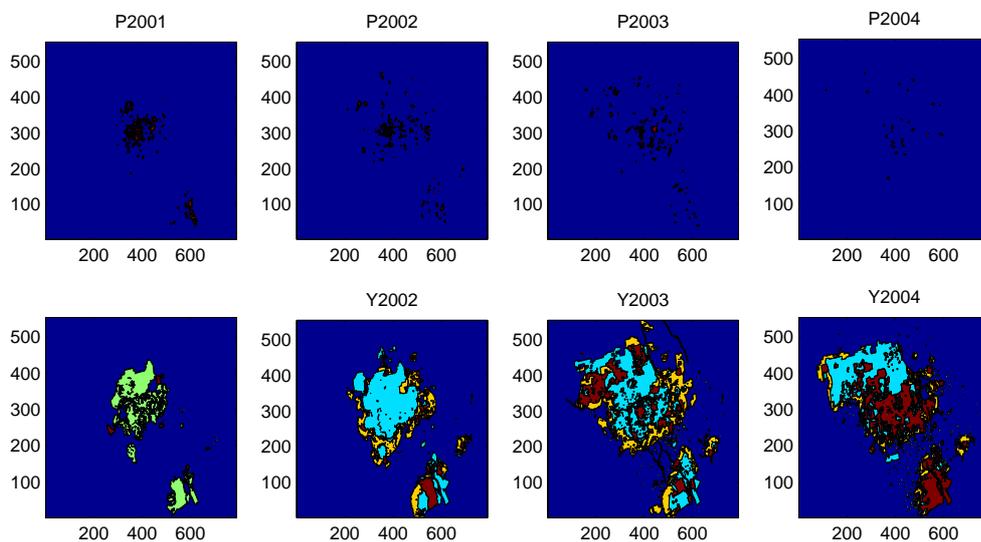


FIG. 6.10 – Haut : données (grille d'environ 800×600 pixels) de présence/absence de nids de fourmis. Bas : actions de diagnostic et/ou éradication effectuées. Données pour les années 2001 à 2004.

Quatrième partie

Conclusion

Chapitre 7

Conclusion et perspectives

7.1 Résumé de mes travaux et de mes perspectives

7.1.1 Décision qualitative possibiliste

Ces dix dernières années, mes travaux concernant la décision qualitative ont été focalisés sur les fondations axiomatiques de la décision qualitative possibiliste et l’algorithmique pour la planification possibiliste dans l’incertain.

Mes travaux sur l’axiomatisation des critères de décision qualitative possibiliste et monotone ont visé en premier lieu à expliciter les hypothèses sur lesquelles se basent ces critères de décision (commensurabilité entre préférences et incertitude, non-compensation entre les conséquences des actes...). Ils ont également contribué à proposer des procédures d’élicitation de préférences et de connaissances dans un cadre possibiliste, ou plus généralement d’utilité qualitative monotone, à partir de comparaison d’actes élémentaires [3], [45]. Mes travaux sur les liens entre utilité qualitative possibiliste et utilité espérée ont permis de montrer que ces deux théories n’étaient pas complètement inconciliables. La théorie qualitative admet des raffinements basés sur des ordres de grandeur d’utilité espérée.

Le second axe de mes travaux en décision possibiliste concerne plus précisément le domaine de la planification dans l’incertain. Mes travaux sur la programmation dynamique possibiliste et les PDM possibilistes ont constitué un premier pas vers une approche qualitative de la planification dans l’incertain. Ces travaux ont été poursuivis par l’étude de langages de représentation structurée, basés sur la logique possibiliste ou les diagrammes d’influence possibilistes. Ceux-ci permettent de représenter de manière concise des problèmes de planification dans l’incertain possibiliste, et de les résoudre à l’aide d’algorithmes spécifiques.

7.1.2 Processus décisionnels de Markov

Au cours de ces dernières années, mes travaux concernant les processus décisionnels de Markov se sont focalisés sur les deux axes suivants :

- processus décisionnels de Markov factorisés et,
- processus décisionnels de Markov épistémiques.

J’ai tout d’abord travaillé sur des méthodes de décomposition de PDM de grande taille. J’ai proposé, en particulier, une méthode de décomposition *sérielle* de PDM de grande taille basée sur la *bipartition*

spectrale [38]. Ensuite, j'ai travaillé sur un modèle de représentation de PDM factorisés basé sur les graphes, et j'ai participé à l'élaboration d'algorithmes de résolution approchée de ces *PDM sur graphes*, utilisant des méthodes de *programmation linéaire approchée* [21], ou *d'approximation en champ moyen* [28].

Plus récemment, j'ai commencé à étudier la classe des problèmes de *planification épistémique*. Nous avons choisi de modéliser ces problèmes d'optimisation d'acquisition de connaissances dans un cadre proche de celui des processus décisionnels de Markov partiellement observables [41].

Ces travaux méthodologiques sur les processus décisionnels de Markov ont été illustrés par un certain nombre d'applications, dans les domaines de la gestion des ressources naturelles (forêts, biodiversité) ou de l'épidémiologie.

7.1.3 Résumé de mes perspectives

En une phrase, mes perspectives de recherche se situent dans le domaine du développement de modèles mathématiques et d'algorithmes pour résoudre des problèmes de gestion agri-environnementale. Plus précisément, les problèmes en question sont liés à la protection de la biodiversité ou à la lutte contre les ravageurs (virus, parasites, mauvaises herbes... "Pests" en anglais) affectant cultures, élevages et forêts.

Ce sont donc clairement des problématiques de recherche finalisée qui motivent mes choix de recherche pour les années qui viennent. Mais ces problématiques finalisées possèdent des caractéristiques communes qui justifient l'exploration d'un certain nombre de pistes méthodologiques en intelligence artificielle :

- Les données sont spatialisées et temporelles. Les données sont souvent disponibles sous la forme de cartes (SIG) d'éléments d'intérêt observés à différentes dates : cas relevés de maladies, présence avérée d'individus d'espèces en danger, etc. Ces données sont utilisées afin de reconstituer des dynamiques de processus spatialisés, sous l'influence d'actions de contrôle.
- Les données sont en général incomplètes et/ou imprécises. Souvent, les cartes disponibles ne représentent pas directement l'état du processus d'intérêt, mais sa manifestation visible, qui ne permet pas une reconstitution directe du processus "caché". C'est le cas en épidémiologie en particulier, où les cas de maladie avérés sont bien recensés, mais où la dynamique réelle du pathogène n'est pas observée directement.

Afin de modéliser et de résoudre des problèmes de gestion agri-environnementale de ce type, je suis persuadé qu'il est utile de continuer à développer des méthodes étendant le cadre classique de la planification sous incertitude ou l'adaptant à ces problématiques. Les perspectives de recherche que cela entraîne pour moi sont les suivantes :

- Progrès autour du cadre des PDM sur graphe, à la fois sur le plan algorithmique (algorithmes de résolution approchée plus "exacts" ou plus efficaces) et sur la spécialisation du cadre des PDMG aux problèmes rencontrés en épidémiologie.
- Progrès autour de la prise en compte de données réelles, à la fois incomplètes et peu fiables. Cela implique de développer le modèle des PDMG dans un cadre partiellement observable : développement de méthodes d'échantillonnage spatial efficaces afin d'améliorer la qualité des données ou développement d'outils de type PDMPO sur graphes. Mais cela peut également impliquer de poursuivre mes travaux sur les approches qualitatives de la décision dans l'incertain, afin de les adapter à des problèmes de décision spatialisée. Ceci devrait permettre de raisonner directement à partir de données éparses et de mauvaise qualité.

Je vais détailler ces perspectives de recherche dans les sections suivantes. Elles se situent bien entendu dans la continuité de mes recherches actuelles. Ces perspectives sont nombreuses et il m'est difficile de dégager des priorités nettes entre ces voies de recherche relativement différentes (et importantes, me semble-t-il). Néanmoins, je me suis clairement engagé pour les prochaines années dans la poursuite de travaux sur l'adaptation du cadre des PDMG aux problèmes de gestion d'épidémies et d'échantillonnage spatial. Cet engagement est formalisé par de nombreuses collaborations, en cours ou en construction, avec des partenaires écologiques ou agronomes.

7.2 Perspectives détaillées autour de la décision qualitative possibiliste

7.2.1 Axiomatisation des raffinements des utilités qualitatives possibilistes

Nous avons montré que les utilités qualitatives possibilistes admettaient des raffinements prenant la forme de critères d'utilité espérée très particuliers, de type *lexicographiques*. En tant que critères d'utilité espérée, ces raffinements lexicographiques respectent les axiomes de Savage¹. La prochaine étape consiste à construire les axiomes supplémentaires permettant de caractériser exactement ces critères lexicographiques.

7.2.2 Raffinements de la décision séquentielle possibiliste, apprentissage par renforcement

J'ai décrit dans la section (3.3) nos travaux visant à relier les théories de *l'utilité espérée* et des *utilités qualitatives possibilistes* par des relations de raffinement. Il serait pertinent d'étendre ces travaux au cas de la décision séquentielle afin d'établir un lien entre PDM stochastiques et possibilistes. En effet, les liens existant dans le cadre stochastique entre probabilités *objectives* et *subjectives* permettent la simulation des modèles de transition des PDM, et permettent d'offrir des approches de type *apprentissage par renforcement* (Sutton and Barto, 2002) pour résoudre des problèmes de décision séquentielle dans l'incertain. Ces approches utilisent des simulateurs (éventuellement complexes), pour lesquels le modèle de transition est inconnu et n'est pas forcément une probabilité markovienne. Or, si l'expression de la vraisemblance des transitions en termes de degrés de possibilité est intuitive et permet de représenter avec moins de biais que des probabilités numériques la connaissance d'experts, il n'existe pas de moyen formel universellement reconnu de *simuler* des transitions possibilistes ou de construire les degrés de possibilité de transitions à partir d'un simulateur.

Si nous parvenions à étendre les travaux entamés avec Hélène Fargier sur les raffinements d'utilités possibilistes par des utilités espérées, au cas séquentiel², nous pourrions construire des PDM stochastiques induisant un ordre sur les politiques qui *raffine* l'ordre induit sur les mêmes politiques dans un PDM possibiliste donné. Il deviendrait alors possible d'appliquer des algorithmes *d'apprentissage par renforcement*, basés sur la simulation des transitions du PDM stochastique, afin de calculer des politiques optimales pour le PDM possibiliste dont le modèle serait décrit par un expert.

¹Les cinq premiers seulement, puisque nous considérons un espace d'état fini.

²En passant à des ordres du type *leximin*(*leximax*(*leximin*)) sur des tableaux à trois dimensions pour tenir compte de l'agrégation des degrés de satisfaction au cours du temps.

7.2.3 Processus décisionnels de Markov possibilistes structurés

Il reste encore à faire la synthèse entre nos travaux sur les Processus décisionnels de Markov possibilistes à horizon infini et nos travaux sur les diagrammes d'influence possibilistes. Dans le cadre des PDM classiques, (Boutilier et al., 2000) ont proposé des algorithmes pour la résolution de PDM stationnaires à horizon infini *structurés*, du type *itération de la valeur* et *itération de la politique*.

Dans le cadre possibiliste, la même démarche pourrait être entreprise afin de construire des politiques stationnaires optimales pour des PDM possibilistes structurés. Dans ce cas, une forme de *réseau bayésien homogène stationnaire possibiliste* serait utilisé pour représenter les indépendances dans les possibilités de transition et les récompenses liées aux transitions. La question de la complexité de la résolution de tels problèmes³ est ouverte : dans le cas stochastique, elle passe de PSPACE à EXP-TIME, qu'en est-il dans le cas possibiliste ?

7.3 Perspectives détaillées autour des PDM structurés

7.3.1 Processus décisionnels de Markov sur graphe

Algorithmique des processus décisionnels de Markov sur graphe

Nos travaux sur la résolution approchée des processus décisionnels de Markov sur graphe doivent être poursuivis. Il n'existe pour l'instant aucune borne théorique satisfaisante sur la perte d'optimalité des politiques fournies par les algorithmes de programmation linéaire approchée (PLA) ou d'approximation en champ moyen (CM). Nous sommes en train en ce moment, avec Nicklas Forsell et Nathalie Peyrard, de proposer une comparaison expérimentale des deux approches avec l'approche heuristique "gloutonne" qui consiste à choisir la politique (locale) maximisant la fonction de récompense locale immédiate. Cette politique locale donne en général, mais pas toujours, de bons résultats.

Dans le même temps, nous cherchons à améliorer l'efficacité des algorithmes précédent, en proposant des schémas d'approximation plus complexes (consistant à accroître la taille des voisinages, ou à inclure les variables d'action dans les voisinages). Ces schémas d'approximation plus complexes conduisent également à envisager des méthodes de résolution par simulation de type *apprentissage par renforcement multi-agents* (travaux en cours avec Nicklas Forsell et Frédéric Garcia).

Classes de processus décisionnels de Markov sur graphe de complexité réduite

Puisque la résolution exacte de PDMG est hors de portée, et puisque nous n'avons pas non plus à disposition d'algorithmes de résolution approchée retournant des politiques de qualité garantie dans le cas général, il est logique de rechercher des classes de PDMG, pertinentes en tant que modèles mais plus "simples" à résoudre.

Dans le domaine de l'épidémiologie, les *processus de contacts sur graphe* peuvent être utilisées pour modéliser l'évolution d'épidémies (végétales, animales, humaines). Nous avons commencé à étudier de manière expérimentale, avec Marc Choisy (IRD / Montpellier), Alain Franc (EFPA-INRA / Bordeaux) et Nathalie Peyrard (MIA-INRA / Toulouse), le contrôle de tels processus de contacts sur graphe, et nous avons montré qu'ils pouvaient se modéliser dans le cadre des PDMG. Dans le même temps, j'ai débuté une étude plus théorique de ces différentes sous-classes intéressantes de PDMG d'un type particulier. Suivant les cas de processus de contact et de modes de contrôle considérés, le problème d'existence de politique de contrôle de valeur supérieure à un certain seuil peut appartenir

³De l'existence de politique d'utilité dépassant un certain seuil.

aux classes de complexité PSPACE , voire NP-complet . Il nous reste maintenant à rechercher des algorithmes de résolution (exacte ou approchée) spécifiques à ces problèmes et garantissant une solution de qualité bornée.

7.3.2 Processus décisionnels de Markov partiellement observables sur graphe

Processus décisionnels de Markov épistémiques sur graphe

Le cadre des processus décisionnels de Markov épistémiques peut être généralisé afin de prendre en compte des croyances (*belief states*) exprimées sous la forme de *champs de Markov* (Guyon, 1995). Une telle généralisation peut être utile pour modéliser des problèmes de diagnostic d'espèces invasives, comme celui (présenté dans le chapitre précédent) que nous étudions dans le cadre du projet ARC "control of invasive species".

L'extension envisagée des processus décisionnels de Markov épistémiques engendre des questions méthodologiques intéressantes à la fois en *intelligence artificielle* (optimisation de stratégies épistémiques), en *statistiques spatiales* (estimation / reconstruction de cartes) et autour des *modèles graphiques* (approximation de lois marginales / modes de distributions). Cette extension, envisagée avec Nathalie Peyrard et Jérôme Lang, pourrait prendre la forme d'une thèse méthodologique coencadrée, motivée par des applications réelles (le problème des *fire ants*, ou un problème de gestion d'aventices, formalisé avec Sabrina Gaba, INRA-Dijon).

Processus décisionnels de Markov partiellement observables sur graphes

Le cadre général des processus décisionnels de Markov structurés partiellement observables⁴ étend à la fois les cadres des PDM sur graphe et des PDM épistémiques sur lesquels j'ai travaillé. Ce cadre est extrêmement général, et les classes de problèmes représentés peuvent être indécidables. Néanmoins, il est fort probable qu'un certain nombre de sous-classes de ce modèle, pertinentes pour représenter des problèmes réels de conduite de systèmes agri-environnementaux, soient de complexité "raisonnable". La recherche de telles sous-classes et d'algorithmes de résolution approchée adaptés fait également partie de mes perspectives de recherche.

⁴Pour être plus précis, il s'agit du cadre des *belief MDP* structurés, un peu plus général.

Bibliographie

- Akström, K. J. (1965). Optimal control of markov decision processes with incomplete state estimation. *Journal of Mathematical Analysis and Applications* 10, 174–205.
- Allais, M. (1953). Le comportement de l'homme rationnel devant le risque : critique des postulats et axiomes de l'école américaine. *Econometrica* 21, 503–546.
- Altman, E. (1999). *Constrained Markov Decision Processes*. Chapman & Hall / CRC.
- Anscombe, A. and R. J. Aumann (1963). A definition of subjective probability. *Annals of Mathematical Statistics* 34, 199–205.
- Bellman, R. E. (1957). *Dynamic Programming*. Princeton University Press, Princeton.
- Bernstein, D., R. Givan, N. Immerman, and S. Zilberstein (2002). The complexity of decentralized control of markov decision processes. *Mathematics of Operations Research* 27(° 4), 819–840.
- Bertsekas, D. P. (1987). *Dynamic Programming : Deterministic and Stochastic Models*. Englewood Cliffs : Prentice-Hall.
- Bertsekas, D. P. and J. N. Tsitsiklis (1996). *Neuro-Dynamic Programming*. Belmont (MA) : Athena Scientific.
- Blythe, J. (1999). An overview of planning under uncertainty. *AI magazine* 20(2), 37–54.
- Boutilier, C. (1994, 24-27 mai). Toward a logic for qualitative decision theory. In P. T. e. J. Doyle, E. Sandewall (Ed.), *Proc. 4th Inter. Conf. on Principles of Knowledge Representation and Reasoning (KR'94)*, Bonn, Allemagne, pp. 75–86.
- Boutilier, C. (1996). Planning, learning and coordination in multiagent decision processes. In *Sixth Conference on Theoretical Aspects of Rationality and Knowledge (TARK'96)*, Amsterdam, pp. 195–210.
- Boutilier, C. (2002). A POMDP formulation of preference elicitation problems. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI 2002)*, pp. 239–246. AAAI.
- Boutilier, C., T. Dean, and S. Hanks (1999). Decision-theoretic planning : Structural assumptions and computational leverage. *Journal of Artificial Intelligence Research* 11, 1–94.
- Boutilier, C., R. Dearden, and M. Goldszmidt (1995). Exploiting structure in policy construction. In *Proc. of the 14th International Joint Conference on Artificial Intelligence (IJCAI'95)*, Montreal, Canada, pp. 1104–1111. Morgan Kaufman.

- Boutilier, C., R. Dearden, and M. Goldszmidt (2000). Stochastic dynamic programming with factored representations. *Artificial Intelligence* 121(1), 49–107.
- Brafman, R. and M. Tennenholtz (1997). Modelling agents as qualitative decision makers. *Artificial Intelligence* 94, 217–268.
- Cassandra, A. R. (1998). *Exact and Approximate Algorithms for Partially Observable Markov Decision Processes*. Ph. D. thesis, Dept. of Computer Science, Brown University.
- Cassandra, A. R., L. P. Kaelbling, and M. L. Littman (1994). Acting optimally in partially observable stochastic domains. In *Proc. 11th Nat. Conf. on Artificial Intelligence (AAAI'94)*, pp. 1023–1028. AAAI Press.
- Chandler, D. (1987). *Introduction to Modern Statistical Mechanics*. Oxford University Press.
- Chateauneuf, A. (1999). Comonotonicity axioms and rank-dependent utility theory for arbitrary consequences. *J. of Mathematical Economics* 32, 21–45.
- Chateauneuf, A. and M. Cohen (2006). *Concepts et méthodes pour l'aide à la décision : risque et incertain*, Volume 2 of *Traité IC2*, Chapter Extensions cardinales du modèle EU basées sur l'intégrale de Choquet. Lavoisier.
- Choquet, G. (1953-1954). Theory of capacities. *Annales de l'Institut Fourier* (5), 131–295.
- Chornei, R. K., H. Daduna, and P. S. Knopov (2006). *Control of Spatially Structured Random Processes and Random Fields with Applications*. Springer.
- Conitzer, V. and T. Sandholm (2003). Definition and complexity of some basic metareasoning problems. In *Proc. of 18th Inter. Joint Conf. on Artificial Intelligence (IJCAI'03)*, Acapulco, Mexico, pp. 1099–1106.
- Costello, C. and S. Polaski (2004). Dynamic reserve site selection. *Resource and Energy Economics* 26, 157–174.
- da Costa, C. (1998). *Planification d'actions en environnement incertain : une approche fondée sur la théorie des possibilités*. Ph. D. thesis, Université Paul Sabatier.
- da Costa Pereira, C., F. Garcia, J. Lang, and R. Martin-Clouaire (1997a). Planning with graded non deterministic actions : a possibilistic approach. *Int. Journal of Intelligent Systems* 12, 935–962.
- da Costa Pereira, C., F. Garcia, J. Lang, and R. Martin-Clouaire (1997b). *Recent advances in AI planning*, Chapter Possibilistic planning : representation and complexity. Lecture Notes in AI. Springer Verlag.
- Darwiche, A. and M. Goldszmidt (1994). On the relation between kappa calculus and probabilistic reasoning. In *Proc. 10th conf. on Uncertainty in Art. Int. (UAI'94)*, Seattle, WA, pp. 145–153. Morgan Kaufmann.
- de Farias, D. P. and B. Van Roy (2003). The linear programming approach to approximate dynamic programming. *Oper. Res.* 51(6), 850–865.
- de Farias, D. P. and B. Van Roy (2004). On constraint sampling in the linear programming approach to approximate dynamic programming. *Mathematics of Op. Research* 29(3), 462–478.

