HAL open science

Définition et exploration des propriétés formelles des logiciels auto-organisateurs à fonctionnalité émergente.

Simon Stuker

To cite this version:

Simon Stuker. Définition et exploration des propriétés formelles des logiciels auto-organisateurs à fonctionnalité émergente.. Mathématiques [math]. Université Toulouse 3 - Paul Sabatier, 2014. Français. NNT : . tel-01117928

HAL Id: tel-01117928

https://hal.science/tel-01117928v1

Submitted on 18 Feb 2015

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

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

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le *17/12/2014* **par :**

Simon STUKER

Définition et exploration des propriétés formelles des logiciels auto-organisateurs à fonctionnalité émergente.

JURY

Xavier BRESSAUD	Professeur d'Université	Président du jury
Jean-Marc COUVEIGNES	Professeur d'Université	Directeur
Françoise ADREIT	Maître de conférences	Encadrant
Bruno GAUJAL	Directeur de Recherche	Rapporteur
Pierre DE LOOR	Professeur d'Université	Rapporteur
Nazim FATES	Chargé de Recherche	Examinateur

École doctorale et spécialité :
MITT : Domaine Mathématiques : Mathématiques appliquées
Unité de Recherche :
Institut de Recherche en Informatique de Toulouse (UMR 5505)
Directeur(s) de Thèse :
Jean-Marc COUVEIGNES et *Françoise ADREIT*
Rapporteurs :
Bruno GAUJAL et *Pierre DE LOOR*

Simon Stuker

DÉFINITION ET EXPLORATION DES PROPRIÉTÉS FORMELLES DES LOGICIELS AUTO-ORGANISATEURS À FONCTIONNALITÉ ÉMERGENTE.

Directeurs de thèse :
Jean-Marc Couveignes, Professeur
Université de Bordeaux
Françoise Adreit, Maître de conférences
Université Paul Sabatier

Résumé

Dans cette thèse, nous nous intéressons à l'étude formelle des systèmes multi-agents à fonctionnalité émergente. Ces systèmes ont la particularité d'être formés d'un grand nombre d'entités, appelées agents, dotés d'objectifs individuels et disposant généralement de capacités de perception, de raisonnement et d'action limitées. Le fonctionnement global du système émerge de leurs interactions.

Dans de nombreuses applications ces systèmes ont montré des propriétés globales intéressantes, comme la convergence rapide vers un régime intéressant ou la stabilité autour de ce régime. L'objectif de cette thèse est d'utiliser des méthodes mathématiques pour démontrer et explorer ces propriétés de manière formelle.

Une étape importante dans cette démarche est la recherche des méthodes mathématiques les plus adéquates pour étudier les systèmes multi-agents. Les qualités et limites de divers modèles pertinents sont analysées, et aboutissent à l'étude approfondie des processus décisionnels Markoviens et leurs variantes multi-agents, d'une part, et des méthodes à champ moyen d'autre part.

Ensuite, nous nous intéressons à des systèmes localisés et dérivons rigoureusement un modèle continu. À l'aide d'outils d'analyse mathématique nous établissons des propriétés dynamiques, à savoir la convergence vers un équilibre intéressant et la stabilité de cet équilibre. Ce raisonnement est ensuite prolongé à un système localisé bien connu dans le domaine des systèmes multi-agents, la collecte de ressources, et permet d'illustrer un phénomène intéressant à l'aide de simulations numériques.

Mots clés : Systèmes multi-agents, Systèmes complexes, Modélisation formelle de systèmes multi-agents, Vérification et validation des systèmes multi-agents, Simulation multi-agents.

Institut de Recherche en Informatique de Toulouse - UMR 5505
Université Paul Sabatier, 118 route de Narbonne, 31062 TOULOUSE cedex 4

Je dédie cette thèse
à Farah-Anaïs, mon épouse.

JE tiens à remercier...

... Jean-Marc COUVEIGNES, Professeur à l'Université de Bordeaux, mon directeur de thèse. Son soutien et sa disponibilité au cours de ces trois années (et demie) ont été sans faille. Nos discussions m'ont toujours donné beaucoup de motivation pour avancer avec des idées fraîches et intéressantes, et je l'en remercie chaleureusement.

... Françoise ADREIT, Maître de conférences à l'Université de Toulouse, ma co-encadrante de thèse. Elle a toujours été très à l'écoute (surtout aux moments de doute), et s'est beaucoup investie pour le bon déroulement de cette thèse. Nos échanges scientifiques m'ont énormément aidé pour fixer mes idées et pour guider mon travail, mais m'ont aussi beaucoup appris. Pour ces raisons je lui adresse toute ma reconnaissance.

... Bruno GAUJAL, Directeur de Recherche à l'Inria. Tout d'abord, pour m'avoir fait l'honneur de rapporter mon manuscrit de thèse. Ensuite, pour son accueil chaleureux à Grenobles. Nos échanges ont été très constructifs pour moi, et m'ont permis de comprendre de plus près les liens entre ma thèse et l'optimisation en champ moyen.

... Pierre DE LOOR, professeur à l'École Nationale d'Ingénieurs de Brest, pour avoir accepté de rapporter mon manuscrit, malgré la difficulté due à la différence de nos disciplines. Ses commentaires et suggestions m'ont permis d'avoir un angle de vue nouveau sur mon travail, et de renforcer son aspect pluridisciplinaire.

... Nazim FATÈS, Chargé de recherche à l'Inria. Comme en témoigne sa lecture très attentive du manuscrit, mais aussi ses diverses remarques et précisions, Nazim a montré un grand intérêt pour mon travail. Chacune de ses visites à Toulouse m'a permis d'éclairer de nouvelles questions, et je lui en suis reconnaissant.

... Pierre GLIZE, Ingénieur CNRS Habilité, et Marie-Pierre GLEIZES, Professeur à l'Université de Toulouse. Tout d'abord pour l'accueil chaleureux qu'ils m'ont réservé au sein de leur équipe, mais aussi pour m'avoir transmis leur passion pour les systèmes multi-agents. Avec leur bonne humeur et leur dynamisme, ils sont largement responsables de l'excellente ambiance qui règne dans l'équipe SMAC.

... Julien MARTIN et Raja BOUJBEL, doctorants dans l'équipe SMAC, tout les deux sur le point de soutenir. Partager mon bureau avec eux a été un réel plaisir. J'ai pu partager de nombreux centres d'intérêt avec Julien, et Raja a été d'une aide précieuse (indispensable) au cours de la rédaction de cette thèse. Je la remercie de m'avoir fait profiter - de façon répétée - de sa grande maîtrise de LaTeX.

... les autres membres, permanents ou non, de l'équipe SMAC. Je garde un souvenir très agréable de mon doctorat. Malgré la différence entre nos disciplines, les SMACkers ont montré beaucoup d'intérêt pour mon travail.

Je garderai également un souvenir fort des discussions passionnantes, sur des sujets divers et variés, qui ont lieu quotidiennement au cours des pauses café. Je retiendrai aussi les nombreuses soirées, dont certaines consacrées aux jeux de société avec mes collègues passionnés.

... les divers mathématiciens de l'Institut de Mathématiques de Toulouse que j'ai ren-

contré au cours de ma thèse. Parmi eux, je remercie particulièrement Xavier BRESSAUD, Professeur à l'Université de Toulouse. Ses enseignements, les discussions que nous avons eues, mais aussi les rencontres scientifiques que j'ai faites grâce à Xavier ont contribué au résultat final de ma thèse.

J'adresse également des remerciements particuliers à Philippe LAURENÇOT, Directeur de Recherche à l'Université de Toulouse. Il m'a apporté une grande aide sur une question précise, alors qu'il ignorait totalement qui j'étais et quel était mon but. Je le remercie de m'avoir montré une telle générosité, et une telle passion pour son travail.

Avant-Propos

L'objectif de cette thèse est de mobiliser des méthodes mathématiques pour démontrer des propriétés formelles d'objets informatiques, et vise donc deux publics provenant de disciplines différentes.

Dans ce manuscrit, nous essayons d'adopter un langage accessible aux deux publics. Les résultats mathématiques sont donc accompagnés d'interprétations concrètes, afin de permettre au lecteur sans formation mathématique approfondie de les comprendre facilement. La lecture de ce manuscrit peut s'effectuer à différentes vitesses selon la familiarité du lecteur avec les notions mathématiques présentées.

Table des matières

Introduction **1**

I Contexte et positionnement scientifique **5**

1 Les systèmes multi-agents à fonctionnalité émergente **7**

1.1 Caractérisation d'un système multi-agent 7

1.1.1 Les agents 8

1.1.2 L'environnement 8

1.1.3 Les interactions des agents 9

1.2 La complexité dans les systèmes multi-agents 9

1.2.1 Les systèmes complexes 9

1.2.2 Propriétés émergentes 11

1.2.3 Les systèmes multi-agents à fonctionnalité émergente 12

1.3 Quelques applications 12

1.3.1 Le génie logiciel 12

1.3.2 La simulation à base d'agents 13

1.3.3 La résolution collective de problèmes 13

1.4 Conclusion et positionnement de la thèse 14

2 Modélisation mathématique des systèmes multi-agents **17**

2.1 Quelques réflexions sur la modélisation 18

2.1.1 Deux points de vue d'une modélisation : la description Lagrangienne et la description Eulérienne 18

2.1.2 L'aléatoire dans la modélisation des systèmes multi-agents 21

2.1.3 Distinction entre les actions des agents et le résultat global : le modèle influence-réaction 23

2.2 Tour d'horizon de modèles envisageables 24

2.2.1 Les modèles descriptifs 24
2.2.1.1 Les systèmes dynamiques 24
2.2.1.2 Les systèmes de particules en interaction 26
2.2.1.3 Les automates cellulaires 27
2.2.1.4 Les méthodes de champ moyen 28
2.2.1.5 Équations à réaction-diffusion 29
2.2.2 Les modèles prescriptifs 29
2.2.2.1 La commande optimale 30
2.2.2.2 La théorie des jeux 31
2.3 Conclusion 33

II Étude approfondie de formalismes possibles pour représenter les systèmes multi-agents 35

3 Les Processus Décisionnels Markoviens 37
3.1 Introduction 37
3.2 Les Processus Décisionnels Markoviens 37
3.2.1 Formalisme 37
3.2.2 Règle de décision, stratégie et gain 39
3.2.3 L'objectif : recherche d'une stratégie optimale 40
3.3 Un exemple simple résolu «à la main» 40
3.4 Méthodes de résolution classiques d'un MDP 46
3.4.1 Backward Induction 46
3.4.2 Value Iteration 50
3.4.3 Policy Iteration 52
3.4.4 Programmation linéaire 54
3.4.5 Bilan 55
3.5 Les Processus Décisionnels Markoviens Multi-Agents 55
3.5.1 Les MMDP 56
3.5.2 Discussion 57
3.5.2.1 À propos de la récompense globale 57
3.5.2.2 À propos de la stratégie optimale 57
3.5.2.3 À propos de la connaissance du système 57
3.5.2.4 Limites mathématiques 58
3.5.3 MMDP et jeux de Markov 59

3.5.4 Les variantes partiellement observées : DEC-POMDP et DEC-MDP . . 60
3.5.4.1 Définition formelle 60
3.5.4.2 Complexités algorithmiques 61
3.5.4.3 Discussion 62
3.5.5 Autres variantes 62
3.5.5.1 GMDP 63
3.5.5.2 GO-DEC-MDP 63
3.5.5.3 COM-TMDP 64
3.6 Retour à l'exemple des couloirs : extension au cas partiellement observé . . . 65
3.6.1 Définition du DEC-POMDP 65
3.6.2 Stratégies et problème d'optimisation 65
3.6.3 Calcul d'une stratégie optimale 66
3.6.4 Quelques stratégies empiriques basées sur des connaissances partielles 67
3.6.4.1 Estimation moyenne du nombre de personnes restantes . . . 67
3.6.4.2 Estimation de l'état global par maximum de vraisemblance . 69
3.6.4.3 Une stratégie empirique adaptative 70
3.6.5 Conclusion 72
3.7 Résolution d'un MDP Multi-Agents 73
3.7.1 Une méthode de programmation linéaire dans le cas factorisé 73
3.7.1.1 Hypothèses 73
3.7.1.2 Présentation de la méthode 73
3.7.1.3 Discussion 75
3.7.2 Une méthode de distribution des équations d'optimalité 75
3.7.2.1 Hypothèses 75
3.7.2.2 Deux exemples d'application 77
3.7.2.3 Présentation de la méthode 78
3.7.2.4 Discussion 79
3.8 Conclusion 80

4 Méthodes de champ moyen 81
4.1 Introduction 81
4.1.1 Présentation de l'approche 81
4.1.2 Systèmes de particules en interaction moyenne 81
4.1.3 Approximation par un champ déterministe 82
4.2 Modèle stochastique avec interactions à travers la densité 83

4.2.1 Formalisme 83
4.2.2 Bilan 87
4.3 Mise en oeuvre sur des exemples 87
4.3.1 Problème de trafic dans un ensemble de couloirs 88
4.3.2 Situation issue de la robotique 90
4.4 La limite champ moyen 92
4.4.1 Le théorème de convergence vers le champ moyen 92
4.4.2 Évolution d'un individu isolé 94
4.4.3 Convergence du champ moyen 94
4.5 Méthodes de champ moyen pour la résolution d'un MDP 96
4.5.1 Modèle stochastique contrôlé 96
4.5.2 Limite champ moyen 98
4.5.3 Application aux MDP sur un horizon fini 98
4.5.4 Application au cas dévalué 99
4.5.5 Commentaire 100
4.6 Retour sur les exemples 100
4.6.1 Trafic unilatéral dans deux couloirs 101
4.6.1.1 Notations et scaling 101
4.6.1.2 Récompense et gain 102
4.6.1.3 Optimisation du champ moyen 103
4.6.1.4 Analyse numérique et comparaison de la performance 105
4.6.2 Attitude optimale pour les robots extracteurs de bâtons 108
4.6.2.1 Notations, récompense et gain 108
4.6.2.2 Optimisation du champ moyen 109
4.6.2.3 Analyse et comparaison de la performance 112
4.7 Conclusion 115

III Étude de systèmes multi-agents situés réactifs à l'aide d'équations à réaction-diffusion-advection 117

5 Utilisation de l'équation de Fokker-Planck pour la paramétrisation d'un système multi-agent situé réactif 119

5.1 Position du problème et modèle utilisé 119
5.1.1 Mouvement collectif des agents 120
5.1.2 Fonction de satisfaction 122

5.2 Approximation du modèle discret 123
5.2.1 Limite champ moyen 124
5.2.2 Limite spatiotemporelle 125
5.2.3 Commentaires 126
5.2.3.1 Le flux macroscopique 126
5.2.3.2 Les conditions au bord 127
5.2.3.3 À propos des raisonnements asymptotiques 128
5.2.4 Formulation continue du problème étudié 129
5.3 Propriétés de l'équation de Fokker-Planck linéaire 130
5.3.1 Régime stationnaire de l'équation d'évolution 130
5.3.2 L'équation de Fokker-Planck linéaire est bien posée 132
5.3.3 Convergence vers le régime stationnaire 135
5.3.3.1 Un argument d'analyse linéaire 136
5.3.3.2 L'entropie quadratique 137
5.3.3.3 L'entropie relative de Kullback 139
5.4 Stratégies optimales 142
5.4.1 Une stratégie dépendant localement de la répartition ciblée 143
5.4.2 Une stratégie basée sur la satisfaction locale 144
5.4.2.1 Premières propriétés de l'équation d'évolution 145
5.4.2.2 Régime stationnaire et convergence des solutions 147
5.5 Validation numérique 149
5.5.1 Validation du modèle 149
5.5.1.1 Le système discret 150
5.5.1.2 Le champ moyen 151
5.5.1.3 Le système continu 151
5.5.1.4 Analyse des résultats 152
5.5.2 Validation des stratégies optimales 152
5.6 Étude numérique dans le cas bidimensionnel 155
5.6.1 Approximation continue 156
5.6.2 Étude numérique 157
5.6.2.1 Paramètres des simulations 157
5.6.2.2 Stratégie dépendant localement de la répartition ciblée 158
5.6.2.3 Stratégie dépendant de satisfaction locale 159
5.6.2.4 Analyse de la performance 160
5.7 Conclusion 161

6 Équations à réaction - diffusion - advection pour étudier une collecte de ressources 165

6.1 Introduction et position du problème 165

6.2 Modèle formel 166

6.3 Dérivation de l'équation à réaction-diffusion-advection 169

6.3.1 Limite champ moyen 170

6.3.2 Loi d'évolution macroscopique continue 171

6.3.3 Conditions au bord et condition initiale 172

6.4 À propos des équations à réaction-diffusion-advection 173

6.5 Étude numérique 174

6.5.1 Paramètres des simulations 174

6.5.2 Simulations numériques dans le cas unidimensionnel 177

6.6 Simulations numériques dans le cas bidimensionnel 180

6.6.1 Le système d'équations à réaction-diffusion-advection 180

6.6.2 Les simulations numériques 183

6.7 Conclusion 187

Conclusion 189

IV Annexe 195

A Recherche de structure dans les graphes pour la certification d'un circuit électrique avec le plan spécifié. 197

A.1 Présentation du problème 197

A.2 Architecture d'un fichier de spécification 199

A.3 Lecture d'un fichier de spécification 201

A.4 Analyse du fichier de spécification 202

A.5 Conclusion 210

Bibliographie 211

Introduction

Les logiciels à fonctionnalité émergente

Le développement de l'outil informatique au cours de ces dernières décennies a apporté de nouvelles problématiques au domaine de l'informatique. Les applications actuelles exigent des systèmes informatiques de plus en plus sophistiqués, difficiles à concevoir et à maintenir.

Les progrès technologiques ont permis de repousser les limites des processeurs, en miniaturisant les composants de manière importante. Cette évolution a permis d'augmenter les capacités calculatoires des systèmes informatiques, mais apporte de nouvelles difficultés au cours du processus de production.

Parallèlement, les appareils informatiques sont de plus en plus interconnectés avec l'évolution de l'internet et des réseaux mobiles. De ce fait, les logiciels qu'ils contiennent sont confrontés à un environnement virtuel dans lequel ils doivent trouver leur place. L'ouverture et l'adaptation à un environnement dynamique sont donc des exigences nouvelles et importantes pour un logiciel.

Ces constats ont contribué à une rupture conceptuelle de l'informatique. La vision monolithique des logiciels a progressivement été abandonnée au profit de systèmes virtuels formés de composants autonomes, qui opèrent en interaction les uns avec les autres. Ce point de vue permet de rapprocher le domaine du logiciel de l'intelligence artificielle distribuée, en considérant un logiciel comme un ensemble d'entités autonomes en interaction. Ce paradigme a trouvé sa place en informatique depuis la fin des années 80 avec les systèmes multi-agents.

Dans cette thèse, nous nous intéressons plus précisément aux systèmes informatiques *complexes*. Ces systèmes sont formés d'un *grand nombre d'entités en interaction*, dont le comportement collectif *émerge* des interactions. La fonctionnalité du système n'est pas prédéfinie par le concepteur ; il s'agit d'un système à fonctionnalité émergente. Cette approche offre un intérêt particulier pour la conception de logiciels de complexité importante.

La conception de tels systèmes est donc un enjeu important en informatique. Pour ce faire, la démarche scientifique classique, qui s'efforce de représenter le système dans sa totalité, n'est pas pertinente. Le nombre important d'entités et la haute non-linéarité de leurs interactions conduisent à des difficultés importantes pour prévoir ou expliquer le comportement du système.

Une démarche intéressante pour concevoir ces systèmes consiste à s'inspirer de certains systèmes biologiques, comme les insectes sociaux. Un exemple typique de cette inspiration est l'algorithme d'optimisation de type colonie de fourmis, qui sera expliqué plus en détail dans la section 1.3.3.

Une autre démarche consiste à faire appel à des *méta-heuristiques*. Il s'agit d'une méthode qui est capable de produire des solutions heuristiques à un ensemble de problèmes, et non un problème spécifique. Il faut souligner le caractère *heuristique* de ces solutions : elles ne sont pas fondées sur un modèle formel, et ne sont pas nécessairement optimales.

Ces deux démarches sont guidées par un ensemble de principes assez génériques. Elles ont généralement le défaut d'être relativement peu formalisées, et leur validité n'est pas toujours garantie par des arguments formels.

Par conséquent, l'adéquation d'un système informatique complexe à fonctionnalité émergente au problème pour lequel il a été conçu est généralement vérifiée de manière expérimentale. Une telle validation est clairement limitée en termes de garantie des résultats, et ne fournit que des réponses d'ordre statistique. De plus, cette méthode de validation s'effectue toujours dans un cadre expérimental complètement spécifié. Par conséquent, elle ne peut pas prendre en compte d'éventuelles perturbations imprévisibles de l'environnement, et garantir la robustesse du système en toutes circonstances.

Objectif de la thèse

Dans l'optique de dépasser le limites de la validation expérimentale, nous nous intéressons dans cette thèse aux outils mathématiques qui pourraient aider à étudier les systèmes informatiques complexes à fonctionnalité émergente. La validation formelle de l'adéquation de ces systèmes à leur tâche est l'objectif de cette thèse.

Pour ce faire, nous nous sommes limités à une classe de systèmes bien définie, et nous nous sommes intéressés à des propriétés dynamiques telles que la convergence, la stabilité de l'équilibre, mais aussi à l'optimisation d'une certaine fonction de satisfaction qui mesure la réussite du système.

Une étape importante de ce travail a consisté à explorer différentes modélisations possibles, et à évaluer leur pertinence pour décrire la classe de systèmes étudiés. Ensuite nous avons utilisé les modèles les plus pertinents pour étudier des systèmes spécifiques, et démontré les propriétés intéressantes de ces systèmes de manière formelle.

Organisation du manuscrit et contributions

Cette thèse est composée de six chapitres et d'une annexe :

Chapitre 1 Le premier chapitre présente le contexte de cette thèse, les *systèmes multi-agents complexes à fonctionnalité émergente*. Après avoir présenté ces systèmes, nous exposons quelques domaines d'application variés. Nous terminons par la définition du cadre de cette

thèse, qui précise la classe de systèmes étudiés.

Chapitre 2 Dans le deuxième chapitre, nous nous intéressons à la modélisation des systèmes étudiés. Après quelques considérations générales sur la modélisation, nous présentons rapidement un ensemble de modèles que nous avons étudiés. Ces modèles sont divisés en deux parties : ceux qui sont purement descriptifs, et ceux qui intègrent la finalité du système dans la description. Nous évaluons la pertinence des modèles présentés vis-à-vis de la classe des systèmes étudiés.

Chapitre 3 Les *Processus Décisionnels de Markov* sont une classe de modèles pertinents de plusieurs points de vue. Pour cette raison, nous étudions leur utilité pour l'étude des systèmes multi-agents à fonctionnalité émergente de manière approfondie dans le chapitre 3. Nous illustrons leur applicabilité, mais aussi leurs limites sur divers exemples au fur et à mesure du chapitre.

Chapitre 4 Le quatrième chapitre est consacré à l'utilisation de *méthodes à champ moyen* pour l'étude formelle des systèmes multi-agents. Ces méthodes sont utilisées en physique statistique pour simplifier la description de grands systèmes de particules en interaction. À travers des exemples, et en faisant appel à des résultats récents, nous illustrons les apports et les limites de cette approche, dans le contexte des systèmes étudiés.

Chapitre 5 Le cinquième chapitre est consacré à un problème spécifique, qui est d'orienter un système d'agents localisés vers une distribution ciblée. À l'aide d'une approximation continue, qui fait appel à des résultats du chapitre 4, nous aboutissons à une équation aux dérivées partielles appelée *équation de Fokker-Planck*. Cette équation réalise un lien intéressant avec plusieurs domaines tels que la thermodynamique ou la mécanique statistique.

En étudiant l'équation de Fokker-Planck à l'aide d'outils d'analyse continue, nous déterminons deux stratégies basées sur des perceptions locales qui permettent aux agents d'atteindre collectivement la distribution souhaitée. Les résultats théoriques énoncés sont confirmés par des simulations numériques.

Chapitre 6 Le sixième chapitre est également tourné vers l'application. En prolongeant le raisonnement du chapitre 5, nous étudions un système d'agents localisés dont l'objectif est de collecter une ressource.

À l'aide d'une approximation continue similaire à celle du chapitre 5, nous aboutissons à un système d'équations à réaction-diffusion-advection. Ces équations suscitent un intérêt mathématique croissant depuis quelques années, mais leur étude formelle reste délicate. Par conséquent, nous proposons des simulations numériques afin d'étudier la dynamique des solutions. Les résultats mettent en avant un phénomène émergent bien connu dans le contexte du problème étudié.

Suite au chapitre 6, la conclusion dresse le bilan des principaux résultats de cette thèse, et propose des perspectives scientifiques à plus ou moins long terme.

Annexe : Certification de circuits intégrés Le chapitre annexe à cette thèse contient le résultat du groupe de travail SEME 2012 auquel nous avons participé.

En raison de la miniaturisation des circuits électriques, il est difficile de réaliser une vérification physique de l'adéquation des circuits à l'issue de la production. En nous basant sur un exemple concret, nous proposons des méthodes basées sur les propriétés d'un graphe de connexion sous-jacent, qui permettent de guider le processus de vérification physique.

Davantage d'explications sur la connexion de ce travail avec les systèmes multi-agents seront données dans cette annexe.

Première partie

Contexte et positionnement scientifique

1 Les systèmes multi-agents à fonctionnalité émergente

L'approche par système multi-agent est un paradigme qui place l'individu au centre du problème, et non le collectif. L'idée de cette approche est que des agents simples avec des règles d'interaction bien choisies vont évoluer de manière autonome vers une solution.

Un avantage important de cette approche est que le système multi-agent peut avoir des capacités qui dépassent de loin la simple somme des capacités des individus. Cela impose une rupture avec la science réductionniste classique, qui tend à expliquer tout phénomène à partir de sa décomposition en constituants élémentaires. Cette approche classique est insuffisante pour certains systèmes dont le fonctionnement ne s'explique pas par la connaissance des constituants, et qui présentent des résultats inattendus.