- de Finetti, B. (1937). La prévision : ses lois logiques, ses sources subjectives. *Annales de l'Institut H. Poincaré* 7, 1–68.
- de Ghellinck, G. (1960). Les problèmes de décision séquentielle. *Cahiers du Centre d'Études de Recherche Opérationnelle* 2, 161–179.
- Dean, T. and R. Givan (1997). Model minimization in markov decision processes. In *Proc. of the 14th National Conf. on Artificial Intelligence (AAAI'97)*, Providence, RI, pp. 106–111. AAAI Press.
- Dean, T. and K. Kanazawa (1989). A model for reasoning about persistence and causation. *Computational Intelligence* 5(3), 142–150.
- Dean, T. and S. H. Lin (1995). Decomposition techniques for planning in stochastic domains. In *Proc. of the 14th International Joint Conference on Artificial Intelligence (IJCAI'95)*, Montreal, Canada, pp. 1121–1127. Morgan Kaufman.
- Dearden, R. and C. Boutilier (1997). Abstraction and approximate decision theoretic planning. *Artificial Intelligence* 89, 219–283.
- Dietterich, T. and N. Flann (1995). Explanation-based learning and reinforcement learning : A unified view. In *Proc. 12th International Conf. on Machine Learning (ICML'95)*, pp. 176–184.
- Doyle, J. and R. H. Thomason (1999). Background to qualitative decision theory. *AI magazine* 20(2), 55–68.
- Dubois, D. (1986). Belief structures, possibility theory and decomposable confidence measures on finite sets. *Computers and Artificial Intelligence* 5, 404–416.
- Dubois, D. and H. Fargier (2004). An axiomatic framework for order of magnitude confidence relations. In *Proc. of UAI'04*, pp. 138–145.
- Dubois, D. and H. Fargier (2007). Lexicographic refinements of sugeno integrals. In *9th European Conf on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'07)*, pp. 611–622, 2007.
- Dubois, D., H. Fargier, and P. Perny (2003). Qualitative models for decision under uncertainty : An axiomatic approach. *Artificial Intelligence* 148, 219–260.
- Dubois, D., H. Fargier, and H. Prade (1996). Possibility theory in constraint satisfaction problems. *Applied Intelligence* 6, 287–309.
- Dubois, D., L. Godo, H. Prade, and A. Zapico (2000). Advances in qualitative decision theory : refined rankings. In *Lecture Notes in AI*, Volume 1952, pp. 427–436. Springer Verlag.
- Dubois, D., J. Lang, and H. Prade (1994). Automated reasoning using possibilistic logic : Semantics, belief revision and variable certainty weights. *IEEE Trans. on Knowledge and Data Engineering* 1(6), 64–69.
- Dubois, D. and H. Prade (1988). *Possibility theory*. Plenum Press.
- Dubois, D. and H. Prade (1994). A survey of belief revision and updating rules in various uncertainty models. *Int. Journal of Intelligent Systems* 9, 61–100.

- Dubois, D. and H. Prade (1995). Possibility theory as a basis for qualitative decision theory. In *Proc. of IJCAI'95*, pp. 1925–1930.
- Dubois, D., H. Prade, and S. Sandri (1993). *Fuzzy Logic : State of the Art*, Chapter On possibility/probability transformations, pp. 103–112. Kluwer Academic Publishers.
- Ellsberg, D. (1961). Risk, ambiguity and the savage axioms. *Quarterly Journal of Economics* (75), 643–669.
- Fikes, R. E. and N. J. Nilsson (1971). Strips : a new approach to the application of theorem proving to problem solving. *Artificial Intelligence* 2.
- Filar, J. and K. Vrieze (1997). *Competitive Markov Decision Processes*. Springer Verlag.
- Forbes, F., G. Celeux, and N. Peyrard (2003). EM procedures using mean field-like approximations for markov model-based image segmentation. *Pattern Recognition* 36(1).
- Gallhab, M., D. Nau, and P. Traverso (2004). *Automated planning : theory and practice*. Morgan Kaufmann.
- Giang, P. H. and P. P. Shenoy (1999). On transformations between probability and spohnian disbelief functions. In H. Prade (Ed.), *Proc. 15th Conf. on Uncertainty in AI (UAI'99)*, Stockholm, Sweden, pp. 236–244. Morgan Kaufmann.
- Giang, P. H. and P. P. Shenoy (2000). A qualitative utility theory for spohn's theory of epistemic beliefs. In *Proc. of 16th Conf. on Uncertainty in AI (UAI'00)*, pp. 220–227.
- Giang, P. H. and P. P. Shenoy (2001). A comparison of axiomatic approaches to qualitative decision making using possibility theory. In *Proc. of 17th Conf. on Uncertainty in AI (UAI'01)*, pp. 162–170.
- Giang, P. H. and P. P. Shenoy (2005). Two axiomatic approaches to decision making using possibility theory. *European Journal of Operations Research* 162(2), 450–467.
- Gilboa, I. (1987). Expected utility with purely subjective non-additive probabilities. *Journal of Mathematical Economics* 16, 65–88.
- Gilboa, I. (Ed.) (2004). *Uncertainty in Economic Theory, Essays in Honor of David Schmeidler's 65th Birthday*. Routledge.
- Givan, R., T. Dean, and M. Greig (2003). Equivalence notions and model minimization in markov decision processes. *Artificial Intelligence* 147(1-2), 163–223.
- Godo, L. and A. Zapico (2001). On the possibilistic-based decision model : Characterisation of preference relations under partial inconsistency. *Applied Intelligence* 3(14), 319–333.
- Guestrin, C., D. Koller, R. Parr, and S. Venkataraman (2003). Efficient solution algorithms for factored MDPs. *Journal of Artificial Intelligence Research* 19, 399–468.
- Guyon, X. (1995). *Random Fields on a Network – Modeling, Statistics and Applications*. Probability and its Applications. Springer Verlag.
- Hansen, E. A., D. S. Bernstein, and S. Zilberstein (2004). Dynamic programming for partially observable stochastic games. In *Proceedings of the Nineteenth National Conference on Artificial Intelligence (AAAI'04)*, pp. 709–715.

- Hauskrecht, M. (2000). Value-function approximations for partially observable markov decision processes. *Journal of Artificial Intelligence Research* 13, 33–94.
- Hauskrecht, M., N. Meuleau, L. Kaelbling, T. Dean, and C. Boutilier (1998). Hierarchical solution of markov decision processes using macro-actions. In *UAI'98*.
- Hendrickson, B. and R. Leland (1995). An improved spectral graph partitioning algorithm for mapping parallel computations. *SIAM J. Stat. and Comput.* 16, 452–469.
- Henrion, M., G. Provan, B. Del Faverol, and G. Sanders (1999). An experimental comparison of numerical and qualitative probabilistic reasoning. In *10th Conf. on Uncertainty in AI. (UAI'94)*, Seattle, WA, pp. 319–326. Morgan Kaufmann.
- Howard, R. A. (1960). *Dynamic Programming and Markov Processes*. Cambridge : MIT Press.
- Jaffray, J. Y. (1989). Linear utility for belief functions. *Operations Research Letters* 24(2), 237–267.
- Jaffray, J.-Y. and P. Wakker (1994). Decision making with belief functions : Compatibility and incompatibility with the sure-thing principle. *Journal of Risk and Uncertainty* (8), 1994. Kluwer Academic Publishers.
- Jensen, F. V. (2001). *Bayesian networks and decision graphs*. Springer.
- Kaelbling, L. P., M. L. Littman, and A. R. Cassandra (1998). Planning and acting in partially observable domains. *Artificial Intelligence* 101, 99–134.
- Kahneman, D. and A. Tversky (1979). Prospect theory : An analysis of decision under risk. *Econometrica* 47, 263–291.
- Karp, R. M. (1972). *Complexity of Computer Computations*, Chapter Reducibility among combinatorial problems. R. E. Miller and J. W. Thatcher (eds.). Plenum Press, New York.
- Kearns, M., Y. Mansour, and A. Ng (1999). A sparse sampling algorithm for near-optimal planning in large mdps. In *15th Conf. on Uncertainty in Artificial Intelligence (UAI'1999)*, pp. 21–30.
- Kernighan, B. and S. Lin (1970). An efficient heuristic procedure for partitioning graphs. *Bell System Technical Journal* 29, 291–307.
- Koller, D. and R. Parr (1999). Computing factored value functions for factored mdps. In *IJCAI'99*.
- Koller, D. and R. Parr (2000). Policy iteration for factored mdps. In *16th Conf. on Uncertainty in Artificial Intelligence (UAI'2000)*, pp. 326–334.
- Kraft, C. H., J. Pratt, and A. Seidenberg (1959). Intuitive probability on finite sets. *Annals of Mathematical Statistics* 30, 408–419.
- Kumar, P. R. and P. P. Varaiya (1986). *Stochastic Systems : Estimation, Identification and Adaptive Control*. Englewood Cliffs, New Jersey : Prentice Hall.
- Ladet, S. (2000). Modélisation de la réponse d'espèces d'oiseaux à la fragmentation forestière dans les coteaux du sud-ouest. Master's thesis, DEA d'écologie des systèmes continentaux, Université Paul Sabatier.

- Lang, J. (1991). *Fundamentals of Artificial Intelligence Research*, Chapter Possibilistic logic as a logical framework for min-max discrete optimization problems and prioritized constraints, pp. 112–126. LNAI- vol. 535. P. Jorrand, J. Kelemen eds.
- Lang, J. (1996). Conditional desires and utilities. In *Proc. of ECAI'96*, Budapest, Hongrie, pp. 318–322.
- Laurance, W., A. Albernaz, P. Fearnside, H. Vasconcelos, and L. Ferreira (2004). Deforestation in amazonia. *Science* 304(5674), 1109–1111.
- Littman, M. L. (1994). Memoryless policies : Theoretical limitations and practical results. In *Proc. of 3rd International Confernece on Simulation and Adaptive Behavior*, Cambridge, MA. MIT Press.
- Luce, R. and H. Raiffa (1957). *Games and decision*. New-York : Wiley and Sons.
- Madani, O., S. Hanks, and A. Condon (2003). On the undecidability of probabilistic planning and related stochastic optimization problems. *Artificial Intelligence* 147, 5–34.
- Mair, E., S. Andelman, and H. Possingham (2004). Does conservation planning matter in a dynamic and uncertain world ? *Ecology letters* 7, 615–622.
- Meuleau, N., M. Hauskrecht, K. Kim, L. Peshkin, L. Kaelbling, T. Dean, and C. Boutilier (1998). Solving very large weakly coupled markov decision processes. In *AAAI'98*.
- Moore, A. W. and C. G. Atkeson (1993). Prioritized sweeping : Reinforcement learning with less data and less real time. *Machine Learning* 13.
- Moulin, H. (1988). *Axioms of Cooperative Decision Making*. New-York : Wiley.
- Mundhenk, M., J. Goldsmith, C. Lusena, and E. Allender (2000, July). Complexity of finite-horizon markov decision process problems. *Journal of the ACM* 47, Issue 4, 681–720.
- Olofsson, E. and E. Blennow (2005). Decision support for identifying spruce forest stand edges with high probability of wind damage. *Forest Ecology and Management* 207, 87–98.
- Papadimitriou, C. H. and J. N. Tsitsiklis (1987, August). The complexity of Markov decision processes. *Mathematics of operations research* 12(3).
- Parr, R. (1998). Flexible decomposition algorithms for weakly coupled MDPs. In *UAI'98*.
- Pearl, J. (1988a). *Probabilistic Reasoning in Intelligent Systems*. San Mateo, CA : Morgan Kaufmann.
- Pearl, J. (1988b). *Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference*. Morgan Kaufmann.
- Pearl, J. (1993). From conditional oughts to qualitative decision theory. In *Proc. of UAI'93*, pp. 12–20.
- Peng, J. and R. J. Williams (1993). Efficient learning and planning within the dyna framework. *Adaptive Behavior* 1(4), 437–454.
- Pineau, J., G. Gordon, and S. Thrun (2003). Point-based value iteration : An anytime algorithm for pomdps. In *Proc. of IJCAI 2003*, Acapulco, Mexico, pp. 1025–1032.

- Pralet, C., G. Verfaillie, and T. Schiex (2006a). Decision with uncertainties feasibilities and utilities : towards a unified algebraic framework. In *Proc' of ECAI'06*, pp. 427–431.
- Pralet, C., G. Verfaillie, and T. Schiex (2006b). Decomposition of Multi-Operator Queries on Semiring-based Graphical Models. In *Proc' of CP'06*, Nantes, France.
- Pralet, C., G. Verfaillie, and T. Schiex (2006c). From Influence Diagrams to Multioperator Cluster DAGs. In *Proc' of UAI'06*, Cambridge, MA, USA.
- Precup, D., R. Sutton, and S. Singh (1998). Theoretical results on reinforcement learning with temporally abstract behaviors. In *Proc. 10th European Conference on Machine Learning (ECML'98)*, Chemnitz, Allemagne, pp. 382–393.
- Puterman, M. L. (1994). *Markov Decision Processes*. New York : John Wiley and Sons.
- Quiggin, J. (1982). A theory of anticipated utility. *J. of Economic Behavior and Organisation* 3, 323–343.
- Sarin, R. and P. P. Wakker (1992). A simple axiomatization of nonadditive expected utility. *Econometrica* 60(6), 1255–1272.
- Savage, L. (1951). The theory of statistical decision. *Journal of the American Statistical Association* 46, 55–67.
- Savage, L. (1954). *The Foundations of Statistics*. New York : Wiley.
- Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica* 57, 571–587.
- Segal, U. (1987). Anticipated utility : A measure representation approach. *Annals of Operations Research* 19, 359–374.
- Shachter, R. (1986). Evaluating influence diagrams. *Operations research* 34(6), 871–882.
- Singh, S. and D. Cohn (1998). How to dynamically merge markov decision processes. In *Advances in Neural Information Processing Systems*, Volume 10, Cambridge, pp. 1057–1063. MIT Press.
- Smallwood, R. D. and E. J. Sondik (1973). The optimal control of partially observable markov processes over a finite horizon. *Operations Research* (21), 1071–1088.
- Smith, T. and R. Simmons (2005). Point-based POMDP algorithms : Improved analysis and implementation. In *Proc. of IJCAI 2005*.
- Sondik, E. J. (1971). *The Optimal Control of Partially Observable Markov Processes*. Ph. D. thesis, Stanford University.
- Sondik, E. J. (1978). The optimal control of partially observable markov processes over the infinite horizon : Discounted costs. *Operations Research* 26(2), 282–304.
- Spohn, W. (1990). A general non-probabilistic theory of inductive reasoning. In *Proc. of UAI'90*, pp. 315–322. Elsevier Science.

- Sugeno, M. (1977). *Fuzzy Automata and Decision Processes*, Chapter Fuzzy measures and fuzzy integrals-a survey, pp. 89–102. North Holland, Amsterdam : M.M. Gupta, G.N. Saridis and B.R. Gaines (Eds.).
- Sutton, R. (1988). Learning to predict by the method of temporal differences. *Machine Learning* 3(1), 9–44.
- Sutton, R. (1991). Planning by incremental dynamic programming. In M. Kaufmann (Ed.), *Proc. of the 8th International Workshop on Machine Learning*, pp. 353–357.
- Sutton, R. S. and A. G. Barto (2002). *Introduction to Reinforcement Learning*. Cambridge, MA : MIT Press.
- Tsitsiklis, J. N. and B. van Roy (1996). Feature-based methods for large-scale dynamic programming. *Machine Learning* 22, 59–94.
- Valinger, E. and J. Fridman (1997). Modelling probability of snow and wind damage in scots pine stands using tree characteristics. *Forest Ecology and Management* 97, 215–222.
- von Neumann, J. and O. Morgenstern (1944). *Theory of Games and Economic Behavior*. Princeton University Press.
- Wakker, P. (1994). Separating marginal utility and risk aversion. *Theory and Decision* 36, 1–44.
- Wald, A. (1950). *Statistical Decision Functions*. New York : John Wiley and Sons.
- Watkins, C. J. (1989). *Learning from Delayed Rewards*. Ph. D. thesis, King's College, Cambridge, UK.
- Watkins, C. J. and P. Dayan (1992). Q-learning. *Machine Learning* 3(8), 279–292.
- Weng, P. (2005). Qualitative decision making under possibilistic uncertainty : Toward more discriminating criteria. In *Proc. of the 21th Int. Conf. on Uncertainty in Artificial Intelligence (UAI-05)*, pp. 615–622.
- Williams, R. and L. Baird (1993). Tight performance bounds on greedy policies based on imperfect value functions. Technical report, College of Computer Science, Northeastern University, Boston, MA.
- Wilson, N. (1995). An order of magnitude calculus. In *Proc. of UAI'95*, pp. 548–555.
- Wu, F.-Y. (1982). The potts model. *Reviews of Modern Physics* 54, 235–268.
- Yaari, M. (1987). The dual theory of choice under risk. *Econometrica* 55, 95–115.

Annexe A

Travaux et publications

A.1 Revues à comité de lecture

A.1.1 Revues internationales

- [1] Dubois (D.), Le Berre (D.), Prade (H.) et Sabbadin (R.). – Using possibilistic logic for modeling qualitative decision : ATMS-based algorithms. *Fundamenta Informaticae*, vol. 37, 1999, pp. 1–30.
- [2] Dubois (D.), Marichal (J. L.), Prade (H.), Roubens (M.) et Sabbadin (R.). – The use of the discrete Sugeno integral in decision-making : a survey. *International Journal of Uncertainty, fuzziness and knowledge-based systems*, vol. 9, n° 5, 2001, pp. 539–561.
- [3] Dubois (D.), Prade (H.) et Sabbadin (R.). – Decision-theoretic foundations of qualitative possibility theory. *European Journal of Operations Research*, vol. 128, 2001, pp. 459–478.
- [4] Fargier (H.) et Sabbadin (R.). – Qualitative decision under uncertainty : back to expected utility. *Artificial Intelligence*, vol. 165, 2005, pp. 245–280.
- [5] Garcia (L.) et Sabbadin (R.). – Complexity results and algorithms for possibilistic influence diagrams. *Artificial Intelligence*, vol. 172, 2008, pp. 1018-1044.
- [6] Sabbadin (R.), Spring (D.) et Rabier (C.-E.). – Dynamic reserve site selection under contagion risk of deforestation. *Ecological Modelling*, vol. 201, 2007, pp. 75–81.
- [7] Sabbadin (R.). – Possibilistic Markov Decision Processes. *Engineering Applications of Artificial Intelligence*, vol. 14, 2001, pp. 287–300.
- [8] Sabbadin (R.), Dubois (D.), Grenier (P.) et Prade (H.). – A fuzzy constraint satisfaction problem in the wine industry. *International Journal of Intelligent and Fuzzy Systems*, vol. 6, 1998, pp. 361–374.
- [9] Sabbadin (R.), Fargier (H.) et Lang (J.). – Towards qualitative approaches to multi-stage decision making. *International Journal of Approximate Reasoning*, vol. 19, 1998, pp. 441–471.

A.1.2 Revues nationales

- [10] Benferhat (S.), Dubois (D.), Prade (H.) et Sabbadin (R.). – Exception-tolerant reasoning and decision under uncertainty using qualitative possibility theory. *Künstliche Intelligenz (KI)*, vol. 3, 2001, pp. 9–15.
- [11] Garcia (L.) et Sabbadin (R.). – Diagrammes d’influence possibilistes : complexité et algorithmes. *Revue d’Intelligence Artificielle*, vol.21, 2007, pp. 521–554.

A.1.3 Article en cours de soumission/révision

- [12] Forsell (N.), Wikström (P.), Garcia (F.), Sabbadin (R.), Blennow (K.) et Eriksson (L.-O.). – Management of the risk of wind damage in forestry : a graph-based Markov decision process approach, 2008, 15 pages. *Soumis à Annals of Operations Research. En cours de révision, nouvelle soumission avant mars 2009.*

A.2 Communications dans congrès avec sélection par comité de lecture sur texte intégral

- [13] Balent (G.), Deconchat (M.), Ladet (S.), Martin-Clouaire (R.) et Sabbadin (R.). – Towards an environmental DSS based on spatio-temporal Markov chain approximation. *International Environmental Modelling and Software Society Conference on Complexity and Integrated Resources Management (IEMSS'04)*, – Osnabrück, Allemagne, 14-17 Juin 2004.
- [14] Cucala (L.), Garcia (F.) et Sabbadin (R.). – Résolution approchée de processus décisionnels markoviens de grande taille. *13ème congrès Reconnaissance des Formes et Intelligence Artificielle (RFIA'02)*, pp. 181–190. – Angers, 2002.
- [15] Dubois (D.), Fargier (H.) et Sabbadin (R.). – Qualitative decision rules under uncertainty. *7th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQA-RU'03)*, pp. 1–21. – Aalborg, Danemark, Juillet 2003.
- [16] Forsell (N.), Eriksson (L. O.), Garcia (F.), Sabbadin (R.) et Wikström (P.). – A graph-based Markov Decision Process approach for managing forests under windthrow risk. *MODSIM'07 International Congress on Modelling and Simulation (MODSIM'07)*. pp. 2168-2174 – Christchurch, New-Zealand, 2007.
- [17] Fargier (H.) et Sabbadin (R.). – Can qualitative utility criteria obey the sure thing principle ? *8th international conference on Information Processing and Management of Uncertainty (IPMU'00)*. pp. 821–826. – Madrid, 2000.
- [18] Fargier (H.) et Sabbadin (R.). – Additive refinements of qualitative decision criteria. *4th International Workshop on Preferences and Decisions*, pp. 23–29. – Trento, Italie, Septembre 2003.
- [19] Fargier (H.) et Sabbadin (R.). – Qualitative decision under uncertainty : Back to expected utility. *18th International Joint Conference on Artificial Intelligence (IJCAI'03)*, pp. 303–308. – Acapulco, Mexique, 9-15 Août 2003.
- [20] Forsell (N.) et Sabbadin (R.). – Algorithme de résolution approchée basé sur la programmation linéaire pour les processus décisionnels de Markov sur graphe. *Premières Journées Françaises de Planification, Décision et Apprentissage (JFPDA'06)*, pp. 115–122. – Toulouse, Mai 2006.
- [21] Forsell (N.) et Sabbadin (R.). – Approximate linear-programming algorithms for graph-based Markov decision processes. *17th European Conference on Artificial Intelligence (ECAI'06)*, pp. 590–594. – Riva del Garda, Italy, Août 2006.
- [22] Garcia (F.) et Sabbadin (R.). – Solving large weakly coupled markov decision processes : Application to forest management. *International congress on modelling and simulation (MODSIM'01)*, Eds. F. Ghassemi, M. Mc Aleer (L. Oxley) et Scoccimarro (M.), pp. 1707–1712. – Canberra, 2001.
- [23] Garcia (L.) et Sabbadin (R.). – Possibilistic influence diagrams. *In : 17th European Conference on Artificial Intelligence (ECAI'06)*, pp. 372–376. – Riva del Garda, Italy, Août 2006.

- [24] Kennedy (J.) et Sabbadin (R.). – Qualitative dynamic programming for optimal timber cutting and conservation of leadbeater's possum. *3rd Conference of the European Federation for Information Technology in Agriculture, Food and the Environment (EFITA'01)*, éd. par Steffe (Ed. J.), pp. 373–378. – Montpellier, 2001.
- [25] Kennedy (J.O.S.) et Sabbadin (R.). – Possibilistic Markov decision processes - application to timber cutting and conservation of leadbeater's possum. *Colloque Automatique et Agronomie*, pp. 35–46. – Montpellier, janvier 2003.
- [26] Mouterfi (M.) et Sabbadin (R.). – Graph-based multidimensional markov decision processes. *ECAI'04, workshop "Multi-Agent Markov Decision Processes : Theories and Models"*. – Valence, Espagne, août 2004.
- [27] Peyrard (N.) et Sabbadin (R.). – Itération de la politique approchée pour processus décisionnels de Markov sur graphe. *15ème congrès Reconnaissance des Formes et Intelligence Artificielle (RFIA 2006)*. – Tours, janvier 2006.
- [28] Peyrard (N.) et Sabbadin (R.). – Mean field approximation of the policy iteration algorithm for graph-based Markov decision processes. *17th European Conference on Artificial Intelligence (ECAI'06)*, pp. 595–599. – Riva del Garda, Italy, Août 2006.
- [29] Peyrard (N.), Sabbadin (R.), Lô-Pelzer (E.) et Aubertot (J. N.). – A graph-based Markov decision process framework for optimising integrated management of diseases in agriculture. *MODSIM'07 International Congress on Modelling and Simulation (MODSIM'07)*. pp. 2175-2181 – Christchurch, New-Zealand, 2007.
- [30] Ravoanjanahary (N.) et Sabbadin (R.). – Processus de décision séquentielle purement épistémique en environnement incertain et partiellement observable. *5èmes journées nationales autour des PDM en IA (PDMIA'05)*. – Lille, France, Juin 2005.
- [31] Ravoanjanahary (N.), Sabbadin (R.) et Lang (J.). – Processus de décision markovien purement épistémique. *2èmes journées francophones en planification, décision et apprentissage (JFPDA'07)*. – Grenoble, France, Juillet 2007.
- [32] Rabier (C. E.), Sabbadin (R.) et Spring (D.). – Dynamic reserve site selection under contagion risk of deforestation. *16th Inter. Congress on Modelling and Simulation (MODSIM'05)*, pp. 2102–2108. – Melbourne, Australia, Décembre 2005.
- [33] Sabbadin (R.). – A possibilistic model for qualitative sequential decision problems under uncertainty in partially observable environments. *15th Conference on Uncertainty in Artificial Intelligence (UAI'99)*. pp. 567–574. – Stockholm, 1999.
- [34] Sabbadin (R.). – Décision séquentielle possibiliste en environnement partiellement observable. *Actes du 12ème congrès Reconnaissance des Formes et Intelligence Artificielle (RFIA'00)*, pp. 257–266. – Paris, 2000.
- [35] Sabbadin (R.). – Empirical comparison of probabilistic and possibilistic Markov decision processes algorithms. *13th European Conference on Artificial Intelligence (ECAI'00)*. pp. 586–590. – Berlin, 2000.
- [36] Sabbadin (R.). – Towards possibilistic reinforcement learning algorithms. *10th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'01)*, pp. 404–407. – Melbourne, 2001.
- [37] Sabbadin (R.). – Approximating spatial Markov decision processes for environmental management. *2nd European Workshop on Sequential Decisions under Uncertainty in Agriculture and Natural Resources (EWDA'02)*, pp. 65–70. – Toulouse, Septembre 2002.

- [38] Sabbadin (R.). – Graph partitioning techniques for Markov decision processes decomposition. *14th European Conf. on Artificial Intelligence (ECAI'02)*, pp. 670–674. – Lyon, France, 21-26 Juillet 2002.
- [39] Sabbadin (R.). – Approximating spatial Markov decision processes for environmental management. *15th International Congress on Modelling and Simulation (MODSIM'03)*, pp. 1868–1873. – Townsville, Australia, 14-17 Juillet 2003.
- [40] Sabbadin (R.), Spring (D.) et Bergonnier (E.). – A reinforcement-learning application to biodiversity conservation in costa-rican forest. *MODSIM'07 International Congress on Modelling and Simulation (MODSIM'07)*. pp. 2189-2195 – Christchurch, New-Zealand, 2007.
- [41] Sabbadin (R.), Lang (J.) et Ravoanjanahy (N.). – Purely epistemic Markov decision processes. *22nd AAAI Conference on Artificial Intelligence (AAAI'07)*, pp. 1057–1062. – Vancouver, Canada, 2007.

A.3 Contributions à des ouvrages collectifs

- [42] Benferhat (S.), Dubois (D.), Fargier (H.), Prade (H.) et Sabbadin (R.). – *Logic-based Artificial Intelligence*, chap. Decision, non monotonic reasoning and possibilistic logic - an introductory survey of recent results, pp. 333–358. – Kluwer International, *Series in Engineering and Computer Science*, volume 597. Ed. J. Minker, 2001.
- [43] Boussard (M.), Bouzid (M.), Mouaddib (A.I.), Sabbadin (R.) et Weng (P.). – *Processus décisionnels de Markov en Intelligence Artificielle (vol. 1)*, chap. Critères non classiques, Hermès Sciences, *Traité IC2, série Informatique et Systèmes d'Information*, Eds. O. sigaud et O. Buffet, 2008.
- [44] Dubois (D.), Fargier (H.), Prade (H.) et Sabbadin (R.). – *Concepts et méthodes pour l'aide à la décision*, chap. Critères qualitatifs pour la décision dans l'incertain, pp. 99–137. – Hermès Sciences, *Traité IC2, série Informatique et Systèmes d'Information*, volume 2. Eds. D. Bouyssou, D. Dubois, M. Pirlot et H. Prade, 2006.
- [45] Dubois (D.), Prade (H.) et Sabbadin (R.). – *Fuzzy Measures and Integrals : Theory and Applications*, chap. Qualitative decision theory with Sugeno integrals, pp. 314–322. – Physica-Verlag. Eds. M. Grabisch, T. Murofushi and M. Sugeno, 2000.
- [46] Fargier (H) et Sabbadin (R.). – *Technologies for Constructing Intelligent Systems*, chap. Can qualitative Utility Criteria obey the Sure Thing Principle ?, pp. 167–178. – Heidelberg, New-York, Physica Verlag, *Studies in Fuzzyness and Soft Computing*, volume 1. Eds. B. Bouchon-Meunier, J. Gutierrez Rios, L. Magdalena et R. R. Yager, 2002.
- [47] Lafage (C.), Lang (J.) et Sabbadin (R.). – *Information, uncertainty and fusion*, chap. A logic of supporters, pp. 381–392. – Kluwer Academic Publishers. Eds. B. Bouchon-Meunier, R.R. Yager and L. Zadeh, 1999.
- [48] Sabbadin (R.). – *Concepts et méthodes pour l'aide à la décision*, chap. Planification dans l'incertain et processus décisionnels de Markov, pp. 203–252. – Hermès Sciences, *Traité IC2, série Informatique et Systèmes d'Information*, volume 2. Eds. D. Bouyssou, D. Dubois, M. Pirlot et H. Prade, 2006.

A.4 Rapports diplômants

- [49] Sabbadin (R.). Une approche ordinaire de la décision dans l'incertain : axiomatisation, représentation logique et application à la décision séquentielle. *Thèse de Doctorat de l'Université Paul Sabatier*, Toulouse, décembre 1998.

A.5 Développements informatiques

- [50] M.-J. Cros, F. Garcia et R. Sabbadin. *MDP Toolbox, v2.0*. Il s'agit d'un ensemble de fonctions MATLAB implémentant les algorithmes classiques de résolution de PDM : *itération de la valeur*, *itération de la politique...*, ainsi qu'un ensemble d'exemples didactiques.
[http ://www.inra.fr/bia/produits/logiciels/Page_home.php/](http://www.inra.fr/bia/produits/logiciels/Page_home.php/).

Annexe B

Décision dans l'incertain

B.1 Décision dans l'incertain probabilisé : l'axiomatique de Savage

L'axiomatisation du critère de l'utilité espérée proposée par Savage (Savage, 1954) est très générale car elle permet, à partir d'une relation de préférence entre les actes, de déterminer à la fois la fonction d'utilité du décideur sur les conséquences, mais aussi "sa" distribution de probabilité *subjective* sur les états possibles du monde. Contrairement à l'approche de (von Neumann and Morgenstern, 1944), la nature probabiliste de l'incertitude n'est pas supposée a priori mais découle d'axiomes portant sur les préférences entre les actes. Du fait que les axiomes proposés par Savage concernent directement les actes, ils peuvent être vérifiés par l'observation du comportement de décideurs en situation de décision concrète.

Savage suppose que l'ensemble \mathcal{S} des états possibles du monde est infini. Cette hypothèse est nécessaire afin de pouvoir construire une distribution de probabilité unique représentant l'incertitude de l'agent. Nous supposons que l'ensemble \mathcal{X} des conséquences est fini, mais il est possible de considérer \mathcal{X} infini, au prix de l'ajout d'un axiome technique. $\mathcal{A} = \mathcal{X}^{\mathcal{S}}$ est l'ensemble des actes "potentiels", applications associant une conséquence à chaque état possible du monde. Dans un problème de décision, les actes réellement disponibles ne constituent en général qu'un sous-ensemble de \mathcal{A} ; néanmoins nous supposons que le décideur est capable d'ordonner tous les actes potentiels via une relation de préférence représentée par le préordre \succeq sur \mathcal{A} .

La première hypothèse posée par Savage est que \succeq est un préordre *complet* sur \mathcal{A} : tous les actes peuvent être comparés (rangés).

Sav1 Rangement.

(\mathcal{A}, \succeq) est un préordre complet, i.e. \succeq est réflexif, transitif et complet.

Deux sous-familles d'actes sont particulièrement utiles pour déterminer, d'une part, la fonction d'utilité du décideur sur les conséquences et, d'autre part, sa distribution de probabilité subjective sur les états possibles du monde : les actes *constants* et les actes *binaires*, respectivement.

Les conséquences des actes sont généralement notées x, y, x', y' (éléments de \mathcal{X}). Les actes constants donnent la même conséquence, quel que soit l'état réel du monde. Pour distinguer un acte constant de son unique conséquence x , on le note en gras (\mathbf{x}) : pour tout $x \in \mathcal{X}$, \mathbf{x} représente l'acte défini par $\forall s \in \mathcal{S}, \mathbf{x}(s) = x$. L'ensemble des actes constants est noté \mathcal{C} . Puisque \succeq est un préordre complet sur \mathcal{A} , il en est aussi un sur $\mathcal{C} \subseteq \mathcal{A}$. Comme \mathcal{C} et \mathcal{X} sont bijectivement reliés, on peut définir, à partir de

\succeq , un préordre complet sur $\mathcal{X} : \geq_P$.

Définition B.1.1 *Préférences sur \mathcal{X} induites par les préférences sur \mathcal{A} .*

$\forall x, y \in X$, si pour tout $s \in \mathcal{S}$, $\mathbf{x}(s) = x$ et $\mathbf{y}(s) = y$, alors $x \geq_P y \Leftrightarrow \mathbf{x} \succeq \mathbf{y}$.

Afin d'éviter le cas dégénéré où toutes les conséquences sont équivalentes, auquel cas tous les actes le sont également, l'axiome suivant est introduit¹ :

Sav5 Non trivialité.

$\exists x, y \in \mathcal{X}, x >_P y$, où $>_P$ représente la partie stricte de \geq_P .

La notion d'*acte binaire* est utile afin de déterminer les probabilités attachées par le décideur aux différents états possibles du monde. Un acte binaire est un acte qui a au plus deux conséquences possibles, x et y , telles que $x >_P y$.

Définition B.1.2 *Acte binaire.*

Un acte binaire, noté xAy , $x, y \in \mathcal{X}, x >_P y, A \subseteq \mathcal{S}$ peut être vu comme un pari sur l'événement A , donnant la conséquence x si A "arrive", et y sinon :

$$\forall s \in A, xAy(s) = x \text{ et } \forall s \in \bar{A}, xAy(s) = y.$$

Définissons maintenant \geq_V , relation de *vraisemblance* entre événements, par :

$$A \geq_V B \Leftrightarrow \left(\forall x, y \in X, (x >_P y) \Rightarrow (xAy \succeq xBy) \right). \quad (\text{B.1})$$

$A \geq_V B$ revient à dire que tout pari sur A est préféré à un pari aux conséquences identiques sur B .

Il va sans dire que, dans le cas général, rien ne garantit que \geq_V soit un préordre complet : changer les conséquences peut modifier l'ordre entre les paris. Afin de rendre \geq_V complet, Savage propose l'axiome suivant :

Sav4 Projection sur l'ensemble des événements.

$\forall x, y, x', y' \in X, x >_P y, x' >_P y', \forall A, B \subseteq \mathcal{S}, (xAy \succeq xBy) \Leftrightarrow (x'Ay' \succeq x'By')$.

Cet axiome assure que la restriction de la relation de préférence sur les actes aux paris sur les événements est bien définie et ne dépend que de l'ordre entre les conséquences des paris.

La notion d'*acte binaire* est un cas particulier de la notion d'*acte mixte* :

Définition B.1.3 *Acte mixte.*

Si f et g sont deux actes, et $A \subseteq \mathcal{S}$ est un événement, l'acte mixte fAg est défini par :

$\forall s \in A, fAg(s) = f(s)$ et $\forall s \in \bar{A}, fAg(s) = g(s)$.

Un acte binaire est donc un acte mixte impliquant deux actes constants.

L'axiome suivant est un axiome clé de la théorie de Savage. C'est lui qui permet de définir le conditionnement des préférences entre actes par rapport à un événement.

¹Les axiomes ne sont pas présentés dans le même ordre que dans (Savage, 1954) mais par souci de cohérence nous conservons la numérotation de Savage.

Sav2 Principe de la chose certaine.²

$$\forall f, g, h, h' \in \mathcal{A}, \forall A \subseteq S, (fAh \succeq gAh) \Leftrightarrow (fAh' \succeq gAh').$$

Cet axiome stipule que si deux actes fAh et gAh donnent les mêmes conséquences sur un événement donné (\bar{A}) alors leur classement relatif ne dépend pas de ces conséquences identiques. Le principe de la chose certaine permet de définir une relation de *préférence conditionnelle* par rapport à un événement A :

Définition B.1.4 *Préférence conditionnelle.*

f est préféré à g conditionnellement à A , noté $(f \succeq g)_A$, si et seulement si :

$$\forall h \in \mathcal{A}, fAh \succeq gAh.$$

Cette définition est bien fondée, grâce au “principe de la chose certaine”. Qui plus est, la propriété $(f \succeq g)_A$ ne dépend pas du choix de h , et la relation de préférence conditionnelle est un préordre complet sur \mathcal{A} . Un ensemble d'événements peut être distingué, celui conditionnellement auquel tous les actes sont équivalents : les *événements nuls*.

Définition B.1.5 *Événement nul.*

Un événement A est dit nul si et seulement si, pour tout triplet d'actes f, g et h , $fAh \succeq gAh$.

On peut montrer que A est nul si et seulement si $A \equiv_V \emptyset$.

Si on restreint la relation de préférence conditionnelle par rapport à un événement aux actes constants, le préordre obtenu doit coïncider avec le préordre sur les conséquences (sauf si on conditionne par rapport à un événement nul, bien entendu). Cette propriété découle de l'axiome suivant :

Sav3 Conditionnement restreint aux actes constants.

Soient $x, y \in X$ et A un événement *non nul*. Soient les actes constants $\mathbf{x} \equiv x$ et $\mathbf{y} \equiv y$.

$$\text{Alors, } (\mathbf{x} \succeq \mathbf{y})_A \Leftrightarrow x \geq_P y.$$

L'approche de Savage vise à proposer une justification axiomatique de la théorie des probabilités, basée sur les préférences entre actes. Il est donc utile d'examiner les propriétés de la relation \geq_V sur les événements induite par une relation \succeq sur les actes satisfaisant les axiomes **Sav1** à **Sav5**. Savage a montré qu'une telle relation sur les événements est une *relation de probabilité qualitative* (de Finetti, 1937). Elle vérifie les propriétés suivantes :

A1 \geq_V est un préordre complet,

A2 $\mathcal{S} >_V \emptyset$ (non trivialité),

A3 $\forall A \subseteq S, A \geq_V \emptyset$ (consistance),

P Si $A \cap (B \cup C) = \emptyset$, alors $B \geq_V C$ si et seulement si $A \cup B \geq_V A \cup C$ (additivité).

Une relation d'ordre \geq_V sur les événements basée sur une mesure de probabilité P ($A \geq_V B \Leftrightarrow P(A) \geq P(B)$) est une relation de probabilité qualitative. Cependant, une relation de probabilité qualitative n'est pas forcément représentable par une distribution de probabilité, comme l'ont montré (Kraft et al., 1959). Afin de déterminer la mesure de probabilité sur \mathcal{S} mesurant l'incertitude d'un décideur à partir de ses préférences sur les actes, il est nécessaire d'ajouter un axiome supplémentaire :

²Le nom original donné par Savage à cet axiome est *the Sure Thing principle*.

Sav6 Probabilité quantitative.

Soient f et $g \in \mathcal{A}$ tels que $f \succ g$, soit $x \in X$. Il existe une partition $\{A_i\}$ de S telle que pour tout i , $x A_i f \succ g$ et $f \succ x A_i g$.

Sav6 décrit une forme de “continuité” de l’espace d’états : il implique qu’il est possible de partitionner S en de nombreux sous-ensembles de probabilité très faible. Si \succeq est une relation de préférence sur \mathcal{A} satisfaisant **Sav1** à **Sav6**, elle est représentable par une fonction d’utilité EU sur \mathcal{A} . Pour tout acte f , $EU(f)$ est l’espérance mathématique d’une fonction d’utilité u sur les conséquences, au sens d’une distribution de probabilité p sur S :

Théorème B.1.1 *Axiomatisation de l’utilité espérée.*

Soit S un ensemble d’états du monde et \mathcal{X} un ensemble fini de conséquences. Soit également \succeq , une relation de préférence sur \mathcal{X}^S vérifiant les axiomes **Sav1** à **Sav6**. Alors il existe une distribution de probabilité p , unique, sur S et une fonction d’utilité u (non unique³) sur \mathcal{X} , telles que :

$$\forall f, g \in \mathcal{X}^S, f \succeq g \Leftrightarrow \sum_{s \in S} p(s) \cdot u(f(s)) \geq \sum_{s \in S} p(s) \cdot u(g(s)).$$

B.2 Algorithmes classiques de résolution des PDM

Recherche arrière

Dans le cas où l’horizon T est fini, l’utilité espérée EU_δ d’une politique $\delta = \{\delta_t\}_{t=0 \dots T-1}$ dépend de l’étape t considérée et peut être calculée récursivement en partant de EU_δ^{T-1} , jusqu’à EU_δ^0 :

$$\begin{aligned} EU_\delta^{T-1}(x_{T-1}) &= \sum_{x_T} p_{T-1}(x_T | x_{T-1}, \delta_{T-1}(x_{T-1})) \cdot r_{T-1}(x_{T-1}, \delta_{T-1}(x_{T-1}), x_T), \\ &\forall x_{T-1} \in \mathcal{X}, \\ EU_\delta^t(x_t) &= \sum_{x_{t+1}} p_t(x_{t+1} | x_t, \delta_t(x_t)) \cdot \left[r_t(x_t, \delta_t(x_t), x_{t+1}) + EU_\delta^{t+1}(x_{t+1}) \right], \\ &\forall t < T - 1, x_t \in \mathcal{X}. \end{aligned} \tag{B.2}$$

Une politique optimale $\delta^* = (\delta_0^*, \dots, \delta_{T-1}^*)$ peut être obtenue en alternant l’étape de maximisation suivante, avec les phases de calcul des fonctions $EU_{\delta^*}^t$ décrites par (B.2) :

$$\begin{aligned} \delta_{T-1}^*(x_{T-1}) &= \arg \max_{a_{T-1} \in \mathcal{A}_{T-1}} \sum_{x_T} p_{T-1}(x_T | x_{T-1}, a_{T-1}) \cdot r_{T-1}(x_{T-1}, a_{T-1}, x_T), \\ &\forall x_{T-1} \in \mathcal{X}, \\ \delta_t^*(x_t) &= \arg \max_{a_t \in \mathcal{A}_t} \left\{ \sum_{x_{t+1}} p_t(x_{t+1} | x_t, a_t) \cdot \left[r_t(x_t, a_t, x_{t+1}) + EU_{\delta_{t+1}^*}^{t+1}(x_{t+1}) \right] \right\}, \\ &\forall t < T - 1, x_t \in \mathcal{X}. \end{aligned}$$

³ u est unique, à une transformation affine positive près.

Itération de la politique

Lorsqu'on passe à un horizon infini (pour des problèmes stationnaires), on peut montrer que la fonction de valeur EU_δ d'une politique stationnaire δ est indépendante du temps et qu'elle est un point fixe du système d'équations suivant (Howard, 1960) (ici, dans le cas du critère γ -pondéré) :

$$EU_\delta(x) = \sum_{x' \in \mathcal{X}} p(x'|x, \delta(x)) \cdot [r(x, \delta(x), x') + \gamma EU_\delta(x')], \forall x \in \mathcal{X}.$$

Ce point fixe peut être calculé par une méthode de *programmation linéaire* ou par un algorithme itératif de type *approximations successives* (Puterman, 1994) :

$$\begin{aligned} v_\delta^0(x) &= 0, \forall x \in \mathcal{X}, \\ v_\delta^n(x) &= \sum_{x' \in \mathcal{X}} p(x'|x, \delta(x)) \cdot [r(x, \delta(x), x') + \gamma v_\delta^{n-1}(x')], \forall x \in \mathcal{X}, \forall n \geq 1. \end{aligned} \quad (\text{B.3})$$

Où l'on peut montrer que : $\lim_{n \rightarrow \infty} v_\delta^n = EU_\delta$.

L'algorithme d'*Itération de la politique* (Howard, 1960) consiste en l'alternance de phases d'*évaluation* et d'*amélioration* de la politique courante pour calculer une politique optimale δ^* . La phase d'*évaluation* calcule la valeur de la politique courante δ , par exemple en appliquant le schéma itératif (B.3), alors que la phase d'*amélioration* transforme la politique courante δ en une politique *améliorée* δ' ($EU_{\delta'}(x) \geq EU_\delta(x), \forall x \in \mathcal{X}$) définie par :

$$\delta'(x) = \arg \max_{a \in A} \left\{ \sum_{x' \in \mathcal{X}} p(x'|x, a) \cdot [r(x, a, x') + \gamma EU_\delta(x')] \right\}, \forall x \in \mathcal{X}.$$

Lorsque $EU_\delta = EU_{\delta'}$, nous avons l'assurance d'avoir trouvé une politique optimale (Howard, 1960).

Itération de la valeur

L'algorithme d'*Itération de la valeur* (Bellman, 1957) s'inspire également du schéma itératif (B.3) pour calculer une politique optimale. Néanmoins, comme dans le cas de l'horizon fini, la politique courante est mise à jour pendant le calcul itératif de la fonction de valeur optimale. Une phase de maximisation est ainsi introduite dans le schéma itératif (B.3) :

$$\begin{aligned} v^0(x) &= 0, \forall x \in \mathcal{X}, \\ v^n(x) &= \max_{a \in A} \left\{ \sum_{x' \in \mathcal{X}} p(x'|x, a) \cdot [r(x, a, x') + \gamma v^{n-1}(x')] \right\}, \forall x \in \mathcal{X}. \end{aligned}$$

La suite (v^n) converge (pour peu que $0 \leq \gamma < 1$) et $\lim_{n \rightarrow \infty} v^n = v^* = EU_{\delta^*}$.