Le but de l'approche multi-agent à fonctionnalité émergente est de mettre ces résultats à profit pour résoudre des problèmes. Le gain en termes de complexité est potentiellement énorme : des agents très simples peuvent résoudre des problèmes difficiles. Utiliser la complexité pour résoudre des problèmes a cependant un sérieux coût : le système conçu a une performance imprévisible, et il est difficile de le contrôler.

Afin de fixer le sujet d'étude de cette thèse, ce premier chapitre contient une présentation des systèmes multi-agents à fonctionnalité émergente. La présentation commence par une définition informelle des systèmes multi-agents (section 1.1). La partie suivante (section 1.2) présente la notion remarquable de *complexité* et de propriétés *émergentes* pour les systèmes multi-agents. Ensuite, nous donnons diverses applications (section 1.3) avant de conclure (section 1.4). Au cours de cette conclusion, nous fixons le cadre de cette thèse en précisant les caractéristiques des systèmes multi-agents étudiés.

1.1 Caractérisation d'un système multi-agent

Les systèmes multi-agents sont caractérisés par trois ingrédients : les agents, l'environnement et les interactions.

Malgré une littérature extensive sur les systèmes multi-agents, aucune définition de ces trois ingrédients ne fait l'unanimité. Dans ce paragraphe, nous donnons une liste de caractéristiques récurrentes, afin de donner une idée générale de ces notions.

1.1.1 Les agents

La première notion à définir est celle d'*agent*, qui est au centre de l'approche multi-agent. Les différentes définitions rencontrées dans la littérature [58, 155, 151, 130] montrent qu'à ce jour il n'y a pas de consensus sur ce qu'est exactement un agent.

Il y a toutefois des points communs à ces définitions. Un *agent* est généralement caractérisé par

- **son autonomie :** les décisions d'un agent ne sont soumises à aucun contrôle extérieur. Il choisit lui-même son action à partir des informations dont il dispose et de sa capacité de raisonnement.
- **son immersion dans un environnement :** l'agent évolue dans un certain *environnement*, qui peut être l'espace physique dans lequel les agents sont positionnés, un graphe social qui représente des interconnexions entre agents, ou encore un objet extérieur sur lequel le système agit,
- **sa connaissance locale et incomplète :** l'agent possède des informations limitées sur lui-même, sur les autres agents et sur l'environnement dans lequel il évolue.
- **sa nature sociale :** chaque agent a la possibilité d'interagir avec d'autres agents. Cette capacité d'interaction peut être limitée à un *voisinage* d'agents défini par un certain graphe (statique ou dynamique),
- **son objectif local :** le comportement de chaque agent est motivé par un certain objectif individuel. Cet objectif peut être statique, ou dynamique et être ajusté en cours d'exécution en fonction de l'expérience de l'agent.

Ces caractéristiques donnent une idée globale de la notion d'agent, et ne sont pas limitées aux systèmes informatiques. Par exemple, tout individu d'une population animale qui vit en société peut être qualifié d'*agent*. Toutes les caractéristiques énoncées ci-dessus sont présentes : autonomie des individus, l'environnement naturel, les connaissances et perceptions limitées sur l'état du monde, la communication et l'objectif local qui est la survie.

1.1.2 L'environnement

La définition du terme agent fait intervenir l'*environnement* des agents. Cet environnement est souvent compris comme l'*espace* dans lequel évoluent les agents [152].

L'environnement peut être un espace métrique (comme l'espace euclidien $\mathbb{R}^2$) dans lequel des agents se déplacent. Dans ce cas le système multi-agent est qualifié de *situé*. Cette situation est typiquement rencontrée en robotique [79] avec des robots en mouvement dans une certaine zone.

Un autre exemple d'environnement est un graphe, dont les noeuds sont les agents et dont les arêtes définissent les voisinages. Un exemple simple est celui des *smart-grid*, réseaux électriques intelligents : les noeuds du graphe (les agents) sont les fournisseurs, utilisateurs et points de stockage d'énergie électrique, tandis que les arêtes sont les connexions électriques entre ces entités.

L'environnement peut aussi être un objet extérieur sur lequel le système agit. Par exemple, l'environnement peut être un moteur à combustion [17]. Les agents du système sont les différents effecteurs et capteurs, et ont pour objectif de calibrer le moteur en mini-

misant la pollution, tout en optimisant le couple moteur.

De manière générale l'environnement peut être *dynamique*, sujet à des modifications dues aux agents, mais aussi à des perturbations extérieures. Dans le cas où l'environnement est dynamique, son évolution peut être déterministe ou aléatoire. Une classification plus complète sur les différentes caractéristiques de l'environnement est donnée dans [151].

1.1.3 Les interactions des agents

Les *interactions* désignent, de manière générale, les influences réciproques qui existent entre les agents. Ces interactions peuvent prendre diverses formes, comme la simple perception d'un autre agent, le communication explicite sous forme de messages ou la communication implicite qui consiste à modifier un environnement commun pour partager des informations. Ce dernier type d'interaction implicite est appelé *stigmergie*, et fera son apparition à plusieurs reprises au cours de cette thèse.

Les effets de ces interactions sur la réalisation des objectifs des agents sont classifiés en plusieurs types. Si les interactions des agents nuisent à l'accomplissement de leurs objectifs personnels, on dit que les agents sont *non-coopératifs*. Si, au contraire, les agents interagissent de manière collaborative de sorte que chaque agent puisse parvenir à son objectif individuel, on dit qu'ils sont en *coopération* [58].

La notion de coopération s'oppose à celle de *compétition*. Dans le domaine des systèmes multi-agents, la compétition désigne le fait que les objectifs des agents sont antinomiques, c'est-à-dire qu'ils ne peuvent pas être réalisés simultanément [58].

1.2 La complexité dans les systèmes multi-agents

Les multiples interactions entre les agents ont pour conséquence de produire des effets non linéaires : le comportement collectif n'est pas une simple somme de comportements individuels.

Si ces non-linéarités sont fortes et si elles sont présentes en grand nombre, il peut être délicat voire impossible de prévoir la dynamique du système par un raisonnement *réductionniste*, c'est-à-dire à partir de la connaissance complète de chaque agent. Dans ce cas le système est dit *complexe*.

Les systèmes complexes constituent un domaine scientifique vaste et multidisciplinaire. Très modestement, nous présentons dans ce paragraphe quelques aspects des systèmes complexes. On pourra trouver dans [77, 148] une réflexion plus approfondie sur le sujet.

1.2.1 Les systèmes complexes

Les systèmes complexes partagent une première particularité avec les systèmes multi-agents : ils n'ont pas de définition formelle qui fasse l'unanimité dans la communauté scientifique. Dans les propositions de définition et exemples cités dans [109, 148] nous retiendrons :

- qu'un système complexe est formé d'**un grand nombre d'entités en interaction**, appelés *agents*, avec des **comportements simples**,
- que les individus sont **autonomes**, **dépourvus de hiérarchie ou de contrôle centralisé**,
- que le **comportement collectif est une conséquence non triviale du comportement des agents**.

Les systèmes complexes se distinguent principalement des systèmes multi-agents par deux caractéristiques :

- un système multi-agent n'a pas nécessairement un grand nombre d'entités,
- les agents d'un système complexe n'ont pas nécessairement des objectifs individuels.

Le lien non trivial entre les individus et le collectif, aussi appelé *lien micro-macro*, joue un rôle important dans les systèmes complexes. Il entraîne que, malgré une connaissance parfaite des comportements des agents, l'évolution du système peut être imprévisible par le calcul.

L'impossibilité de prévoir l'évolution du système n'est pas liée à des limites mathématiques, et peut avoir plusieurs raisons :

- les calculs prédictifs ont un grand coût calculatoire, et ne sont pas réalisables en temps raisonnable,
- le système possède un comportement chaotique, c'est-à-dire une forte sensibilité aux conditions initiales.

Dans le cas où le système est chaotique, la moindre incertitude sur la donnée initiale entraîne une grande incertitude sur les états futurs du système. Par conséquent, il n'est pas possible de prédire l'évolution du système avec précision.

Le problème des n corps Afin d'illustrer l'imprévisibilité d'un système, nous exposons le problème historique des n corps [129].

Un ensemble de n points matériels (de masses $\{m_i\}$) placés dans l'espace sont en interaction gravitationnelle selon la loi de Newton. Le point i exerce une force d'attraction sur le point j donnée par :

$$\vec{F_{i,j}} = -G\frac{m_i m_j}{d(i,j)^2}\,\vec{e}_{i,j}$$

où

- G est la constante de gravitation universelle,
- $d(i,j)$ est la distance séparant les points i et j,
- $\vec{e}_{i,j}$ est le vecteur unitaire orientée de i vers j.

La question est de savoir si l'on peut prévoir les trajectoires des points, en appliquant les principes de la mécanique classique.

Newton a démontré que le cas $n = 2$ peut être résolu de manière analytique, mais le cas $n = 3$ s'est avéré beaucoup plus difficile. La contribution historique de H. Poincaré [1] a montré que le problème des trois corps n'admet généralement [2] pas de solution analytique [6], et par extension il en est de même pour $n \geq 4$.

Des approches numériques ont permis d'étudier certaines solutions d'un système à trois corps. Il s'est avéré que ce système est chaotique, c'est-à-dire que de minuscules variations

1. Qui est une des origines de sa reconnaissance scientifique.
2. On exclut la solution approchée dans le cas où l'un des trois corps a une masse négligeable [99].

dans les conditions initiales peuvent donner lieu à des trajectoires complètement différentes. De ce fait, il est impossible de réaliser des prédictions de manière précise.

On trouve toutefois une solution formelle au problème des trois corps dans [76]. Cette solution se présente comme une série entière qui converge très lentement, et qui est inutilisable en pratique.

Cet exemple montre que même pour un très petit système, formé de trois entités seulement, on peut avoir une évolution imprévisible. De plus, l'impossibilité de prédire l'état futur du système est essentiellement due à deux raisons : le système est chaotique, et le calcul explicite des états futurs n'est pas réalisable en temps raisonnable.

1.2.2 Propriétés émergentes

Un système complexe peut avoir des propriétés qui ne peuvent être déduites de la connaissance des agents. Ces propriétés, dites *émergentes* [134], peuvent toutefois être observées sur un système en cours d'évolution.

Dans cette partie, nous présentons deux propriétés émergentes en détail : l'auto-organisation et la robustesse.

Auto-organisation Une propriété émergente importante (et souvent souhaitée) d'un système multi-agent complexe est l'*auto-organisation* [3] [49]. Cette notion désigne l'apparition de structures organisées (des motifs) dans le système qui ne sont pas programmées dans les agents.

Un exemple simple rencontré régulièrement dans la nature est celui des oiseaux migrateurs. Au cours de leurs trajets, ces oiseaux ont pour habitude de s'organiser « en V », qui est une formation efficace pour se déplacer dans les airs. Il est cependant fort probable que les oiseaux migrateurs n'aient pas de notion d'aérodynamisme.

Par conséquent, les oiseaux migrateurs n'ont pas pour objectif spécifique de créer une formation en V. Cette formation est le résultat d'un processus d'auto-organisation, dont nous ignorons les mécanismes et les motivations [4].

De manière générale, l'auto-organisation est comprise comme une tendance naturelle à faire croître l'ordre du système. Cet notion d'*ordre* peut être mesurée à l'aide une quantité appelée *entropie* retrouvée en thermodynamique ou en théorie de l'information [80], qui sera expliquée et mise en oeuvre dans un cas particulier traité dans le chapitre 5.

Robustesse ou sensibilité aux perturbations Dans le cadre des systèmes multi-agents, la robustesse est définie comme la capacité du système à maintenir une activité satisfaisante en présence de perturbations [132, 34].

Pour un système complexe, une perturbation peut avoir un effet important et complètement imprévisible sur son évolution. Même une modification locale peut rapidement se

3. Elle est parfois exigée pour qualifier un système de *complexe*.
4. Il est fort possible qu'il s'agit simplement d'une tendance vers le moindre effort.

propager dans le système, en vertu des interactions locales, et avoir un impact radical sur le fonctionnement du système.

Prenons l'exemple d'un organisme affecté par un virus. Deux cas de figure sont envisageables :

- l'organisme succombe au virus, qui a complètement perturbé son évolution par une modification locale,
- l'organisme survit, grâce à son système immunitaire.

Dans le second cas, le système immunitaire assure la robustesse de l'organisme. La robustesse n'est cependant pas une exigence de tous les systèmes, puisqu'elle réduit le degré de contrôle local qu'un intervenant extérieur peut exercer sur le système.

Pour revenir à l'exemple du virus, si l'on souhaite agir de manière locale sur l'organisme par une intervention médicale, cette défense peut causer un choc opératoire et ainsi entraver l'intervention. La robustesse du système a donc pour effet d'en réduire la contrôlabilité.

1.2.3 Les systèmes multi-agents à fonctionnalité émergente

Un système multi-agent à fonctionnalité émergente est un système multi-agent complexe dont la fonctionnalité n'est pas prédéfinie par le concepteur, mais qui émerge dynamiquement des interactions des agents. Cette émergence est en général accompagnée de propriétés émergentes comme l'auto-adaptation et la robustesse.

1.3 Quelques applications

Dans cette partie nous présentons quelques domaines d'application des système multi-agents complexes.

1.3.1 Le génie logiciel

Depuis la fin des années 80, la vision monolithique des logiciels informatiques a été progressivement abandonnée au profit de logiciels formés de composants autonomes. Ces briques élémentaires de programme produisent le résultat souhaité à travers des interactions virtuelles.

Selon [58], l'approche multi-agent permet à un concepteur de logiciels de se concentrer sur des composantes humainement appréhendables. Ainsi le choix des bonnes règles d'interactions entre les composants s'apparente au choix d'un contrat social, et permet de transposer la réflexion dans le domaine social, qui est parfois plus simple à appréhender pour l'esprit humain.

Des méthodologies comme ADELFE [14] ou TROPOS [25] permettent de guider le processus de conception d'un système multi-agent à partir d'une spécification de besoins. Il existe également de nombreuses plateformes techniques, comme CORMAS [22] ou JADE [8], qui permettent de développer des logiciels multi-agents ; une liste plus complète figure dans [114].

1.3.2 La simulation à base d'agents

La simulation multi-agent permet de reproduire des systèmes complexes sur des supports virtuels afin de mieux les comprendre et d'en prédire l'évolution. Elle permet de simuler ces systèmes alors que leur mise en oeuvre réelle serait difficile, pour des raisons techinques ou financières.

Les applications de la simulation à base d'agents sont diverses et multidisciplinaires :
- en **robotique**, elle permet de réaliser des expérimentations virtuelles peu coûteuses par rapport au coût d'une simulation réelle qui demande des moyens techniques importants,
- en **sociologie**, elle permet de comprendre et de prévoir l'organisation d'une communauté,
- en **dynamique des populations**, elle permet d'anticiper et prévenir la formation d'embouteillages,
- en **épidémiologie**, on peut l'utiliser pour prévoir la propagation d'un virus.

Des plateformes adaptées comme GAMA [140], REPAST [81] ou NetLogo [141] permettent de réaliser des simulations multi-agents.

1.3.3 La résolution collective de problèmes

Lorsqu'un système multi-agents produit un comportement approprié à l'échelle macroscopique, alors que les agents ont des comportements simples, on parle d'*intelligence distribuée* [5]. Cette intelligence collective peut alors être mise à profit pour résoudre des problèmes conséquents, qui font intervenir de nombreuses variables et contraintes.

La résolution collective par systèmes multi-agents a été utilisée pour des problèmes industriels, comme la gestion d'un stock de ressources [12] ou le contrôle d'un moteur à combustion [17], mais aussi pour des problèmes plus académiques comme des problèmes de satisfaction de contraintes [104, 120], ou l'optimisation sous contraintes d'une fonction mathématique [96, 82].

Un avantage important de l'approche par systèmes multi-agents est qu'elle donne un aspect social au système. Par conséquent il est possible de s'inspirer de systèmes complexes sociaux rencontrés dans la nature pour créer un système complexe virtuel.

Les algorithmes de type «*colonie de fourmis*» sont un exemple de cette inspiration. En observant une colonie de fourmis à la recherche de nourriture, des biologistes ont observé que les fourmis finissent par établir un itinéraire commun, qui est le chemin le plus court de la base à la nourriture.

L'explication qui a été donnée est que les fourmis chargées (qui retournent au nid) laissent derrière elles une trace de phéromones qui attire les fourmis à la recherche de nourriture. Comme ces phéromones s'évaporent, leur intensité diminue sur les pistes les plus longues, contrairement aux pistes les plus courtes où l'intensité des phéromones est renforcée grâce aux multiples passages.

Cette stratégie peut être utilisée pour établir le chemin le plus court entre deux noeuds

5. *Intelligence artificielle distribuée* si le système multi-agent est virtuel.

d'un graphe. Des fourmis virtuelles sont déposées sur un des noeuds, et parcourent le graphe de façon aléatoire jusqu'à arriver à l'autre noeud, en déposant des phéromones virtuelles. À cause de l'évaporation des phéromones virtuelles, un chemin finit par s'établir : il s'agit du chemin le plus court entre les noeuds.

L'exemple montre qu'un système d'agents très simples peut parvenir à résoudre un problème relativement difficile [6]. Les fourmis virtuelles n'ont pas de mémoire, ne communiquent que de manière indirecte (par *stigmergie*) et ont des perceptions locales très limitées (la quantité de phéromones sur les arêtes voisines).

Il est important de noter la différence de l'approche multi-agent avec la méthode cartésienne. Cette dernière consiste à diviser un problème donné en sous-problèmes indépendants, de manière répétée, jusqu'à ce que chacun des sous-problèmes soit suffisamment simple pour être résolu. La résolution de tous les sous-problèmes apporte alors une solution au problème initial.

Dans une approche multi-agent on n'altère absolument pas le problème considéré, mais on essaye de construire un mécanisme de résolution adapté. Par ailleurs, le mécanisme de résolution est obtenu comme l'action collective d'agents en interaction et peut donc être complexe. De ce fait il est possible qu'un problème difficile puisse être résolu à l'aide d'un système formé d'agents très simples.

Les méthodes de résolution collective par systèmes multi-agents restent relativement heuristiques à ce jour. La théorie AMAS [31, 64] donne des guides pour concevoir de tels systèmes, en basant les interactions des agents sur la notion de *coopération*. Cette notion de coopération, tout comme la méthode de conception qui en découle, n'a cependant pas de formulation mathématique précise et n'admet - pour l'instant - aucune justification formelle.

1.4 Conclusion et positionnement de la thèse

À défaut de donner une définition précise des systèmes multi-agents, qui fait toujours débat, ce chapitre expose un certain nombre de caractéristiques propres à ces systèmes. La présence d'individus en interaction nous rapproche du domaine des systèmes complexes, qui est une science relativement jeune dont la frontière n'est pas clairement tracée. Un aspect important de ces systèmes est que le comportement du collectif ne se déduit pas simplement du comportement des agents. Les systèmes multi-agents à fonctionnalité émergente sont des systèmes multi-agents complexes dont la fonctionnalité n'est pas prédéfinie par le concepteur, mais qui émerge des interactions des agents.

Un défaut inhérent à ces méthodes est la perte de prévisibilité et de contrôle du système. Il peut être impossible de savoir si un système multi-agent a un fonctionnement satisfaisant sans l'exécuter. De la même manière, s'il s'avère que la performance du système est insuffisante, il est difficile de savoir comment les agents doivent être ajustés pour réaliser une meilleure performance.

La validation et le calibrage des systèmes multi-agents à fonctionnalité émergente passe

6. Le problème du chemin le plus court peut être résolu en temps polynomial à l'aide de l'algorithme de Dijkstra [50].

donc principalement par l'expérimentation. Or cette démarche ne fournit que des réponses partielles, d'ordre statistique. De tels résultats limitent l'utilisation aux systèmes non-critiques, qui ne font pas intervenir de vies humaines, ni de matériel technologique coûteux. Pour cette raison, l'étude formelle des systèmes multi-agents à fonctionnalité émergente est un enjeu majeur.

L'objectif de cette thèse est précisément d'explorer les méthodes mathématiques pertinentes pour l'étude formelle des systèmes multi-agents à fonctionnalité émergente. Cependant, les systèmes multi-agents étant d'une grande diversité, à la fois en termes de types d'agents, de types d'environnement, et en termes de modes d'interaction, nous avons limité notre travail à une classe de systèmes spécifiques.

Les systèmes multi-agents auxquels cette thèse s'intéresse plus particulièrement ont les caractéristiques suivantes :

- ils sont formés d'**un grand nombre d'agents simples**, qui agissent selon un cycle de la forme **perception-décision-action** [7],
- l'évolution des agents est **synchrone**, déterminée par une horloge centrale,
- l'**environnement** est assimilable à un ensemble d'**agents passifs**,
- les **interactions** des agents sont **locales et simples**,
- les agents sont **coopératifs**, et partagent un **objectif commun**.

Ces caractéristiques déterminent une classe de système multi-agents précise, qui va conditionner le choix du modèle mathématique.

Afin de procéder à l'étude formelle des systèmes multi-agents ciblés, il faut leur donner une représentation mathématique convenable. Le prochain chapitre est consacré à une discussion sur cette étape de modélisation, ainsi qu'à une présentation de modèles envisageables.

7. Dans la littérature on parle d'*agents réactifs* [58, 151].

2 Modélisation mathématique des systèmes multi-agents

L'utilisation de systèmes multi-agents à fonctionnalité émergente s'est montrée fructueuse dans divers domaines, et prouve qu'il est possible de construire des mécanismes efficaces pour des problèmes difficiles à partir d'un ensemble d'entités simples en interaction.

Utiliser les systèmes multi-agents à fonctionnalité émergente afin de résoudre des problèmes est une démarche relativement nouvelle, et permet de voir les problèmes sous un angle nouveau. Cette démarche suit une progression *bottom-up* : dans une premier temps, le système est construit. Ensuite, il est exécuté afin de vérifier son adéquation au problème. Cette progression s'oppose à l'approche *top-down*, où le concepteur part du problème afin de construire la solution [1] et doit anticiper l'effet de ses décisions sur le résultat.

Au cours de la conception d'un système multi-agents à fonctionnalité émergente, la progression bottom-up est intrinsèque. En conséquence, l'adéquation du système au problème pour lequel il a été conçu ne peut être vérifiée qu'à l'issu du processus conception.

L'adéquation d'un système à sa tâche est donc généralement validée de façon expérimentale et pose des limites évidentes en termes de garantie des résultats et de coût en temps. L'objectif de cette thèse est d'explorer des approches formelles pour permettre la vérification des propriétés de ces systèmes, telles que la convergence, l'existence et la stabilité de certains régimes stationnaires...

Pour ce faire, il faut donner au système une représentation formelle convenable. Par cela nous entendons que le modèle est fidèle au système qu'il représente, mais aussi qu'il doit permettre de démontrer les propriétés souhaitées du système.

La modélisation d'un système complexe est une tâche délicate, car la simple donnée des règles des agents ne permet, par définition, pas de prévoir son évolution. Nous avons constaté qu'il n'existe que très peu de travaux mathématiques consacrées aux systèmes multi-agents, et il est donc difficile de dresser un état de l'art précis. Une partie importante de ce travail de thèse a donc été d'explorer les modèles mathématiques envisageables et d'évaluer leur pertinence pour la représentation des systèmes multi-agents.

L'apport des modèles mathématiques pour le domaine des systèmes multi-agents ne s'arrête pas à l'étude des propriétés formelles d'un système donné. En effet, pour étudier

1. C'est notamment le cas dans une démarche cartésienne.

le modèle mathématique d'un système multi-agent il n'est pas nécessaire que ce système soit effectivement construit. En modélisant un système qu'on cherche à construire, l'étude mathématique peut donc apporter des réponses afin de guider le processus de conception.

Pour ce faire, l'on peut modéliser un système multi-agent en cours de conception comme une famille de modèles, paramétrée par des inconnues qui représentent les parties non déterminées du système. Ces inconnues peuvent être des paramètres internes des agents, ou même leur règle de comportement complète. Le choix de leurs valeurs optimales, vis-à-vis de l'objectif global du système, revient alors à un problème d'optimisation. Il est possible qu'une approche mathématique permette d'apporter des réponses (éventuellement partielles) sur ce problème d'optimisation, et ainsi, de contribuer au processus de conception du système multi-agent.

La première partie du chapitre est dédiée à une discussion générale sur la modélisation des systèmes multi-agents à fonctionnalité émergente. La seconde partie contient une présentation d'un certain nombre de modèles mathématiques envisagés. Nous évaluons la pertinence de chacun de ces modèles dans le cadre des systèmes multi-agents étudiés dans cette thèse.

2.1 Quelques réflexions sur la modélisation

L'étude d'un système complexe, en vue d'établir des propriétés formelles, passe par une étape non-triviale de modélisation.

Il faut alors comprendre un modèle comme un objet mathématique qui représente un système existant ou un système qu'on veut construire. Par « représenter » nous entendons qu'il partage un certain nombre de propriétés avec le système considéré. En particulier, un modèle peut ignorer une partie des caractéristiques du système qu'il représente afin de se concentrer sur des caractéristiques jugées importantes par le modélisateur.

La qualité d'un modèle n'est pas uniquement mesurée par la similarité avec l'objet qu'il représente, mais aussi par ce qu'il nous apprend sur cet objet. Cette expressivité est évidemment conditionnée par le nombre de caractéristiques qui sont intégrés dans le modèle, mais aussi (et surtout) par la maîtrise mathématique que nous avons du modèle. Si celle-ci est insuffisante, il est impossible d'exploiter le modèle de manière satisfaisante.

2.1.1 Deux points de vue d'une modélisation : la description Lagrangienne et la description Eulérienne

La *description Lagrangienne* [7] est une méthode de représentation qui consiste à décrire le système à l'aide de grandeurs qui dépendent des positions des particules. Une grandeur X dépendant d'une particule située à la position $P(t)$ dans l'espace à l'instant t est représentée à l'aide d'une formule de la forme

$$X(t) = f(P(t), t).$$

À l'opposé, dans une *description Eulérienne* [7] le système est décrit avec des grandeurs qui dépendent d'un point géométrique considéré. Soient X une grandeur et P un certain

point de l'espace. Dans une description Eulérienne la grandeur X s'exprime à l'aide d'une formule de la forme

$$X(t) = f(P, t).$$

Notons que contrairement au cas Lagrangien le point P est fixe, ce qui permet en particulier d'avoir un référentiel fixe dans la description.