Une politique optimale δ^* est alors calculable à partir de v^* :

$$\delta^*(x) = \arg \max_{a \in A} \left\{ \sum_{x' \in \mathcal{X}} p(x'|x, a) \cdot [r(x, a, x') + \gamma v^*(x')] \right\}, \forall x \in \mathcal{X}.$$

B.3 Mesures d'incertitude monotones

Distribution de possibilité

Une *distribution de possibilité* π sur un ensemble d'états \mathcal{S} est une application de \mathcal{S} dans $(L, >)$, une échelle ordonnée finie, équipée d'une fonction de renversement n . 1_L et 0_L représentent respectivement le plus grand et plus petit élément de L . Dans le cadre de la théorie qualitative de la décision, une distribution de possibilité est utilisée pour modéliser la connaissance imparfaite que l'on a du monde dans un problème de décision dans l'incertain, distinguant les états "plausibles" ou normaux, des états peu vraisemblables ou surprenants. La fonction $\pi : \mathcal{S} \rightarrow L$ modélise cette connaissance, avec les conventions suivantes : $\pi(s) = 0_L$ signifie que s est considéré comme impossible alors que $\pi(s) = 1_L$ signifie que s est un état "normal" ou totalement possible. $\pi(s) > \pi(s')$ signifie que s est plus vraisemblable que s' . Il se peut que plusieurs états aient une possibilité de 1_L : cela signifie que tous ces états sont également plausibles et sont plus plausibles que tous les autres. Le cas extrême d'ignorance est celui où tous les états partagent une possibilité de 1_L : tous les états sont possibles et rien ne permet de les départager. Au contraire, si un seul état a une possibilité de 1_L et tous les autres sont impossibles (possibilité 0_L), alors on est dans un état de connaissance parfaite. Cette description de la connaissance en termes de distribution de possibilité est flexible dans la mesure où tous les degrés de l'échelle L compris entre 0_L et 1_L sont utilisables pour modéliser le degré de possibilité $\pi(s)$ d'un état. La seule contrainte sur la distribution π est qu'il existe au moins un état s^* de possibilité 1_L (normalisation) : quelle que soit notre connaissance sur l'état du monde, il existe au moins un état accepté comme "normal".

Mesures de possibilité et de nécessité

Une distribution de possibilité peut être vue comme la contrepartie qualitative d'une distribution de probabilité, à partir de laquelle peut être définie une *mesure d'incertitude* sur les événements (sous-ensembles de \mathcal{S}). Néanmoins, contrairement à la théorie des probabilités pour laquelle il n'existe qu'une seule mesure d'incertitude attachée à une distribution de probabilité p ,

$$P(A) = \sum_{s \in A} p(s), \forall A \subseteq \mathcal{S},$$

deux mesures d'incertitude différentes ont été définies dans le cadre de la théorie des possibilités (Dubois and Prade, 1988) :

– Mesure de possibilité :

$$\Pi(A) = \max_{s \in A} \pi(s), \forall A \subseteq \mathcal{S}. \quad (\text{B.4})$$

– Mesure de nécessité :

$$N(A) = \min_{s \in \bar{A}} n(\pi(s)), \forall A \subseteq \mathcal{S}. \quad (\text{B.5})$$

$\Pi(A)$ est appelé *degré de possibilité* de A et mesure à quel point A est un événement "normal". Pour évaluer le degré de normalité d'un événement A , $\Pi(A)$ se focalise sur le degré de possibilité de l'état du monde "le plus plausible", compatible avec l'événement A . Par convention, $\Pi(\emptyset) = 0_L$. Le *degré de nécessité* $N(A)$ au contraire, se focalise sur l'état du monde le plus plausible non compatible avec A pour évaluer à quel point A est "certain". A est d'autant plus certain que son complémentaire \bar{A} est peu plausible. Cette dernière remarque liant possibilité et nécessité est justifiée par la relation suivante :

$$N(A) = n(\Pi(\bar{A})), \forall A \subseteq \mathcal{S}. \quad (\text{B.6})$$

La mesure de nécessité N est *duale* de la mesure de possibilité Π . Notons qu'en théorie des probabilités, la mesure de probabilité est *auto-duale* ($P(A) = 1 - P(\overline{A}), \forall A \subseteq \mathcal{S}$).

Les mesures d'incertitude possibilistes vérifient respectivement deux axiomes de décomposabilité duaux, contreparties de l'axiome d'additivité de la théorie des probabilités :

$$\Pi(A \cup B) = \max\{\Pi(A), \Pi(B)\}, \forall A, B \subseteq \mathcal{S} \quad (\text{B.7})$$

$$N(A \cap B) = \min\{N(A), N(B)\}, \forall A, B \subseteq \mathcal{S}. \quad (\text{B.8})$$

A partir de ce constat, il est possible de définir deux relations de *possibilité comparative* et *nécessité comparative* (Dubois, 1986) analogues de la relation de probabilité qualitative entre les événements de (de Finetti, 1937), à partir des axiomes suivants :

A1 \geq_V est un préordre complet,

A2 $\mathcal{S} >_V \emptyset$ (non trivialité),

A3 $\forall A \subseteq \mathcal{S}, A \geq_V \emptyset$ (consistance),

A3' $\forall A \subseteq \mathcal{S}, \mathcal{S} \geq_V A$ (consistance),

Pos $\forall A, B, C \subseteq \mathcal{S}$, si $A \geq_V B$ alors $A \cup C \geq_V B \cup C$ (décomposabilité disjonctive).

Nec $\forall A, B, C \subseteq \mathcal{S}$, si $A \geq_V B$ alors $A \cap C \geq_V B \cap C$ (décomposabilité conjonctive).

Une relation de possibilité comparative (resp. nécessité comparative) est une relation de préférence \succeq sur $2^{\mathcal{S}}$ vérifiant les axiomes **A1**, **A2**, **A3**, **Pos** (resp. **A1**, **A2**, **A3'**, **Nec**). (Dubois, 1986) a montré que contrairement au cas probabiliste, une relation de possibilité (resp. nécessité) comparative est toujours représentable par une mesure de possibilité (resp. nécessité) lorsque \mathcal{S} est fini.

Mesures monotones

Les mesures de possibilité et de nécessité sont des cas particuliers de *mesures monotones*. Une mesure monotone est une fonction $\sigma : 2^{\mathcal{S}} \rightarrow L$ vérifiant les axiomes suivants :

M1 $\sigma(\mathcal{S}) >_L \sigma(\emptyset)$ (non trivialité),

M2 $\sigma(A) \geq_L \sigma(B), \forall B \subseteq A$ (monotonie).

Une relation comparative monotone \succeq sur $2^{\mathcal{S}}$ satisfait les axiomes **A1**, **A2**, ainsi que :

Mon : $\forall A \subseteq B \subseteq \mathcal{S}, B \geq_V A$.

Les relations comparatives monotones sont extrêmement générales. En particulier, il est facile de constater que toute relation de possibilité, nécessité ou probabilité qualitative est une relation comparative monotone (Dubois, 1986).

B.4 Axiomatisation de l'utilité qualitative monotone

Les relations de préférence entre actes construites à partir du critère d'utilité qualitative monotone satisfont trivialement les axiomes **Sav 1** et **Sav5**. Par contre, on peut aisément vérifier qu'elles ne satisfont pas les axiomes **Sav 2**, **Sav 3** et **Sav 4**.

On peut montrer que l'axiome suivant est vérifié :

MON Monotonie.

si \mathbf{x} ($\equiv x$) et \mathbf{y} ($\equiv y$) sont des actes constants, alors $\forall A \subseteq \mathcal{S}, \forall h \in \mathcal{X}^{\mathcal{S}}, (x \geq_P y) \Rightarrow (\mathbf{x}Ah \succeq \mathbf{y}Ah)$.

L'axiome de monotonie peut sembler faible, et on peut vérifier en particulier qu'il est satisfait par toute relation de préférence satisfaisant **Sav 1**, **Sav 5** et l'un quelconque des axiomes **Sav 2**, **Sav 3** et **Sav 4**. Toutefois, il a d'importantes conséquences. En particulier, il permet d'assurer la *monotonie* de la relation de préférence, au sens suivant :

Définition B.4.1 (Relation de préférence monotone) Soit \succeq une relation de préférence sur $\mathcal{X}^{\mathcal{S}}$ réflexive, transitive et complète et soit \geq_P la relation de préférence sur \mathcal{X} associée⁴.

\succeq est dite monotone, si et seulement si elle satisfait la relation suivante :

$$\forall f, g \in \mathcal{X}^{\mathcal{S}}, (f(x) \geq_P g(x), \forall x \in \mathcal{X}) \Rightarrow f \succeq g.$$

La monotonie de la relation de préférence entre actions est une propriété souhaitable, en décision dans l'incertain mais aussi dans le domaine du choix social, où elle est connue sous le terme de *règle d'unanimité* (Moulin, 1988). Mais l'axiome de monotonie a d'autres conséquences importantes. Tout d'abord, il est possible de construire une relation de vraisemblance \geq_V à partir d'une relation de préférence entre actions \succeq , vérifiant **Sav 1**, **MON** et **Sav 5**, par :

$$\forall A, B \subseteq \mathcal{S}, A \geq_V B \Leftrightarrow x^\top Ax_\perp \succeq x^\top Bx_\perp. \quad (\text{B.9})$$

On peut montrer que \geq_V est une relation comparative monotone.

Proposition B.4.1 Soient \mathcal{S} et \mathcal{X} deux ensembles finis, et \succeq une relation de préférence sur $\mathcal{X}^{\mathcal{S}}$ satisfaisant les axiomes **Sav 1**, **MON** et **Sav 5**. Alors \geq_V est une relation comparative monotone. Donc, il est possible de construire une échelle finie L_V et une mesure monotone $\sigma : 2^{\mathcal{S}} \rightarrow L_V$ telles que la relation de vraisemblance \geq_V sur les événements soit représentable par la mesure monotone σ :

$$A \geq_V B \Leftrightarrow \sigma(A) \geq_{L_V} \sigma(B), \forall A, B \subseteq \mathcal{S}.$$

Revenons maintenant à l'hypothèse de commensurabilité entre incertitude et préférence dans les approches qualitatives. Nous venons de voir que si une relation de préférence sur les actes \succeq satisfait les axiomes **Sav 1**, **MON** et **Sav 5**, on peut construire une échelle L_V et une mesure d'incertitude monotone σ sur \mathcal{S} à valeurs dans L_V représentant les connaissances d'un décideur, compatibles avec \succeq . Mais, comme nous l'avons vu dans le chapitre précédent, l'axiome **Sav 1** suffit à "ranger" les conséquences, et donc à construire une échelle L_P , et une fonction d'utilité qualitative $\mu : \mathcal{X} \rightarrow L_P$ représentant les préférences du décideur.

La question de la commensurabilité entre incertitude et préférence revient simplement à se demander si l'on peut comparer les degrés des échelles L_V et L_P . Or, une nouvelle application de l'axiome **Sav 1** à la relation de préférence \succeq suffit pour répondre positivement à cette question. En effet, puisque tous les actes sont comparables, les actes constants sont comparables aux actes binaires. Chaque élément $\alpha \in L_V$, représentatif d'au moins un acte binaire, est donc comparable à tout élément $\alpha' \in L_P$ représentatif d'au moins un acte constant. Il est donc possible de construire une échelle L à partir de l'union des échelles L_V et L_P d'où ont été retirés les éléments "redondants", telle que la restriction de \succeq aux actes constants soit représentable par une fonction d'utilité qualitative $\mu : \mathcal{X} \rightarrow L$ et la restriction de \succeq aux paris sur les événements soit représentable par une mesure monotone à valeurs dans la même échelle $\sigma : \mathcal{S} \rightarrow L$. En d'autres termes,

⁴ \geq_P est bien définie puisque \succeq est réflexive, transitive et complète, donc satisfait **Sav 1**.

Proposition B.4.2 Soient \mathcal{S} et \mathcal{X} deux ensembles finis, et \succeq une relation de préférence sur $\mathcal{X}^{\mathcal{S}}$ satisfaisant les axiomes **Sav 1**, **MON** et **Sav 5**. Il est possible de construire une échelle finie L , une fonction d'utilité qualitative $\mu : \mathcal{X} \rightarrow L$ représentant la restriction de \succeq aux actes constants et une mesure d'incertitude monotone $\sigma : 2^{\mathcal{S}} \rightarrow L$ représentant la restriction de \succeq aux paris sur les événements.

Mais les axiomes **Sav 1**, **MON** et **Sav 5** ne suffisent pas à axiomatiser le critère de l'utilité qualitative monotone, même s'ils permettent d'en construire les ingrédients L , σ et μ . Nous avons suggéré [45] l'ajout des deux axiomes supplémentaires suivants :

DCR Dominance Conjonctive Restreinte.

Soient f et g deux actes quelconques, et $\mathbf{y} \equiv y$ un acte constant :

$$(g \succ f \text{ et } \mathbf{y} \succ f) \Rightarrow (g \wedge \mathbf{y} \succ f), \text{ où}$$

$$(g \wedge \mathbf{y})(s) = \begin{cases} y & \text{si } g(s) >_P y \\ g(s) & \text{sinon.} \end{cases}$$

DDR Dominance Disjonctive Restreinte.

Soient f et g deux actes quelconques, et $\mathbf{y} \equiv y$ un acte constant :

$$(f \succ g \text{ et } f \succ \mathbf{y}) \Rightarrow (f \succ g \vee \mathbf{y}), \text{ où}$$

$$(g \vee \mathbf{y})(s) = \begin{cases} y & \text{si } y >_P g(s) \\ g(s) & \text{sinon.} \end{cases}$$

A partir de ces cinq axiomes, nous avons démontré un théorème de représentation de l'utilité qualitative monotone [45] :

Théorème B.4.1 (Axiomatisation de l'utilité qualitative monotone)

Si \succeq est une relation de préférence sur les actes satisfaisant **Sav 1**, **MON**, **Sav 5**, **DCR** et **DDR** alors on peut construire une échelle finie L , une fonction d'utilité qualitative $\mu : \mathcal{X} \rightarrow L$ et une mesure d'incertitude monotone $\sigma : 2^{\mathcal{S}} \rightarrow L$, telles que

$$\forall f, g \in \mathcal{X}^{\mathcal{S}}, (f \succeq g) \Leftrightarrow (U_{SUG, \sigma, \mu}(f) \geq_L U_{SUG, \sigma, \mu}(g)).$$

DCR et **DDR** autorisent une décomposition partielle de la fonction d'utilité qualitative monotone de certains actes mixtes :

$$\begin{aligned} - U_{SUG, \sigma, \mu}(f \wedge \mathbf{y}) &= \min\{U_{SUG, \sigma, \mu}(f), U_{SUG, \sigma, \mu}(\mathbf{y})\}, \\ - U_{SUG, \sigma, \mu}(f \vee \mathbf{y}) &= \max\{U_{SUG, \sigma, \mu}(f), U_{SUG, \sigma, \mu}(\mathbf{y})\}, \end{aligned}$$

si \mathbf{y} est un acte constant (et $U_{SUG, \sigma, \mu}(\mathbf{y}) = y$).

B.5 Axiomatisation des utilités qualitatives possibilistes

Les deux axiomes **DDR** et **DCR** imposent une forme de *décomposabilité restreinte* du critère de décision représentant les préférences d'un décideur, comme nous venons de le remarquer. Les mesures de possibilité et de nécessité présentent elles aussi une forme de décomposabilité, puisque nous savons que :

$$\begin{aligned} \forall A, B \subseteq \mathcal{S}, \Pi(A \cup B) &= \max\{\Pi(A), \Pi(B)\}, \\ \forall A, B \subseteq \mathcal{S}, N(A \cap B) &= \min\{N(A), N(B)\}. \end{aligned}$$

Il est facile de vérifier que si $\succeq_{OPT,\pi,\mu}$ est une relation de préférence basée sur l'utilité qualitative possibiliste $U_{OPT,\pi,\mu}$, elle vérifie la propriété suivante :

Si g et h sont des paris sur des événements et f est un acte quelconque,

$$\left((f \succ g) \text{ et } (f \succ h) \right) \Rightarrow (f \succ g \vee h).$$

En fait, nous avons démontré [3] que cette relation est vérifiée pour des actes quelconques, pas seulement pour des paris. Ceci nous a conduit à proposer l'axiome suivant, qui renforce l'axiome **DDR** :

DD Dominance Disjonctive.

Soient f , g et h , trois actes quelconques :

$$\begin{aligned} (f \succ g \text{ et } f \succ h) &\Rightarrow (f \succ g \vee h), \text{ où} \\ (f \vee g)(s) &= \begin{cases} f(s) & \text{si } f(s) >_P g(s) \\ g(s) & \text{sinon.} \end{cases} \end{aligned}$$

Nous avons montré que **Sav 1**, **MON**, **Sav 5**, **DCR** et **DD** permettent d'axiomatiser l'utilité qualitative possibiliste optimiste :

Théorème B.5.1 (Axiomatisation de l'utilité qualitative possibiliste optimiste)

Soit \succeq , une relation de préférence sur $\mathcal{X}^{\mathcal{S}}$ satisfaisant **Sav 1**, **MON**, **Sav 5**, **DCR** et **DD**. Alors, il existe une échelle finie L , une fonction d'utilité qualitative $\mu : \mathcal{X} \rightarrow L$ et une distribution de possibilité $\pi : \mathcal{S} \rightarrow L$, telles que \succeq soit représentable par une fonction d'utilité qualitative possibiliste optimiste, $U_{OPT,\pi,\mu}$:

$$\forall f, g \in \mathcal{X}^{\mathcal{S}}, (f \succeq g) \Leftrightarrow \left(U_{OPT,\pi,\mu}(f) \geq U_{OPT,\pi,\mu}(g) \right).$$

De la même manière, nous pouvons définir une justification axiomatique du critère pessimiste à partir de l'axiome suivant, "dual" de **DD** :

DC Dominance Conjonctive.

Soient f , g et h , trois actes quelconques :

$$\begin{aligned} (g \succ f \text{ et } h \succ f) &\Rightarrow (g \wedge h \succ f), \text{ où} \\ (g \wedge h)(s) &= \begin{cases} g(s) & \text{si } h(s) >_P g(s) \\ h(s) & \text{sinon.} \end{cases} \end{aligned}$$

Et comme pour le critère optimiste, nous avons montré que :

Théorème B.5.2 (Axiomatisation de l'utilité qualitative possibiliste pessimiste)

Soit \succeq , une relation de préférence sur $\mathcal{X}^{\mathcal{S}}$ satisfaisant **Sav 1**, **MON**, **Sav 5**, **DDR** et **DC**. Alors, il existe une échelle finie L , une fonction d'utilité qualitative $\mu : \mathcal{X} \rightarrow L$ et une distribution de possibilité $\pi : \mathcal{S} \rightarrow L$, telles que \succeq soit représentable par une fonction d'utilité qualitative possibiliste pessimiste, $U_{PES,\pi,\mu}$:

$$\forall f, g \in \mathcal{X}^{\mathcal{S}}, (f \succeq g) \Leftrightarrow \left(U_{PES,\pi,\mu}(f) \geq U_{PES,\pi,\mu}(g) \right).$$

B.6 Raffinements discri-monotones

Nous avons étudié tout d'abord des raffinements *discri-monotones* des utilités qualitatives monotones et possibilistes [17], [46]. Le principe de la chose certaine implique que les états pour lesquels deux actes donnent des conséquences identiques ne doivent pas jouer de rôle dans la comparaison de ces actes. Une manière simple de restaurer ce principe est donc de modifier la relation de préférence $\succeq_{SUG,\sigma,\mu}$ de manière à ne comparer des actes que sur les états dans lesquels leurs conséquences diffèrent. Cela conduit à définir une relation de préférence *discri-monotone* sur les actes :

Définition B.6.1 (Relation de préférence discri-monotone)

Nous notons abusivement α une conséquence arbitrairement fixée, dont le degré d'utilité μ est $\alpha \in L$. Nous définissons $\succeq_{SUG,\sigma,\mu}^{\Delta\alpha}$ par :

$$f \succeq_{SUG,\sigma,\mu}^{\Delta\alpha} g \Leftrightarrow U_{SUG,\sigma,\mu}(f\Delta\alpha) \geq U_{SUG,\sigma,\mu}(g\Delta\alpha)$$

où $\Delta = \Delta(f, g) = \{s \in S, f(s) \neq g(s)\}$ et $f\Delta\alpha(s) = f(s)$ si $s \in \Delta$ et $f\Delta\alpha(s) = \alpha$ si $s \in \bar{\Delta}$.

Le paramètre α de la relation de préférence discri-monotone permet de définir toute une gamme de critères. Néanmoins, ils sont tous exprimables en fonction de deux critères particuliers :

Proposition B.6.1 $\forall f \in \mathcal{X}^S, \forall \alpha \in L$ et $\forall \Delta \subseteq S$,

$$U_{SUG,\sigma,\mu}(f\Delta\alpha) = \max \left\{ \min(\alpha, U_{SUG,\sigma,\mu}(f\Delta x^\top)), U_{SUG,\sigma,\mu}(f\Delta x_\perp) \right\}.$$

Puisque pour tout sous-ensemble $A \subseteq S$, $\Delta(fAh, gAh) \subseteq A$, il est facile de montrer que :

Proposition B.6.2 $\succeq_{SUG,\sigma,\mu}^{\Delta\alpha}$ satisfait le principe de la chose certaine.

Malheureusement, si $\succeq_{SUG,\sigma,\mu}^{\Delta\alpha}$ satisfait le principe de la chose certaine, nous avons montré [46] qu'il violait deux propriétés hautement souhaitables pour une relation de préférence entre actions :

1. $\succeq_{SUG,\sigma,\mu}^{\Delta\alpha}$ n'est pas forcément transitif et même sa partie stricte $\succ_{SUG,\sigma,\mu}^{\Delta\alpha}$ peut ne pas être transitive.
2. $\succeq_{SUG,\sigma,\mu}^{\Delta\alpha}$ n'est pas forcément un raffinement de $\succeq_{SUG,\sigma,\mu}$.

Cette méthode de restauration du principe de la chose certaine ne semble pas très intéressante dans le cas de l'utilité qualitative monotone en général, mais sa restriction aux utilités possibilistes l'est plus. Dans [46], nous avons étudié la sous-classe des critères de décision *discri-possibilistes*, $\succeq_{OPT,\pi,\mu}^{\Delta\alpha}$ et $\succeq_{PES,\pi,\mu}^{\Delta\alpha}$. Là encore, les propriétés citées ci-dessus ne sont pas forcément valables pour tout α . Néanmoins, nous avons montré qu'elles le sont dans certains cas. Nous avons défini de la manière suivante les critères *discri-optimiste* et *discri-pessimiste* :

Définition B.6.2 (Critère discri-optimiste)

$$f \succeq_{OPT,\pi,\mu}^{\Delta} g \Leftrightarrow f\Delta x_\perp \succeq_{OPT,\pi,\mu} g\Delta x_\perp.$$

Définition B.6.3 (Critère discri-pessimiste)

$$f \succeq_{PES,\pi,\mu}^{\Delta} g \Leftrightarrow f\Delta x^\top \succeq_{PES,\pi,\mu} g\Delta x^\top.$$

Nous avons montré que ces critères étaient des raffinements non biaisés des utilités possibilistes.

Proposition B.6.3 (Raffinements non biaisés)

$$\forall f, g \in \mathcal{L}_{OPT}, f \succeq_{OPT, \pi, \mu} g \Leftrightarrow f \succeq_{\Delta_{OPT, \pi, \mu}} g.$$

$$\forall f, g \in \mathcal{L}_{PES}, f \succeq_{PES, \pi, \mu} g \Leftrightarrow f \succeq_{\Delta_{PES, \pi, \mu}} g.$$

Ces deux critères semblent donc posséder toutes les propriétés requises pour être des raffinements des utilités possibilistes satisfaisant le principe de la chose certaine. Il leur manque néanmoins une propriété souhaitable : *la transitivité de la relation d'indifférence entre actions*. Il est possible de trouver trois actes f, g et h tels que :

$$f \sim_{\Delta_{OPT, \pi, \mu}} g \text{ et } g \sim_{\Delta_{OPT, \pi, \mu}} h \text{ et } f \succ_{\Delta_{OPT, \pi, \mu}} h.$$

Exemple B.6.1 Soient $\mathcal{S} = \{s_1, s_2, s_3\}$ et $L = \{0, 1, 2, 3, 4, 5\}$. Soient f, g et h , trois actes dont les utilités des conséquences dans les états s_1, s_2 et s_3 sont listées dans la table suivante, de même que les degrés de possibilité de s_1, s_2 et s_3 :

| | s_1 | s_2 | s_3 |
|-------------|-------|-------|-------|
| $\mu(f(s))$ | 5 | 4 | 3 |
| $\mu(g(s))$ | 5 | 2 | 4 |
| $\mu(h(s))$ | 5 | 4 | 2 |
| π | 5 | 4 | 4 |

On vérifie tout d'abord que $\Delta(f, g) = \Delta(g, h) = \{s_2, s_3\}$ et $\Delta(f, h) = \{s_3\}$, puis que $f \sim_{\Delta_{OPT, \pi, \mu}} g$, $g \sim_{\Delta_{OPT, \pi, \mu}} h$ et $f \succ_{\Delta_{OPT, \pi, \mu}} h$.

La perte de transitivité de la relation d'indifférence est moins ennuyeuse que la perte de transitivité de la relation de préférence stricte, mais elle empêche toute représentation des relations de préférence $\succeq_{\Delta_{OPT, \pi, \mu}}$ et $\succeq_{\Delta_{PES, \pi, \mu}}$ par des fonctions d'utilité à valeurs dans une échelle complètement ordonnée. C'est pourquoi, nous avons décidé d'explorer une autre piste pour le raffinement des utilités possibilistes. Nous avons cherché explicitement une représentation *numérique* d'un critère raffinant les utilités possibilistes satisfaisant le principe de la chose certaine. Plus précisément, nous avons opté pour la recherche d'un raffinement non biaisé basé sur le critère de *l'utilité espérée*.

B.7 Raffinements des utilités qualitatives possibilistes basés sur l'utilité espérée

Afin de proposer des raffinements non biaisés des critères de décision possibilistes basés sur l'utilité espérée, nous avons défini la notion de *transformation probabiliste d'une échelle L* ([18], [19], [4]) :

Définition B.7.1 (Transformation probabiliste d'une échelle ordinale)

Soit $\langle \mathcal{S}, \mathcal{X}, \mathcal{A}, \succeq_{\cdot, \pi, \mu} \rangle$ un modèle de décision possibiliste (optimiste ou pessimiste) et soit L , l'échelle des degrés d'incertitude et de préférence associée. Une transformation probabiliste de l'échelle L par rapport à la distribution de possibilité π est une fonction strictement croissante $\chi : L \rightarrow [0, 1]$ telle que $\chi(0_L) = 0$ et telle que :

$$\sum_{s \in \mathcal{S}} \chi(\pi(s)) = 1.$$

A partir de χ et d'un modèle de décision possibiliste optimiste $(\mathcal{S}, \mathcal{X}, \mathcal{A}, \succeq_{OPT, \pi, \mu})$, il est possible de définir un modèle de décision de type utilité espérée, transformé du précédent :

Définition B.7.2 (Transformation probabiliste d'un modèle possibiliste optimiste)

Soit $(\mathcal{S}, \mathcal{X}, \mathcal{A}, \succeq_{OPT, \pi, \mu})$ un modèle de décision qualitative possibiliste optimiste et χ une transformation probabiliste de l'échelle L associée.

$(\mathcal{S}, \mathcal{X}, \mathcal{A}, \succeq_{EU, \chi \circ \pi, \chi \circ \mu})$ est un modèle de décision de type utilité espérée transformé du modèle possibiliste optimiste par χ .

La raison pour laquelle ce type de transformation probabiliste est intéressante est qu'il est possible d'exhiber une condition suffisante sur la fonction χ pour que la relation de préférence de type utilité espérée obtenue *raffine* la relation de préférence possibiliste optimiste [4] :

$$\mathbf{H} : \forall \alpha, \alpha' \in L \text{ tels que } \alpha > \alpha' : \chi(\alpha)^2 > |\mathcal{S}| \cdot \chi(1_L) \cdot \chi(\alpha').$$

Si χ respecte **H**, la distribution de probabilité $p = \chi \circ \pi$ obtenue par transformation de π est une *probabilité à grandes marches* (Dubois and Fargier, 2004). Elle vérifie donc :

$$\forall s \in \mathcal{S}, p(s) > \sum_{s', p(s) > p(s')} p(s').$$

Notons que la condition **H** ne porte que sur l'échelle L et sur la taille de l'espace d'états et pas sur π ou μ . Nous avons ensuite montré le résultat suivant :

Théorème B.7.1 (Raffinement du critère optimiste par transformation probabiliste)

Soit $(\mathcal{S}, \mathcal{X}, \mathcal{A}, \succeq_{OPT, \pi, \mu})$ un modèle de décision qualitative possibiliste optimiste et soit χ une transformation probabiliste de L à partir de π satisfaisant **H**. Alors, $\succeq_{EU, \chi \circ \pi, \chi \circ \mu}$ *raffine* $\succeq_{OPT, \pi, \mu}$.

Nous avons montré qu'un résultat du même type (mais un peu plus complexe à exprimer) pouvait être obtenu dans le cas du critère pessimiste :

Théorème B.7.2 (Raffinement du critère pessimiste par transformation probabiliste)

Soit $(\mathcal{S}, \mathcal{X}, \mathcal{A}, \succeq_{PES, \pi, \mu})$ un modèle de décision qualitative possibiliste pessimiste et soit χ une transformation probabiliste de L à partir de π satisfaisant **H**. Soient $p(s) = \chi \circ \pi(s)$ et $u(x) = \chi \circ \mu(x^\top) - \chi \circ \mu(x)$, pour tout $s \in \mathcal{S}, x \in \mathcal{X}$. Alors, $\succeq_{EU, p, u}$ *raffine* $\succeq_{PES, \pi, \mu}$.

Le raffinement du critère pessimiste nécessite deux transformations des degrés de possibilité et d'utilité différentes. La forme obtenue résulte de l'égalité des deux relations de préférence possibilistes suivantes :

Proposition B.7.1 Soit $(\mathcal{S}, \mathcal{X}, \mathcal{A}, \succeq_{OPT, \pi, \mu})$ un modèle de décision qualitative possibiliste optimiste, et soit $\succeq_{PES, \pi, \mu'}$ une relation de préférence entre actions construite à partir de π et $\mu' = n \circ \mu$. Alors,

$$\succeq_{OPT, \pi, \mu} \equiv \preceq_{PES, \pi, \mu'}.$$

Exemple B.7.1 Reprenons l'exemple (3.3.1), et notons χ^* la transformation probabiliste définie par $\chi^*(\alpha_i) = \frac{v}{N^{2^i+1}}$ où $v = 1/(\sum_{\alpha_i \in L} \frac{n_i}{N^{2^i+1}})$ et $N = |\mathcal{S}|$. Dans l'exemple (3.3.1), $N = 2$ et $L = \{0, 1, 2, 3, 4, 5\}$. Nous avons déjà montré que les deux actes suivants n'étaient pas discriminés par le critère optimiste :

| | | |
|----------|-------|-------|
| | s_1 | s_2 |
| $\mu(f)$ | 3 | 4 |
| $\mu(g)$ | 3 | 1 |
| π | 5 | 2 |

$\chi^*(L)$ est la suite $\frac{v}{N^{32}}, \frac{v}{N^{16}}, \frac{v}{N^8}, \frac{v}{N^4}, \frac{v}{N^2}, 0$ où $v = \frac{1}{(\frac{1}{N^2} + \frac{1}{N^{16}})}$. Donc :

$$EU(f) = \chi^*(5) \cdot \chi^*(3) + \chi^*(4) \cdot \chi^*(2) = \frac{v^2}{N^{10}} + \frac{v^2}{N^{20}},$$

$$EU(g) = \chi^*(5) \cdot \chi^*(3) + \chi^*(2) \cdot \chi^*(1) = \frac{v^2}{N^{10}} + \frac{v^2}{N^{48}}.$$

D'où, $f \succ_{EU, \chi^* \circ \pi, \chi^* \circ \mu} g$.

La question suivante se pose naturellement : *peut-on trouver plusieurs raffinements non biaisés différents ?* Nous avons montré que non, en tous cas pour des raffinements satisfaisant la condition suffisante **H** [4] :

Théorème B.7.3 (Équivalence des raffinements)

Soit $(\mathcal{S}, \mathcal{X}, \mathcal{A}, \succeq, \pi, \mu)$ un modèle possibiliste (optimiste ou pessimiste). Toutes les relations de préférence de type utilité espérée définies à partir de χ satisfaisant **H**, comme dans le théorème (B.7.1) ou comme dans le théorème (B.7.2) respectivement, sont équivalentes.

B.8 Raffinement et ordre lexicographique

Ainsi, nous avons montré que tous les raffinements d'une relation de préférence possibiliste optimiste ou pessimiste, basés sur une transformation probabiliste satisfaisant **H** (que nous appellerons transformation *robuste* à partir de maintenant), sont équivalents. Mais nous sommes allés plus loin, en montrant que ces relations de préférence de type utilité espérée pouvaient être modélisées par une simple comparaison des éléments de matrices représentatives des actes.

A cette fin, nous avons défini la représentation matricielle d'un acte dans le cadre de la théorie possibiliste de la décision :

Définition B.8.1 (Représentation matricielle d'un acte) Soient $(\mathcal{S}, \mathcal{X}, \mathcal{A}, \succeq, \pi, \mu)$ un cadre de décision possibiliste (optimiste ou pessimiste) et un acte $f \in \mathcal{A}$.

La représentation matricielle de f est la matrice suivante :

$$F = \begin{pmatrix} \pi_1 & \mu_1 \\ \vdots & \vdots \\ \pi_N & \mu_N \end{pmatrix}$$

où $N = |\mathcal{S}|$, $\pi_i = \pi(s_i)$ et $\mu_i = \mu(f(s_i))$.

On peut voir les relations de préférence possibilistes optimistes et pessimistes comme des relations de préférence sur des matrices $N \times 2$ d'éléments de L . Plus précisément, dans le cas optimiste, les matrices représentatives des actes sont comparées suivant la relation de préférence *Maxmin* :

Définition B.8.2 (Relation de préférence Maxmin) Soient F et G deux matrices $N \times M$ d'éléments de L .

$$F \succeq_{Maxmin} G \Leftrightarrow \max_{i=1, N} \min_{j=1, M} F(i, j) \geq \max_{i=1, N} \min_{j=1, M} G(i, j).$$

Le résultat suivant, reliant les relations de préférences $\succeq_{OPT,\pi,\mu}$ et \succeq_{Maxmin} , est évident [4] :

Proposition B.8.1 (Utilité optimiste et relation de préférence Maxmin)

$$\forall f, g \in \mathcal{X}^S, f \succeq_{OPT,\pi,\mu} g \Leftrightarrow F \succeq_{Maxmin} G.$$

Il s'agit maintenant de trouver un raffinement de la relation \succeq_{Maxmin} . Dans un cadre monodimensionnel ($M = 1$), cela revient à trouver un raffinement de la relation max , pour lequel l'ordre lexicographique $lmax$ suivant est un candidat naturel (Moulin, 1988), de même que l'ordre lexicographique $lmin$ peut raffiner une relation basée sur l'opérateur min .

Définition B.8.3 (Relations de préférence Leximax, Leximin) Soient deux vecteurs $\vec{u}, \vec{v} \in L^N$.

$$\begin{aligned} \vec{u} \succeq_{lmax} \vec{v} &\Leftrightarrow (\forall j, u_{(j)} = v_{(j)}) \text{ ou } (\exists i, \forall j < i, u_{(j)} = v_{(j)} \text{ et } u_{(i)} > v_{(i)}) \\ \vec{u} \succeq_{lmin} \vec{v} &\Leftrightarrow (\forall j, u_{(j)} = v_{(j)}) \text{ ou } (\exists i, \forall j > i, u_{(j)} = v_{(j)} \text{ et } u_{(i)} > v_{(i)}). \end{aligned}$$

Où, pour tout vecteur $\vec{w} \in L^N$, $w_{(k)}$ est le k -ième plus grand élément de \vec{w} (i.e. $w_{(1)} \geq \dots \geq w_{(N)}$).

En pratique, les comparaisons $leximin$ et $leximax$ consistent à ordonner deux vecteurs par ordre croissant ou décroissant, puis à les comparer suivant l'ordre lexicographique.

Exemple B.8.1 Soient $\vec{u} = (3, 2, 4)$ et $\vec{v} = (2, 2, 4)$.

$\vec{u} \succ_{lmax} \vec{v}$ puisque $u_{(1)} = v_{(1)} = 4$ et $u_{(2)} = 3 > v_{(2)} = 2$.

$\vec{u} \succ_{lmin} \vec{v}$ puisque $u_{(3)} = v_{(3)} = 2$ et $u_{(2)} = 3 > v_{(2)} = 2$.

Il est évident que \succeq_{lmax} raffine \succeq_{max} et \succeq_{lmin} raffine \succeq_{min} . De plus, ces deux relations sont très "décisives" : deux vecteurs ne sont équivalents que s'ils sont une permutation l'un de l'autre. Nous avons étendu les relations d'ordre lexicographique aux matrices, afin de pouvoir comparer les représentations matricielles de deux actes. A cette fin, il suffit de constater que les ordres $lmax$ et $lmin$ comparent des vecteurs d'éléments quelconques, appartenant à un ensemble totalement ordonné. Ces éléments peuvent eux-mêmes être des vecteurs d'éléments de L :

Définition B.8.4 (leximax(\triangleright), leximin(\triangleright))

Soit \triangleright un préordre complet sur L^M , \triangleright sa partie stricte et \cong sa partie symétrique.

Soient F, G , deux matrices d'éléments de L , de taille $N \times M$.

$$\begin{aligned} - F \succeq_{lmax(\triangleright)} G &\Leftrightarrow (\forall j, F_{(\triangleright,j)} \cong G_{(\triangleright,j)}) \text{ ou } (\exists i \text{ t.q. } \forall j < i, F_{(\triangleright,j)} \cong G_{(\triangleright,j)} \text{ et } F_{(\triangleright,i)} \triangleright G_{(\triangleright,i)}). \\ - F \succeq_{lmin(\triangleright)} G &\Leftrightarrow (\forall j, F_{(\triangleright,j)} \cong G_{(\triangleright,j)}) \text{ ou } (\exists i \text{ t.q. } \forall j > i, F_{(\triangleright,j)} \cong G_{(\triangleright,j)} \text{ et } F_{(\triangleright,i)} \triangleright G_{(\triangleright,i)}). \end{aligned}$$

Où, pour toute matrice H d'éléments de L , $H_{(\triangleright,i)}$ est la i ème plus "grande" ligne de H , suivant l'ordre \triangleright .

La relation d'ordre lexicographique sur les matrices peut être appliquée au cas où $\triangleright = \succeq_{lmin}$ ou $\triangleright = \succeq_{lmax}$. En particulier, on peut définir l'ordre $\succeq_{lmax(\succeq_{lmin})}$, et montrer qu'il vérifie les propriétés suivantes :

Proposition B.8.2 (Relation de préférence Leximaxmin sur les matrices)

- i) $\succeq_{lmax(\succeq_{lmin})}$ est un préordre complet ;
- ii) $\succeq_{lmax(\succeq_{lmin})}$ raffine \succeq_{maxmin} ;
- iii) Si $N = 1$, alors $\succeq_{lmax(\succeq_{lmin})} \equiv \succeq_{lmin}$;
- iv) Si $M = 1$, alors $\succeq_{lmax(\succeq_{lmin})} \equiv \succeq_{lmax}$.

À partir de cette relation $\succeq_{lmax(\succeq_{lmin})}$ sur les matrices, on peut définir une relation sur les actes, notée identiquement, par :

Définition B.8.5 (Relation de préférence Leximaxmin sur les actes)

Soient f et g deux actes,

$$f \succeq_{lmax(\succeq_{lmin})} g \Leftrightarrow F \succeq_{lmax(\succeq_{lmin})} G,$$

où F et G sont les représentations matricielles de f et g .

Exemple B.8.2 Les matrices représentatives des actes f et g de l'exemple (3.3.1) sont :

$$F = \begin{pmatrix} 5 & 3 \\ 2 & 4 \end{pmatrix} \text{ et } G = \begin{pmatrix} 5 & 3 \\ 2 & 1 \end{pmatrix}.$$

Donc, puisque $(5, 3) \cong_{lmin} (5, 3)$ et $(4, 2) \succ_{lmin} (2, 1)$, nous trouvons que $f \succ_{lmax(\succeq_{lmin})} g$.

Il est alors possible de montrer que $\succeq_{lmax(\succeq_{lmin})}$ raffine $\succeq_{OPT,\pi,\mu}$:

Proposition B.8.3 (Raffinement Leximaxmin de l'utilité optimiste)

$$\succeq_{lmax(\succeq_{lmin})} \text{ raffine } \succeq_{OPT,\pi,\mu} \text{ sur } \mathcal{A}.$$

La relation de préférence $\succeq_{lmax(\succeq_{lmin})}$ qui raffine $\succeq_{OPT,\pi,\mu}$ n'implique que des "comparaisons" de degrés de possibilité et d'utilité. Il s'agit donc encore d'une relation de préférence qualitative. Néanmoins, nous avons montré que ces raffinements lexicographiques sont équivalents aux raffinements par utilité espérée, définis précédemment. Paradoxalement, les raffinements non biaisés et basés sur l'utilité espérée du critère de décision possibiliste optimiste restent donc qualitatifs !

Théorème B.8.1 (Raffinement basé sur l'utilité espérée et relation de préférence Leximaxmin)

Soit $(\mathcal{S}, \mathcal{X}, \mathcal{A}, \succeq_{OPT,\pi,\mu})$, un modèle d'utilité qualitative optimiste et soit χ une transformation probabiliste de l'échelle L satisfaisant **H**. Nous avons :

$$\succeq_{EU,\chi \circ \pi, \chi \circ \mu} \equiv \succeq_{lmax(\succeq_{lmin})}.$$

Dans [4] nous avons démontré un résultat similaire dans le cas pessimiste. Un raffinement basé sur l'Utilité Espérée de la relation de préférence possibiliste pessimiste peut s'exprimer comme une relation de préférence de type *leximinmax* sur des matrices. Néanmoins les matrices en question sont légèrement différentes de celles utilisées dans le cas optimiste. Les éléments $\{\pi(s)\}_{s \in S}$ sont remplacés par $\{n(\pi(s))\}_{s \in S}$.

B.9 Algorithmes de programmation dynamique possibiliste

Algorithme de recherche arrière

Les contreparties possibilistes (optimiste et pessimiste) des équations de Bellman (équations (B.2)) sont :

– Dans le cas optimiste :

$$\begin{aligned} U_{OPT,\delta}^T(x_T) &= \mu_T(x_T) \text{ et pour tout } t = T - 1, \dots, 0 : \\ U_{OPT,\delta}^t(x_t) &= \max_{x_{t+1} \in \mathcal{X}} \min \left\{ \pi_t(x_{t+1}|x_t, \delta_t(x_t)), \mu_t(x_t, \delta_t(x_t), x_{t+1}), U_{OPT,\delta}^{t+1}(x_{t+1}) \right\} \end{aligned} \quad (\text{B.10})$$

– Dans le cas pessimiste :

$$\begin{aligned} U_{PES,\delta}^T(x_T) &= \mu_T(x_T) \text{ et pour tout } t = T-1, \dots, 0 : & (B.11) \\ U_{PES,\delta}^t(x_t) &= \min_{x_{t+1} \in \mathcal{X}} \max \left\{ n(\pi_t(x_{t+1}|x_t, \delta_t(x_t))), \min\{\mu_t(x_t, \delta_t(x_t), x_{t+1}), U_{PES,\delta}^{t+1}(x_{t+1})\} \right\} \end{aligned}$$

Alors, tout comme dans le cas stochastique (équations (B.3)), une politique optimale peut être calculée récursivement, comme suit :

– Dans le cas optimiste :

$$\begin{aligned} \delta_t^*(x_t) &= \arg \max_{a_t \in \mathcal{A}} \left\{ \max_{x_{t+1} \in \mathcal{X}} \min \left\{ \pi_t(x_{t+1}|x_t, a_t), \mu_t(x_t, a_t, x_{t+1}), U_{OPT,\delta^*}^{t+1}(x_{t+1}) \right\} \right\} \\ U_{OPT,\delta^*}^t(x_t) &= \max_{a_t \in \mathcal{A}} \left\{ \max_{x_{t+1} \in \mathcal{X}} \min \left\{ \pi_t(x_{t+1}|x_t, a_t), \mu_t(x_t, a_t, x_{t+1}), U_{OPT,\delta^*}^{t+1}(x_{t+1}) \right\} \right\} \\ &\quad \forall t = T-1, \dots, 1, \forall x_t \in \mathcal{X} & (B.12) \end{aligned}$$

– Dans le cas pessimiste :

$$\begin{aligned} \delta_t^*(x_t) &= \arg \max_{a_t \in \mathcal{A}} \left\{ \min_{x_{t+1} \in \mathcal{X}} \max \left\{ n(\pi_t(x_{t+1}|x_t, a_t)), \min\{\mu_t(x_t, a_t, x_{t+1}), U_{PES,\delta^*}^{t+1}(x_{t+1})\} \right\} \right\} \\ U_{PES,\delta^*}^t(x_t) &= \max_{a_t \in \mathcal{A}} \left\{ \min_{x_{t+1} \in \mathcal{X}} \max \left\{ n(\pi_t(x_{t+1}|x_t, a_t)), \min\{\mu_t(x_t, a_t, x_{t+1}), U_{PES,\delta^*}^{t+1}(x_{t+1})\} \right\} \right\} \\ &\quad \forall t = T-1, \dots, 1, \forall x_t \in \mathcal{X} & (B.13) \end{aligned}$$

Ces équations sont constructives, car les expressions de U_{OPT,δ^*}^{t+1} et U_{PES,δ^*}^{t+1} ne font intervenir que les politiques $\{\delta_{t+1}^*, \dots, \delta_{T-1}^*\}$ déjà calculées.