Nous illustrons cette opposition avec un exemple (issu de [39]) qui se prête à chacune de ces descriptions, connu dans la littérature comme *la fourmi de Langton*. Considérons le réseau infini $\mathbb{Z}^2$ représenté sous la forme d'un quadrillage, où chaque cellule a un état parmi $\{\text{noir}, \text{blanc}\}$. Une fourmi est positionnée sur une de ces cases et évolue selon la règle suivante :

- Si la fourmi est sur une case noire, elle tourne de 90° vers la droite, change la couleur de la case en blanc et avance d'une case.
- Si la fourmi est sur une case blanche, elle tourne de 90° vers la gauche, change la couleur de la case en noir et avance d'une case.

La figure 2.1 montre quelques itérations de cet algorithme.

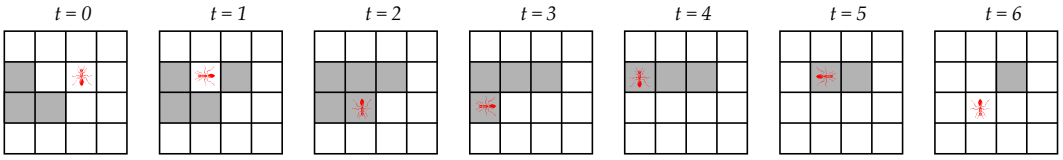


Figure 2.1 — Évolution d'une fourmi de Langton.

Une *description Lagrangienne* de cette situation est centrée sur le comportement de la fourmi, et a les caractéristiques suivantes :

- **États :** La position de la fourmi.
- **Environnement :** Coloriage du réseau.
- **Transitions :** Selon la couleur de la case occupée, rotation et déplacement vers la gauche ou la droite, et inversion de la couleur de la cellule (voir fig. 2.2).

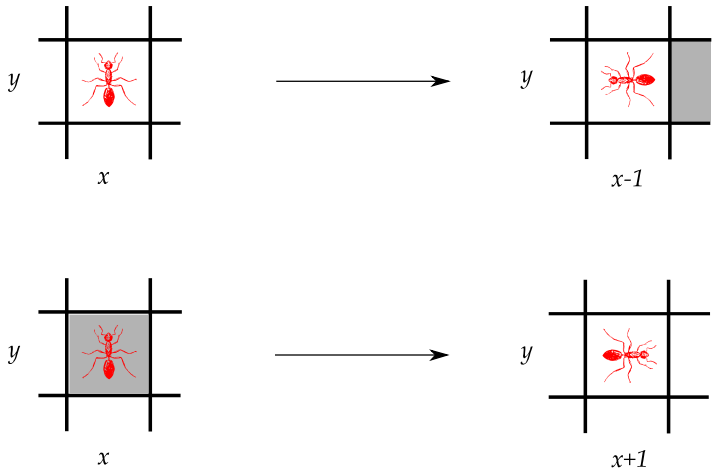


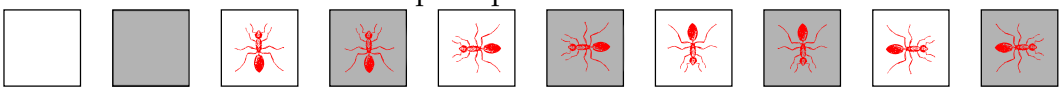
Figure 2.2 — Transitions de la fourmi de Langton.

La figure 2.2 ne représente que le cas où la fourmi est orientée vers le haut. Les règles de

transition complètes sont obtenues en effectuant des rotations de 90, 180 et 270.

Une *description Eulérienne* équivalente à celle que nous venons de donner est centrée sur les différentes cellules, et a pour caractéristiques :

- **États :** Noire (ou blanche) inoccupée, ou noire (ou blanche) avec une fourmi qui peut être orientée dans les directions principales.



- **Environnement :** États des quatre cellules adjacentes.
- **Transitions :** Si parmi la cellule et ses voisines, aucune ne contient la fourmi, ne rien faire. Sinon évoluer selon les règles représentées sur la figure 2.3.

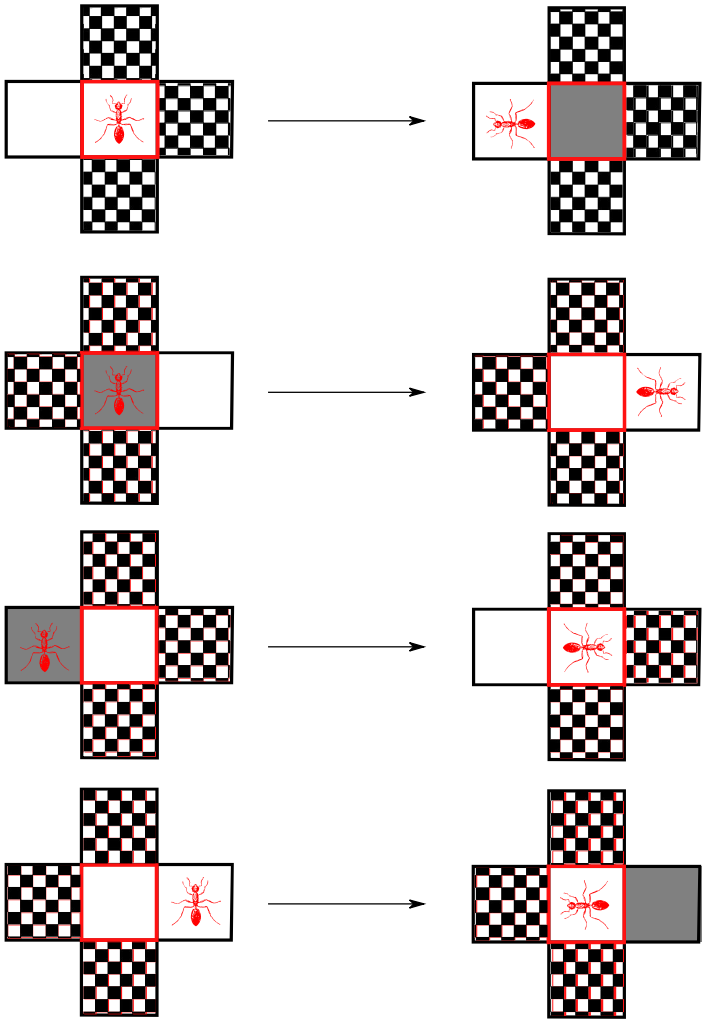


Figure 2.3 — Transitions de la cellule centrale (entourée de rouge).

Remarque 1. *En ce qui concerne la figure 2.3 :*

- *L'état des cellules quadrillées n'a pas d'impact sur la transition (et peut donc indifféremment être noir ou blanc, avec ou sans fourmi).*
- *Les règles de transition complètes sont en réalité quatre fois plus nombreuses. Il faut ajouter à ces quatre transitions, celles obtenues en effectuant des rotations (du voisinage complet) de*

90°, 180° et 270°.

La fourmi de Langton peut donc être décrite de manière équivalente de façon Eulérienne ou Lagrangienne.

Voyons ce qui advient de ces descriptions en présence de plusieurs fourmis de Langton. Si l'on n'exclut pas le fait d'avoir plusieurs fourmis sur une même cellule, la description Lagrangienne correspond simplement à l'évolution simultanée de plusieurs fourmis identiques.

En revanche, la description Eulérienne devient plus délicate. Chaque cellule doit intégrer dans son état le nombre de fourmis qui s'y trouvent, ainsi que leurs orientations. Le nombre d'états potentiel d'une cellule est donc polynomial vis-à-vis nombre total de fourmis.

Dans une situation avec beaucoup de fourmis, la description Eulérienne s'avère donc moins avantageuse qu'une description Lagrangienne. Elle possède toutefois une qualité importante en mécanique statistique : le fait de considérer une position fixe permet d'avoir un référentiel immobile. Pour la fourmi de Langton, cela signifie que l'on peut utiliser un système de coordonnées absolues qui est invariant au cours du temps.

À l'opposé, une description centrée en un point mobile (comme la description Lagrangienne) impose l'introduction d'un référentiel mobile. En conséquence les variations des grandeurs caractéristiques du système ne peuvent être décrites qu'en tenant compte de l'évolution du point, et nécessitent par exemple d'introduire la notion de *dérivée particulaire*.

Pour conclure sur ces considérations, nous indiquons que le passage de la description Lagrangienne vers la description Eulérienne est utilisé de manière intensive à partir du chapitre 4. La raison est que les agents considérés sont échangeables, et que le système est essentiellement décrit par le nombre d'agents en chaque état. L'évolution du système est ainsi décrite par l'évolution du nombre d'individus en chaque état, et fera intervenir autant d'équations que d'états. Dans un contexte où il y a beaucoup d'agents par rapport au nombre d'états, une telle description sera plus compacte qu'une description centrée en l'individu et donc plus avantageuse pour une étude formelle.

Par exemple, dans le chapitre 5 les déplacements d'un groupe d'individus seront représentés par des *équations bilan* qui décrivent le taux d'entrée et de sortie dans chaque *élément de volume*. Cette représentation est donc attachée à chaque lieu de l'espace et non aux individus.

2.1.2 L'aléatoire dans la modélisation des systèmes multi-agents

Quelles que soient les décisions prises sur la représentation du système, tout modèle mathématique comprendra une *variable d'état* qui représente l'état du système. Notons $X(t)$ la valeur de la *variable d'état* à l'instant t. Cette variable va évoluer au cours du temps, selon une certaine règle qui varie en fonction de la représentation du temps choisie.

Il est possible que pour un même état initial du système, représenté par la valeur $X(0)$, on obtienne des états différents à un moment ultérieur[2] (i.e. différentes valeurs de $X(t_0)$

2. Dans ce cas on dit que l'évolution du système est *non déterministe*.

pour un certain $t_0 > 0$). Les *modèles probabilistes* permettent de représenter ces phénomènes, en attribuant différentes *probabilités* [3] à ces états possibles.

Pour les systèmes multi-agents, le choix d'un modèle probabiliste peut se justifier par le fait que l'environnement est incertain, et exerce des influences imprévisibles sur le système. Dans ce cas, le non-déterminisme est attribué à une perturbation extérieure.

Ce raisonnement peut même être appliqué de manière locale, en considérant que chaque particule du système est isolée, et subit les influences du système comme des perturbations extérieures. Un exemple historique, utilisé en mécanique statistique, est l'*équation de Langevin*. Cette équation décrit le mouvement d'une grosse particule qui est plongée dans un liquide visqueux, en collision permanente avec un grand nombre de petites particules, par :

$$m\dot{v}(t) = -k\,v(t) + \Gamma(t)$$

où

- m est la masse de la particule,
- v est la vitesse de la particule,
- k est la constante de viscosité du milieu,
- Γ une force fluctuante (aléatoire).

Il s'agit du principe fondamental de la dynamique [4] avec une composante aléatoire. La force aléatoire Γ (parfois appelée *force de Langevin*) représente l'effet des nombreuses collisions avec les petites particules, difficile, voire impossible, à mettre en équation.

Une autre motivation pour considérer des modèles probabilistes est que dans certains systèmes les agents ont eux-mêmes un comportement aléatoire. Dans certains problèmes, les comportements déterministes ne permettent pas d'aboutir à un comportement satisfaisant pour le système. Dans ces problèmes, il nécessaire de considérer des comportements aléatoires (qui englobent les comportements déterministes [5]) pour arriver à des performances satisfaisantes.

Considérons un exemple très simple où les comportements déterministes sont largement sous-optimaux. Cet exemple sera largement étudié dans les chapitres 3 et 4. Un grand groupe de personnes identiques se trouve dans une pièce, et souhaite la quitter, mais l'unique porte qui permet de sortir ne peut laisser passer qu'une personne à la fois. À chaque instant, toutes les personnes doivent simultanément indiquer si elles souhaitent quitter la pièce ou non. Si une et une seule personne souhaite quitter la pièce, elle peut le faire. Sinon, personne ne quitte la pièce.

Si toutes les personnes raisonnent de manière identique, et ont une information parfaite (et identique) de la situation, alors une règle de décision déterministe ne peut mener qu'à deux issues : toutes voudront sortir ou personne ne voudra le faire. Par conséquent personne ne peut quitter la pièce.

Comme les personnes raisonnent de manière identique, avec une règle déterministe basée sur les mêmes informations, elles sont condamnées à rester dans la pièce. Pour éviter ce

3. C'est-à-dire des mesures de la chance qu'un évènement ou un autre se produise.
4. « Somme des forces = masse × accélération ».
5. Pour comprendre cela, il suffit de constater que les comportements aléatoires qui affectent la probabilité 1 à une certaine action parcourent complètement les comportements déterministes.

problème, on peut envisager des stratégies aléatoires simples, qui permettent de débloquer la situation [6].

Pour l'instant, la discussion a principalement porté sur le rôle et la pertinence de l'aléatoire au niveau microscopique (c'est-à-dire au niveau des agents). L'aléatoire peut également exister au niveau macroscopique dans un système, mais ces situations sont souvent plus difficiles à cerner. Par exemple, un système asynchrone où les actions des agents sont exécutées dans un ordre choisi aléatoirement peut donner des résultats complètement différents d'une exécution à une autre.

Dans le cadre de cette thèse, nous privilégions les systèmes multi-agents dont le comportement global est déterministe, et écartons donc la possibilité de perturbations aléatoires au niveau global.

2.1.3 Distinction entre les actions des agents et le résultat global : le modèle influence-réaction

L'évolution du système est modélisée comme une transformation qui décrit les changements de l'état du système au cours du temps. Cette transformation résulte des actions exercées par les agents, qui peuvent - notamment dans le cas d'un système complexe - avoir des effets inattendus. Il convient donc de distinguer l'action des agents et le résultat global de ces actions. Cette nuance est généralement négligée dans les modèles dynamiques, qui réduisent l'évolution du système à une succession de transitions d'état.

Le modèle influence-réaction, initialement proposé dans [58], permet de distinguer l'action des agents et le résultat de ces actions. Selon ce modèle les transitions du système s'effectuent en deux temps :

- **Influence :** Les agents exercent des *influences*, en exprimant des intentions d'action. Ces actions inabouties peuvent se présenter comme des projets, ou des souhaits [7].
- **Réaction :** La phase de *réaction* calcule le résultat des influences exercées par les agents. Cette réaction peut être vue comme la *réponse* de l'environnement, ou encore comme le résultat des actions en appliquant « *les lois de l'univers* » ([58] p. 176).

Le diagramme suivant illustre cette factorisation de la loi d'évolution du système :

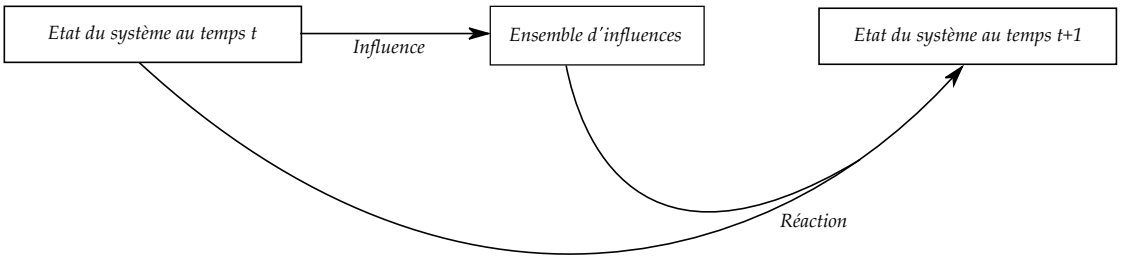


Figure 2.4 — Le modèle influence/réaction

Le principe d'influences et réactions a été utilisé dans [39] pour une famille de fourmis de Langton, en déplacement simultanée un quadrillage avec la contrainte que chaque cellule ne

6. Ce constat a également été fait sur le problème similaire du *Bar d'El Farol* [4].
7. Par exemple : «Je souhaite me déplacer dans cette direction.»

peut contenir (au plus) qu'une seule fourmi. Avant de se déplacer, chaque fourmi exerce une influence, en indiquant la cellule où elle souhaite se déplacer. La réaction à ces influences revient simplement à autoriser le déplacement, si la cellule sollicitée est inoccupée, et si elle n'est sollicitée que par une seule fourmi. De cette manière aucune cellule n'est occupée par plus d'une fourmi.

Un autre exemple est celui de la mécanique classique, pour un système de corps en mouvement. Lorsque deux corps se rencontrent chacun exerce une influence sur l'autre, sous la forme d'une force de collision. La réaction de ces influences est obtenue en appliquant les lois de la mécanique à ces forces, et a pour effet de modifier les trajectoires des corps.

Le modèle influence-réaction permet donc de distinguer les actions des agents et le résultat de ces actions. Cette distinction est nécessaire parce que les types agents auxquels cette thèse s'intéresse ont des connaissances limitées, et sont généralement incapables de déterminer le résultat de leurs actions.

2.2 Tour d'horizon de modèles envisageables

Suite à ces considérations sur la modélisation, nous exposons plus en détail un ensemble de modèles qui prennent en compte certaines caractéristiques des systèmes multi-agent ciblés dans cette thèse. Nous analysons en particulier les forces et faiblesses de ces modèles mathématiques.

Les modèles seront partagés en deux familles. Dans un premier temps nous présentons des modèles purement descriptifs, qui considèrent un système donné et en étudient les propriétés formelles. Ensuite, nous présentons quelques modèles prescriptifs, qui sont munis d'une certaine fonction de satisfaction et la tâche implicite de la maximiser.

2.2.1 Les modèles descriptifs

Les modèles descriptifs permettent d'étudier les propriétés formelles d'un système multi-agents donné, afin d'en prévoir ou expliquer le déroulement. Dans cette section, nous présentons rapidement quelques modèles descriptifs possibles.

2.2.1.1 Les systèmes dynamiques

Les systèmes dynamiques permettent, comme leur nom l'indique, de décrire l'aspect dynamique d'un système multi-agent. Cette famille de modèles est essentiellement caractérisée par

- **la causalité :** l'état futur du système ne dépend que de son état présent et ses états passés.
- **le déterminisme :** un état présent du système ne donne lieu qu'à un et un seul état «futur».

Les systèmes dynamiques se présentent sous deux formes :

- **Les systèmes dynamiques à temps discret :** Dans ce cas, l'évolution est donnée par

une règle de récurrence de la forme

$$X(t+1) = F(t, X(t))$$

qui décrit comment la variable d'état du système à l'instant $t+1$ est obtenue à partir de sa valeur à l'instant précédent.

– **Les systèmes dynamiques à temps continu :** Dans ce cas, l'évolution est donnée par une équation différentielle de la forme

$$X'(t) = F(t, X(t))$$

qui décrit les variations de la variable d'état du système à chaque instant.

Comme la variable $X(t)$ peut prendre des valeurs vectorielles, un système dynamique peut se présenter comme un système d'équations.

Ce formalisme est encore très général, et n'exclut par exemple pas la présence de mémoire. En effet, la variable d'état $X(t)$ n'a aucune sémantique et peut par exemple comprendre l'état actuel du système, ainsi qu'un historique d'états passés.

Dans le domaine des systèmes dynamiques, un objectif omniprésent est l'étude qualitative du comportement à long terme. Quelques aspects de cette étude sont :

– **La recherche des points fixes.** Il s'agit des états dans lesquels le système n'évolue plus.
– **La détermination de la nature des points fixes.** Un point fixe peut être
  - **attractif :** lorsque le système s'en approche suffisamment, il converge vers ce point fixe[8].
  - **répulsif :** si le système est initialement positionné en ce point alors il y reste. Par contre, si l'état du système varie, même légèrement, il va s'éloigner de ce point.
– **La recherche de points périodiques.** Il est possible que l'évolution de la variable d'état du système se répète au bout d'un certain temps (la période). Ces trajectoires périodiques peuvent également être attractifs ou répulsifs.
– **L'étude des dynamiques chaotiques.** Certains systèmes dynamiques (comme le système des n corps présenté à la section 1.2.1) ont une dynamique particulière, et sont très sensibles aux conditions initiales. Ces systèmes *chaotiques* possèdent des trajectoires qui ne sont ni périodiques, ni convergentes.

Pour déterminer la nature des points fixes, par exemple, il existe des outils spécifiques comme l'étude des valeurs propres de la matrice jacobienne de la fonction F. Cette méthode est cependant limitée par la taille de la variable d'état. En effet, la matrice jacobienne de F a la même taille que la variable d'état. Or l'étude des valeurs propres d'une matrice devient difficile si cette matrice est grande.

Pour cette raison les systèmes dynamiques utilisés en pratique sont généralement petits (à moins de 10 équations), bien qu'ils exigent des simplifications à la limite du raisonnable. Un exemple classique de telles simplifications est le système de Lotka-Volterra, décrivant l'évolution d'une population de proies et prédateurs en interaction :

$$\begin{cases} x'(t) = ax(t) - bx(t)y(t) \\ y'(t) = -cy(t) + dx(t)y(t) \end{cases} \tag{2.1}$$

8. Lorsque la variable d'état $X(t)$ prend ses valeurs dans un espace métrique, il est possible de donner un sens précis mathématique à la notion de convergence.

où
- $x(t)$ et $y(t)$ représentent les effectifs de proies et prédateurs respectivement,
- a représente le taux de croissance naturel des proies,
- b représente le taux de mortalité des proies dû à la prédation,
- c représente le taux de mortalité naturel des prédateurs,
- d représente le rendement de la prédation.

La réalité est simplifiée de diverses manières : les populations sont supposées parfaitement homogènes, la «prédation» est supposée proportionnelle au produit des effectifs des population, la croissance des proies en l'absence de prédateurs est supposée exponentielle, ... Le modèle de Lotka-Volterra permet toutefois d'expliquer des phénomène périodiques constatés sur des données réelles [111], ou même des simulations de proies et prédateurs à base d'agents [9].

Dans la littérature, les travaux [98, 79] sont consacrés à la modélisation des systèmes multi-agents par les systèmes dynamiques. Malgré les comparaisons expérimentales effectuées dans ces travaux, spécifiques aux problèmes étudiés, il n'existe pas de méthode générale pour mesurer l'erreur commise au cours de la modélisation à l'aide de systèmes dynamiques.

Étant donné une famille paramétrée de systèmes dynamiques [10], il est possible de calculer les paramètres pour modéliser au mieux un système donné, en se basant sur un échantillon de simulations. Des méthodes statistiques précises, comme l'inférence bayesienne [30], permettent de réaliser cet ajustement en commettant une erreur minimale.

Il est cependant difficile de savoir comment les paramètres du modèle (comme les coefficient a, b, c, d de l'équation (2.1)) sont liés à la description des agents. Par conséquent, il est difficile d'utiliser les systèmes dynamiques pour calibrer les agents.

2.2.1.2 Les systèmes de particules en interaction

Les systèmes de particules en interaction [102] sont une famille de modèles aléatoires définis par :
- **Un espace de sites** S, qui est généralement le réseau $\mathbb{Z}^2$ ou un graphe non orienté. L'ensemble S est toujours supposé dénombrable.
- **Un ensemble d'états locaux** E, qui est généralement l'ensemble $\{0, 1\}$, et toujours un ensemble compact.

L'état global (ou la *configuration globale*) est la donnée d'un état local en chaque site, c'est-à-dire un élément de E^S.

La dynamique d'un système de particules en interaction est donnée par une loi d'évolution
- **à temps continu :** le temps est représenté par l'ensemble continu $\mathbb{R}^+$.
- **locale :** le changement d'état de chaque cellule ne dépend que d'un certain nombre de cellules voisines.
- **aléatoire :** plusieurs transitions peuvent avoir lieu, selon des probabilités spécifiques.

9. Un modèle de ce type figure dans la librairie de Netlogo.
10. Comme l'équation de Lotka-Volterra (2.1) qui est paramétrée par a, b, c, d.

– **à intervalles de temps aléatoires :** les instants $t_0, t_1, \ldots$ auxquels les transitions des sites ont lieu sont aléatoires, généralement représentés par des processus de Poisson locaux.

De manière intuitive, chaque site possède un réveil à durée aléatoire. Lorsque ce réveil «sonne» le site effectue une transition aléatoire.

Le fait de temporiser les actions des cellules par des processus de Poisson a pour conséquence que le système est presque sûrement asynchrone : la probabilité que deux sites évoluent simultanément est nulle.

Lorsque l'espace des positions est le réseau $\mathbb{Z}^2$, il est généralement supposé que les lois d'évolution des cellules sont invariantes par translation. Par conséquent tous les processus de Poisson ont la même intensité : la perception du temps est - en moyenne - identique pour toutes les cellules. On peut toutefois obtenir des fortes disparités au cours d'une simulation : avec une probabilité strictement positive, une cellule peut agir 100 fois pendant qu'une cellule voisine est inactive.

Comme nous nous intéressons aux systèmes multi-agents synchrones, nous réduisons l'étude aux systèmes de particules synchrones. Ces systèmes de particules particuliers sont aussi connus sous le nom d'*automates cellulaires*. Nous les présentons dans le paragraphe suivant.

2.2.1.3 Les automates cellulaires

Les *automates cellulaires* [154] sont, comme les systèmes de particules en interaction, définies par :

– **un espace de sites** (ou **cellules**) S (généralement $\mathbb{Z}^2$),
– **un espace d'états locaux** E (souvent $\{0, 1\}$),
– **un ensemble de voisinages** $V(s)$, c'est-à-dire une fonction V qui à chaque site s donne un ensemble de sites voisins $V(s)$.
– **une règle de transition locale**, qui décrit le changement d'état de chaque site $s \in S$ en fonction des états des sites voisins $V(s)$:

Lorsque l'espace des sites est $\mathbb{Z}^2$, on suppose généralement que les voisinages, ainsi que la loi d'évolution sont invariants par translation.