Itération de la politique possibiliste

Tout d'abord, on peut montrer que pour toute politique stationnaire δ , les fonctions d'utilité possibilistes $U_{OPT,\delta}$ et $U_{PES,\delta}$ sont respectivement des points fixes des systèmes d'équations suivants, contreparties qualitatives des systèmes définis dans le cas stochastique :

$$U(x) = \max_{x' \in \mathcal{X}} \min \left\{ \pi(x'|x, \delta(x)), \mu(x, \delta(x), x'), U(x') \right\}, \forall x \in \mathcal{X}, \quad (B.14)$$

$$U(x) = \min_{x' \in \mathcal{X}} \max \left\{ n(\pi(x'|x, \delta(x))), \min\{\mu(x, \delta(x), x'), U(x')\} \right\}, \forall x \in \mathcal{X}, \quad (B.15)$$

où les inconnues sont $\{U(x), x \in \mathcal{X}\}$.

Contrairement au cas stochastique, les systèmes d'équations (B.14) et (B.15) n'admettent pas de solution unique, à cause de la propriété d'idempotence des opérateurs min et max. Néanmoins, les fonctions d'utilité de la Définition 4.2.2 sont les *plus grands* points fixes de ces systèmes.

Proposition B.9.1 (Solutions des équations de point fixe possibilistes)

- Cas optimiste : $U_{OPT,\delta}$ est un point fixe de (B.14) et $\forall U^*$, point fixe de (B.14), $\forall x \in \mathcal{X}, U^*(x) \leq U_{OPT,\delta}(x)$.
- Cas pessimiste : $U_{PES,\delta}$ est un point fixe de (B.15) et $\forall U^*$, point fixe de (B.15), $\forall x \in \mathcal{X}, U^*(x) \leq U_{PES,\delta}(x)$.

Nous avons montré que les fonctions d'utilité $U_{OPT,\delta}$ et $U_{PES,\delta}$ pouvaient être calculées par les schémas itératifs suivants :

– Cas optimiste :

$$\begin{aligned} U_{OPT,\delta}^0(x) &= 1_L, \forall x \in \mathcal{X} \\ U_{OPT,\delta}^{t+1}(x) &= \max_{x' \in \mathcal{X}} \min \left\{ \pi(x'|x, \delta(x)), \mu(x, \delta(x), x'), U_{OPT,\delta}^t(x') \right\}, \forall x \in \mathcal{X}. \end{aligned} \quad (\text{B.16})$$

– Cas pessimiste :

$$\begin{aligned} U_{PES,\delta}^0(x) &= 1_L, \forall x \in \mathcal{X} \\ U_{PES,\delta}^{t+1}(x) &= \min_{x' \in \mathcal{X}} \max \left\{ n(\pi(x'|x, \delta(x))), \min \{ \mu(x, \delta(x), x'), U_{PES,\delta}^t(x') \} \right\}, \forall x \in \mathcal{X}. \end{aligned} \quad (\text{B.17})$$

On peut montrer que ces deux schémas itératifs convergent en un nombre borné d'itérations et que l'initialisation des fonctions $U_{OPT,\delta}^0$ et $U_{PES,\delta}^0$ à la valeur 1_L en tout x permet de garantir que la fonction retournée est bien le *plus grand* point fixe des systèmes d'équations (B.14) et (B.15) :

Proposition B.9.2 (Calcul itératif de l'utilité possibiliste d'une politique) *Soit $(\mathcal{X}, \mathcal{A}, \pi, \mu)$ un modèle de PDM possibiliste et δ une politique stationnaire. $U_{OPT,\delta}$ et $U_{PES,\delta}$ peuvent être calculées par les schémas itératifs (B.16) et (B.17) :*

$$\begin{aligned} U_{OPT,\delta} &= \lim_{t \rightarrow \infty} U_{OPT,\delta}^t, \\ U_{PES,\delta} &= \lim_{t \rightarrow \infty} U_{PES,\delta}^t. \end{aligned}$$

De plus, la convergence est obtenue après au plus $T = |\mathcal{X}| \times |L|$ itérations.

L'algorithme *d'itération de la politique possibiliste* alterne, comme dans le cas stochastique, des phases *d'évaluation* de la politique courante par le schéma itératif que nous venons de décrire et des phases *d'amélioration* de la politique courante :

– Cas optimiste :

$$\delta'(x) = \arg \max_{a \in \mathcal{A}} \left\{ \max_{x' \in \mathcal{X}} \min \left\{ \pi(x'|x, \delta(x)), \mu(x, \delta(x), x'), U_{OPT,\delta}(x') \right\} \right\}, \forall x \in \mathcal{X}. \quad (\text{B.18})$$

– Cas pessimiste :

$$\delta'(x) = \arg \max_{a \in \mathcal{A}} \left\{ \min_{x' \in \mathcal{X}} \max \left\{ n(\pi(x'|x, \delta(x))), \min \{ \mu(x, \delta(x), x'), U_{PES,\delta}(x') \} \right\} \right\}, \forall x \in \mathcal{X}. \quad (\text{B.19})$$

Il est possible de montrer que :

- $U_{OPT,\delta'}(x) \geq U_{OPT,\delta}(x)$ et
- $U_{PES,\delta'}(x) \geq U_{PES,\delta}(x), \forall x \in \mathcal{X}$,

et que le nombre d'itérations de l'algorithme d'itération de la politique possibiliste est borné à la fois dans les cas optimiste et pessimiste. Les deux politiques limites obtenues, δ_{OPT}^* et δ_{PES}^* sont optimales vis-à-vis des deux critères, respectivement.

Itération de la valeur possibiliste

Un schéma itératif de type *itération de la valeur possibiliste* a également été mis au point pour le calcul des politiques optimales δ_{OPT}^* et δ_{PES}^* . Comme dans le cas stochastique, les itérations portent sur une fonction de valeur courante :

– Cas optimiste :

$$\begin{aligned} U_{OPT}^0(x) &= \max_{a \in \mathcal{A}} \max_{x' \in \mathcal{X}} \min \left\{ \pi(x'|x, a), \mu(x, a, x') \right\}, \forall x \in \mathcal{X}, \\ U_{OPT}^{t+1}(x) &= \max_{a \in \mathcal{A}} \max_{x' \in \mathcal{X}} \min \left\{ \pi(x'|x, a), \mu(x, a, x'), U_{OPT}^t(x') \right\}, \forall x \in \mathcal{X}, \forall t \geq 1. \end{aligned} \quad (\text{B.20})$$

– Cas pessimiste :

$$\begin{aligned} U_{PES}^0(x) &= \max_{a \in \mathcal{A}} \min_{x' \in \mathcal{X}} \max \left\{ n(\pi(x'|x, a)), \mu(x, a, x') \right\}, \forall x \in \mathcal{X}, \\ U_{PES}^{t+1}(x) &= \max_{a \in \mathcal{A}} \min_{x' \in \mathcal{X}} \max \left\{ n(\pi(x'|x, a)), \min \{ \mu(x, a, x'), U_{PES}^t(x') \} \right\}, \forall x \in \mathcal{X}, \forall t \geq 1. \end{aligned} \quad (\text{B.21})$$

Au bout d'un nombre fini d'itérations, ces deux schémas convergent vers deux fonctions U_{OPT}^* et U_{PES}^* . Une politique possibiliste optimale peut alors être calculée à partir de U_{OPT}^* et U_{PES}^* :

– Cas optimiste :

$$\delta_{OPT}^*(x) = \arg \max_{a \in \mathcal{A}} \max_{x' \in \mathcal{X}} \min \left\{ \pi(x'|x, a), \mu(x, a, x'), U_{OPT}^*(x') \right\}, \forall x \in \mathcal{X}. \quad (\text{B.22})$$

– Cas pessimiste :

$$\delta_{PES}^*(x) = \arg \max_{a \in \mathcal{A}} \min_{x' \in \mathcal{X}} \max \left\{ n(\pi(x'|x, a)), \min \{ \mu(x, a, x'), U_{PES}^*(x') \} \right\}, \forall x \in \mathcal{X}. \quad (\text{B.23})$$

B.10 Décision et logique possibiliste

Syntaxe de la décision en logique possibiliste

A partir de ce modèle logique élémentaire $(\mathcal{X}, \mathcal{D}, \mathcal{Y}, K, P)$, il est possible d'exprimer des problèmes simples de décision dans l'incertain, pour un décideur ayant soit une attitude optimiste, soit une attitude pessimiste.

D'un point de vue *optimiste*, un décideur recherche une décision qui, étant données ses connaissances exprimées par la base $K = \{\varphi_i\}$, est *compatible* avec la satisfaction de ses préférences, exprimées par la base $P = \{\psi_j\}$. Formellement, on recherche une instanciation d des variables de décision, telle que :

$$\left(\bigwedge \varphi_i \right) \wedge d \wedge \left(\bigwedge \psi_j \right) \text{ soit cohérent.} \quad (\text{B.24})$$

D'un point de vue *pessimiste*, par contre, un décideur va rechercher une décision d qui, étant données ses connaissances, va lui permettre la satisfaction à *coup sûr* de toutes ses préférences. Cela se traduit par l'implication suivante :

$$\left(\bigwedge \varphi_i \right) \wedge d \vdash \left(\bigwedge \psi_j \right). \quad (\text{B.25})$$

Exemple B.10.1 Dans l'exemple (4.3.2), on constate que $(\bigwedge \varphi_i) \wedge d \wedge (\bigwedge \psi_j)$ est cohérente pour $d = IDO$, car satisfaite par l'interprétation $\omega = IDO \wedge F \wedge O \wedge \neg G$. La décision $d = IDO$ est donc

satisfaisante, d'un point de vue optimiste. Mais la décision $d = \neg IDO$ est également satisfaisante, car l'interprétation $\omega' = \neg IDO \wedge \neg F \wedge O \wedge \neg G$ satisfait $(\wedge \varphi_i) \wedge d \wedge (\wedge \psi_j)$.

Par contre, dans le cas pessimiste, aucune de ces deux décisions n'est satisfaisante, car on peut trouver dans les deux cas une instanciation des variables de $\mathcal{X} \cup \mathcal{Y}$ satisfaisant $(\wedge \varphi_i) \wedge d \wedge \neg(\wedge \psi_j)$:
 $\omega = IDO \wedge \neg F \wedge \neg O \wedge \neg G$ et $\omega' = \neg IDO \wedge F \wedge O \wedge G$.

Cette forme de raisonnement logique peut s'étendre dans le formalisme de la logique possibiliste (Lang, 1991), permettant ainsi d'exprimer des connaissances et préférences graduelles. Dans ce cadre, K et P sont deux bases logiques stratifiées. $K = \{(\varphi_i, \alpha_i)\}$ est un ensemble de formules valuées, où $\alpha_i \in L$ exprime le degré de certitude de la formule φ_i . $P = \{(\psi_j, \beta_j)\}$ est également un ensemble de clauses valuées, où $\beta_j \in L$ représente un degré de priorité de la préférence exprimée par ψ_j .

A partir de ce nouvel ensemble $(\mathcal{X}, \mathcal{D}, \mathcal{Y}, K, P)$ dans lequel les bases K et P sont stratifiées, on peut définir l'utilité syntaxique d'une décision d en étendant naturellement, dans les cas optimiste et pessimiste, les formules (B.24) et (B.25) :

Définition B.10.1 (Utilité syntaxique d'une décision) Etant donné un modèle de décision logique stratifié $(\mathcal{X}, \mathcal{D}, \mathcal{Y}, K, P)$ et une décision d sur \mathcal{D} , on définit :

– Utilité syntaxique optimiste de d :

$$U_{OPT,K,P}(d) = \max_{\alpha \in L} \left\{ \varphi_{n(\alpha)} \wedge d \wedge \psi_{n(\alpha)} \text{ cohérent} \right\}.$$

– Utilité syntaxique pessimiste de d :

$$U_{PES,K,P}(d) = \max_{\alpha \in L} \left\{ \varphi_{\bar{\alpha}} \wedge d \vdash \psi_{n(\alpha)} \right\}.$$

Où, par définition, $\varphi_{\gamma} = \left(\bigwedge_{\alpha_i > \gamma} \varphi_{\alpha_i} \right)$, $\varphi_{\bar{\gamma}} = \left(\bigwedge_{\alpha_i \geq \gamma} \varphi_{\alpha_i} \right)$ et $\psi_{\gamma} = \left(\bigwedge_{\beta_j > \gamma} \psi_{\beta_j} \right)$.

L'interprétation de ces utilités est la suivante :

- Dans le cas optimiste, une décision est entièrement satisfaisante ($U_{OPT,K,P}(d) = 1_L$) si et seulement si elle est cohérente avec l'ensemble des connaissances ($\varphi_{0_L} = \varphi$) et des préférences ($\psi_{0_L} = \psi$). Le degré de satisfaction décroît au fur et à mesure que des connaissances (φ_i, α_i) de fort degré de certitude ou des préférences (ψ_j, β_j) de haut degré de priorité doivent être retirées afin de restaurer la cohérence. Dans le pire des cas, $\varphi_{\overline{1_L}} \wedge d \wedge \psi_{\overline{1_L}}$ est incohérent, ce qui implique que $U_{OPT,K,P}(d) = 0_L$.
- Dans le cas pessimiste, idéalement les connaissances les plus certaines ($\varphi_{\overline{1_L}}$) suffisent à assurer, avec la décision d , la satisfaction de tous les buts (ψ_{0_L}), soit :

$$U_{PES,K,P}(d) = 1_L \Leftrightarrow \varphi_{\overline{1_L}} \wedge d \vdash \psi_{0_L}.$$

Le degré de satisfaction décroît au fur et à mesure qu'il devient nécessaire de considérer de nouvelles connaissances (moins certaines) pour prouver moins de buts. Dans le pire des cas, l'ensemble des connaissances (φ_{0_L}) ne suffit pas à prouver, avec la décision d , même les buts les plus prioritaires ($\psi_{\overline{1_L}}$) :

$$U_{PES,K,P}(d) = 0_L \Leftrightarrow \varphi_{0_L} \wedge d \not\vdash \psi_{\overline{1_L}}.$$

Sémantique de la décision en logique possibiliste

Nous avons montré [1] que la sémantique correspondant à l'approche syntaxique était celle des utilités qualitatives possibilistes. En effet, soient $\pi_{K,d}$ et μ_P les distributions de possibilité et fonction d'utilité qualitatives construites respectivement à partir des formules de la logique possibiliste $\{(\varphi_i, \alpha_i)\} \cup \{(d, 1_L)\}$ et $\{(\psi_j, \beta_j)\}$ suivant la méthode classique (Dubois et al., 1994) :

$$\begin{aligned} \forall d \in \mathcal{D}, \forall (x, y) \in \mathcal{X} \times \mathcal{Y}, \pi_{K,d}(x, y) &= \min_{\{(\varphi_i, \alpha_i), (x, y, d) \models \neg \varphi_i\}} n(\alpha_i), \\ \forall y \in \mathcal{Y}, \mu_P(y) &= \min_{\{(\psi_j, \beta_j), y \models \neg \psi_j\}} n(\beta_j). \end{aligned}$$

Nous avons montré [1] le résultat suivant :

Théorème B.10.1 (Utilité sémantique d'une décision d) *Les définitions syntaxiques et sémantiques de l'utilité qualitative optimiste (resp. pessimiste) sont équivalentes. Soit $(\mathcal{X}, \mathcal{D}, \mathcal{Y}, K, P)$ un modèle de décision logique possibiliste et d une décision quelconque :*

$$\begin{aligned} U_{OPT,K,P}(d) = U_{OPT,\pi_{K,d},\mu_P}(d) &= \max_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \min\{\pi_{K,d}(x, y), \mu_P(y)\}. \\ U_{PES,K,P}(d) = U_{PES,\pi_{K,d},\mu_P}(d) &= \min_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \max\{n(\pi_{K,d}(x, y)), \mu_P(y)\}. \end{aligned}$$

Annexe C

Extensions du cadre des processus décisionnels de Markov

C.1 Programmation linéaire approchée

Une méthode de programmation linéaire approchée spécifique aux PDMG

Nous avons choisi une approche un peu différente de celle de (de Farias and Van Roy, 2003), mais qui présente l'avantage de fournir facilement une borne supérieure de la perte d'optimalité au pire cas de la politique approchée.

Dans notre approche, w^* est solution du problème de minimisation suivant :

$$\begin{aligned} \min \quad & \phi \\ \text{t.q.} \quad & \begin{cases} \phi \geq \|v_w - \mathcal{B}(v_w)\|_\infty, \\ w = (w_1, \dots, w_k) \text{ libre, } \phi \text{ libre,} \end{cases} \end{aligned} \quad (\text{C.1})$$

où \mathcal{B} est l'opérateur de Bellman, transformant une fonction de valeur v quelconque, défini par :

$$v'(x) = \mathcal{B}(v)(x) = \max_{a \in \mathcal{A}} \left\{ r(x, a) + \gamma \sum_{x' \in \mathcal{X}} p(x'|x, a) \cdot v(x') \right\}, \forall x \in \mathcal{X}.$$

Le problème (C.1) peut s'écrire de manière équivalente :

$$\begin{aligned} \min \quad & \phi \\ \text{t.q.} \quad & \begin{cases} \phi \geq v_w(x) - \mathcal{B}(v_w)(x), \forall x \in \mathcal{X}, \\ \phi \geq \mathcal{B}(v_w)(x) - v_w(x), \forall x \in \mathcal{X}, \\ w = (w_1, \dots, w_k) \text{ libre, } \phi \text{ libre.} \end{cases} \end{aligned} \quad (\text{C.2})$$

Malheureusement, le problème (C.2) n'est pas un programme linéaire. Le deuxième ensemble de contraintes ($\phi \geq \mathcal{B}(v_w)(x) - v_w(x), \forall x$) se ramène bien à un ensemble de contraintes linéaires (une par couple (x, a)), mais le premier ($\phi \geq v_w(x) - \mathcal{B}(v_w)(x), \forall x$) ne s'y ramène pas, à cause de la présence du signe $-$ devant la transformée de Bellman $\mathcal{B}(v_w)(x)$. Nous avons suggéré de mettre à profit la structure d'un PDMG afin de remplacer le premier ensemble de contraintes non-linéaires par un ensemble de contraintes linéaires plus "contraignant" (restreignant l'ensemble des solutions admissibles¹). Le programme linéaire obtenu n'est pas équivalent au premier, mais ses solutions satisfont toutes les contraintes du programme initial. De plus, un choix judicieux des fonctions de base

¹Le premier ensemble de contraintes se réécrit : $\min_{a \in \mathcal{A}} \{v_w(x) - r(x, a) - \gamma \sum_{x' \in \mathcal{X}} p(x'|x, a) \cdot v_w(x')\}, \forall x \in \mathcal{X}$. L'approximation suggérée consiste à choisir aléatoirement une action a , ce qui permet de supprimer l'opérateur min et de retrouver un programme linéaire.

de l'espace d'approximation de la fonction de valeur permet la décomposition du nouveau système en un ensemble de systèmes linéaires indépendants, de taille réduite.

Plus précisément, considérons $\langle \mathcal{X}, \mathcal{A}, p, r, G \rangle$ un PDMG, avec $\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_n$, et $\mathcal{A} = \mathcal{A}_1 \times \dots \times \mathcal{A}_n$. Dans la suite de cette section, l'espace d'états de la variable \mathcal{X}_i est identifié avec l'ensemble $\{1, \dots, |\mathcal{X}_i|\} \subseteq \mathbb{N}$. Nous considérerons une approximation paramétrée de la fonction de valeur, similaire à celle de (5.2) dans laquelle les fonctions de base sont les suivantes (il y en a exactement $\sum_{i=1}^n |\mathcal{X}_i|$) :

Définition C.1.1 (Fonctions de base du PDMG)

Soit $\langle \mathcal{X}, \mathcal{A}, p, r, G \rangle$ un PDMG dont l'espace d'états est $\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_n$, et l'espace d'actions $\mathcal{A} = \mathcal{A}_1 \times \dots \times \mathcal{A}_n$. Les fonctions de base : $h_{ij}, i \in 1 \dots n, j \in 1 \dots |\mathcal{X}_i|$, sont définies par :

$$h_{ij}(x_i) = \begin{cases} 1 & \text{si } x_i = j \text{ et} \\ 0 & \text{sinon.} \end{cases}$$

La fonction de valeur paramétrée v_w s'écrit alors : $v_w(x) = \sum_{i=1}^n v_w^i(x_i), \forall x \in \mathcal{X}$,

où les fonctions élémentaires v_w^i sont définies par $v_w^i(x_i) = \sum_{j=1}^{|\mathcal{X}_i|} w_{ij} h_{ij}(x_i), \forall x_i \in \mathcal{X}_i$.

La transformée de Bellman $\mathcal{B}(v_w)(x)$ peut alors se décomposer en :

Proposition C.1.1 (Transformée de Bellman)

$$\begin{aligned} \mathcal{B}(v_w)(x) &= \sum_{i=1}^n \mathcal{B}_i(v_w^i)(x_{N(i)}), \forall x \in \mathcal{X} \text{ où} \\ \mathcal{B}_i(v_w^i)(x_{N(i)}) &= \max_{a_i \in \mathcal{A}_i} \left\{ r_i(x_{N(i)}, a_i) + \gamma \sum_{x'_i \in \mathcal{X}_i} p_i(x'_i | x_{N(i)}, a_i) v_w^i(x'_i) \right\}, \forall x_{N(i)} \in \mathcal{X}_{N(i)}. \end{aligned}$$

Cette décomposition de $\mathcal{B}(v_w)(x)$ en une somme de termes $\mathcal{B}_i(v_w^i)(x_{N(i)})$, ne dépendant chacun que de la valeur des variables du voisinage du noeud i , permet de calculer en un temps raisonnable une approximation de la solution du système (C.2). Plus précisément, au lieu de résoudre le système (C.2), non-linéaire et comportant $|\mathcal{X}| \times |\mathcal{A}|$ contraintes, il suffit de résoudre les n problèmes non-linéaires indépendants suivants :

$$\begin{aligned} \min \quad & \phi_i \\ \text{t.q.} \quad & \begin{cases} \phi_i \geq v_w^i(x_i) - \mathcal{B}_i(v_w^i)(x_{N(i)}), \forall x_{N(i)} \in \mathcal{X}_{N(i)}, \\ \phi_i \geq \mathcal{B}_i(v_w^i)(x_{N(i)}) - v_w^i(x_i), \forall x_{N(i)} \in \mathcal{X}_{N(i)}, \\ w_i = (w_{i1}, \dots, w_{i|\mathcal{X}_i|}) \text{ libre, } (\phi_1, \dots, \phi_n) \text{ libre} \end{cases} \end{aligned} \quad (\text{C.3})$$

Comme le système (C.2), ces n problèmes sont non-linéaires. Nous avons donc suggéré de remplacer les contraintes non-linéaires par l'ensemble de contraintes linéaires suivant :

$$\phi_i \geq v_w^i(x_i) - \psi(x_{N(i)}, \tilde{a}_i), \forall x_{N(i)},$$

où \tilde{a}_i est une action aléatoirement choisie.