Contrairement aux systèmes de particules généraux, les automates cellulaires ont un fonctionnement synchrone à temps discret. Cela signifie qu'on peut noter les instants auxquels le système(la totalité des cellules) évolue $0, 1, 2, \ldots$. À chaque instant, tous les sites évoluent en appliquant simultanément leur règle d'évolution. Si cette règle de transition est *aléatoire* (resp. *déterministe*) on dit que l'*automate cellulaire* est *aléatoire* (resp. *déterministe*).

La description Eulérienne de la fourmi de Langton (section. 2.1.1) correspond exactement à un automate cellulaire déterministe. De manière générale, les automates cellulaires permettent de décrire des systèmes d'agents situés avec des règles d'interaction locales. Cette approche est étudiée plus en détail dans [137], qui intègre le principe d'influence-réaction dans les règles de transition locales.

Les automates cellulaires constituent par ailleurs un excellent exemple de systèmes complexes «minimaux». Le recueil historique de Wolfram [154] montre que les automates cellu-

laires les plus simples [11] exhibent déjà des comportements inattendus. Il va sans dire que les automates cellulaires plus sophistiqués (à cellules dans $\mathbb{Z}^2$ et à plus de deux états locaux) vont pouvoir montrer des phénomènes complexes encore plus surprenants.

La limite des automates cellulaires dans le cadre de la description des systèmes multi-agents se situe au niveau de l'étude formelle. Il est parfaitement possible d'exécuter un automate cellulaire de nombreuses fois et de constater des phénomènes émergents. En revanche, il est généralement difficile, voire impossible, de démontrer ces propriétés.

Pour l'exemple particulier de la fourmi de Langton, il a été démontré que la trajectoire de la fourmi est non bornée [138], mais la preuve est ad-hoc et difficilement généralisable.

Une question courante sur la dynamique d'un système aléatoire est de savoir s'il existe une unique distribution invariante [12] attractive. Or, pour un automate cellulaire (déterministe ou probabiliste) sur $\mathbb{Z}$ cette question est *indécidable* [105] : il n'existe aucun algorithme général qui permet de répondre à cette question, en un nombre fini d'étapes !

Par ailleurs, on trouve dans [70] une collection de problèmes NP-complets [13] sur des automates cellulaires spécifiques. Ces constats compromettent notre objectif, qui est d'établir des propriétés de grands systèmes multi-agents par des arguments formels.

2.2.1.4 Les méthodes de champ moyen

Les *méthodes de champ moyen* [118] ne sont pas réellement un famille de modèles, mais plutôt une simplification statistique pour décrire un grand système d'entités en interaction. On retrouve ces méthodes en magnétisme et en physique statistique mais aussi dans des applications récentes comme les réseaux de communication ou la théorie des jeux. Un chapitre entier est consacré à cette approche (chapitre 4), et pour cette raison nous ne présentons ici que les principes généraux.

Modéliser un système à l'aide d'une approche de type champ moyen revient essentiellement à remplacer les interactions des agents par une interaction moyenne. Du point de vue d'un agent, les interactions avec les autres agents sont modélisées par l'interaction avec une entité macroscopique, appelée *champ moyen*. Ce champ moyen résume les interactions locales par une seule interaction « moyenne ».

En plus de simplifier la représentation des interactions, l'approche champ moyen permet de réaliser une autre simplification importante : si les agents évoluent de manière aléatoire, et si leur « interaction moyenne » converge vers une quantité déterministe [14] lorsque le nombre d'agents tend vers l'infini, alors le champ moyen est déterministe. Les agents évoluent donc de manière aléatoire, tout en interagissant avec une grandeur déterministe sur laquelle ils ont un effet négligeable.

11. Les « *automates élémentaires* » qu'il étudie sont formés de cellules positionnées en ligne droite ($S = \mathbb{Z}$), en interaction avec leurs deux cellules adjacentes ($V(s) = (s-1, s+1)$), et à états dans $\{0, 1\}$.

12. C'est-à-dire une unique distribution de probabilités sur les états du système qui est invariante au cours du temps.

13. Auxquels on ne peut répondre en temps polynomial à l'aide d'un algorithme séquentiel déterministe.

14. À l'image de loi forte des grands nombres, qui montre la moyenne d'une famille de variables aléatoires identiques et indépendantes converge vers une quantité déterministe (l'espérance).

Nous aurons une discussion plus approfondie sur la pertinence de ce modèle au chapitre 4.

2.2.1.5 Équations à réaction-diffusion

Les équations à réaction-diffusion sont une famille d'équations aux dérivées partielles qui modélisent l'évolution de plusieurs substances chimiques sous l'effet de deux phénomènes :

- **la diffusion :** la dispersion spatiale des substances.
- **la réaction :** les transformations des différentes substances en contact.

Les équations à réaction-diffusion ont la forme suivante

$$\frac{\partial u}{\partial t} = d\Delta u + F(u)$$

où

- u désigne la concentration locale des substances en jeu, sous forme vectorielle.
- $\frac{\partial u}{\partial t}$ représente les variations de concentration par unité de temps.
- Le terme $d\Delta u$ modélise la diffusion spatiale des substances [15], dont l'intensité est quantifiée par *le coefficient de diffusion* d.
- La fonction F représente les réactions, en quantifiant les changements de concentration des différentes substances dus aux réactions.

Ce formalisme est naturellement adapté à la description des réactions chimiques, mais il peut être utilisé pour décrire des systèmes d'agents en interaction. En effet, il intègre des caractéristiques importantes comme la localité des interactions. De plus, il permet de prendre les positions des agents en considération, ce qui augmente fortement la taille de l'espace d'états et a posé problème pour beaucoup d'autres modèles. De plus, l'étude formelle de ces équations aux dérivées partielles est réalisée à l'aide de méthodes d'analyse mathématique. Ces méthodes ne sont absolument pas influencées par la taille du système, contrairement aux systèmes dynamiques ou les automates cellulaires.

Le chapitre 6 est entièrement consacré à ces équations dans un cadre plus général, avec des déplacements plus riches que la diffusion. Pour cette raison, nous remettons une discussion plus approfondie au chapitre 6.

2.2.2 Les modèles prescriptifs

À ce stade, nous avons présenté un certain nombre de modèles qui permettent de représenter et étudier un système donné. Ces modèles peuvent aider à comprendre et prévoir la dynamique d'un système donné.

Dans cette partie, nous exposons des modèles qui permettent de guider la conception des systèmes multi-agents en représentant ce processus comme un problème d'optimisation. Tous ces modèles prescriptifs diffèrent des modèles descriptifs présentés ci-dessus par les faits suivants :

- ils sont munis d'une ou plusieurs **fonctions de satisfaction**, qui représente la réussite du système dans sa tâche.

15. Le symbole Δ représente l'opérateur Laplacien.

- ils sont **définis de manière partielle**, i.e. un certain nombre de leurs paramètres ne sont pas fixés.
- ils ont pour **problème implicite** d'ajuster ces paramètres en vue d'optimiser la (ou les) fonction(s) de satisfaction.

2.2.2.1 La commande optimale

On considère un système dynamique classique (Cf. section 2.2.1.1) dont l'évolution est paramétrée par une **commande**. Cela signifie que sa loi d'évolution est de la forme

$$X(t+1) = F(t, X(t), u(t)) \tag{2.2}$$

si le système est discret, et de la forme

$$X'(t) = F(t, X(t), u(t)) \tag{2.3}$$

si le système dynamique est continu. Concrètement, l'évolution du système est paramétrée par un certain paramètre u appelée *commande*. Dans le cas d'un système multi-agent, la commande peut représenter n'importe quel paramètre microscopique (dans la règle d'évolution des agents) ou macroscopique indéterminé.

Le système dynamique décrit ci-dessus est muni d'un certaine *fonction de satisfaction* (ou *critère de performance*) de la forme

$$G(u) = g_f(T, x(T)) + \int_{t=0}^{T} g(t, x(t), u(t))dt \tag{2.4}$$

où

- $G(u)$ représente la satisfaction associée à la commande u,
- T représente l'instant final de la durée sur laquelle on souhaite exécuter le système,
- $g_f(T, x(T))$ représente le gain final (à l'instant T) associé à la commande u,
- $g(t, x(t), u(t))$ représente la satisfaction instantanée à l'instant t associée à la commande u.

Dans le cas d'un système discret, l'intégrale dans (2.4) est à remplacer par la somme discrète $\sum_{t=0}^{T}$. Dans tous les cas, le critère de performance G représente la satisfaction cumulée sur la période $[0, T]$. Une **commande** est dite **optimale** si elle maximise la fonction G.

Le système présenté ci-dessus est donc muni du problème d'optimisation implicite :

Maximiser $G(u)$**, sous la contrainte donnée par l'équation d'évolution** (2.2) **(ou** (2.3)**).**

La recherche d'une commande optimale peut être difficile, et les méthodes utilisées exigent des hypothèses de régularité relativement fortes sur les fonctions F, g_f et g. Parmi les techniques de résolution les plus utilisées, on peut citer l'équation de Hamilton-Jacobi-Bellmann et le principe de Pontryagin [85]. Les équations qui interviennent dans ces deux méthodes sont techniques, et n'admettent généralement pas de solutions régulières. De plus, leur application semble très délicate en dimensions élevées [78].

Les systèmes multi-agents étudiés dans cette thèse sont formés d'un grand nombre d'agents et leur description formelle passe typiquement par un espace d'états de dimension élevée. De ce fait, les méthodes de contrôle optimal exposés ici ne semblent pas adéquates

pour leur étude formelle en toute généralité. Il n'est toutefois pas exclu que, lorsqu'un système multi-agent admet une représentation compacte [16], ces méthodes fournissent des réponses qualitatives.

2.2.2.2 La théorie des jeux

La théorie des jeux [112] n'est pas un modèle, ni une famille de modèles spécifiques, mais plutôt un ensemble d'outils pour étudier des situations dans laquelle plusieurs acteurs (des *joueurs*) ont des intérêts qui dépendent des décisions des autres acteurs et de leur décision personnelle. Elle possède divers domaines d'application tels que l'économie, la politique et la psychologie.

L'objectif principal de cette théorie est de trouver des stratégies qui mènent aux situations les plus satisfaisantes pour l'ensemble des joueurs. Ces situations, appelées *équilibres de Nash*, sont telles qu'aucun joueur ne puisse être plus satisfait en changeant seul de stratégie.

Formellement, un *jeu* est défini par
- **un ensemble de joueurs**, noté $\{1, \ldots, N\}$.
- **un ensemble de stratégies individuelles** pour chacun de ces joueurs (la stratégie du joueur n sera notée π_n),
- **un ensemble de fonctions d'utilité individuelles** qui évaluent la satisfaction obtenue par chacun des joueurs lorsque tous les joueurs ont choisi leur stratégie. On notera $g_n(\pi_1, \ldots, \pi_N)$ le gain obtenu par le joueur n lorsque les joueurs choisissent collectivement d'appliquer la stratégie $(\pi_1, \ldots, \pi_N)$.

Un *équilibre de Nash* est alors une stratégie collective $(\pi_1^*, \ldots, \pi_N^*)$ telle que pour joueur n et toute stratégie individuelle π_n de ce joueur

$$g_n(\pi_1^*, \ldots, \pi_n, \ldots, \pi_N^*) \leq g_n(\pi_1^*, \ldots, \pi_n^*, \ldots, \pi_N^*)$$

où il faut comprendre que dans le membre de gauche, la n-ième stratégie a été remplacée par π_n. Ainsi, dans un équilibre de Nash aucun joueur n'a intérêt à dévier unilatéralement de sa stratégie.

Un défi omniprésent dans la théorie des jeux est de trouver et caractériser les équilibres de Nash. Il est montré qu'un grand nombre de jeux possèdent des équilibres de Nash [115]. Par contre, déterminer les équilibres de Nash d'un jeu est généralement un tâche difficile, NP-complète pour de nombreux jeux [115].

Au cours de la conception d'un système multi-agent, la théorie des jeux peut apporter des éléments de réflexion utiles. Tout d'abord, elle apporte une vision formelle et sans ambiguïté sur les mécanismes de raisonnement des agents. Cela permet de définir de manière précise les stratégies et attitudes des agents, en fonction de leurs capacités d'anticipation et de réflexion, mais aussi en fonction de la connaissance qu'ils ont sur les autres agents.

Quelques attitudes simples pour un jeu à deux joueurs sont :
- un joueur qui joue contre un autre joueur débutant, complètement imprévisible, peut opter pour une attitude appelée *raison insuffisante* : il va jouer une stratégie qui per-

16. Par exemple sous forme d'un système dynamique de petite dimension.

met d'obtenir un gain moyen maximal, en considérant que son adversaire choisit sa stratégie de manière aléatoire et uniforme.
- un joueur *rationnel* [17] ne joue que des stratégies contenues dans un équilibre de Nash, i.e. il élimine toutes les stratégies pour lesquelles son adversaire aurait intérêt à choisir une stratégie sous-optimale (au sens de Nash).
- un joueur *super-rationnel* considère que son adversaire raisonne de manière identique à lui-même, et tient compte de ce fait dans le choix de sa stratégie.

Une utilisation intéressante de la théorie des jeux pour la conception d'un système multi-agent est de transformer le problème auquel le concepteur est confronté en un jeu pour lequel des joueurs rationnels vont naturellement choisir une solution.

Plus précisément, le concepteur possède une fonction de récompense globale G qui évalue la réussite du système qu'il cherche à construire. L'idée est de donner un comportement rationnel simple à tous les agents, et de choisir les fonctions de gain g_n individuelles de sorte que les équilibres de Nash du jeu correspondant soient optimales d'un point de vue global, i.e. pour la fonction G. Cette démarche est aussi appelée *mécanisme d'incitation* (ou *reverse game theory*) et a fait l'objet du prix Nobel d'économie en 2007 [18].

La décomposition de la récompense globale en récompenses individuelles

$$G \mapsto (g_1, \ldots, g_N) \tag{2.5}$$

est une étape qui peut être réalisée de plusieurs manières. La plus simple est sans doute de donner la récompense globale à chacun des joueurs, c'est-à-dire $g_1 = g_2 = \ldots = g_N = G$. Le défaut de cette méthode est qu'elle ne localise pas réellement le gain. Chaque joueur doit avoir connaissance de l'objectif global, et coordonner sa stratégie avec celles de tous les autres joueurs.

Une méthode plus intéressante pour réaliser la décomposition (2.5) est de fixer une stratégie « nulle » π_n^0 pour chaque joueur, et de poser

$$g_n(\pi_1, \ldots, \pi_N) = G(\pi_1, \ldots, \pi_n, \ldots, \pi_N) - G(\pi_1, \ldots, \pi_n^0, \ldots, \pi_N).$$

Ce gain individuel g_n peut être vu comme la contribution du joueur n au gain global, et en pratique ce gain ne dépend (généralement) que d'un petit nombre de stratégies. La propriété remarquable de cette décomposition est que toute stratégie réalisant le maximum de la fonction G est un équilibre de Nash pour le jeu obtenu [19], ce qui garantit en partie l'adéquation du jeu avec l'objectif global. Une étudie plus détaillée de cette approche, dans le cadre d'un certaine classe de jeux, peut être trouvée dans [106, 100].

Dans certains systèmes multi-agents, les agents peuvent avoir des objectifs partiellement spécifiés, qui sont ajustés en cours d'exécution en fonction de l'expérience des agents. La vision fixe et complètement spécifiée des gains de la théorie des jeux n'est pas compatible avec cet aspect. De plus, l'ensemble des stratégies possibles est également susceptible d'évoluer en cours d'exécution, ce qui n'est pas le cas dans le cadre des jeux classiques.

17. Au sens économique.
18. `http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2007/press.html`.
19. La réciproque n'est cependant pas vraie.

D'autre part, l'aspect social (la communication) et local des agents n'est pas pris en compte, et il n'est pas certain que ces points soient facilement intégrés dans un jeu classique. En effet, les interactions des agents dans un système multi-agent peuvent avoir une topologie variable, avec des voisinages dynamiques.

Pour finir cette discussion, nous mentionnons que la définition proposée correspond aux *jeux statiques*, qui se terminent directement lorsque chacun des joueurs a choisi de sa stratégie. Cette définition n'intègre pas l'aspect dynamique des systèmes multi-agents étudiés dans cette thèse. Un formalisme qui permet d'intégrer cet aspect est celui des *Processus Décisionnels Markoviens*. Ce modèle est positionné à mi-chemin entre la théorie des jeux et le contrôle optimal. Le chapitre 3 est consacré à ce modèle, et contient une discussion plus approfondie.

2.3 Conclusion

Modéliser un système multi-agent complexe est une tâche difficile qui exige un certain nombre de simplifications. Le modèle choisi doit partager suffisamment de caractéristiques du système étudié. Cette similarité n'est cependant pas le seul critère, si l'on souhaite que le modèle apporte des réponses. Il est également important que l'étude mathématique puisse être menée à terme, et que les résultats qu'elle permet d'établir soient suffisamment précis pour apporter des réponses.

L'exemple des automates cellulaires est peut-être le plus parlant, dans le cas où l'on souhaite modéliser un système d'agents en déplacement sur un quadrillage (comme un ensemble de fourmis de Langton). En procédant comme dans la section 2.1.1, il est possible de traduire les règles d'évolution des agents en termes de règles d'évolution des cellules. L'automate cellulaire obtenu décrit précisément le système étudié, et est capable de mettre des phénomènes complexes en évidence sur des simulations. En revanche, démontrer ces phénomènes complexes est potentiellement très difficile.

Complètement à l'opposé, les systèmes dynamiques utilisés en écologie décrivent les systèmes étudiés parfois avec très peu de précision. Reprenons le système d'équations de Lotka-Volterra : les populations de proies et de prédateurs ne sont représentées que par leurs effectifs, et leur évolution n'est paramétrée que par 4 grandeurs. Il est clair que de fortes simplifications, difficilement justifiables, ont eu lieu. Ce modèle fournit toutefois des réponses satisfaisantes par une analyse mathématique relativement élémentaire, et permet par exemple d'expliquer l'apparition d'oscillations périodiques dans l'évolution des deux populations.

Dans la partie suivante et les deux chapitres qu'elle contient, nous nous concentrons sur deux modèles particuliers : les **Processus Décisionnels Markoviens** et les modèles de type **Champ Moyen**.

Deuxième partie

Étude approfondie de formalismes possibles pour représenter les systèmes multi-agents

3 Les Processus Décisionnels Markoviens

3.1 Introduction

Dans ce chapitre, nous étudions une famille de modèles mathématiques appelés *Processus Décisionnels Markoviens* (en anglais : *Markov Decision Processes* ou *MDP*) et leur utilité pour la représentation des systèmes multi-agents. Grossièrement, un MDP est un processus stochastique soumis à un contrôle, c'est-à-dire un ensemble de paramètres qui gouvernent son évolution. Ce processus est également muni d'une fonction de satisfaction évaluant numériquement les états et valeurs prises au cours du temps. Traditionnellement [123], les paramètres d'un MDP sont appelés *actions*, et la fonction de satisfaction *récompense*. À chaque MDP est associé un problème qui consiste à choisir les actions qui maximisent les récompenses acquises au cours du temps. Cela fixe le contexte : ce formalisme représente de manière formelle un système qui doit prendre une succession de décisions, récompensées ou pénalisées par une certaine fonction.

L'objet de ce chapitre est d'étudier la pertinence de ce formalisme pour l'étude des systèmes multi-agents. La partie 3.2 présenté les MDP classiques. Ensuite, dans la partie 3.3 un exemple concret est modélisé à l'aide d'un MDP classique. La partie 3.4 qui suit présente rapidement quelques techniques de résolution de MDP classiques. Dans la partie 3.5 sont discutées les principales adaptations des MDP aux SMA. Ces adaptations sont mises en oeuvre à l'aide d'un exemple concret dans la partie 3.6. La partie 3.7 discute ensuite quelques techniques de résolution de ces MDP multi-agents.

3.2 Les Processus Décisionnels Markoviens

Dans cette partie sont définis les MDP classiques [123]. À ce stade on n'essaie pas encore de décrire des systèmes et on ne parle donc que de variable d'état, sans préciser s'il s'agit d'un système ou d'un agent.

3.2.1 Formalisme

Dans ce premier paragraphe, nous dressons la liste des ingrédients qui définissent un MDP.

Instants Les instants auxquels un changement d'état peut avoir lieu sont supposés discrets, et fixés à l'avance. On les notera $0, 1, \ldots$ dans l'ordre chronologique et on désignera par t un instant générique.

États L'espace d'états est noté $\mathcal{S}$. On le suppose fini et invariant au cours du temps. On notera $S := Card(\mathcal{S})$ son cardinal, et s_t l'état à l'instant t.

Actions À chaque instant, un certain nombre d'actions est disponible. On notera $\mathcal{A}$ cet ensemble d'actions, qui est supposé fini et invariant au cours du temps. On notera $A := Card(\mathcal{A})$ son cardinal, et a_t l'action à l'instant t.

Loi d'évolution À chaque instant, la variable d'état passe dans un nouvel état selon une loi de probabilités qui dépend de l'état actuel, ainsi que de l'action choisie. La probabilité de passer de l'état s à l'état s' en choisissant l'action a est notée $p_{s,s'}(a)$. La matrice de transition $P(a) = (p_{s,s'}(a))_{s,s'}$ est une matrice stochastique paramétrée par l'action a.

Récompense À chaque instant, l'état actuel et l'action choisie sont récompensés à l'aide d'une fonction $r : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$. Le réel $r(s, a)$ représente le gain enregistré pour le choix de l'action a dans l'état s.

Un MDP est défini par l'ensemble de ces ingrédients, soit un quadruplet (S, A, P, r) où P désigne la fonction $a \mapsto P(a)$. Le schéma suivant montre son l'évolution.

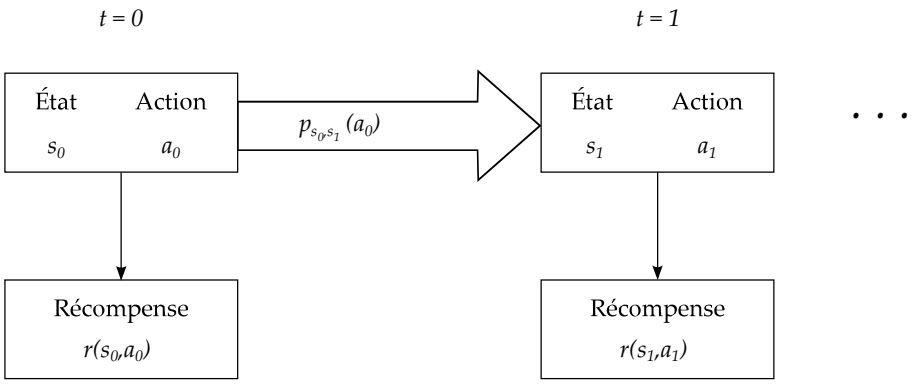


Figure 3.1 — Évolution d'un MDP avec une suite d'actions $(a_0, a_1, \ldots)$ fixée

À $t = 0$ la variable est dans un état initial s_0 et "choisit" l'action a_0. L'état s_0 et l'action a_0 génèrent une récompense $r(s_0, a_0)$. Ensuite, la variable passe à l'état s_1 selon une distribution de probabilités $p_{s_0,s_1}(a_0)$ dépendant de l'action a_0. Le processus se répète ensuite : on choisit une action - l'état et l'action génèrent une récompense - on obtient un nouvel état avec une distribution de probabilités dépendant de l'action.

Pour l'instant, les actions sont supposées fixées à l'avance. Ainsi, pour une succession d'actions $(a_0, a_1, \ldots)$ donnée, le processus $(s_t)_t$ évolue comme une chaîne de Markov à valeurs dans $\mathcal{S}$. Dans le paragraphe suivant, on prolonge cette définition aux situations où

l'action choisie dépend de l'état courant. On y définit également le problème d'optimisation associé à un MDP.

3.2.2 Règle de décision, stratégie et gain

Un Processus Décisionnel Markovien avec une suite d'actions fixée est une chaîne de Markov munie d'une suite de récompenses $(r(s_t, a_t))_t$. Supposer la suite d'actions $(a_t)_t$ fixée revient à planifier la dynamique du système, et les récompenses obtenues, à l'avance. Lorsqu'on considère un système adaptatif, muni d'une certaine autonomie, il est nécessaire de considérer que ces actions résultent d'un choix. Ainsi les actions ne sont plus des données fixes, mais le fruit d'une décision prise en vue d'obtenir un certain résultat. Prendre une décision revient à définir une fonction

$$\Pi_t : \mathcal{S} \to \mathcal{A}$$

qui indique l'action qui est prise dans chaque état à l'instant t. La fonction Π_t est appelée *règle de décision à l'instant t*. Une suite de règles de décision à chaque instant $\Pi = (\Pi_t)_t$ est appelée *stratégie* (ou *politique*). L'évolution de la variable d'état sous une politique $\Pi = (\Pi_t)_t$ donnée est représentée sur la figure 3.2.

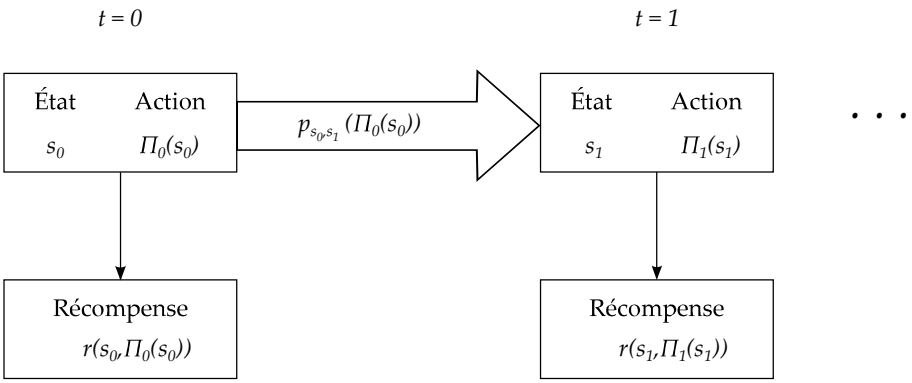


Figure 3.2 — Évolution d'un MDP avec une politique $\Pi = (\Pi_0, \Pi_1, \ldots)$ fixée

Parfois une politique est composée de règles de décision toutes identiques. Dans ce cas, on dira qu'elle est *stationnaire*, et on la notera $\Pi = (\Pi, \Pi, \ldots)$. Cette situation se présente lorsque les décisions sont basées uniquement sur l'observation de l'état courant, et ignorent l'instant auquel on se trouve.

Lorsqu'une politique $\Pi = (\Pi_t)_t$ est donnée, le couple état-action $(s_t, \Pi_t(s_t))_t$ évolue comme une chaîne de Markov, et les récompenses successives sont données par la suite aléatoire $\Big(r\big(s_t, \Pi_t(s_t)\big)\Big)_t$.

On peut alors définir plusieurs types de gains :

– *Le gain cumulé sur une période finie* $[0, T]$:

$$G_\Pi = \mathbb{E}\left(\sum_{t=0}^{T} r\big(s_t, \Pi_t(s_t)\big)\right)$$

où $\mathbb{E}$ désigne l'espérance mathématique sur l'ensemble des états possibles $\{s_0, \ldots, s_T\}$. Ce gain représente simplement la somme des récompenses espérées jusqu'à l'instant T.

– *Le gain moyen sur une période infinie :*

$$G_\Pi = \limsup_{T\to\infty} \mathbb{E}\left(\frac{1}{T}\sum_{t=0}^{T-1} r\big(s_t, \Pi_t(s_t)\big)\right).$$

Celui-ci exprime le gain moyen par instant, sur une longue durée.

– *Le gain dévalué sur une longue période :*

$$G_\Pi = \lim_{T\to\infty} \mathbb{E}\left(\sum_{t=0}^{T} \delta^t r\big(s_t, \Pi_t(s_t)\big)\right)$$

où le facteur $\delta \in]0,1[$ est un taux de dévaluation : une récompense ultérieure a moins de valeur, et est réduite par un facteur δ pour chaque instant de latence.

Remarque 2. *Le gain cumulé sur une période infinie*

$$G_\Pi = \lim_{T\to\infty} \mathbb{E}\left(\sum_{t=0}^{T} r\big(s_t, \Pi_t(s_t)\big)\right)$$

ne figure pas dans cette liste de gains. La raison est qu'il se peut que cette somme ne converge pas. On peut toutefois trouver des résultats sur ce gain dans [123].

3.2.3 L'objectif : recherche d'une stratégie optimale

Une fois que nous avons choisi le gain qui exprime au mieux la réussite dans le problème étudié, nous pouvons simplement exprimer la recherche d'une stratégie optimale comme le problème d'optimisation suivant : chercher les stratégies Π rendant maximal le gain G_Π. On notera :

$$G^* = \sup_\Pi G_\Pi \tag{3.1}$$

le gain optimal, où Π parcourt l'ensemble des stratégies. Dans de nombreux cas qui nous intéressent, l'ensemble des stratégies optimales est non vide. Un abus courant consiste à noter Π^* l'une de ces stratégies, sans préciser laquelle.

L'équation (3.1) est un problème d'optimisation, dont la fonction objectif est le gain G_Π et la variable est la stratégie Π. Il n'est cependant pas facile de le traiter en tant que tel, car :

– l'espace de recherche est l'ensemble des stratégies, qui peut être immense.
– étant donnée une stratégie Π, le gain associé G_Π peut être difficile à évaluer.

Ces difficultés sont essentiellement calculatoires.

3.3 Un exemple simple résolu «à la main»

Pour illustrer l'expressivité du modèle MDP introduit, nous présentons un exemple simple qui est décliné au fur et à mesure de ce chapitre. On considère deux pièces reliées

par un couloir. La première pièce contient un groupe de personnes qui souhaite se rendre dans la deuxième pièce, mais le couloir a une certaine capacité qui ne permet pas à tous de passer simultanément.

Dans cette section, nous inscrivons ce problème dans le cadre des MDP classiques afin de déterminer, par le calcul, l'attitude optimale des personnes. Les résultats des calculs sont confirmés par des simulations numériques.

États Sous l'hypothèse où les personnes sont parfaitement identiques, et où le nombre total de personnes N est connu, on peut complètement décrire le système par le nombre de personnes dans la première pièce. Ce nombre définit donc l'état du système $s_t \in \{0, \ldots, N\}$.

Actions À chaque instant les personnes de la première pièce peuvent essayer de traverser, ou ne rien faire. Les personnes qui sont parvenues à traverser le couloir n'ont plus aucune activité. Les actions possibles pour chaque individu sont donc : essayer de traverser le couloir et ne rien faire. Comme les personnes sont identiques, l'action du système a_t correspond au nombre de personnes qui essayent de traverser le couloir.

Loi d'évolution La capacité du couloir est un entier naturel $c > 0$. Lorsque le nombre de personnes qui essayent de traverser le couloir est inférieur à cette capacité, toutes parviennent à traverser. Dans le cas contraire, aucune personne ne traverse et toutes restent dans la première pièce.

La loi d'évolution du système découle de cette règle. D'un instant à un autre, le nombre de personnes dans la première pièce diminue de k si k personnes ont demandé à traverser, et si $k \leq c$. Dans le cas contraire le nombre de personnes restantes est inchangé.

Récompense et gain L'objectif étant de traverser le couloir, nous définissons la récompense r comme le nombre de personnes qui y parviennent à chaque instant :

$$r(s_t, a_t) = \begin{cases} a_t & \text{si } a_t \leq c \\ 0 & \text{sinon} \end{cases}$$

Avec cette récompense, le gain cumulé sur une période finie $[0, T]$ est le nombre total de personnes dans la seconde pièce :

$$\sum_{t=0}^{T} r(s_t, a_t) = N - s_T$$

Stratégies Il est nécessaire de prendre une précaution afin de définir l'espace des stratégies. En effet, les personnes sont identiques et observent toutes l'état complet du système s_t. Si elles décident de traverser ou non le couloir à l'aide d'une règle commune déterministe, alors il n'y a que deux cas de figure : toutes essayent de traverser ou personne ne le fait.

Pour cette raison, nous nous intéressons aux stratégies individuelles aléatoires : chaque individu observe le nombre de personnes restantes s_t, et demande à traverser le couloir avec

une probabilité dépendant de cette information. Ainsi, chaque individu choisit de traverser selon une loi de type Bernoulli avec un paramètre $\pi(s_t)$ dépendant du nombre de personnes restantes. Les personnes prennent cette décision de manière identique et indépendante, ce qui entraîne que l'action collective est une loi binomiale de paramètres $(s_t, \pi(s_t))$.

Lorsque toutes les personnes utilisent une stratégie aléatoire et stationnaire[1] $\pi : s_t \mapsto \pi(s_t)$, la loi d'évolution s'écrit

$$p_{s,s'}(\pi(s)) = \begin{cases} 0 & \text{si } s < s' \text{ ou } s' < s - c \\ \mathbb{P}\Big(\mathcal{B}(s, \pi(s)) = s - s'\Big) & \text{si } s - c \leq s' < s \\ 1 - \sum_{z=1}^{c} \mathbb{P}\Big(\mathcal{B}(s, \pi(s)) = z\Big) & \text{si } s = s' \end{cases} \tag{3.2}$$

où $\mathcal{B}(s, \pi(s))$ désigne une variable aléatoire qui suit une loi binomiale avec les paramètres indiqués.

Dans cette formule :
- s désigne le nombre de personnes restantes avant la transition,
- s' un nombre de personnes possible après transition,
- $p_{s,s'}(\pi(s))$ est la probabilité de passer de s personnes restantes à s' personnes restantes, sous la stratégie π,
- c désigne la capacité du couloir.

La sollicitation totale du couloir est aléatoire, comme chaque personne restante demande à traverser avec probabilité $\pi(s)$, et le nombre de personnes restantes évolue donc selon une certaine loi de probabilités. La formule (3.2) évalue précisément les probabilités des transitions possibles.

La matrice de transition du système est $P = \Big(p_{s,s'}(\pi(s))\Big)_{s,s'}$.

Le problème d'optimisation Pour une stratégie stationnaire π donnée, le gain cumulé sur une période finie $[0, T]$ s'écrit :

$$G_\pi = \mathbb{E}\left(\sum_{t=0}^{T} R(s_t, a_t)\right) = N - \mathbb{E}\left(s_T\right) = N - \sum_{k=1}^{N} k \cdot \mathbb{P}(s_T = k)$$

où G_π désigne le gain total associé à la stratégie π. Or, d'après la propriété de Markov, la distribution de probabilités de l'état du système s_T à l'instant T est donnée par

$$\Big(\mathbb{P}(s_T = 0), \ldots, \mathbb{P}(s_T = N)\Big) = (0, \ldots, 0, 1) \cdot P^T,$$

où nous avons utilisé le fait qu'initialement tous les individus se trouvent dans la première pièce. Le gain associé à la stratégie π s'écrit donc

$$G_\pi = N - (0, \ldots, 0, 1) \cdot P^T \cdot \begin{pmatrix} 0 \\ 1 \\ \vdots \\ N \end{pmatrix} \tag{3.3}$$

1. Indépendante de l'instant où le système se trouve.

On observe que cette dernière expression est une fonction polynomiale de $(\pi(0), \pi(1), \ldots, \pi(N))$. La recherche du gain optimal revient à maximiser cette fonction, sous les contraintes $0 \leq \pi(s) \leq 1, 1 \leq s \leq N$.

Résolution de quelques cas particuliers Dans ce paragraphe, nous proposons de calculer le gain optimal et la stratégie optimale dans deux cas particuliers, à savoir le cas où la capacité du couloir est limitée à une personne ($c = 1$) et le cas où cette capacité est de deux personnes ($c = 2$).

Dans le cas où la capacité du couloir est limitée à une seule personne, c'est-à-dire $c = 1$, la matrice de transition P s'écrit

$$P = \begin{pmatrix} 1 & 0 & 0 \\ \pi(1) & 1-\pi(1) & 0 \\ 0 & 2\pi(2)\,(1-\pi(2)) & 1-2\pi(2)\,(1-\pi(2)) \\ 0 & & \ddots \\ \vdots & & \end{pmatrix}.$$

et la récompense moyenne instantanée, c'est-à-dire le nombre moyen de personnes qui traverse pour la stratégie π, est donnée par

$$r_1(s, \pi(s)) = s\pi(s)(1-\pi(s))^{s-1},\ s \geq 1. \tag{3.4}$$

La stratégie optimale est obtenue en maximisant toutes ces récompenses instantanées.

Remarque 3. *Nous souhaitons que la chaîne de Markov (s_t) converge le plus rapidement vers la mesure stationnaire $(1, 0, \ldots)$. Pour ce faire on peut essayer de minimiser le gap spectral de la matrice de transition, c'est-à-dire l'écart qui sépare les deux plus grandes valeurs propres. En lisant les valeurs propres sur la diagonale, on constate que ce gap spectral vaut :*

$$\max\left(\pi(1)\,,\, 2\pi(2)\,(1-\pi(2))\,, \ldots, N\pi(N)(1-\pi(N))^{N-1}\right).$$

Pour cet exemple, maximiser la plus grande récompense est donc équivalent à maximiser le gap spectral.

On note $\pi^*(s)$ l'argument p qui maximise la fonction $(s, p) \mapsto r_1(s, p)$ et on vérifie que $\pi^*(s) = 1/s$. Pour $c = 1$ la stratégie optimale consiste donc à solliciter le couloir avec probabilité $1/s$, où s est le nombre de personnes restantes, ce qui est conforme à l'intuition. La figure 3.3 montre l'évolution du nombre de personnes restantes avec cette stratégie, partant d'un effectif initial de $s_0 = N = 100$.

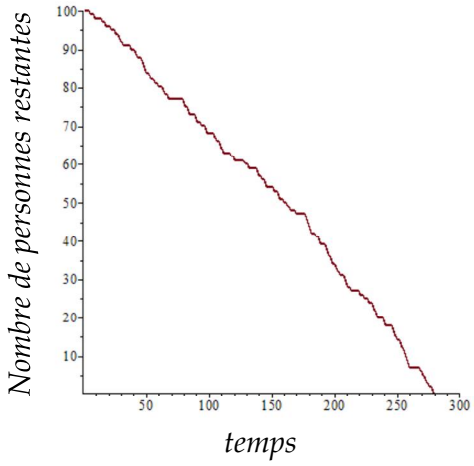


Figure 3.3 — Nombre de personnes restantes en fonction du temps (sous la stratégie optimale)

On constate que le nombre de personnes décroît de manière quasi-linéaire vers 0. L'évolution des probabilités $\pi^*(s_t)$ est représentée sur la figure suivante :

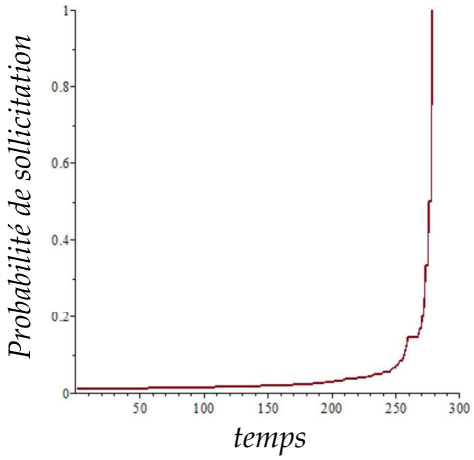


Figure 3.4 — Évolution de la probabilité avec laquelle les personnes essayent de traverser (sous la stratégie optimale)

À mesure que la pièce se vide la probabilité avec laquelle les personnes restantes s'engagent augmente. Cette évolution est conforme à l'intuition : lorsqu'il reste moins de personnes, elles doivent se montrer plus audacieuses, car les conflits (sollicitations multiples) sont plus rares.

Dans le cas où la capacité du couloir est limitée à deux personnes ($c = 2$), la récompense moyenne instantanée vaut

$$r_2(s, \pi(s)) = s\pi(s)(1 - \pi(s))^{s-1},\ s \geq 1.$$

Cette récompense correspond au nombre moyen de personnes qui parvient à traverser lorsqu'il reste s personnes dans la pièce initiale, et que ces personnes choisissent de s'engager

avec probabilité $\pi(s)$. La stratégie optimale est obtenue en maximisant cette récompense instantanée vis-à-vis de π.

Nous maximisons cette fonction en utilisant un argument standard d'analyse, qui consiste à étudier les dérivées partielles. Un calcul montre que ces dérivées partielles valent :

$$\frac{\partial r_2}{\partial p}(s,p) = s \cdot (1-p)^{s-3} \left[(-s^2+2s)p^2 + (s-3)p + 1\right], \; s \geq 3.$$

En étudiant le signe de cette expression, on trouve l'argument qui maximise la fonction partielle $p \mapsto r_2(s,p)$. On vérifie que cet argument est donné par :

$$\pi^* : s \mapsto \begin{cases} 1 & \text{si } s \in \{0,1,2\} \\ \frac{-(s-3)-\sqrt{(s-3)^2+4(s-1)^2-4}}{2(-s^2+2s)} & \text{sinon} \end{cases}.$$

La fonction π^* définit donc la stratégie optimale, c'est-à-dire la probabilité optimale avec laquelle il faut solliciter le couloir lorsqu'il reste s personnes, et lorsque la capacité du couloir est de 2 personnes.

La figure 3.5 montre l'évolution du nombre de personnes restantes et de la probabilité $\pi^*(s_t)$, partant d'un effectif initial de $s_0 = N = 100$.

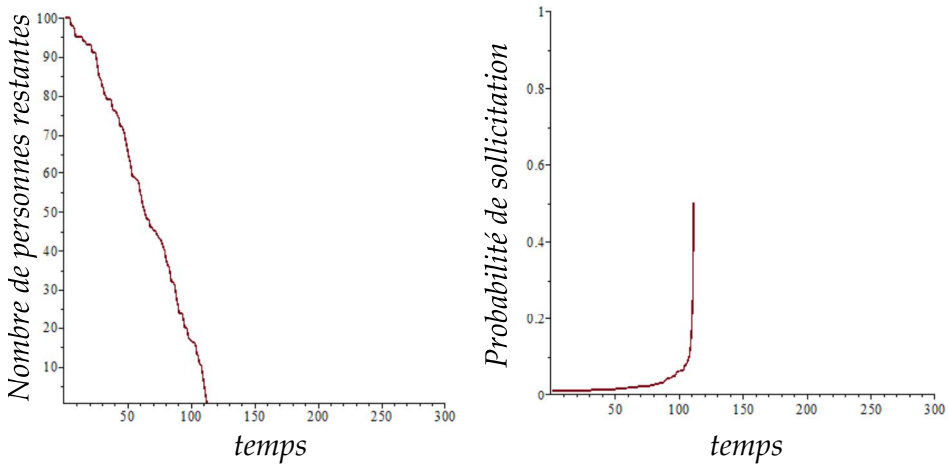


Figure 3.5 — Nombre de personnes restantes en fonction du temps (à gauche) et de la probabilité avec laquelle elles s'engagent (à droite)

Ces courbes sont similaires aux précédentes : le nombre de personnes décroît plus ou moins linéairement vers 0, tandis que la probabilité avec laquelle elles s'engagent croît à mesure que la pièce se vide.

De manière générale, pour c quelconque, la récompense moyenne instantanée est donnée par :

$$r_c(s,\pi(s)) = \sum_{k=1}^{c} k \begin{pmatrix} s \\ k \end{pmatrix} \pi(s)^k (1-\pi(s))^{s-k}$$

On note $\pi^*(s)$ l'argument p qui minimise la fonction $(s,p) \mapsto r_c(s,p)$. La stratégie optimale consiste à choisir la probabilité $\pi^*(s)$ lorsqu'il reste s personnes dans la pièce.

Conclusion Cet exemple montre comment une situation simple s'étudie à l'aide d'un MDP. Le résultat final exprime le gain associé à une politique comme une fonction polynomiale, qu'il convient de maximiser pour résoudre le MDP.

L'approche utilisée a cependant ses limites. Tout d'abord, nous avons complètement évalué la dynamique du système dans (3.3). Les algorithmes présentés dans la partie suivante permettent de se passer de cette évaluation complète.

Ensuite, la difficulté de résolution de notre exemple dépend du degré du gain (qui est une fonction polynomiale), et semble donc augmenter avec la durée T de problème. Or, dans la partie 3.4.1 qui suit, nous verrons que la durée d'un MDP sur une période finie n'a pas réellement d'importance pour sa résolution.

D'autre part, la description du système considéré a pu être grandement simplifiée grâce à l'échangeabilité des agents, et le fait qu'ils n'occupent que deux états. Dans la partie 3.5.2 nous verrons que sans ces deux hypothèses, la difficulté de résolution d'un MDP multi-agents croît de manière exponentielle avec sa taille.

Finalement, la maximisation de (3.3) fait intervenir de manière explicite la capacité c du couloir. Dans un contexte multi-agent, il est possible que les agents n'aient pas cette information.

La résolution du MDP proposé pour l'exemple est un peu lourde, en raison des observations faites ci-dessus. Il existe des méthodes de résolution génériques plus performantes pour les MDP dont quelques-unes, classiques, sont présentées dans la partie suivante.

3.4 Méthodes de résolution classiques d'un MDP

Avec les notations introduites dans la section 3.2.1, on s'intéresse au problème de trouver

$$G^* = \sup_{\Pi} G_{\Pi} \tag{3.5}$$

où G_Π est le *gain* associé à la politique Π. Comme nous l'avons vu, il est possible de choisir différents types de gains, et les méthodes de résolution correspondantes sont naturellement différentes. Cette partie est consacrée à la présentation des techniques de résolution classiques suivantes : *Backward Induction*, *Value Iteration*, *Policy Iteration* et la *programmation linéaire* (pour les MDP).

3.4.1 Backward Induction

Nous commençons par la méthode la plus simple. Elle s'applique au gain cumulé sur une période finie $[0, T]$

$$G_\Pi = \mathbb{E}\left(\sum_{t=0}^{T} r\big(s_t, \Pi_t(s_t)\big)\right).$$

L'idée de cette méthode est élémentaire, et donnera un algorithme clé en main pour l'appliquer. Dans un premier temps, le maximum (3.5) est exprimé en termes d'actions

$$G^* = \max_{a_0, a_1, \ldots a_T} \mathbb{E}\left(\sum_{t=0}^{T} r\big(s_t, a_t\big)\right). \tag{3.6}$$

où s_t évolue selon la dynamique représentée sur la figure 3.1. L'équation (3.6) est ensuite réécrite sous la forme

$$G^* = \max_{a_0} \mathbb{E}\left[r(s_0, a_0) + \max_{a_1} \mathbb{E}\left[r(s_1, a_1) + \ldots + \max_{a_T} \mathbb{E}\left[r(s_T, a_T)\right]\right]\right] \tag{3.7}$$

moyennant une légère imprécision sur le sens de ces espérances. La t-ième espérance de cette expression porte sur les états $\{s_t, s_{t+1}, \ldots, s_T\}$. Cette réécriture vient du fait que l'action a_t, prise à l'instant t, n'a d'impact que sur les états ultérieurs $s_{t+1}, s_{t+2}, \ldots$ et les récompenses ultérieures $r(s_t, a_t), r(s_{t+1}, a_{t+1}), \ldots$

L'idée de cette méthode est donc d'effectuer ces calculs de maximum dans l'ordre inverse, en commençant par $\max_{a_T} \mathbb{E}\left[r(s_T, a_T)\right]$, qui se calcule de manière isolée. En effet, l'état s_T ne dépend pas de l'action a_T, et on peut donc définir la fonction :

$$V_T(s) = \max_{a_T} \mathbb{E}\left[r(s_T, a_T) \mid s_T = s\right] = \max_{a_T} r(s, a_T).$$

qui correspond au meilleur gain qu'on peut obtenir à l'instant final.

Cette fonction permet ensuite de calculer :

$$\begin{aligned} V_{T-1}(s) = \max_{a_{T-1}} \mathbb{E}\left[r(s_{T-1}, a_{T-1}) + \max_{a_T} \mathbb{E}\left[r(s_T, a_T)\right] \mid s_{T-1} = s\right] \\ = \max_{a_{T-1}} \left[r(s, a_{T-1}) + \sum_{s' \in \mathcal{S}} p_{s,s'}(a_{T-1})\, V_T(s')\right] \end{aligned}$$

qui ne dépend que de a_{T-1}. La fonction $V_{T-1}(s)$ correspond au meilleur gain moyen qu'on peut obtenir en partant de l'état s à l'instant $T-1$.

On réitère ensuite cette opération jusqu'à obtenir l'expression complète (3.7). Au cours de ces itérations, il faut définir les fonctions V_t par

$$V_t(s) = \max_{a \in \mathcal{A}} \left(r(s, a) + \sum_{s' \in \mathcal{S}} p_{s,s'}(a) V_{t+1}(s')\right),\ 0 \leq t \leq T. \tag{3.8}$$

La valeur $V_t(s)$ est appelée *Valeur* (en anglais *Value*) et représente le meilleur gain moyen que l'on peut obtenir en partant de l'état s à l'instant t.

L'algorithme suivant correspond à l'exécution de ces différents calculs :

Algorithme 3.1 — Backward Induction

1 **pour** $s \in \mathcal{S}$ **faire**
2 $\quad V_{T+1}(s) \leftarrow 0$
3 **fin**
4 **pour** $t = T$ ***à*** $t = 0$ **faire**
5 $\quad$ **pour** $s \in \mathcal{S}$ **faire**
6 $\quad\quad V_t(s) \leftarrow \max_{a \in \mathcal{A}} \left(r(s,a) + \sum_{s' \in \mathcal{S}} p_{s,s'}(a) V_{t+1}(s') \right)$
7 $\quad\quad a_t^*(s) \leftarrow \operatorname{argmax}_{a \in \mathcal{A}} \left(r(s,a) + \sum_{s' \in \mathcal{S}} p_{s,s'}(a) V_{t+1}(s') \right)$
8 $\quad$ **fin**
9 **fin**
10 **retourner** $a_t^*(s) \quad s \in \mathcal{S}, 0 \leq t \leq T$

L'opérateur $\operatorname{argmax}_{a \in \mathcal{A}}$ désigne un argument a réalisant le maximum de l'expression considérée. Lorsque plusieurs actions réalisent ce maximum, l'opérateur $\operatorname{argmax}_{a \in \mathcal{A}}$ est mal défini. Une manière de remédier à ce problème est de supposer les actions ordonnées, et de définir l'opérateur $\operatorname{argmax}_{a \in \mathcal{A}}$ comme *la première action* qui réalise le maximum recherché.

Remarque 4. *Malheureusement, pour un système multi-agent il n'est pas toujours possible d'établir un ordre sur les actions qui soit le même par tous les agents. En effet, cela exige que les agents soient capables de parvenir à un consensus sur l'ordre choisi. Plus généralement, lorsque plusieurs actions sont optimales dans un état donné, la sélection collective de l'une de ces actions par les agents peut être problématique. Nous reviendrons sur cette remarque dans la section 3.5.2 consacrée aux MDP multi-agents.*

L'algorithme 3.1 retourne une suite d'actions optimales $(a_t^*(s))_{s,t}$. Il suffit alors de définir les règles de décision

$$\Pi_t : s \mapsto a_t^*(s). \tag{3.9}$$

La stratégie $\Pi = (\Pi_t)_{0 \leq t \leq T}$ est optimale au sens de (3.5). De plus, le gain maximal est donné par la valeur $V_0(s_0)$ où s_0 est l'état initial. Ce gain est calculé au cours de l'algorithme et peut être extrait en cours d'exécution pour être donné en sortie.

Avant d'appliquer la *Backward Induction* à l'exemple nous donnons sa complexité algorithmique, qui montre les limites d'applicabilité de la méthode.

Proposition 1 (Complexité de la Backward Induction)**.** *L'algorithme 3.1 s'exécute en temps* $O(T\ A\ S^2)$.