Celui-ci est plus contraignant car il implique que chaque contrainte soit satisfaite pour l'action \tilde{a}_i fixée. Ce remplacement permet de transformer le système (C.3) en un ensemble de n problèmes $LP_i, i \in \{1 \dots n\}$, *linéaires* et *indépendants*, dont l'aggrégation des solutions forme une solution *faisable* de (C.3) :

$$\begin{array}{l} \min \quad \phi_i \\ \text{tq.} \quad \left\{ \begin{array}{l} \phi_i \geq \sum_{j=1}^{|\mathcal{X}_i|} w_{ij} \left(h_{ij}(x_i) - \gamma p_i(x'_i = j | x_{N(i)}, a_i) \right) - r_i(x_{N(i)}, a_i), \\ \forall x_{N(i)} \in \mathcal{X}_{N(i)}, \forall a_i \in \mathcal{A}_i, \\ \phi_i \geq r_i(x_{N(i)}, \tilde{a}_i) - \sum_{j=1}^{|\mathcal{X}_i|} w_{ij} \left(h_{ij}(x_i) - \gamma p_i(x'_i = j | x_{N(i)}, \tilde{a}_i) \right), \\ \forall x_{N(i)} \in \mathcal{X}_{N(i)}, \\ (w_{ij})_{j=1 \dots |\mathcal{X}_i|} \text{ libres, } (\phi_1, \dots, \phi_n) \text{ libres.} \end{array} \right. \end{array} \quad (\text{C.4})$$

Une approximation v_{w^*} de la fonction de valeur optimale v^* peut être calculée en résolvant les n systèmes linéaires LP_i (C.4), comportant chacun $|\mathcal{X}_i| + 1$ variables et $|\mathcal{X}_{N(i)}| \times (|\mathcal{A}_i| + 1)$ contraintes, au lieu de résoudre un système non linéaire unique comportant $|\mathcal{X}| + 1$ variables et $2 \times |\mathcal{X}| \times |\mathcal{A}|$ contraintes. La solution de chaque sous-système LP_i permet de calculer la valeur des paramètres $(w_{ij}^*)_{j=1 \dots |\mathcal{X}_i|}$ et, une fois l'ensemble des n sous-systèmes résolus, de définir une solution approchée à partir de ces valeurs. La politique "gloutonne" δ_{w^*} associée à la fonction de valeur v_{w^*} est *locale* : $\delta_{w^*}(x)$ peut être obtenue comme une réunion de n politiques $(\delta_{w^*}^1(x_{N(1)}), \dots, \delta_{w^*}^n(x_{N(n)}))$ indépendantes, définies par :

$$\delta_{w^*}^i(x_{N(i)}) = \arg \max_{a_i \in \mathcal{A}_i} \left\{ r_i(x_{N(i)}, a_i) + \gamma \sum_{x'_i \in \mathcal{X}_i} p_i(x'_i | x_{N(i)}, a_i) v_{w^*}^i(x'_i) \right\}, \\ \forall x_{N(i)} \in \mathcal{X}_{N(i)}, \forall i \in 1, \dots, n.$$

Borne sur l'erreur d'approximation

Notons $\phi_w = \|v_w(x) - \mathcal{B}(v_w)(x)\|_\infty$, l'*erreur de Bellman* associée à v_w et $\varepsilon_w = \|v^* - v_w\|_\infty = \max_{x \in \mathcal{X}} |v^*(x) - v_w(x)|$, la perte d'optimalité au pire cas associée au choix de paramètres w pour la politique δ_w . Cette expression de la perte d'optimalité ε_w nécessite de connaître v^* pour être opérationnelle. Néanmoins, il est possible de calculer une borne sur ε_w à partir d'une borne sur ϕ_w , grâce au résultat suivant (Williams and Baird, 1993) :

$$\|v^* - v\|_\infty \leq \frac{2\gamma}{1-\gamma} \|v - \mathcal{B}(v)\|_\infty, \forall v : \mathcal{X} \rightarrow \mathbb{R}$$

Il est donc possible d'obtenir la borne suivante sur la perte d'optimalité :

$$\varepsilon_w \leq \frac{2\gamma}{1-\gamma} \phi_w.$$

Une erreur de Bellman *locale*, $\phi_{w^*}^i = \|v_{w^*}^i - \mathcal{B}(v_{w^*}^i)\|_\infty$ peut être facilement calculée pour chacun des programmes linéaires indépendants, ce qui permet de borner l'erreur de Bellman totale ϕ_{w^*} associée à v_{w^*} :

$$\phi_{w^*} \leq \sum_{i=1}^n \phi_{w^*}^i.$$

A partir de cette borne, nous obtenons la borne supérieure suivante sur la perte d'optimalité au pire cas ε_{w^*} engendrée par l'utilisation de la politique approchée δ_{w^*} à la place de la politique optimale

δ^* :

$$\varepsilon_{w^*} \leq \frac{2\gamma}{1-\gamma} \sum_{i=1}^n \phi_{w^*}^i. \quad (\text{C.5})$$

Notons que cette borne surestime très largement la perte d'optimalité dans le cas général et n'est donc malheureusement pas d'un très grand intérêt pratique...

C.2 Itération de la politique approchée en “champ moyen”

Fonction de valeur d'une politique et mesure d'occupation

La mesure d'occupation est reliée au temps qu'un processus stochastique passe dans un état donné, sachant l'état initial du processus. Par extension, les mesures d'occupation peuvent aussi être utilisées dans le contexte des PDM avec facteur d'amortissement. Si X^t représente la variable aléatoire correspondant à l'état du processus généré par une politique δ dans un PDM au temps t , $P_\delta(X^t = y | X^0 = x)$ représente la probabilité que la valeur de X^t soit y , sachant que l'état initial est x et que la politique δ est appliquée.

Définition C.2.1 (Mesure d'occupation dans un PDM)

Soit $\langle \mathcal{X}, \mathcal{A}, p, r \rangle$ un PDM stationnaire et γ un facteur d'amortissement ($0 < \gamma < 1$). Soit $\delta : \mathcal{X} \rightarrow \mathcal{A}$ une politique fixée et $x \in \mathcal{X}$ un état initial. La mesure d'occupation $P_{x,\delta,\gamma} : \mathcal{X} \rightarrow [0, 1]$ est définie par :

$$\forall y \in \mathcal{X}, P_{x,\delta,\gamma}(y) = (1-\gamma) \sum_{t=0}^{+\infty} \gamma^t \cdot P_\delta(X^t = y | X^0 = x). \quad (\text{C.6})$$

La fonction $P_{x,\delta,\gamma}$ est une distribution de probabilité sur \mathcal{X} , car :

1. la somme infinie converge ($\gamma < 1$) pour tous les couples (δ, x) ,
2. $\sum_{y \in \mathcal{X}} P_\delta(X^t = y | X^0 = x) = 1$ et,
3. $(1-\gamma) \sum_{t=0}^{+\infty} \gamma^t = 1$.

(Altman, 1999) a montré que la fonction de valeur v_δ d'une politique δ pouvait être calculée à partir de la distribution $P_{x,\delta,\gamma}$:

Proposition C.2.1 (Fonction de valeur et mesure d'occupation)

$$\forall \delta \in \mathcal{A}^{\mathcal{X}}, \forall x \in \mathcal{X}, v_\delta(x) = EU_\delta(x) = \frac{1}{1-\gamma} \cdot \sum_{y \in \mathcal{X}} P_{x,\delta,\gamma}(y) \cdot r(y, \delta(y)). \quad (\text{C.7})$$

Dans [27], [28], nous avons proposé une méthode de résolution approchée de PDMG basée sur l'expression (C.7). Plus précisément, l'expression (C.7) se réécrit de la manière suivante lorsque les récompenses sont locales, comme dans le cas d'un PDMG :

Proposition C.2.2 (Décomposition de la fonction de valeur d'une politique)

Soit $\langle \mathcal{X}, \mathcal{A}, p, r, G \rangle$ un PDMG. Soit δ , une politique quelconque :

$$\begin{aligned} v_\delta(x) &= \sum_{i=1}^n v_\delta^i(x), \forall x, \text{ où} \\ v_\delta^i(x) &= \sum_{t=0}^{+\infty} \gamma^t \cdot \sum_{y \in \mathcal{X}} P_\delta(X^t = y | X^0 = x) \cdot r_i(y_{N(i)}, \delta_i(y_{N(i)})). \end{aligned} \quad (\text{C.8})$$

L'idée principale de la méthode de décomposition que nous avons proposée est d'approcher la distribution de probabilité $P_\delta(X^t = y | X^0 = x)$ par une distribution factorisée (produit de lois marginales), afin de simplifier le calcul de l'expression de $v_\delta^i(x)$. En effet, si $P_\delta(X^t = y | X^0 = x)$ peut être réécrite sous la forme d'un produit $\prod_{i=1}^n Q_\delta^i(X_i^t = y_i | X_i^0 = x_i)$, alors (C.8) peut se réécrire sous la forme suivante, qui permet une expression de v_δ en espace polynomial (à la condition que la largeur du graphe G soit bornée) :

$$v_\delta^i(x_{N(i)}) = \sum_{t=0}^{+\infty} \gamma^t \cdot \left(\sum_{y_{N(i)} \in \mathcal{X}_{N(i)}} \left(\prod_{j \in N(i)} Q_\delta^j(X_j^t = y_j | X_j^0 = x_j) \right) \cdot r_i(y_{N(i)}, \delta_i(y_{N(i)})) \right).$$

Approximation de type champ moyen de la fonction de valeur

Malheureusement, il n'y a aucune raison pour que P_δ soit factorisable. Cependant, nous avons proposé d'approcher la distribution de probabilité $P_\delta(X^t = y | X^0 = x)$ par une distribution factorisée, afin de simplifier le calcul de l'expression des termes $v_\delta^i(x)$. Cette approximation est de type *champ moyen*. Le principe du champ moyen vient, à l'origine, du domaine de la mécanique statistique (Chandler, 1987) et de l'étude des systèmes de particules en interaction et des distributions de Gibbs. Ce principe a depuis été largement repris dans d'autres domaines tels que l'analyse d'image, les modèles graphiques, l'épidémiologie, etc. L'idée sous-jacente est d'approcher un système de variables en interaction par un système de variables indépendantes, obtenu en approchant l'influence des variables voisines par leur influence “moyenne”.

L'approximation est dérivée d'un critère de distance entre distributions jointes. L'approximation en champ moyen d'une distribution multidimensionnelle $P(u_1, \dots, u_m)$ est définie comme la loi jointe Q de variables *indépendantes* u_1, \dots, u_m la plus proche de P , au sens de la divergence de Kullback-Leibler,

$$KL(Q|P) = E_Q \left[\log \left(\frac{Q}{P} \right) \right]. \quad (\text{C.9})$$

Le minimum de $KL(Q|P)$ est recherché parmi les distributions jointes $Q(u_1, \dots, u_m) = \prod_{i=1}^m Q_i(u_i)$.

Dans le cadre PDMG, la difficulté du calcul de la loi jointe $P_\delta(X^t = y | X^0 = x)$ provient de la dépendance spatio-temporelle entre les états des sites. En effet, si $X^t = (X_1^t, \dots, X_n^t)$ est la variable aléatoire représentant l'état des sites au temps t , pour une politique donnée δ chaque variable X_i^t dépend de l'état de toutes les variables “voisines” à l'instant précédent, $X_{N(i)}^{t-1}$ à travers les probabilités de transition locales. Nous avons proposé de nous affranchir de la dépendance temporelle en appliquant le principe du champ moyen à la distribution jointe de X^t et X^0 , complètement déterminée par la loi de X^0 et la loi conditionnelle de X^t sachant X^0 . Nous recherchons donc une famille d'approximations $\{Q_\delta^t(x^0, x^t)\}_{t=1, \dots, \infty}$ des lois jointes $\{P_\delta^t(x^0, x^t)\}_{t=1, \dots, \infty}$. Suivant le principe du champ moyen, les lois $Q_\delta^t(x^0, x^t)$ sont des lois produits de lois jointes indépendantes, $Q_\delta^{t,i}(x_i^0, x_i^t)$.

La procédure d'approximation que nous avons proposée est itérative. Puisqu'aucune information a priori n'est disponible en général, nous supposons que $P^0(x^0)$ est factorisable. Ensuite, au pas de temps $t = 1$, la distribution jointe exacte à la forme suivante (en se servant de la propriété de Bayes, $P_\delta^1(x^0, x^1) = P_\delta^1(x^1 | x^0) \cdot P^0(x^0)$) :

$$P^0(x^0) = \prod_{i=1}^n P^{0,i}(x_i^0),$$

$$P_\delta^1(x^1|x^0) = \prod_{i=1}^n p_i(x_i^1|x_{N(i)}^0, \delta_i(x_{N(i)}^0)) \text{ et}$$

$$P_\delta^1(x^0, x^1) = \prod_{i=1}^n P^{0,i}(x_i^0) \cdot p_i(x_i^1|x_{N(i)}^0, \delta_i(x_{N(i)}^0)).$$

On recherche une approximation en champ moyen de la loi jointe P_δ^1 sous la forme :

$$Q_\delta^0(x^0) = P^0(x^0) = \prod_{i=1}^n Q^{0,i}(x_i^0),$$

$$Q_\delta^1(x^1|x^0) = \prod_{i=1}^n Q_\delta^{1,i}(x_i^1|x_i^0) \quad (\text{C.10})$$

$$Q_\delta^1(x^0, x^1) = \prod_{i=1}^n Q^{0,i}(x_i^0) \cdot Q_\delta^{1,i}(x_i^1|x_i^0).$$

Dans le modèle approché, X_i^1 ne dépend que de X_i^0 , pour toute politique fixée δ . Notons \mathcal{Q} l'ensemble de toutes les distributions jointes de X^0 et X^1 pouvant être décomposées suivant (C.10). Parmi tous les éléments de \mathcal{Q} , nous recherchons la plus proche de la loi jointe exacte, au sens de la divergence de Kullback-Leibler. Puisque $Q^0 = P^0$, la seule approximation dans le calcul de la loi jointe provient du calcul de $Q_\delta^1(x^1|x^0)$. Nous avons montré [27], [28] le résultat suivant :

Proposition C.2.3 (Approximation en champ moyen de la loi conditionnelle)

L'approximation en champ moyen de P_δ^1 est définie par :

$$Q_\delta^1(x^1|x^0) = \arg \min_{Q \in \mathcal{Q}} KL(Q(x^1|x^0) \cdot P^0(x^0) | P_\delta(x^1|x^0) \cdot P^0(x^0)), \forall x^0, x^1 \in \mathcal{X}.$$

On peut montrer que :

$$Q_\delta^1(x^1|x^0) = \prod_{i=1}^n Q_\delta^{1,i}(x_i^1|x_i^0), \text{ où}$$

$$Q_\delta^{1,i}(x_i^1|x_i^0) \propto \exp \left(E_{P^0} \left[\log p_i \left(x_i^1|x_i^0, X_{N(i)-\{i\}}^0, \delta_i(x_i^0, X_{N(i)-\{i\}}^0) \right) \right] \right). \quad (\text{C.11})$$

Il est possible de calculer exactement les termes $Q_\delta^{1,i}(x_i^1|x_i^0)$, mais il est courant d'utiliser, afin de simplifier les calculs, l'approximation supplémentaire de $Q_\delta^{1,i}$, obtenue en échangeant les opérateurs "espérance" et "logarithme" dans l'expression (C.11). Nous obtenons alors l'expression :

$$Q_\delta^{1,i}(x_i^1|x_i^0) = E_{P^0} \left[p_i(x_i^1|x_i^0, X_{N(i)\setminus i}^0, \delta_i(x_i^0, X_{N(i)\setminus i}^0)) \right],$$

$$= \sum_{x_{N(i)-i}^0} \left(\prod_{j \in N(i)-i} P^{0,j}(x_j^0) \right) \cdot p_i(x_i^1|x_{N(i)}^0, \delta_i(x_{N(i)}^0)).$$

Il est facile de vérifier que cette dernière quantité est normalisée, contrairement à l'expression (C.11). De plus, elle est facile à interpréter puisqu'elle consiste simplement à prendre comme probabilité de transition approchée l'espérance des probabilités de transition (par rapport à la loi P^0), pour toutes les valeurs possibles des variables du voisinage.

Connaissant maintenant une approximation en champ moyen de la probabilité de transition $P_\delta(x^1|x^0)$ et puisque cette distribution à l'instant $t = 1$ est factorisable, il est possible de réitérer l'approximation en champ moyen (minimisation de la divergence de Kullback-Leibler) pour tous les instants $t \geq 2$, et d'obtenir ainsi des distributions $Q_\delta^t(x^t|x^0)$ factorisées :

$$\begin{aligned} Q_\delta^{t,i}(x_i^t|x_i^{t-1}) &= E_{Q_\delta^{t-1,i}} \left[p_i(x_i^t|x_i^{t-1}, X_{N(i)\setminus i}^{t-1}, \delta_i(x_i^{t-1}, X_{N(i)\setminus i}^{t-1})) \right], \\ Q_\delta^{t,i}(x_i^t|x_i^0) &= \sum_{x_i^{t-1} \in \mathcal{X}_i} Q_\delta^{t,i}(x_i^t|x_i^{t-1}) \cdot Q_\delta^{t-1,i}(x_i^{t-1}|x_i^0). \end{aligned} \quad (\text{C.12})$$

Il faut noter que les probabilités de transition approchées $Q_\delta^t(x^t|x^{t-1})$ dépendent de t , alors que les probabilités de transition exactes du modèle initial sont indépendantes de t .

Ce schéma nous permet d'obtenir une approximation de la mesure d'occupation $P_{x,\delta,\gamma}$, nécessaire à l'évaluation de la fonction de valeur (C.7). Nous pouvons maintenant utiliser l'approximation en champ moyen de la mesure d'occupation (C.12) pour mettre en œuvre l'étape d'évaluation. Nous obtenons alors directement les approximations suivantes pour chaque terme $v_\delta^i(x)$ de la fonction de valeur :

$$v_\delta^i(x) \approx \hat{v}_\delta^i(x_{N(i)}) = \sum_{t=0}^{+\infty} \gamma^t \cdot \sum_{x_{N(i)}^t \in \mathcal{X}_{N(i)}} \left(\prod_{j \in N(i)} Q_\delta^{t,j}(x_j^t|x_j^0 = x_j) \cdot r_i(x_{N(i)}^t, \delta_i(x_{N(i)}^t)) \right). \quad (\text{C.13})$$

Nous avons défini une approximation similaire de la phase d'amélioration de la politique, puis un algorithme complet d'itération de la politique approchée. Pour plus de détails, le lecteur est invité à se référer à [27], [28].

C.3 Processus décisionnels de Markov épistémiques

Révision de l'état de croyance dans un PDME

Dans un PDME, la mise à jour de l'état de croyance courant après qu'une observation o ait été perçue à la suite de l'application d'une action épistémique a peut être calculée comme dans un PDMPO, à l'aide de la formule suivante :

$$b_a^o(x') = \frac{O(x', a, o) \sum_{x \in \mathcal{X}} (p(x'|x, a) \cdot b(x))}{Pr(o|b, a)}, \quad \forall x' \in \mathcal{X},$$

où $Pr(o|b, a)$ est un facteur de normalisation indépendant de x' .

Cependant, dans un PDME l'état physique du système ne change pas au cours du temps : $p(x|x, a) = 1, \forall x \in \mathcal{X}, \forall a \in \mathcal{A}$. Il en résulte une simplification de la formule de mise à jour de l'état de croyance courant dans le cas d'un PDME :

$$b_a^o(x') = \frac{O(x', a, o) \cdot b(x')}{Pr(o|b, a)}, \quad \forall a \in \mathcal{A}, o \in \mathcal{O}, b \in \mathcal{B} \text{ et } \forall x' \in \mathcal{X}, \quad (\text{C.14})$$

où le facteur de normalisation $Pr(o|a, b)$ est déterminé par :

$$Pr(o|b, a) = \sum_{x' \in \mathcal{X}} O(x', a, o) \cdot b(x').$$

Politiques épistémiques

Dans un PDME à horizon fini $\langle \mathcal{X}, \mathcal{A}, \mathcal{O}, O, \rho, b_0, T \rangle$, comme dans un PDMPO, une politique épistémique δ peut être représentée par un arbre de décision τ_δ de profondeur T , où les nœuds représentent les actions, et les arcs partant d'un nœud action représentent les observations possibles. De manière équivalente, une politique peut être vue comme une fonction partielle δ associant à l'ensemble fini des séquences de paires (action, observation), une action. $\delta()$ est la première action de la politique et $\delta(a_1, o_1, \dots, a_i, o_i)$ est l'action appliquée après avoir exécuté les actions a_1, \dots, a_i et observé o_1, \dots, o_i . Il peut être commode, enfin, de voir une politique δ comme une fonction de l'ensemble continu \mathcal{B} des états de croyance vers \mathcal{A} , même si l'ensemble des états de croyance *atteignables* est fini, lorsque \mathcal{X} , \mathcal{A} et T sont finis.

L'utilité d'une politique δ appliquée à l'état de croyance initial b_0 est l'espérance de la somme des récompenses reçues le long de toutes les branches de l'arbre τ_δ :

$$v^\delta(b_0) = E_{\tau_\delta} \left[\sum_{t=0}^T \rho(b_t, a_t) | b_0 \right].$$

La fonction d'utilité v^δ d'une politique δ peut être calculée récursivement, car elle satisfait les équations de programmation dynamique suivantes :

$$\begin{aligned} v^{\delta,t}(b) &= -C(\delta(b)) + \sum_{o \in \mathcal{O}} Pr(o|b, \delta(b)) \cdot v^{\delta,t+1}(b_{\delta(b)}^o) \text{ si } \delta(b) \neq stop, \forall t < T, \forall b \in \mathcal{B}, \\ v^{\delta,t}(b) &= -\rho(b) \text{ si } t = T \text{ ou si } \delta(b) = stop. \end{aligned} \quad (\text{C.15})$$

Ces équations portent a priori sur l'ensemble continu \mathcal{B} des états de croyance. Néanmoins, l'ensemble des états de croyance effectivement accessibles par δ est fini. En construisant l'arbre τ_δ correspondant à la politique δ , nous pouvons construire itérativement l'ensemble des états de croyance accessibles. L'état de croyance b_0 est attaché à la racine de l'arbre. A chaque nœud de l'arbre, associé à l'état de croyance courant b , on attache l'action $a = \tau_\delta(b)$. Les états de croyance (b_a^o) successeurs de b , correspondant à l'action $a = \tau_\delta(b)$ et l'observation o sont ensuite ajoutés à l'arbre. Une fois l'arbre entièrement construit, l'utilité de la politique δ peut être calculée récursivement *vers l'arrière*, en attachant des récompenses aux états de croyance feuilles de l'arbre, et des coûts à toutes les actions.