Démonstration

Les deux calculs de la ligne 6 peuvent être résumés à un seul, en conservant simplement l'argument qui réalise le maximum. Le calcul $V_t(s) \leftarrow \max_{a \in \mathcal{A}} \left(r(s,a) + \sum_{s' \in \mathcal{S}} p_{s,s'}(a) V_{t+1}(s') \right)$ nécessite le parcours de l'ensemble des états (pour la somme) et des actions (pour l'opérateur max), soit AS opérations. Ce calcul est répété $(T+1)S$ fois, d'où la complexité énoncée. □

Application à l'exemple L'exemple du couloir (partie 3.3) se place dans un contexte un peu différent puisque la stratégie est aléatoire. On peut toutefois utiliser une méthode de type Backward Induction pour retrouver les résultats précédents. On se place dans le cas $c = 1$. Le gain qu'on souhaite optimiser est :

$$G_\Pi = \mathbb{E}\left(\sum_{t=0}^{T} r(s_t, a_t)\right) \tag{3.10}$$

où a_t suit une loi binomiale $\mathcal{B}(s_t, \pi_t(s_t))$. Pour l'instant, nous cherchons une stratégie non stationnaire $\Pi = (\Pi_1, \ldots, \Pi_T)$. Les calculs qui suivent montrent qu'en réalité la stratégie $\Pi_t(s) = 1/s$ trouvée précédemment est optimale et indépendante du temps.

En procédant comme dans (3.7) on trouve l'équation d'optimalité :

$$G^* = \max_{\Pi_0} \mathbb{E}\left[r(s_0, a_0) + \max_{\Pi_1} \mathbb{E}\left[r(s_1, a_1) + \ldots + \max_{\Pi_T} \mathbb{E}\left[r(s_T, a_T)\right]\right]\right]$$

où les différentes espérances sont prises sur les états *et les actions* ultérieurs [2]. La dernière espérance peut toujours se calculer indépendamment des autres. Notons :

$$V_T(s) = \max_{\Pi_T} \mathbb{E}\left[r(s_T, a_T) \mid s_T = s\right] = \max_{\Pi_T}\left[s\,\Pi_T(s)\,(1 - \Pi_T(s))^{s-1}\right].$$

Nous avons déjà vu que $V_T(s) = (1 - \frac{1}{s})^{s-1}$ et que la règle de décision optimale est $\Pi_T^*(s) = 1/s$. Ensuite, on définit :

$$\begin{aligned} V_{T-1}(s) &= \max_{\Pi_{T-1}} \mathbb{E}\left[r(s_{T-1}, a_{T-1}) + \max_{\Pi_T} \mathbb{E}\left[r(s_T, a_T)\right] \mid s_{T-1} = s\right] \\ &= \max_{\Pi_{T-1}} \mathbb{E}\left[r(s, a_{T-1}) + \sum_{s' \in \mathcal{S}} p_{s,s'}(a_{T-1})\, V_T(s')\right] \end{aligned}$$

où l'espérance est prise sur toutes les actions $\{a_{T-1}\}$ possibles. Comme il n'y a que deux transitions possibles $s \to s$ et $s \to s-1$, on voit que si $s \geq 1$:

$$\begin{aligned} V_{T-1}(s) = \max_{\Pi_{T-1}} \Big[& s\,\Pi_{T-1}(s)\,(1 - \Pi_{T-1}(s))^{s-1} + s\,\Pi_{T-1}(s)\,(1 - \Pi_{T-1}(s))^{s-1}\, V_T(s-1) \\ & + \left(1 - s\,\Pi_{T-1}(s)\,(1 - \Pi_{T-1}(s))^{s-1}\right) V_T(s)\Big] \end{aligned}$$

soit

$$V_{T-1}(s) = \max_{\Pi_{T-1}}\left[s\,\Pi_{T-1}(s)\,(1 - \Pi_{T-1}(s))^{s-1}\right](1 + V_T(s-1) - V_T(s)) + V_T(s).$$

Comme $0 \leq V_T(s) \leq 1$, le facteur $1 + V_T(s-1) - V_T(s)$ est positif, et la règle de décision optimale à $T-1$ est $\Pi_{T-1}^*(s) = \frac{1}{s} = \Pi_T^*(s)$. De plus

$$V_{T-1}(s) = V_T(s)\,(1 + V_T(s-1) - V_T(s)) + V_T(s) \leq 2$$

car $s \mapsto V_T(s)$ est décroissante. On peut donc poursuivre le raisonnement, et montrer que toutes les règles de décision optimales sont identiques

$$\Pi_1(s) = \Pi_2(s) = \ldots = \Pi_T(s) = \frac{1}{s}.$$

2. On rappelle que ces actions dépendent de manière aléatoire des états.

Ce résultat coïncide par ailleurs avec celui trouvé à la partie 3.3 qui indique que, lorsque la capacité du couloir est limitée à une personne, la stratégie optimale consiste à s'engager avec probabilité $\frac{1}{s}$ où s est le nombre de personnes restantes.

À présent, nous passons à d'autres techniques classiques de résolution de MDP.

3.4.2 Value Iteration

Nous prolongeons en partie les principes de la *Backward Induction* au cas du gain dévalué, mais en se limitant aux stratégies stationnaires (indépendantes du temps) :

$$G_\Pi = \lim_{T\to\infty} \mathbb{E}\left(\sum_{t=0}^{T} \delta^t\, r(s_t, \Pi(s_t))\right)$$

où $\delta \in]0,1[$ est le facteur de dévaluation des récompenses. Le problème de la recherche d'une stratégie optimale est formulé par (Cf. (3.5))

$$G^* = \max_{\Pi\in\mathcal{A}^{\mathcal{S}}} G_\Pi$$

où Π parcourt l'ensemble des règles de décision. Cette équation se réécrit

$$G^* = \max_{a_t} \lim_{T\to\infty} \mathbb{E}\left(\sum_{t=0}^{T} \delta^t\, r(s_t, a_t)\right) = \max_{a_t} \mathbb{E}\left(\sum_{t=0}^{\infty} \delta^t\, r(s_t, a_t)\right).$$

De manière heuristique, sans être trop précis sur le sens des différentes espérances, on peut écrire

$$G^* = \max_{a_0} \mathbb{E}\left[r(s_0, a_0) + \delta \max_{a_1} \mathbb{E}\left[r(s_1, a_1) + \delta \max_{a_2} \mathbb{E}\left[r(s_2, a_2) + \dots \quad \right]\right]\right].$$

Comme nous cherchons une stratégie stationnaire, les différents maximums de cette expression sont calculés de manière similaire. En s'inspirant de la Backward Induction, on introduit les équations

$$V(s) = \max_{a\in\mathcal{A}} \left[r(s,a) + \delta \sum_{s'\in\mathcal{S}} p_{s,s'}(a)\, V(s')\right],\ s \in \mathcal{S} \tag{3.11}$$

appelées *équations d'optimalité* ou *équations de Bellman*. Elles définissent implicitement une fonction $s \mapsto V(s)$, dont la valeur $V(s)$ représente le meilleur gain que l'on peut obtenir en partant de l'état s. De la même manière, les actions $a(s)$ réalisant ces maximums définissent la règle de décision optimale.

Il n'est pas évident que l'équation (3.11) admette une solution, mais le théorème suivant répond positivement à cette question :

Théorème 1. *Il existe une unique fonction $V : \mathcal{S} \mapsto \mathbb{R}$ solution de l'équation* (3.11).

Démonstration

On considère *l'opérateur de Bellman*

$$\begin{array}{rccc} B: & \mathbb{R}^{\mathcal{S}} & \to & \mathbb{R}^{\mathcal{S}} \\ & V & \mapsto & B(V) \end{array}$$

par

$$\begin{array}{rccl} B(V): & \mathcal{S} & \to & \mathbb{R} \\ & s & \mapsto & \max\limits_{a\in\mathcal{A}} \left(r(s,a) + \delta \sum\limits_{s'\in\mathcal{S}} p_{s,s'}(a) V(s') \right). \end{array}$$

L'existence et l'unicité d'une solution au problème (3.11) équivalent à l'existence d'un unique point fixe pour l'opérateur B. La démonstration consiste à prouver que B est une contraction, et d'appliquer le théorème du point fixe de Banach [26]. Soient V et $\hat{V}$ deux fonctions de $\mathcal{S}$ dans $\mathbb{R}$, et $s \in \mathcal{S}$. Supposons que $B(V)(s) \leq B(\hat{V})(s)$, et notons

$$a^*(s) = \operatorname*{argmax}_{a\in\mathcal{A}} \left(r(s,a) + \delta \sum_{s'\in\mathcal{S}} p_{s,s'}(a) V(s') \right).$$

Alors, par hypothèse

$$0 \leq B(V)(s) - B(\hat{V})(s) = r(s,a^*(s)) + \delta \sum_{s'\in\mathcal{S}} p_{s,s'}(a^*(s)) V(s') - \max_{a\in\mathcal{A}} \left(r(s,a) + \delta \sum_{s'\in\mathcal{S}} p_{s,s'}(a) \hat{V}(s') \right).$$

On en déduit que

$$\begin{aligned} 0 &\leq B(V)(s) - B(\hat{V})(s) \\ &\leq r(s,a^*(s)) + \delta \sum_{s'\in\mathcal{S}} p_{s,s'}(a^*(s)) V(s') - r(s,a^*(s)) - \delta \sum_{s'\in\mathcal{S}} p_{s,s'}(a^*(s)) \hat{V}(s') \\ &\leq \delta \sum_{s'\in\mathcal{S}} p_{s,s'}(a^*(s)) \left(V(s') - \hat{V}(s') \right) \end{aligned}$$

puis, comme $\sum\limits_{s'\in\mathcal{S}} p_{s,s'}(a^*(s)) = 1$,

$$0 \leq B(V)(s) - B(\hat{V})(s) \leq \delta \max_{s'\in\mathcal{S}} |V(s') - \hat{V}(s')|.$$

Lorsque $B(V)(s) \geq B(\hat{V})(s)$ on raisonne de manière similaire, en échangeant les rôles de V et $\hat{V}$, pour obtenir

$$0 \leq B(\hat{V})(s) - B(V)(s) \leq \delta \max_{s'\in\mathcal{S}} |V(s') - \hat{V}(s')|.$$

On conclut que

$$\max_{s\in\mathcal{S}} \left| B(\hat{V})(s) - B(V)(s) \right| \leq \delta \max_{s'\in\mathcal{S}} |V(s') - \hat{V}(s')|,$$

ce qui montre que l'opérateur B est une contraction (de facteur δ) dans l'espace de Banach $\left(\mathbb{R}^{\mathcal{S}}, \|\cdot\|_\infty\right)$. La démonstration est achevée en appliquant le théorème du point fixe de Banach. □

La méthode *Value Iteration* s'inspire directement de cette démonstration et du théorème du point fixe auquel il fait appel. Elle consiste à appliquer de manière répétée l'opérateur B, jusqu'à obtenir un point fixe V^*. Cette fonction permet ensuite de déterminer l'action optimale $a^*(s)$ à partir de

$$a^*(s) = \operatorname*{argmax}_{a\in\mathcal{A}} \left[r(s,a) + \delta \sum_{s'\in\mathcal{S}} p_{s,s'}(a)\, V^*(s') \right],\ s \in \mathcal{S}.$$

La règle de décision optimale est alors donnée par la fonction

$$\Pi : s \mapsto a^*(s).$$

L'algorithme 3.2 décrit l'exécution de k itérations de cette méthode. Le choix de la fonction initiale V_0 est arbitraire, et n'a pas d'incidence sur la limite de la suite $(V_k)_k$.

Algorithme 3.2 — Value Iteration

1 Choisir $V_0 \in \mathbb{R}^{\mathcal{S}}$
2 **pour** $t = 1$ ***à*** k **faire**
3 | **pour** $s \in \mathcal{S}$ **faire**
4 | | $V_t(s) \leftarrow \max_{a \in \mathcal{A}} \left(r(s,a) + \delta \sum_{s' \in \mathcal{S}} p_{s,s'}(a)\, V_{t-1}(s') \right)$
5 | **fin**
6 **fin**
7 **pour** $s \in \mathcal{S}$ **faire**
8 | $a^*(s) \leftarrow \operatorname{argmax}_{a \in \mathcal{A}} \left(r(s,a) + \delta \sum_{s' \in \mathcal{S}} p_{s,s'}(a)\, V_k(s') \right)$
9 **fin**
10 **retourner** $a^*(s),\ s \in \mathcal{S}$

La convergence de cet algorithme est garantie par le théorème du point fixe. Pour limiter le nombre d'itérations, on peut munir cet algorithme d'un critère d'arrêt comme :

$$|V_t(s) - V_{t-1}(s)| \leq \varepsilon,\ s \in \mathcal{S}$$

où ε est un seuil fixé à l'avance.

Remarque 5.
- *Le nom Value Iteration s'explique par le fait que cet algorithme agit itérativement sur la fonction de valeur V_t.*
- *Comme pour la Backward Induction, l'unicité des actions optimales a^* en sortie n'est pas garantie*[3] *(Cf. remarque 4).*

Pour finir, nous donnons la complexité algorithmique de cette méthode.

Proposition 2 (Complexité de la Value Iteration)**.** *L'algorithme 3.2 s'exécute en temps $O(k\,A\,S^2)$.*

Démonstration

À la ligne 4 de l'algorithme on trouve $V_t(s) \leftarrow \max_{a \in \mathcal{A}} \left(r(s,a) + \delta \sum_{s' \in \mathcal{S}} p_{s,s'}(a)\, V_{t-1}(s') \right)$. L'opérateur max et la somme nécessitent le parcours complet de l'ensemble des états et actions, soit AS opérations. Ce calcul est répété kS fois, d'où la proposition énoncée. □

3.4.3 Policy Iteration

Cette méthode classique de résolution d'un MDP est fortement inspirée par la précédente. Nous restons dans le cas dévalué :

$$G_\Pi = \lim_{T \to \infty} \mathbb{E} \left(\sum_{t=0}^{T} \delta^t\, r\big(s_t, \Pi(s_t)\big) \right).$$

3. À cause de l'opérateur *argmax*

Pour les mêmes raisons que celles invoquées dans la partie 3.4.2, on s'intéresse aux équations d'optimalité :

$$V(s) = \max_{a \in \mathcal{A}} \left[r(s,a) + \delta \sum_{s' \in \mathcal{S}} p_{s,s'}(a)\, V(s') \right],\ s \in \mathcal{S}.$$

La *Policy Iteration* consiste à proposer une politique et à l'améliorer au cours des itérations. L'algorithme 3.3 décrit k itérations de cette méthode.

Algorithme 3.3 — Policy Iteration

1 Choisir $a_0 \in \mathcal{A}^{\mathcal{S}}$
2 **pour** $t = 1$ ***à*** k **faire**
3 Résoudre, en $\{V(s)\}_{s \in \mathcal{S}}$,

$$V(s) = r(s, a_{t-1}(s)) + \delta \sum_{s' \in \mathcal{S}} p_{s,s'}(a_{t-1}(s)) V(s'),\ s \in \mathcal{S}$$

4 **pour** $s \in \mathcal{S}$ **faire**
5 $a_t(s) \leftarrow \operatorname{argmax}_{a \in \mathcal{A}} \left(r(s, a_{t-1}(s)) + \delta \sum_{s' \in \mathcal{S}} p_{s,s'}(a_{t-1}(s))\, V(s') \right)$
6 **fin**
7 **fin**
8 **retourner** $a_k(s),\ s \in \mathcal{S}$

Cet algorithme retourne une fonction $s \mapsto a_k(s)$ qui tend vers une stratégie optimale lorsque $k \to \infty$.

Remarque 6.
- *La principale différence avec la Value Iteration réside dans le fait que cette méthode améliore progressivement la stratégie*[4] *(Policy), et non la valeur.*
- *Comme dans la partie précédente, on peut ajouter un critère d'arrêt de la forme*

$$|V_t(s) - V_{t-1}(s)| \leq \varepsilon,\ s \in \mathcal{S}$$

  pour limiter le nombre d'exécutions de l'algorithme. Comme les valeurs sont croissantes au cours de l'exécution [123] on peut montrer qu'elles convergent.
- *La stratégie (Policy) ne converge pas nécessairement, du fait qu'elle n'est pas unique*[5] *(Cf. remarque 4).*
- *Il est montré [156] que le nombre d'itérations est borné par un terme d'ordre* $O(S \times A)$.

On termine cette présentation par la complexité de l'algorithme 3.3.

Proposition 3 (Complexité de la Policy Iteration)**.** *L'algorithme 3.3 s'exécute en temps* $O\left(k(S^3 + A\, S^2)\right)$.

Démonstration

4. Les lignes 4 – 6 sont interprétées comme l'*affinage* de la stratégie.
5. À cause de l'opérateur argmax

La résolution du système linéaire à la ligne 3 s'effectue en $O(S^3)$ opérations . Le calcul de la ligne 5 nécessite la parcours de l'ensemble des actions (pour l'opérateur max) et des états (pour la somme), et est répété S fois. On en déduit la complexité énoncée à la proposition. □

3.4.4 Programmation linéaire

Cette dernière méthode diffère des autres par le fait qu'elle n'est pas réservée aux MDP. Le terme *programmation linéaire* fait référence à une méthode très générale d'optimisation sous contraintes. Nous restons à nouveau dans le cas dévalué et les stratégies stationnaires

$$G_\Pi = \lim_{T\to\infty} \mathbb{E}\left(\sum_{t=0}^{T} \delta^t\, r(s_t, \Pi(s_t))\right).$$

Pour les mêmes raisons que celles invoquées ci-dessus, on s'intéresse à l'équation d'optimalité (Cf. (3.11))

$$V(s) = \max_{a\in\mathcal{A}} \left[r(s,a) + \delta \sum_{s'\in\mathcal{S}} p_{s,s'}(a)\, V(s') \right],\ s \in \mathcal{S}. \tag{3.12}$$

L'idée est de formuler ce problème comme le problème d'optimisation

$$\text{Minimiser} \sum_{s\in\mathcal{S}} V(s) \tag{3.13}$$

sous les contraintes

$$V(s) \geq r(s,a) + \delta \sum_{s'\in\mathcal{S}} p_{s,s'}(a)\, V(s'),\ \forall s \in \mathcal{S}, \forall a \in \mathcal{A}. \tag{3.14}$$

Intuitivement, cette méthode consiste à trouver la plus petite fonction surestimant l'équation d'optimalité (3.12) [6]. On peut montrer que ce problème admet une solution unique ([16], Volume 2, paragraphe 1.3.4), et qu'elle résout le MDP :

Théorème 2. *L'unique solution du problème d'optimisation* (3.13), (3.14) *est la solution de l'équation d'optimalité* (3.12).

Démonstration

Soit V la solution du problème linéaire (3.13), (3.14) et V^* la solution de (3.12). Notons à nouveau B l'opérateur de Bellmann [7]. Sous forme fonctionnelle, (3.13) et (3.14) s'écrivent

$$\text{Minimiser} \sum_{s\in\mathcal{S}} V(s)$$

sous la contrainte

$$V \geq B(V).$$

Comme l'opérateur B est croissant, on a $V \geq B(V) \geq B^2(V) \geq B^3(V) \geq \ldots \geq V^*$ (on rappelle que V^* est le point fixe de l'opérateur B). Par conséquent $\sum_{s\in\mathcal{S}} V(s) \geq \sum_{s\in\mathcal{S}} V^*(s)$, et V^* vérifie trivialement la contrainte (3.14). On conclut par unicité de la solution du problème linéaire que $V^* = V$. □

6. Comme nous avons montré au théorème 1 qu'il existe une solution à cette équation, l'ensemble de fonctions considéré est non vide.

Résolution d'un problème d'optimisation linéaire Il existe plusieurs méthodes de résolution d'un problème d'optimisation linéaire. Une des plus populaires, et sans doute la plus connue, est celle du simplexe. Elle repose sur l'observation suivante : la contrainte linéaire (3.14) définit un polytope convexe (non vide). La fonction objectif $\sum_{s \in \mathcal{S}} V(s)$ est une fonction convexe, et atteint donc son minimum en un des sommets de ce polytope.

L'algorithme classique développé par Dantzig [44] consiste à choisir un sommet initial, puis à parcourir à chaque itération une arête (issue de ce sommet) le long de laquelle la fonction objectif diminue. Le choix de cette arête est appelée *pivotage*, et peut simplement être l'arête de la plus forte diminution. L'algorithme se termine lorsqu'on ne peut plus améliorer la fonction objectif.

Bien que la méthode du simplexe ait fourni des résultats satisfaisants, elle a des limites en termes de temps de calcul. Le problème réside dans le choix de la règle de pivotage. Klee et Minsky ont montré [86] qu'avec un mauvais choix, le nombre d'itérations de l'algorithme peut être exponentiel.

Lorsque toutes les valeurs $p_{s,s'}(a)$ et $r(s,a),\ s,s' \in \mathcal{S}, a \in \mathcal{A}$, sont rationnelles, on peut utiliser l'algorithme décrit dans [84]. Cet algorithme résout le problème d'optimisation linéaire (3.13), (3.14) en temps polynomial, mais peut être très lent en pratique [103].

3.4.5 Bilan

Dans cette partie, nous avons présenté les méthodes de résolution classiques d'un MDP. Les trois premières exploitent fortement le caractère markovien de l'évolution du processus, tandis que la quatrième englobe le problème dans le domaine de l'optimisation linéaire sous contraintes.

Nous mentionnons que ces méthodes fournissent une stratégie optimale *sans mémoire*, c'est-à-dire un ensemble de des règles de décisions qui ne dépendent que de l'état courant. On pourrait considérer les stratégies dont les règles de décision dépendent des actions et états passés [8]. En vérité, dans [123] il est démontré que parmi toutes ces stratégies, celles qui sont sans mémoire sont optimales, ce qui permet de restreindre la recherche d'une stratégie optimale à cette dernière classe.

Les méthodes présentées fournissent une stratégie optimale en un temps (au mieux) polynomial vis-à-vis du nombre d'états et actions. Ces complexités algorithmiques les rendent inadaptées aux grands MDP, et en particulier aux MDP Multi-Agents que nous présentons dans la partie suivante.

3.5 Les Processus Décisionnels Markoviens Multi-Agents

À ce stade nous avons donné la définition formelle d'un MDP classique et passé en revue quelques méthodes de résolution. Cette partie est consacrée à l'utilisation des MDP pour formaliser les systèmes multi-agents.

8. C'est-à-dire un *historique* du passé du système.

3.5.1 Les MMDP

Nous présentons ici la manière la plus naturelle pour représenter un SMA à l'aide d'une MDP [23]. Il s'agit d'une description vectorielle de l'ensemble des états et actions des agents. Pour simplifier, on suppose que les agents partagent le même espace d'états fini $\mathcal{S}_a$ et le même espace d'actions fini $\mathcal{A}_a$. Il sera également supposé que ces agents évoluent de manière synchrone à des instants discrets $1, 2, \ldots$.

États L'état du système à l'instant t est décrit par le vecteur $s_t = \left(s_t^1, s_t^2, \ldots, s_t^N\right)$, où la composante s_t^n représente l'état de l'agent n à l'instant t. La variable d'état s_t prend donc ses valeurs dans l'espace $\mathcal{S} = \mathcal{S}_a^N$.

Actions De la même manière, l'action du système à l'instant t est le vecteur $a_t = \left(a_t^1, a_t^2, \ldots, a_t^N\right)$ qui prend ses valeurs dans l'espace $\mathcal{A} = \mathcal{A}_a^N$. La n-ième coordonnée a_n^t de ce vecteur représente l'action de l'agent n à l'instant t.

Loi d'évolution A priori, la loi d'évolution du système est globale. Cela revient à dire qu'il existe un ensemble de probabilités de transition $\left(p_{s,s'}(a)\right)_{s,s' \in \mathcal{S}}$ qui décrit les changements d'état de l'ensemble des agents, lorsqu'une action collective a est sélectionnée. On notera $P(a) = \left(p_{s,s'}(a)\right)_{s,s' \in \mathcal{S}}$ la matrice de transition correspondante.

Récompense Dans ce contexte la récompense du système est une fonction globale[9] $r : (s, a) \mapsto r(s, a)$.

Le MDP $(\mathcal{S}, \mathcal{A}, P, r)$ formé par ces données, est appelé Multi-Agent MDP (MMDP [23]). Il évolue selon le schéma ci-dessous.

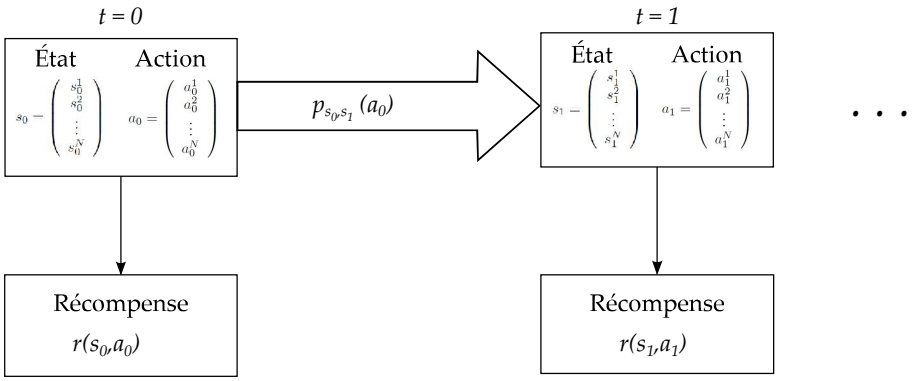


Figure 3.6 — Évolution d'un MMDP avec une suite d'action $(a_0, a_1, \ldots)$

Pour le MDP ainsi défini, une règle de décision $\Pi : s \mapsto a$ est une fonction globale, dans le sens où elle détermine l'action collective en fonction de l'ensemble des états des agents.

9. Par *globale*, nous entendons qu'elle fait intervenir l'état complet du système

La n-ième composante π_n de la fonction Π est la règle de décision individuelle de l'agent n, et dépend a priori de l'état global :

$$\Pi : s \mapsto \Pi(s) = \begin{pmatrix} \pi_1(s) \\ \pi_2(s) \\ \vdots \\ \pi_N(s) \end{pmatrix}.$$

Pour simplifier, on s'intéresse aux stratégies stationnaires $(\Pi, \Pi, \ldots)$, identiques à chaque instant.

Étant donné un gain G_Π choisi parmi les propositions de la partie 3.2.2, le problème d'optimisation

$$\max_{\Pi \in \mathcal{A}^S} G_\Pi \tag{3.15}$$

correspond à la recherche d'une stratégie collective maximisant la réussite globale. Cette formulation présente toutefois quelques difficultés que nous discutons dans la partie 3.5.2.

3.5.2 Discussion

Les techniques de résolution proposées dans la partie 3.4 permettent en théorie de déterminer une stratégie optimale vis-à-vis de l'objectif du système. Dans le cas des systèmes multi-agents, cette approche a cependant quelques limitations, que nous discutons ici.

3.5.2.1 À propos de la récompense globale

Le fait de définir une récompense au niveau global limite l'approche aux systèmes dont les agents poursuivent un but commun, c'est-à-dire les systèmes multi-agents coopératifs. De plus l'ensemble des règles de décision considérées dans le problème (3.15) sont globales, ce qui borne l'application aux systèmes à contrôle centralisé. Ce dernier point est antinomique avec la nature des systèmes multi-agents auxquels s'intéresse cette thèse.

3.5.2.2 À propos de la stratégie optimale

La résolution d'un MDP passe par une équation d'optimalité qui peut avoir plusieurs solutions, qui correspondent à plusieurs comportements optimaux pour les agents. Même lorsque les agents ont connaissance de toutes ces solutions, ils n'ont pas nécessairement les capacités pour tous choisir collectivement la même stratégie optimale. Cette difficulté peut être surmontée en conférant des capacités de communication (dépassant la simple rationalité) aux agents, ou bien en hiérarchisant les différentes stratégies.

3.5.2.3 À propos de la connaissance du système

Par définition, les MDP résolvent des problèmes qui sont fermés, et complètement définis. Certains problèmes rencontrés peuvent avoir des populations d'agents variables et un

ensemble d'états qui peut varier au cours du temps. Il se peut également que les agents aient une connaissance incomplète de l'état du système.

Ces spécifications excluent les méthodes présentées à la section 3.4 qui utilisent de manière cruciale la loi d'évolution du système, ainsi que la récompense du système. Pour appliquer les algorithmes présentés, il faut connaître la fonction de récompense r et les probabilités de transition $p_{s,s'}(a)$. Un explication possible est que les stratégies recherchées sont *sans mémoire* (Cf. section 3.4.5), ce qui induit que la connaissance instantanée de l'état du système doit permettre de choisir une action optimale.

Si la fonction de récompense, les probabilités de transition et l'état du système ne sont que partiellement connus par les agents il est possible que les agents ne soient pas en mesure de déterminer leur action optimale. Afin de pouvoir prendre une décision judicieuse et suffisamment informée dans ce contexte, on peut considérer des stratégies qui utilisent des expériences passées.

Il existe des méthodes d'apprentissage dans lesquelles la récompense et les probabilités de transition sont ignorées par les agents, et apprises à travers leurs valeurs $\{r(s_t, a_t)\}$, $\{p_{s_t,s_{t+1}}(a_t)\}$ prises au cours du temps. On peut consulter [28] et [65] pour ce type de méthodes dans le cadre des MDP.

3.5.2.4 Limites mathématiques

Jusqu'à présent, toutes les objections soulevées concernent la capacité des MDP à représenter correctement les SMA. Même dans le cas où un MMDP représente parfaitement un SMA (connaissance complète, ...) il demeure des difficultés mathématiques, d'ordre calculatoire. La raison essentielle est que les équations d'optimalité deviennent rapidement insolubles.

L'espace d'états du système $\mathcal{S} = \mathcal{S}_a^N$ et l'espace des actions du système $\mathcal{A} = \mathcal{A}_a^N$ ont une taille exponentielle par rapport au nombre d'agents N. La complexité de calcul de la value V et des stratégies optimales croît donc de manière explosive. Reprenons l'équation d'optimalité (3.11) :

$$V(s) = \max_{a \in \mathcal{A}} \left[r(s,a) + \delta \sum_{s' \in \mathcal{S}} p_{s,s'}(a)\, V(s') \right],\ s \in \mathcal{S},$$

il s'agit en réalité d'un système non linéaire formé d'autant d'équations que d'états du système, soit $|\mathcal{S}_a|^N$. Par ailleurs le maximum est réalisé sur l'ensemble des actions, au nombre de $|\mathcal{A}_a|^N$. Les algorithmes de résolution présentés en section 3.4 sont inutilisables à cause de cette croissance exponentielle. Comme leurs complexités sont toutes polynomiales vis-à-vis du nombre d'états et actions du système, elles croissent de manière exponentielle avec le nombre d'agents.

Une dernière objection vient lorsque, en dépit les difficultés invoquées, une stratégie optimale est trouvée. En fixant la stratégie Π, l'état du système $(s_t)_t$ évolue comme une chaîne de Markov. Lorsque l'espace d'états $\mathcal{S} = \mathcal{S}_a^N$ est grand, il peut être difficile de l'étudier formellement.

Ces remarques laissent un peu pessimiste quant à la possibilité de trouver une stratégie optimale pour les agents. Pour l'instant, nous allons les ignorer et présenter quelques variations des MMDP qui permettent d'avoir une représentation plus réaliste des SMA.

3.5.3 MMDP et jeux de Markov

Le formalisme des MMDP est proche de la théorie des jeux, et notamment des jeux de Markov [135]. Un jeu de Markov à N joueurs est défini par un quadruplet $(\mathcal{S}, \mathcal{A}, P, (r_n)_n)$ où

- $\mathcal{S}$ est **l'espace d'états** du jeu. Il peut être *factorisé* : $\mathcal{S} = \times_{n=1}^{N} \mathcal{S}_n$ où $\mathcal{S}_n$ est l'espace d'états du joueur n.
- $\mathcal{A} = \times_{n=1}^{N} \mathcal{A}_n$ est **l'espace des actions** collectives. L'ensemble $\mathcal{A}_n$ est l'espace d'actions de l'agent n.
- P définit les **probabilités de transition** : pour une action collective $a \in \mathcal{A}$, $P(a)$ est la matrice de transition du système sur l'espace $\mathcal{S}$.
- La fonction $r_n : \mathcal{A} \to \mathbb{R}$ représente la **récompense individuelle** du joueur n.

Un jeu de Markov évolue comme un MMDP (figure 3.6), mais la satisfaction d'un état s et une action collective a est représentée par une suite de récompenses individuelles $(r_n(s,a))_n$ attribuées aux différents joueurs. À tout jeu de Markov, on associe un ensemble de problèmes d'optimisation, qui consiste à maximiser les gains individuels (dévalués)

$$\mathbb{E}\left(\sum_{t=0}^{\infty} \delta^t \, r_n(s_t, a_t) \mid s_0 = s\right).$$

Si l'on exprime ce problème en termes de stratégies, on cherche à maximiser

$$V_n^{\Pi}(s) = \mathbb{E}\left(\sum_{t=0}^{\infty} \delta^t \, r_n(s_t, \Pi(s_t)) \mid s_0 = s\right).$$

où Π parcourt l'ensemble des stratégies collectives. Lorsque tous les agents partagent la même récompense r, ce dernier problème équivaut à un MMDP.

Dans le cas général, soit $\Pi = (\pi_1, \ldots, \pi_N)$ une stratégie collective. On note $\Pi^{-n} = (\pi_m)_{m \neq n}$ la stratégie réduite, obtenue en enlevant la stratégie de l'agent n. Une stratégie locale π_n^* est la *meilleure réponse de l'agent n* s'il ne peut pas obtenir un meilleur gain en changeant individuellement de stratégie.

$$V^{(\pi_n^*, \Pi^{-n})} \geq V^{(\pi_n, \Pi^{-n})}, \ \forall \pi_n \in \mathcal{A}_a^{\mathcal{S}}.$$

On appelle *équilibre de Nash* une stratégie $\Pi^* = (\pi_1^*, \ldots, \pi_N^*)$ formée de meilleures réponses.

La définition d'un MMDP en termes de jeu de Markov permet de remplacer la recherche d'une politique collective optimale par la recherche d'un équilibre de Nash. Cette traduction exprime plus précisément des situations où les agents ne sont pas capables de choisir collectivement leur stratégie, et où ils peuvent seulement spéculer sur les stratégies de leurs congénères et répondre de la meilleure manière possible. Le formalisme des jeux de Markov permet donc d'intégrer le problème de la coordination des stratégies.

3.5.4 Les variantes partiellement observées : DEC-POMDP et DEC-MDP

En général, les agents d'un SMA ont une vision partielle et incomplète du système [58]. Les MMDP tels que nous les avons définis à la section 3.5.1 ne permettent pas de prendre en compte cet aspect. Pour cette raison [15] a introduit le modèle DEC-POMDP (*Decentralized Partially Observable Markov Decision Process*) qui permet de distinguer l'état du système et ce que les agents en perçoivent (les observations).

3.5.4.1 Définition formelle

L'essentiel des composants d'un MMDP est conservé :

- On considère toujours une collection de N agents partageant le même **espace d'états** $\mathcal{S}_a$ et d'**actions** $\mathcal{A}_a$, qui évoluent de manière synchrone à des instants discrets notés $1, 2, \ldots$
- Comme dans le cadre des MMDP, l'**espace d'états du système** est le produit cartésien $\mathcal{S} = \mathcal{S}_a^N$, et l'**espace d'actions** est $\mathcal{A} = \mathcal{A}_a^N$.
- La **règle de transition** ne diffère pas de celle d'un MMDP, et est définie comme un ensemble de probabilités de transition $\{p_{s,s'}(a)\}_{s,s' \in \mathcal{S}}$ formant une matrice

$$P(a) = \Big(p_{s,s'}(a)\Big)_{s,s' \in \mathcal{S}}$$

- La **récompense** est une fonction globale $r(a, s)$.

À ces ingrédients, on ajoute une **variable d'observation**, qui est un vecteur

$$o_t = \left(o_t^1, o_t^2, \ldots, o_t^N\right).$$

La n-ième composante o_t^n représente l'**observation de l'agent** n **à l'instant** t.

On peut supposer, pour simplifier, que toutes les observations des agents vivent dans un même espace fini $\mathcal{O}_a$, et noter $\mathcal{O} = \mathcal{O}_a^N$ l'espace des observations du système.

La variable d'observation O dépend de manière probabiliste de l'état et de l'action du système, c'est-à-dire qu'il existe une probabilité $\mathbb{P}_{s,a}(o)$ d'observer o lorsque l'action a est réalisée dans l'état s. On note $\mathbb{P}(o) = (\mathbb{P}_{s,a}(o))_{s,a}$ la matrice des probabilités d'observation.

Le vecteur $(\mathcal{S}, \mathcal{A}, \mathcal{O}, P, r, \mathbb{P})$ définit alors un DEC-POMDP. Le schéma 3.7 montre l'évolution d'un DEC-POMDP avec une suite d'actions fixée.

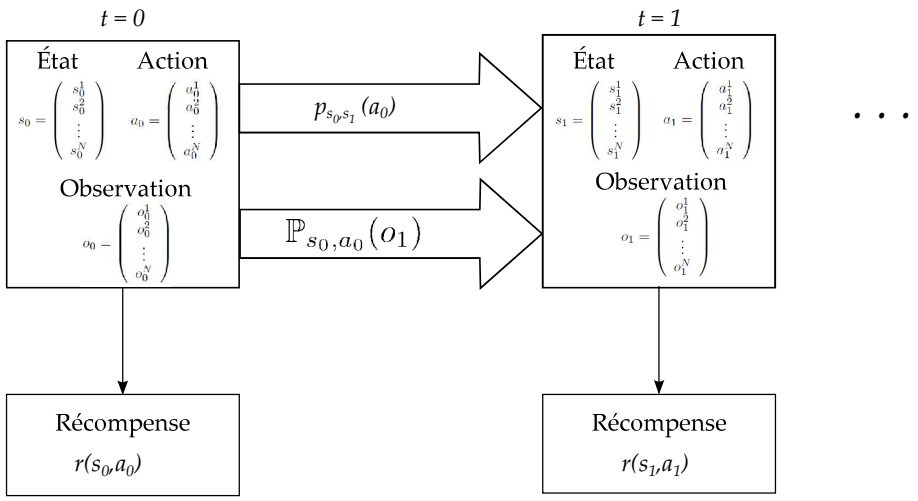


Figure 3.7 — Évolution d'un DEC-POMDP avec une suite d'actions fixée

À la différence d'un MMDP, les agents choisissent leur action en fonction de leurs observations, et non en fonction de l'état du système. Une règle de décision est donc une fonction

$$\Pi : \mathcal{O} \mapsto \mathcal{A}.$$

Nous verrons dans le paragraphe 3.5.4.1 que cette fonction induit une complexité algorithmique élevée.

Et la recherche d'une stratégie optimale stationnaire est formulée par :

$$\max_{\Pi \in \mathcal{A}^{\mathcal{O}}} G_{\Pi} \tag{3.16}$$

où G_{Π} est un gain choisi parmi les propositions de la partie 3.2.2. Ce problème d'optimisation revient à chercher la meilleure décision collective qui optimise le gain du système pour un ensemble d'observations donné.

Cas particulier des DEC-MDP Lorsque l'ensemble des observations permet de retrouver l'état complet du système, alors le système est dit *collectivement observable*. Cette définition équivaut à l'existence d'une fonction :

$$J : \mathcal{O} \mapsto \mathcal{S}$$

de sorte que $\mathbb{P}_{s,a}(o) > 0 \Rightarrow J(o) = s$. En d'autres termes, un ensemble d'observations détermine de manière unique l'état du système. Un DEC-POMDP collectivement observable est appelé DEC-MDP.

3.5.4.2 Complexités algorithmiques

Dans [15] la difficulté pour résoudre un DEC-POMDP ou un DEC-MDP est exprimée en termes de complexité algorithmique. Précisons clairement le contexte de ces résultats :

«Étant donné une durée T et un entier K, on cherche une politique Π telle que $G_{\Pi} > K$ »

où G_Π désigne le gain sur une période finie de durée T. Ce problème est désigné par $\mathcal{P}$.

Théorème 3. *Pour $N \geq 2$, le problème $\mathcal{P}$ pour un DEC-POMDP est NEXP-complet.*

Théorème 4. *Pour $N \geq 3$, le problème $\mathcal{P}$ pour un DEC-MDP est NEXP-complet.*

Ces théorèmes sont démontrés dans [15].

La classe de complexité NEXP est formée par les problèmes auxquels une machine de Turing non déterministe peut répondre en temps exponentiel.

En termes simples, un algorithme séquentiel, capable de choisir le cheminement de ses calculs lorsqu'un choix se présente, a besoin d'un temps au moins exponentiel pour répondre au problème $\mathcal{P}$. Un problème est dit NEXP-complet s'il est dans la classe NEXP, et au moins aussi difficile [10] que tous les autres problèmes qui sont NEXP. Ces résultats montrent donc qu'au-delà de quelques agents, il est calculatoirement impossible de résoudre un DEC-POMDP ou un DEC-MDP en temps raisonnable.

En conclusion, les extensions DEC-POMDP et DEC-MDP présentées dans cette partie permettent de prendre en compte l'observabilité partielle et incomplète du système des agents. Il n'est pas surprenant de voir la difficulté de résolution croître au-delà de celle des MMDP.

3.5.4.3 Discussion

Les théorèmes 3 et 4 ne font aucune hypothèse sur la forme du modèle considéré. En particulier ils n'écartent pas certains cas pathologiques, qui expliquent des temps de calcul élevés, comme :

- des observations extrêmement réduites. Par exemple, lorsque les agents ne perçoivent que leur propre état.
- une loi d'évolution hautement couplée. Par exemple, lorsque la transition d'un agent dépend des états et actions de **tous** les autres agents.

Ces situations ne sont pas pertinentes pour les systèmes multi-agents étudiés dans cette thèse, où les agents ont généralement une perception dépassant leur état personnel, et ne sont affectés que par un nombre d'agents limité.

Ainsi, lorsqu'un DEC-POMDP possède une structure particulière il peut être possible de le résoudre en temps raisonnable. Il existe en effet des méthodes performantes pour certaines classes de DEC-POMDP, dont quelques-unes sont présentées dans la partie 3.7. Avant cela, nous présentons d'autres extensions des MMDP.

3.5.5 Autres variantes

Il existe de nombreuses variantes des MMDP, prenant en compte différents aspects des SMA. Nous en présentons rapidement quelques-unes dans cette partie.

10. Une définition précise de cette notion est donnée dans [119] p.159.

3.5.5.1 GMDP

La localité des interactions entre agents est une caractéristique importante des systèmes multi-agents étudiés, et n'a pas encore été prise en compte à ce stade. Les GMDP (Graph Markov Decision Processes) [40] permettent d'intégrer cet aspect, en joignant au MMDP un graphe G qui représente les dépendances entre agents.

Le graphe de dépendance contient autant de noeuds que d'agents, et représente les interactions locales de la manière suivante :

L'agent i dépend de l'agent j si et seulement si le graphe G contient l'arête (i, j).

De la même façon le **voisinage** $\mathcal{V}_i$ d'un agent i est défini comme l'ensemble des noeuds reliés au noeud i. Ces voisinages sont généralement supposés statiques, c'est-à-dire qu'un agent interagit toujours avec les mêmes agents au cours du temps.

Ces voisinages permettent de définir la notion de localité pour les différents éléments du MMDP.

La loi d'évolution est *locale* et *factorisée* si elle s'écrit :

$$p_{s,s'}(a) = \prod_{n=1}^{N} p^n_{s_n,s'_n}(s_{\mathcal{V}_n}, a_n),$$

c'est-à-dire : les transitions des agents sont indépendantes, et la probabilité de transition d'un agent n ne dépend que des états de ses voisins $s_{\mathcal{V}_n}$ et de son action personnelle.

La récompense d'un GMDP est *locale* si elle s'écrit :

$$r(s, a) = \sum_{n=1}^{N} r_n(s_{\mathcal{V}_n}, a_n),$$

ce qui revient à dire qu'elle est la somme de récompenses individuelles, et que la récompense individuelle de l'agent n ne dépend que de l'état de ses voisins et de son action personnelle.

Une stratégie $\Pi = (\pi_1, \ldots, \pi_N)$ d'un GMDP est *locale* si chacune de ses composantes dépend localement des états des agents. Plus précisément, la règle de décision de l'agent n ne dépend que des états des voisins de n :

$$\pi_n : s_{\mathcal{V}_n} \mapsto a_n$$

Cependant, même dans le cas spécifique où la loi d'évolution et la récompense sont locales et factorisées, il n'existe pas de méthode efficace à ce jour pour résoudre les grands GMDP. Des méthodes approximatives ont été essayées, basées sur des techniques de champ moyen [131] ou de méthodes variationnelles [125]. La méthode de programmation linéaire présentée dans la section 3.7.1 permet toutefois de résoudre certains GMDP de manière approchée.

3.5.5.2 GO-DEC-MDP

C. Goldman et S. Zilberstein définissent les GO-DEC-MDP (*Goal Oriented Decentralized Markov Decision Processes*) [66]. Ce modèle identifie un certain ensemble d'états comme des

objectifs à atteindre. Ces états génèrent une récompense, tandis que tous les autres états donnent lieu à des pénalités.

Le problème de savoir s'il existe une stratégie permettant d'obtenir un gain supérieur à un certain entier donné (voir partie 3.5.4.2) est *NEXP-complet* [66].

Le formalisme des GO-DEC-MDP est donc adapté aux situations où il n'y a pas réellement de gain permanent, associé à chaque état et action, et où l'on souhaite emmener le système dans certains états. La résolution d'un tel MDP permet de planifier les actions que prendra le système pour y parvenir. Cependant, la difficulté calculatoire de cette résolution (NEXP) met les GO-DEC-MDP hors de portée pour les grands systèmes.

3.5.5.3 COM-TMDP

D. V. Pynadath et M. També proposent les COM-TMDP [124]. Cette extension des MMDP introduit deux notions supplémentaires :

- Les agents ont la possibilité d'échanger des messages au moment de choisir leur action. Ces messages peuvent éventuellement engendrer un coût.
- Les agents ont la possibilité de formuler un état de croyance du système, en se basant sur leurs observations personnelles et les messages reçus.

Les auteurs proposent également deux algorithmes permettant de les résoudre, appelées *JESP* (*Joint Equilibrium based Search for Policies*) et *DP-JESP* (*Dynamic Programming Joint Equilibrium Based Search for Policies*). Ces algorithmes utilisent la même idée : tous les agents -sauf un- fixent leur stratégie et le dernier résout un MDP «local» dont il est le seul acteur. Cette démarche sera utilisée, et expliquée plus en détail dans la partie 3.7.2.

Bien que *JESP* et *DP-JESP* convergent théoriquement vers un équilibre de Nash, leur coût calculatoire est immense. Dans [124], les auteurs les mettent en oeuvre sur un problème partiellement observables, avec 2 états, 3 actions, et 2 observations possibles par agent. Il s'avère que *JESP* ne fonctionne en temps raisonnable que pour un MDP sur une durée finie de $T \leq 2$, tandis que *DP-JESP* fonctionne en temps raisonnable pour $T \leq 7$.

Les COM-TMDP permettent donc de prendre en compte la communication entre agents, ce qui leur confère une bonne expressivité pour décrire les systèmes multi-agents étudiés dans cette thèse. Ils sont cependant difficiles à résoudre, même dans des cas très simples.

Pour conclure sur ces extensions, disons qu'il est possible d'introduire des spécifications supplémentaires sans trop s'écarter du modèle DEC-POMDP. Le fait d'augmenter l'expressivité du modèles a cependant un coût en termes de difficulté de résolution, et de façon générale les modèles plus riches sont manifestement plus difficiles à résoudre.

Précisons pour finir qu'il existe une vaste littérature sur les MDP Multi-Agents, et une importante variété de modèles prenant en compte différents aspects. Une liste de variantes supplémentaires est trouvée dans [116, 136, 135].

3.6 Retour à l'exemple des couloirs : extension au cas partiellement observé

Reprenons l'exemple de la partie 3.3. Un groupe de N personnes se trouve dans une pièce, et souhaite traverser un couloir pour se rendre dans une pièce voisine. Le couloir a une certaine capacité c qui limite le nombre de passages simultanés à chaque moment.

3.6.1 Définition du DEC-POMDP

Dans la partie 3.3, nous avons traité ce problème sous l'hypothèse d'une observabilité totale de l'état du système et l'échangeabilité des individus. Dans cette partie, nous supposons que les individus n'observent l'état du système que partiellement, et plaçons l'exemple dans le cadre des DEC-POMDP. Chaque personne dans la pièce initiale perçoit un sous-ensemble des personnes présentes dans cette pièce. Elle observe chaque personne avec probabilité α et s'observe toujours elle-même, soit $1 + (s_t - 1)\alpha$ personnes en moyenne.

On définit alors la variable d'observation $o_n(t)$ de la personne n à l'instant t comme le triplet

$$o_n(t) = (n_0, n_1, Coll)$$

où

- n_0 est le nombre de personnes observées.
- n_1 est le nombre de personnes observées qui ont essayé de traverser le couloir entre les instants $t-1$ et t.
- $Coll$ est un booléen qui indique si ces n_1 personnes sont parvenues à traverser le couloir ou si une collision a eu lieu. La variable $Coll$ n'est définie que si $n_1 > 0$.

On notera $O(t) = \begin{pmatrix} o_1(t) \\ \vdots \\ o_N(t) \end{pmatrix}$ la variable d'observation du système. Elle dépend de manière aléatoire de l'état et de l'action précédente. L'espace des observations de chaque agent est

$$\mathcal{O}_a = \left\{ (n_0, n_1, Coll) \ , \ 1 \leq n_0 \leq N, \ 0 \leq n_1 \leq n_0, \ Coll \in \{0,1\} \right\}$$

et a un cardinal de l'ordre de N^2. L'espace d'observation du système est $\mathcal{O} = \mathcal{O}_a^N$ et a un cardinal de l'ordre de N^{2N}. Avec cette variable d'observation, nous avons placé le problème dans le cadre des DEC-POMDP.

3.6.2 Stratégies et problème d'optimisation

On souhaite que les personnes décident de traverser le couloir en fonction de cette observation (et ignorent en particulier la capacité du couloir). Comme les personnes se trouvent initialement dans la même pièce, il n'est pas exclu qu'elles aient des observations identiques et prennent la même décision à l'unanimité : toutes traversent ou personne ne traverse (si la règle de décision est déterministe). Pour éviter ce genre de situations, nous considérons les

stratégies aléatoires, c'est-à-dire les fonctions

$$\pi_n : o_n \mapsto p_n$$

qui prescrivent à chaque observation la probabilité de s'engager. Comme nous avons supposé les individus identiques et échangeables, les règles de décision individuelles sont identiques : $\pi_1 = \pi_2 = \ldots = \pi_N = \pi$.

Une stratégie collective stationnaire est l'agrégation d'un ensemble de stratégies individuelles

$$\Pi : \begin{pmatrix} o_1 \\ \vdots \\ o_N \end{pmatrix} \mapsto \begin{pmatrix} \pi(o_1) \\ \vdots \\ \pi(o_N) \end{pmatrix}.$$

Pour une stratégie collective donnée, le gain moyen sur une période finie $[0, T]$ correspond au nombre de personnes dans la deuxième pièce (voir 3.3) et s'écrit

$$G_\Pi = N - \mathbb{E}\left(s_T\right)$$

où l'espérance est prise sur tous les états possibles $\{s_1, \ldots, s_T\}$, toutes les observations possibles $\{o_1(0), \ldots, o_N(0), \ldots, o_1(T), \ldots, o_N(T)\}$ et toutes les actions possibles $\{a_1, \ldots, a_T\}$.

3.6.3 Calcul d'une stratégie optimale

On cherche une stratégie stationnaire Π^* maximisant

$$G_\Pi = N - \mathbb{E}(s_T).$$

En s'inspirant des remarques faites à la partie 3.4.1, et en notant

- r la fonction de récompense,
- $p_{s,s'}(a)$ la probabilité de transition,
- $\mathbb{P}_{s,a}(O')$ la probabilité d'observer O' lorsque l'action a est exécutée dans l'état s,

on peut évaluer une stratégie Π donnée à l'aide de la relation de récurrence

$$V_{t-1}^{\Pi}(s, O) = r\left(s, \Pi(O)\right) + \sum_{s' \in \mathcal{S}} \sum_{O' \in \mathcal{O}} p_{s,s'}(\Pi(O))\, \mathbb{P}_{s,\Pi(O)}(O')\, V_t^{\Pi}(s', O') \tag{3.17}$$

pour $t = T$ à $t = 1$. La valeur $V_t^{\Pi}(s, O)$ représente le gain obtenu à partir de l'instant t, lorsque le système est dans l'état s et observe O.