Une *politique optimale* δ^* pour un PDME à horizon fini peut également être calculée récursivement, grâce à des équations similaires aux équations (C.15) :

$$\begin{aligned} v^{*,t}(b) &= \max \left\{ \rho(b), \max_{a \in \mathcal{A}} \left(-C(a) + \sum_{o \in \mathcal{O}} Pr(o|b, a) \cdot v^{*,t+1}(b_a^o) \right) \right\}, \forall t < T, \forall b \in \mathcal{B}, \\ v^{*,T}(b) &= -\rho(b). \\ \delta_t^*(b) &= \begin{cases} \arg \max_{a \in \mathcal{A}} \left(-C(a) + \sum_{o \in \mathcal{O}} Pr(o|b, a) \cdot v^{*,t+1}(b_a^o) \right) & \text{si } v^{*,t}(b) > \rho(b) \text{ et} \\ stop & \text{sinon.} \end{cases} \end{aligned}$$

Cet algorithme récursif de calcul de politiques optimales d'un PDME est de complexité exponentielle (relativement à la taille du problème) en temps et également en espace, si la politique optimale est stockée. Les actions optimales dans les états de croyance effectivement atteints peuvent toutefois être calculées "en ligne", ce qui permet de n'utiliser qu'un espace polynomial. Dans [41], [31], nous nous sommes intéressés plus précisément à la complexité théorique du problème de décision associé à la résolution d'un PDME. La section suivante résume les résultats que nous avons obtenus.

C.4 Complexité du problème de décision associé à la résolution d'un PDME

Cas général

Considérons le problème d'EXISTENCE D'UNE POLITIQUE dans un PDME :

Soit un PDME $\langle \mathcal{X}, \mathcal{A}, \mathcal{O}, O, \rho, b_0, T \rangle$ et un nombre réel v . Existe-t-il une politique δ dont la valeur espérée dans l'état initial, $v_\delta(b_0)$, est au moins égale à v ?

Rappelons que le problème équivalent dans le cadre des PDMPO est PSPACE-complet dans le cas d'un horizon court ($T \leq |\mathcal{X}|^m$), EXPSPACE dans le cas où l'horizon est fini mais long ($T \leq e^{m|\mathcal{X}|}$) et indécidable lorsque l'horizon est infini (Papadimitriou and Tsitsiklis, 1987; Mundhenk et al., 2000; Madani et al., 2003). Dans [41], [31], nous avons prouvé que le problème de décision lié à l'existence d'une politique pour les PDME est également PSPACE-complet lorsque l'horizon est court (et a fortiori PSPACE -difficile dans le cas général de l'horizon fini).

Proposition C.4.1 *Le problème EXISTENCE D'UNE POLITIQUE pour un PDME à horizon court est PSPACE-complet .*

Ce résultat est plutôt négatif, puisqu'il montre qu'a priori un PDME est aussi difficile à résoudre qu'un PDMPO. Cependant, nous avons exhibé dans [41], [31] une classe particulière de PDME plus "faciles" à résoudre.

PDME avec fonction d'observation déterministe

Considérons la sous-classe de PDME dans laquelle la fonction d'observation est déterministe, c'est-à-dire que $o = O(x, a) \in \mathcal{O}, \forall x \in \mathcal{X}, \forall a \in \mathcal{A}$. Puisque la fonction d'observation est déterministe, la même observation o est toujours reçue lorsque l'action a est appliquée en x . Ceci implique que, si nous écrivons $X_a^o = \{x \in \mathcal{X}, O(x, a) = o\}$, alors $\{X_a^o\}_{o \in \mathcal{O}(x, a)}$ forme une partition de \mathcal{X} , pour toute action non-terminante a . Dans ce cas, la mise à jour de l'état de croyance (équation C.14) s'écrit :

$$b_a^o(x) = \frac{b(x)}{Pr(o|a, b)} = \frac{b(x)}{b(X_a^o)} \text{ si } x \in X_a^o \text{ et } b_a^o(x) = 0 \text{ sinon} \quad (\text{C.16})$$

La forme de cette équation permet de montrer que tout état de croyance $b_{a_1, \dots, a_t}^{o_1, \dots, o_t}$ accessible après une séquence de paires action/observation a une expression bien particulière, et facilement calculable :

$$b_{a_1, \dots, a_t}^{o_1, \dots, o_t}(x) = \frac{b(x)}{b(\bigcap_{i=1}^t X_{a_i}^{o_i})} \text{ si } x \in \bigcap_{i=1}^t X_{a_i}^{o_i} \text{ et } b_{a_1, \dots, a_t}^{o_1, \dots, o_t}(x) = 0 \text{ sinon.} \quad (\text{C.17})$$

A partir de la proposition (C.17), nous déduisons que $\forall t \leq T$ et pour toute trajectoire partielle possible $(a_1, o_1, \dots, a_t, o_t)$, l'état de croyance courant $b_{a_1, \dots, a_t}^{o_1, \dots, o_t}$ est uniquement déterminé par l'état de croyance initial b et le sous-ensemble d'états $(\bigcap_{i=1}^t X_{a_i}^{o_i})$.

Si nous considérons une politique δ et si nous définissons $X_\delta^{o_1, \dots, o_t}$ par :

$$X_\delta^\emptyset = \mathcal{X} \text{ et } X_\delta^{o_1, \dots, o_t, o_{t+1}} = X_\delta^{o_1, \dots, o_t} \cap X_{\delta(a_1, o_1, \dots, a_t, o_t)}^{o_{t+1}} = \bigcap_{i=1}^t X_{a_i}^{o_i},$$

où pour tout i , $a_i = \delta(a_1, o_1, \dots, a_{i-1}, o_{i-1})$. Alors nous pouvons montrer la proposition suivante :

Proposition C.4.2 *Pour δ fixée et pour tout $t \leq T$, l'ensemble des sous-ensembles non vides de $\{X_\delta^{o_1, \dots, o_t}\}_{(o_1, \dots, o_t) \in \mathcal{O}^t}$ forme une partition de \mathcal{X} .*

Une conséquence directe de la proposition (C.4.2) est que l'ensemble des *supports* de tous les états de croyance accessibles à un instant t fixé forme une partition de \mathcal{X} . $|\mathcal{X}|$ est donc un majorant du nombre d'états de croyance accessibles à l'instant t , ce qui implique que l'arbre τ_δ représentant une politique δ quelconque a une taille bornée par $|\mathcal{X}| \times T$ dans un PDME à horizon fini.

Il ne nous reste plus qu'à noter que dans un PDME dont la fonction d'observation est déterministe, il n'est jamais utile d'appliquer deux fois la même action épistémique (puisque l'on est certain de recevoir une nouvelle fois la même observation). Ceci permet de réduire la recherche d'une politique optimale aux politiques de profondeur maximale $|\mathcal{A}| + 1$. Par conséquent, tout PDME à horizon fini dont la fonction d'observation est déterministe peut être réduit à un problème équivalent dont l'horizon est borné par $|\mathcal{A}| + 1$.

Nous en déduisons enfin que la politique optimale de tout PDME à horizon fini dont la fonction d'observation est déterministe, peut être représentée en espace $O(|\mathcal{X}| \times |\mathcal{A}|)$.

Dans [41], nous avons montré un résultat plus fort :

Proposition C.4.3 *Le problème EXISTENCE D'UNE POLITIQUE pour un PDME dont la fonction d'observation est déterministe est NP-complet .*

Le résultat de difficulté est obtenu par réduction polynomiale du problème MINIMUM SET COVER (Karp, 1972). Nous avons également montré que le problème d'optimisation visant à trouver la politique optimale d'un PDME dont la fonction d'observation est déterministe n'est pas dans APX. Cela signifie qu'il n'existe pas d'algorithme polynomial retournant une solution approchée d'un PDME avec une perte d'optimalité bornée.

Annexe D

Organigramme du laboratoire

Unité de Biométrie et Intelligence Artificielle

Direction : **R. MARTIN-CLOUAIRE** (DR)

Assistante de Direction : **P. Fauré** (TR)

14 chercheurs dont

1 mis à disposition

17 ITA dont

1 mis à disposition

1 en disponibilité

1 CNRS

4 Doctorants - 1 post-doc - 5 CDD

Secrétariat-Documentation

Génopôle

M. Falières (IE)

P. Fauré (TR)

J. Feau (TR)

M.T. Filippi (TR) 50%

P. Delhotal (CR) 60%

Informatique

A. Boukhari (AI)

M. Grialou (TR)

M. Bouchez (AI)

Modélisation des Agro-écosystèmes et

Décision

P. Chabrier (IE)

I. Chadés (CR)

R. Faivre (DR)

F. Garcia (DR) 90%

R. Martin-Clouaire (DR)

N. Peyrard (CR)

G. Quesnel (CR)

H. Raynal (IR)

JP. Rellier (IR)

R. Sabbadin (CR)

Mise à disposition

F. Guerrin (IR)

M. Goulard (CR)

Post-doc : R. Trepos

Doct : N. Forsell, N. Ravoanjanahary

Statistique et Informatique appliquées à la génétique et à la biologie moléculaire

D. Allouche (IR) 80%

C. Cierco (CR) 70%

M.J. Cros (IE)

C. Gaspin (DR)

S. de Givry (CR)

S. Jasson (IR)

B. Mangin (DR) 80%

A. Moisan (AI)

T. Schiex (DR)

Plates formes Bioinformatiques

GENOPOLE – SIGENAE

C. KLOPP (IR)

J.M. Larré (IE) – D. Laborie (IE) – C. Noirot (IE)

CDD : P. Casel, J. Dupiot, A. Pollidoro

Doct : A. Kozomara, E. Audemard

CDD : B. Escalière, J. Mariette

Autres doctorants encadrés
par un scientifique UBIA

• Rachelson Emmanuel

• Martin Guillaume

• Monnet Yannick

Annexe E

Réseau de collaborations

Le réseau de relations entre les différentes personnes avec qui j'ai collaboré (articles ou projets en cours) est représenté page suivante. Les personnes sont regroupés suivant le thème principal de nos collaborations, ou à proximité de plusieurs thèmes quand nos collaborations ont pris place dans plusieurs thèmes. Les arêtes représentent des co-publications ou des co-participations à des projets en cours.

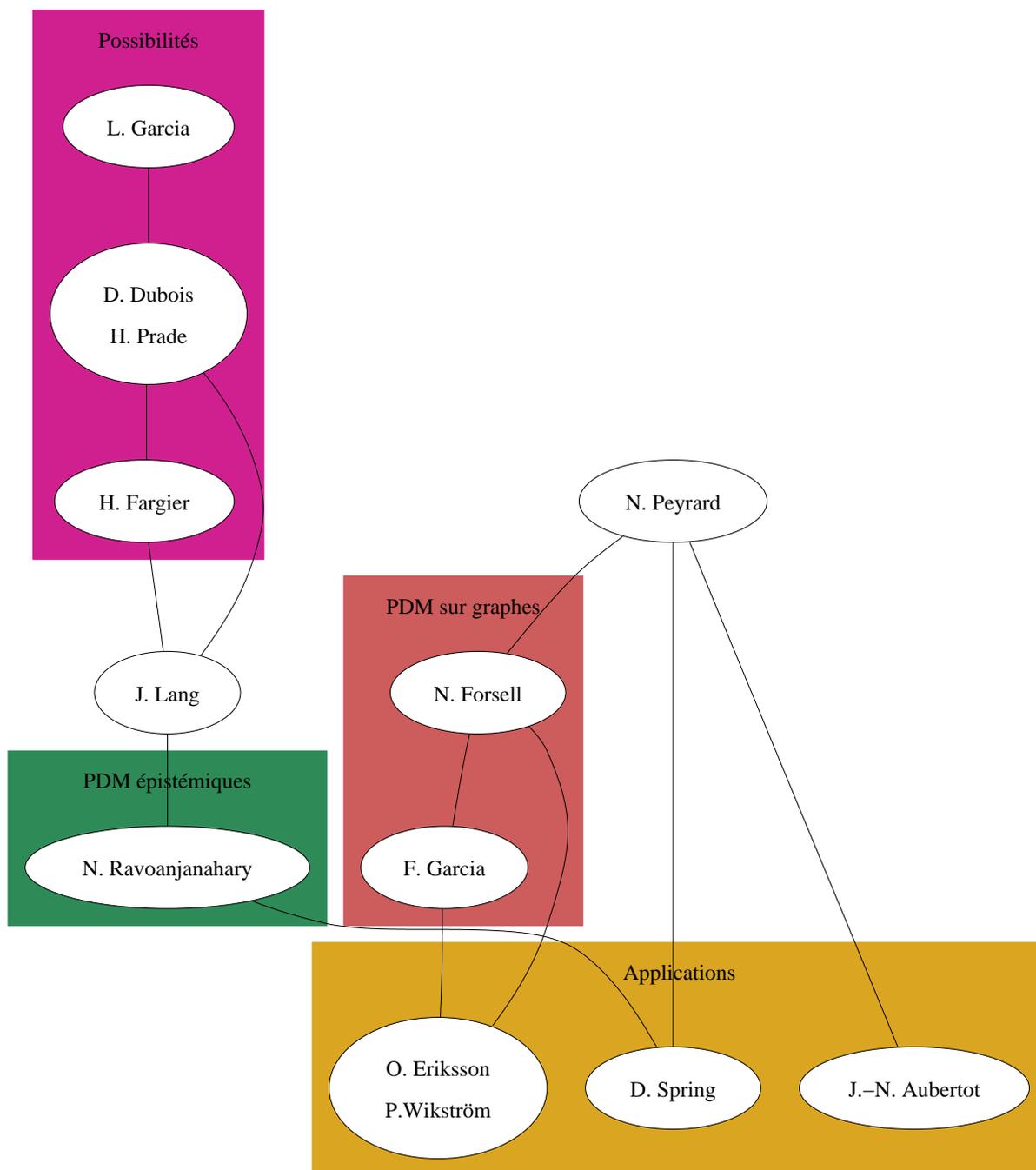


FIG. E.1 – Réseau de relations entre les différentes personnes avec qui j’ai collaboré (articles ou projets en cours).

Annexe F

Animation scientifique

F.1 Jurys et comités de thèse, concours de recrutement

F.1.1 Jury et comité de thèse

- Florent Teichtel, thèse ONERA-CERT, décembre 2005, membre du jury.
- Cédric Pralet, thèse ONERA-CERT, 2004-2007, membre du comité de thèse.

F.1.2 Concours de recrutement

- Concours de recrutement de MCF, ENSAR, 2001 (section 27).
- Concours de recrutement de CR2, INRA-MIA, 2004.
- Concours de recrutement de MCF, ENSAR, 2008 (section 27).

F.2 Comités de programme de conférences, relecture d'articles

F.2.1 Membre de comités de programme et d'organisation de conférences

- *Rencontre des Jeunes Chercheurs en Intelligence Artificielle* (RJCIA), membre du comité de programme en 2002,
- *International Conference on Modeling and Simulation* (MODSIM), coorganisateur d'une session *management of natural resources and biodiversity*, en 2007.

F.2.2 Relecteur pour conférences

- *International Joint Conference on Artificial Intelligence* (IJCAI), 2001, 2003, 2005 et 2007,
- *European Conference on Artificial Intelligence*, (ECAI), 1998, 2000, 2002, 2006, 2008,
- etc.

F.2.3 Reviewer pour revues

- *Fuzzy Sets and Systems*, 2004,
- *Revue d'Intelligence Artificielle*, 2006, 2007,
- *European Journal of Operations Research*, 2006,
- *Ecological Modeling*, 2007,

- *Artificial Intelligence*, 2008,
- etc.

F.3 Expertise

- Programme FAST *France-Australia Science and Technology*, 2002,
- Projets soutenus par les *Haras nationaux*, 2002.

F.4 Groupes de travail

- Participation au groupe de travail *Processus décisionnels de Markov et Intelligence Artificielle* (PD-MIA), depuis sa création (2001),
- co-animateur du groupe de travail *Modèles Spatio-Temporels sur Graphes et Applications* (MSTGA), depuis sa création (2006).

F.5 Participation à la vie scientifique de l'INRA

- Membre du conseil scientifique du département MIA (précédemment, BIA), 2002-2006,
- membre du groupe de travail prospectif, INRA-2020.

Annexe G

Projets de recherche

G.1 Projet FOFRAMIP

- **Intitulé** : *Les forêts fragmentées dans les territoires et les exploitations agricoles de Midi-Pyrénées.*
- **Période** : 2002-2004.
- **Responsables** : Gérard Balent, DR INRA, André Gavaland, IR INRA et Marc Hewison, DR INRA.
- **Financement** : Région Midi-Pyrénées, programme CCRDT,

G.2 Projet CEDRE

- **Intitulé** : *Exploitation durable de résistances aux maladies chez les végétaux.*
- **Période** : 2005-2008.
- **Responsable** : Lydia Bousset, CR1 INRA (responsable du projet), Jean-Noël Aubertot (responsable WP2).
- **Financement** : ANR, programme *Agriculture et Développement Durable.*

G.3 Projet Australian Research Council (ARC)

- **Intitulé** : *Applying search theory for eradicating invasive species.*
- **Période** : 2007-2009.
- **Responsables** : Ralph Mac Nally, Prof., Danny Spring, Dr. Monash University (Australie), Barry Brook, Prof. Charles Darwin University (Australie), Brian Leung, Dr. Mac Gill University (Canada) et Régis Sabbadin, CR INRA.
- **Financement** : Australian Research Council, programme *Discovery Projects.*

G.4 Projet ComEvol

- **Intitulé** : *Dynamique, Evolution et Contrôle des Maladies Infectieuses sur Réseaux d'Interactions*
- **Période** : 2007-2008.
- **Membres du projet** : Marc Choisy, IR IRD, Alain Franc, DR INRA, Nathalie Peyrard, CR INRA et Régis Sabbadin, CR INRA.
- **Financement** : GDR ComEvol, axe *Modèles Systèmes Complexes.*

Annexe H

Activités d'encadrement

H.1 Encadrement de thèses

- **Nicklas Forsell**. 2005-2009. INRA-MIA, Toulouse et SLU, Umea, Suède. *MDP Approaches for Complex Forest Planning Under Uncertainty*. Encadrement : L. O. Eriksson (SLU, Umea, *main supervisor*) 25%, F. Garcia (INRA-MIA, Toulouse) 25%, R. Sabbadin (INRA-MIA, Toulouse) 25%, P. Wikström (SLU, Umea), 25%.
- **Nasolo Ravoanjanahary**. 2005-2008. INRA-MIA, Toulouse et Univertisté Toulouse III. *Processus décisionnels épistémiques et élicitation interactive de préférences*. Encadrement : J. Lang (IRIT-UPS-CNRS, *directeur de thèse*) 50%, R. Sabbadin (INRA-MIA, Toulouse) 50%.

H.2 Encadrement de Master 2 (DEA)

- M. Mouterfi (DEA Systèmes Industriels, Ecole Doctorale Systèmes). Processus Décisionnels de Markov et Graphes : application à un problème de gestion environnementale. Mars-Juin 2004.
- N. Ravoanjanahary (DEA Intelligence Artificielle, Ecole Doctorale Informatique, Université Paul Sabatier). Planification Purement Epistémique. Mai-Août 2004.
- L. Conde (Master 2 recherche Probabilités et Statistiques, Université Paul Sabatier). Extension du cadre des PDMG pour l'optimisation de la structure de réseaux d'interactions dans les processus spatio-temporels. Mars-Juin 2007.
- B. Borgy (Master 2 Biostatistiques et Modélisation, Université Paul Sabatier). Exploitation du cadre des PDMG pour l'aide à la conception de stratégies de gestion collectives de résistances variétales au phoma du colza. Février-Juillet 2008.

H.3 Encadrements de stages

- E. Huard (INSA 4ème année). Approximation d'un processus stochastique multidimensionnel : application à la propagation du feu. Juillet-Août 2002.
- C. Paccard (INSA 4ème année). Modélisation de l'évolution d'une population de rouges-gorges dans la vallée de la Nère. Juin-Septembre 2003.
- P.-M. Pair (INSA 4ème année). Modélisation du problème de course automobile dans le cadre des PDM. Implémentation d'algorithmes de résolution. Juillet-Août 2003.
- C.-E. Rabier (5ème année INSA). Processus Décisionnels de Markov : application à la gestion de réserves naturelles. Mars-Juin 2005.

- D. Arnal (4ème année INSA). Processus Décisionnels de Markov Possibilistes. Juillet-Août 2005.
- E.-A. Guerch (Master 1, Ingénierie Mathématique). Gestion durable des résistances variétales à l'aide des Processus Décisionnels de Markov sur Graphes. Mai-Août 2006.
- E. Bergonnier (4ème année INSA). Modélisation d'un problème de gestion de réserves naturelles au Costa-Rica. Juillet-Août 2006.
- M. Laviro (2ème année SupAéro). Gestion durable de résistances variétales à l'aide des Processus Décisionnels de Markov sur Graphes. Juillet-Août 2007.
- M. A. Rafik (Master 1 Informatique et Gestion, Université Paul Sabatier). Développement d'un simulateur de dynamiques épidémiques sur réseau. Mars-Août 2008.

H.4 Encadrement de CDD

- N. Ravoanjanahary. Développement d'un démonstrateur informatique d'algorithmes d'intelligence Artificielle basés sur les Processus Décisionnels de Markov. Application à un problème de diagnostic d'invasion des cultures par des nuisibles. CDD INRA-Monash University (2008).

Annexe I

Activités d'enseignement

Ecole Supérieure de Commerce de Rennes

Cours et Travaux Dirigés en *Anglais* (vocation internationale de l'école).

- **Sujet** : Recherche Opérationnelle : *Théorie de la décision, Algorithmique des graphes, Programmation Linéaire...*
- **Public** : étudiants de première et deuxième années d'école de commerce.
- **Volume horaire** : environ 40 heures par an entre 2002 et 2007.

Master Biostatistique et Modélisation (Université Toulouse III)

Cours et Travaux Dirigés.

- **Sujet** : *Théorie des Possibilités, Logique possibiliste, Processus Décisionnels de Markov.*
- **Public** : étudiants de Master 2.
- **Volume horaire** : 6 heures en 2005-2006, 9h en 2006-2007, 15h en 2007-2008.

Master Agrofood Chain (UT III - Agricampus)

Cours et Travaux Dirigés en *Anglais* (étudiants majoritairement étrangers).

- **Sujet** : Recherche Opérationnelle : *Algorithmique des graphes, Production planning, Scheduling.*
- **Public** : étudiants de Master 1.
- **Volume horaire** : 6 heures en 2005-2006, 4h en 2006-2007 et 4h en 2007-2008.

NOVA PhD course, SLU-University, Umea, Suède

Cours, Travaux Dirigés et Travaux Pratiques en *Anglais* (cours organisés par le programme NOVA, réseau *Forest planning and inventory*).

- **Sujet** : *Markov Decision Processes : Methods and Applications.*
- **Public** : étudiants doctorants dans des laboratoires scandinaves (Suède, Finlande, Danemark. . .).
- **Volume horaire** : 20 heures en 2008.

Annexe J

Sélection de publications