Pour un système dans un état initial s_0 avec une observation initiale O_0, la stratégie $\Pi^* = (\pi^*, \ldots, \pi^*)$ réalisant le maximum

$$G^* = \max_{\Pi} \left(V_0^{\Pi}(s_0, O_0) \right)$$

constitue la stratégie optimale.

Il n'est pas clair que ce calcul puisse être réalisé de manière efficace. Lorsque l'ensemble d'actions est fini et la stratégie est déterministe, on peut montrer [117] que le nombre de calculs nécessaire pour évaluer une seule stratégie est doublement exponentiel en T. Dans notre

cas, les stratégies aléatoires considérées sont paramétrées par des probabilités $\{\pi(o)\}_{o\in\mathcal{O}_a}$ dans l'ensemble infini $[0,1]$, et le calcul de leurs valeurs optimales est difficile.

Par ailleurs, il est probable que la stratégie optimale Π^* fasse explicitement intervenir la capacité c du couloir, ainsi que la probabilité d'observation α. Il est intéressant d'étudier à quel point l'absence de ces informations limite la performance du système. Dans le paragraphe suivant, nous allons précisément nous intéresser à cette question, dans le cadre du problème de traversée du couloir.

3.6.4 Quelques stratégies empiriques basées sur des connaissances partielles

Le problème de traverser le couloir de manière efficace, avec des agents qui observent de manière aléatoire l'état du système a été formulé comme un DEC-POMDP. En vertu des remarques faites ci-dessus, nous allons nous intéresser à certaines stratégies heuristiques qui sont basées sur des informations partielles.

Tout au long de cette section, les agents ignorent l'état du système, c'est-à-dire le nombre de personnes restantes. Le fait d'ignorer cette information globale nous rapproche des systèmes multi-agents ciblés dans cette thèse.

Dans les deux premières stratégies heuristiques proposées, les agents disposent de la connaissance de la probabilité d'observation α, et de la capacité du couloir. En gardant en tête la résolution globale effectuée à la partie 3.3, les agents vont essayer de deviner le nombre de personnes restantes pour déterminer la probabilité avec laquelle ils décident de s'engager.

Rappelons cette solution. S'il reste s personnes dans la pièce initiale et qu'elles s'engagent toutes avec une probabilité p, la récompense moyenne vaut

$$r_c(s,p) = \sum_{k=1}^{c} k \begin{pmatrix} s \\ k \end{pmatrix} p^k (1-p)^{s-k}.$$

L'argument $p^*(s)$ qui maximise $p \mapsto r_c(s,p)$ correspond à la stratégie aléatoire optimale.

Dans la troisième et dernière stratégie heuristique, les agents ignorent à la fois la capacité du couloir c et la probabilité d'observation α. En se basant sur leurs observations, les agents vont progressivement ajuster la probabilité avec laquelle ils sollicitent le couloir.

3.6.4.1 Estimation moyenne du nombre de personnes restantes

Nous commençons par une stratégie simple où les agents ont connaissance de la probabilité d'observation et de la capacité du couloir, mais où ils ignorent le nombre de personnes restantes.

La stratégie proposée consiste à raisonner de la manière suivante : chaque individu restant (hormis soi-même) est observé avec probabilité α. Chaque agent observe donc $1+(s-1)\alpha$ personnes en moyenne. Ainsi, pour n_0 personnes observées, le nombre de personnes restantes dans la première pièce vaut environ

$$\hat{s} = \frac{n_0 - 1}{\alpha} + 1.$$

En supposant que ce nombre correspond exactement à l'état du système s, l'agent peut utiliser la stratégie optimale calculée à la section 3.3. La probabilité optimale associée à cette estimation $\hat{s}$ est

$$p^* = \underset{p}{\operatorname{argmax}}\Big(r_c(\hat{s}, p)\Big),$$

où

$$r_c(\hat{s}, p) = \sum_{k=1}^{c} k \begin{pmatrix} \hat{s} \\ k \end{pmatrix} p^k (1-p)^{s-k}$$

est la récompense moyenne associée à la capacité c.

Nous avons fixé les paramètres à $N = 100$, $\alpha = 0,05$ et $c = 2$. Il y a donc une population de 100 agents, qui s'observent avec une probabilité de $\alpha = 0,05$ et le couloir ne permet qu'à deux personnes (au maximum) de traverser simultanément. À la section 3.3, nous avons montré que si le nombre de personnes restantes est connu et vaut s, la stratégie optimale consiste à essayer de traverser le couloir avec probabilité

$$\begin{cases} 1 & \text{si } s \in \{0,1,2\} \\ \frac{-(s-3)-\sqrt{(s-3)^2+4(s-1)^2-4}}{2(-s^2+2s)} & \text{sinon} \end{cases}.$$

Les agents vont utiliser cette formule en utilisant l'estimation $\hat{s}$ à la place de la valeur de s, qui est inconnue. Les résultats des simulation sont représentés sur la figure 3.8.

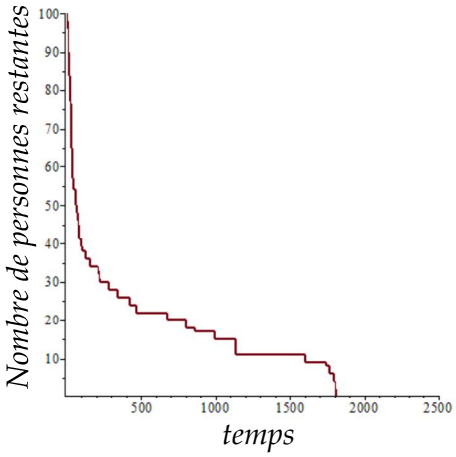


Figure 3.8 — Évolution du nombre de personnes restantes

La figure 3.8 montre que le nombre de personnes restantes décroît au cours du temps jusqu'à atteindre 0. La décroissance est extrêmement lente, il faut attendre $t = 1800$ pour que la pièce se vide. Ce résultat est à comparer au résultat trouvé à la section 3.3, où les personnes avaient une connaissance complète de l'état du système. Nous la rappelons ci-dessous :

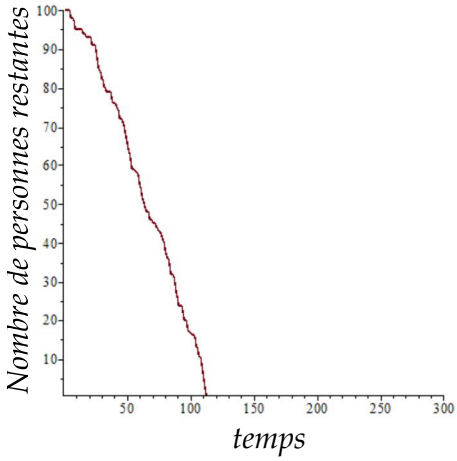


Figure 3.9 — Évolution du nombre de personnes restantes lorsque l'état complet du système est connu

La stratégie proposée a clairement une performance largement inférieure à celle de la stratégie optimale.

La figure 3.10 montre la probabilité moyenne avec laquelle les personnes restantes s'engagent. Cette probabilité évolue de manière très irrégulière, mais semble augmenter de manière chaotique au cours du temps.

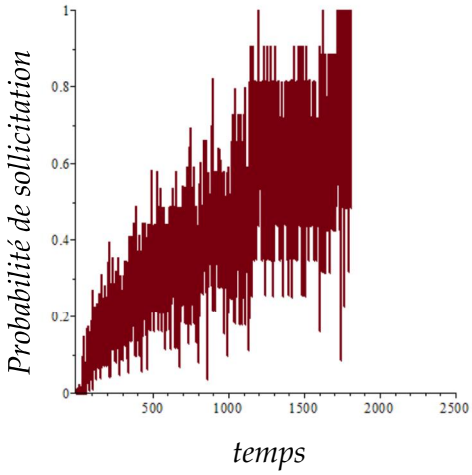


Figure 3.10 — Évolution de la probabilité moyenne de sollicitation

Il s'agit d'un comportement qui était attendu : quand il ne reste que peu de personnes, il y a peu de chances que des collisions aient lieu. Les personnes restantes peuvent donc se montrer plus audacieuses et demander à traverser avec une probabilité plus élevée.

3.6.4.2 Estimation de l'état global par maximum de vraisemblance

Un raisonnement plus fin consiste à faire une estimation statistique (et non moyenne) du nombre de personnes restantes par une maximisation de la vraisemblance. S'il reste s

personnes dans la première pièce, la probabilité d'observer n_0 personnes est

$$\phi(s, n_0, \alpha) = \begin{pmatrix} s \\ n_0 - 1 \end{pmatrix} \alpha^{n_0-1}(1-\alpha)^{s-n_0+1}.$$

La deuxième stratégie consiste à estimer le nombre de personnes restantes par l'argument $\hat{s}$ qui maximise la *vraisemblance* $s \mapsto \phi(s, n_0, \alpha)$ [11]. La probabilité optimale associée à cette estimation est

$$p^* = \underset{p}{\operatorname{argmax}}\Big(r_c(\hat{s}, p)\Big).$$

La figure 3.11 montre l'évolution du système lorsque $N = 100$, $\alpha = 0,05$ et $c = 2$.

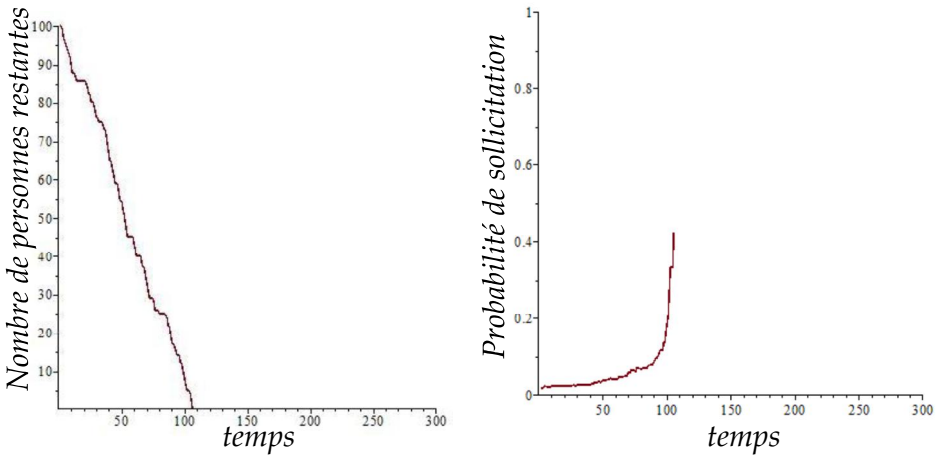


Figure 3.11 — Évolution du nombre de personnes restantes (à gauche) et de la probabilité moyenne (à droite)

Ces courbes ont une allure similaire à celles de la figure 3.8 : le nombre de personnes restantes décroît jusqu'à 0, et la probabilité moyenne de s'engager augmente à mesure que la pièce initiale se vide.

La performance de cette stratégie est meilleure que la précédente. La pièce initiale se vide plus rapidement qu'à la figure 3.8, et presque aussi rapidement que sous information complète (figure 3.9). D'autre part, la probabilité moyenne avec laquelle les personnes restantes s'engagent augmente plus rapidement, et de manière plus stable que dans la figure 3.10.

3.6.4.3 Une stratégie empirique adaptative

La troisième stratégie que nous proposons n'utilise ni la capacité du couloir c, ni la probabilité d'observation α, mais nécessite le réglage au préalable d'un seuil K. Elle consiste à ajuster progressivement la probabilité avec laquelle les agents s'engagent en fonction leurs observations.

11. De manière équivalente, on peut considérer l'argument s qui maximise la *log-vraisemblance* $\ln\left[\begin{pmatrix} s \\ n_0 - 1 \end{pmatrix} \alpha^{n_0-1}(1-\alpha)^{s-n_0+1}\right]$ et qui se prête généralement mieux au calcul.

À partir d'une observation $(n_0, n_1, Coll)$, chaque agent évalue empiriquement la proportion d'agents qui ont essayé de s'engager à l'instant précédent avec le ratio n_1/n_0. L'idée est que si une collision a eu lieu, alors il faut s'engager avec un probabilité p inférieure à n_1/n_0.

La plus petite probabilité empirique observée est donc $1/n_0$, soit $1/\alpha s$ en moyenne. D'autre part, les calculs effectués dans la partie 3.3 (avec observation totale) suggèrent que la probabilité optimale est inférieure (et proche de) c/s. Pour que les observations soient significatives, il faut donc que $1/n_0$ soit inférieur à c/s. Nous supposons donc que $\alpha \geq \frac{1}{c}$.

Cette stratégie consiste à fixer un nombre K à l'avance, faisant office de seuil, et à tenir le raisonnement suivant :

- Si $1 \leq n_0 \leq K$, j'observe trop peu de personnes pour raisonner globalement, et j'essaye de traverser avec probabilité $1/K$.
- Si $n_0 > K$, alors je définis la probabilité empirique par $\bar{p} = n_1/n_0$.
  - Si une collision a eu lieu ($Coll = 1$), la probabilité empirique $\bar{p}$ est trop importante par rapport à la capacité relative c/s. Je vais donc choisir de m'engager avec un probabilité inférieure à ce ratio, en choisissant $p = f(\bar{p})$.
  - Si je n'ai pas observé de collision ou si $n_1 = 0$, la probabilité $\bar{p}$ est trop faible par rapport à la capacité relative c/s. Je vais donc choisir de m'engager avec un probabilité inférieure à ce ratio, en choisissant $p = g(\bar{p})$.

Les fonctions f et g permettent respectivement d'augmenter ou diminuer la probabilité empirique. On les choisit de sorte que $f(1) < 1$, $f'(1) > 1$ et $g(0) = 0$, $g'(0) < 1$. Ces conditions assurent que la probabilité soit suffisamment diminuée (respectivement augmentée) lorsqu'on observe des collisions avec une haute probabilité empirique (respectivement pas de collisions avec une basse probabilité empirique). La figure 3.12 représente les fonctions f et g qui ont été choisies pour les simulations.

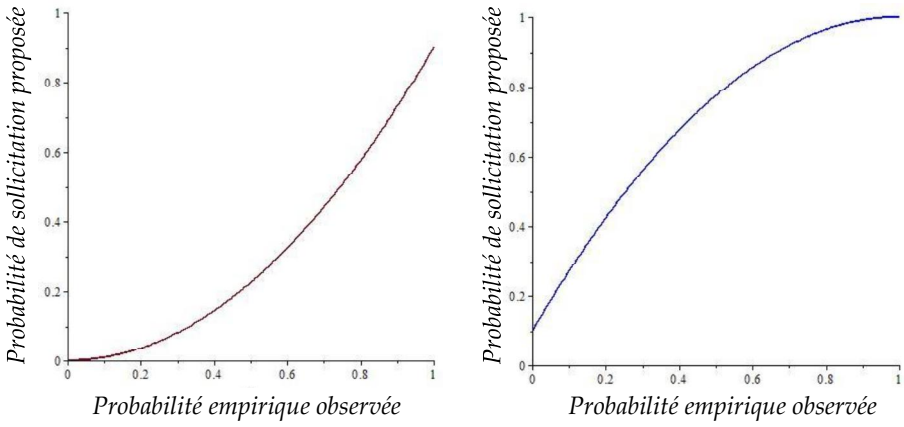


Figure 3.12 — Fonctions d'ajustement f (diminution, à gauche) et g (augmentation, à droite)

La figure 3.13 montre l'évolution du système avec $\alpha = 0,05$, $c = 2$ et $K = 20$. Les probabilités de s'engager sont initialisées à 1 pour tous les agents, signifiant qu'il sont très agressifs en début de simulation.

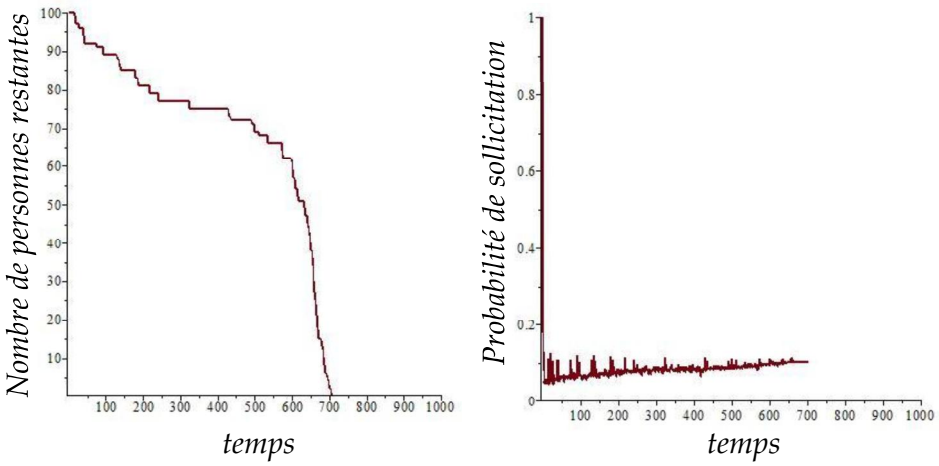


Figure 3.13 — Évolution du nombre de personnes restantes (à gauche) et de la probabilité moyenne de sollicitation (à droite)

La courbe représentant l'évolution du nombre de personnes restantes montre que dans un premier temps (court) personne ne parvient à traverser le couloir. Pendant cette même période, la probabilité moyenne avec laquelle chaque personne sollicite le couloir décroît rapidement sur la courbe de droite. Ensuite, les personnes commencent lentement à traverser le couloir et la probabilité de sollicitation moyenne commence à croître. Au-delà de $t = 600$, la situation se débloque, et le nombre de personnes restantes diminue très rapidement.

Cette stratégie est beaucoup moins performante que la stratégie optimale (calculée à la partie 3.3). Elle possède cependant un avantage majeur : aucune connaissance sur les paramètres α et c du système n'est requise. De plus, elle possède un caractère adaptatif : la probabilité avec laquelle les personnes s'engagent est ajustée en fonction des observations. Il est possible que cette méthode permette de vider la pièce même si les coefficients α et c varient au cours du temps.

Pour justifier la décroissance du nombre de personnes avec cette dernière stratégie, on peut penser à étudier la chaîne de Markov $(s_t, O(t))_t$. Cette chaîne de Markov possède cependant un grand espace d'états, et par conséquent il est difficile d'étudier sa dynamique, en raison de la taille de sa matrice de transition. On peut expliquer la décroissance par le fait que le nombre de personnes restantes dans la première pièce décroît à chaque instant avec une probabilité non nulle dès lors que toutes les probabilités sont différentes de 1. Il est plus délicat de quantifier cette décroissance de manière précise.

3.6.5 Conclusion

Cet exemple montre que la représentation d'un système multi-agent par un DEC-POMDP ne permet pas de le résoudre rapidement. Les équations d'optimalité sont difficiles à traiter, et leur complexité calculatoire est rapidement hors de portée si le nombre d'agents est grand. Ce fait n'exclut pas d'essayer certaines stratégies heuristiques. Ces stratégies peuvent être basées sur une connaissance plus ou moins complète de l'état du système ou de sa loi d'évolution.

Dans la partie qui suit, nous présentons deux méthodes de résolution de MMDP trouvées dans la littérature. Elles partagent la même idée : exploiter la forme particulière du MMDP pour réaliser des simplifications.

3.7 Résolution d'un MDP Multi-Agents

Dans cette partie, nous présentons quelques méthodes de résolution de MDP Multi-Agents trouvées dans la littérature. Au cours de la partie précédente, nous avons mis en évidence une complexité algorithmique exponentielle pour résoudre ce genre de problèmes. Les méthodes que nous présentons ici permettent de réduire cette complexité en exploitant la structure du MDP.

3.7.1 Une méthode de programmation linéaire dans le cas factorisé

Guestrin et al. proposent [74] d'exploiter le caractère local des interactions, en décomposant la fonction de valeur. Leur méthode s'applique en particulier au cas des GMDP factorisés.

3.7.1.1 Hypothèses

Considérons un GMDP $(\mathcal{S}, \mathcal{A}, p, r, G)$, tel que :

- Les transitions des agents sont statistiquement indépendantes, et dépendent de manière locale des états des autres agents. La probabilité de transition du système s'écrit

$$p_{s,s'}(a) = \prod_{n=1}^{N} p^n_{s_n,s'_n}(s_{\mathcal{V}_n}, a)$$

  où $p^n_{s_n,s'_n}(s_{\mathcal{V}_n}, a_n)$ est la probabilité que l'agent n passe de l'état s_n à s'_n lorsque son voisinage est dans l'état $s_{\mathcal{V}_n}$ et si l'action collective est a.
- La récompense est additive, et dépend de manière locale des agents :

$$r(s, a) = \sum_{n=1}^{N} r_n(s_{\mathcal{V}_n}, a_n)$$

  où r_n représente le gain obtenu par l'agent n.

Notons que ces hypothèses ne correspondent pas exactement à la localité au sens des GMDP (Cf. section 3.5.5.1), car la probabilité de transition individuelle d'un agent $p^n_{s_n,s'_n}(s_{\mathcal{V}_n}, a)$ dépend a priori de l'action collective.

3.7.1.2 Présentation de la méthode

La proposition de [74] consiste à approcher la fonction Value sous la forme

$$V(s) \simeq \sum_{n=1}^{N} \alpha_n \, V_n(s_{\mathcal{V}_n}) \tag{3.18}$$

pour un ensemble de fonctions $\{V_n\}_n$ choici à l'avance. Nous commentons le choix de ces fonctions en conclusion de cette section.

Chaque fonction V_n représente la *Value* de l'agent n et ne dépend que des états des voisins de n. Trouver les meilleurs coefficients $\{\alpha_n\}_n$ revient à projeter la fonction V sur l'espace vectoriel engendré par les fonctions $\{V_n\}_n$.

Avec des méthodes trouvées dans [74, 87] on montre, en utilisant l'approximation (3.18) et la structure locale du MMDP, que l'équation d'optimalité associée au gain dévalué

$$V(s) = \max_{a \in \mathcal{A}} \left[r(s,a) + \delta \sum_{s' \in \mathcal{S}} p_{s,s'}(a)\, V(s') \right],\ s \in \mathcal{S}. \tag{3.19}$$

est approchée par un problème d'optimisation linéaire. Contrairement à (3.19), la difficulté de ce problème linaire ne croît pas exponentiellement avec le nombre d'agents et peut être résolu en temps raisonnable avec la méthode du simplexe.

La solution V du problème (3.19) est distribuée sur les agents par une méthode élimination des variables. Elle consiste à reformuler l'équation (3.19)

$$V(s) = \max_{a_1} \left[\max_{a_2} \ldots \left[\max_{a_N} \left[r(s,a) + \delta \sum_{s' \in \mathcal{S}} p_{s,s'}(a)\, V(s') \right] \right] \right],\ s \in \mathcal{S}.$$

et à effectuer les maxima à tour de rôle. L'agent N calcule

$$\mathcal{Q}_N(s, a_1, \ldots, a_{N-1}) = \max_{a_N} \left[r(s,a) + \delta \sum_{s' \in \mathcal{S}} p_{s,s'}(a)\, V(s') \right]$$

et transmet cette fonction à l'agent $N-1$. Cet agent calcule

$$\mathcal{Q}_{N-1}(s, a_1, \ldots, a_{N-2}) = \max_{a_{N-1}} \mathcal{Q}_N(s, a_1, \ldots, a_{N-1})$$

et ainsi de suite, jusqu'à l'agent 1 qui calcule

$$\mathcal{Q}_1(s) = \max_{a_1} \mathcal{Q}_2(s, a_1).$$

Toute action $a_1^*(s)$ réalisant ce dernier maximum est optimale pour l'agent 1.

Remarque 7. *La dépendance vis-à-vis des actions $a_1, \ldots, a_n$ a été progressivement éliminée, et ce dernier calcul de maximum ne fait intervenir que l'action de l'agent* 1.

Ensuite, les autres agents trouvent leurs actions optimales en procédant dans l'ordre inverse $n = 2, \ldots N$. L'agent 1 transmet son action optimale $a_1^*(s)$ à l'agent 2 qui calcule

$$\max_{a_2} \mathcal{Q}_3(s, a_1^*(s), a_2).$$

Toute action $a_2^*(s)$ réalisant ce maximum est optimale pour l'agent 2. En poursuivant ce raisonnement de manière itérative, pour $n = 3, \ldots, N$ on obtient les actions optimales de tous les agents.

3.7.1.3 Discussion

La méthode de résolution de MMDP présentée ci-dessus se passe en deux temps. La première étape consiste à calculer la valeur globale par une méthode de programmation linéaire. La seconde est le calcul des action optimales de chaque agent. La résolution effective du problème est donc réalisée au niveau global, et en ce sens la méthode est en opposition avec les approches multi-agents.

Le calcul des actions optimales est effectué, de manière individuelle, par chacun des agents. On ne peut cependant pas le qualifier de "local" puisqu'il fait intervenir un ordre d'exécution ($N, \ldots, 1$ dans notre cas) établi de manière centralisé. De plus, pour calculer leurs actions optimales les agents communiquent des fonctions de value partielles $\mathcal{Q}_n$. Transmettre de telles fonctions peut s'avérer extrêmement coûteux. Prenons l'exemple de la fonction $\mathcal{Q}_N$ qui dépend de $(s, a_1, \ldots, a_{N-1})$. L'agent N transmet donc un message contenant $S_a \times A_a^{N-1}$ valeurs. Quelle que soit la manière d'encoder ces valeurs, un tel message nécessite une place mémoire trop importante pour que cette méthode puisse être utilisée en pratique.

Par ailleurs, la méthode s'appuie de manière importante sur l'approximation linéaire $V = \sum_{n=1}^{N} \alpha_n V_n$, dont la qualité dépend du choix des fonctions $\{V_n\}_n$. Il n'existe pas de méthode générale pour effectuer ce choix, qui nécessite pour l'instant une bonne connaissance du problème étudié [133, 43]. On peut toutefois trouver quelques guides justifiées sur un plan théorique dans [45].

Pour conclure, cette méthode permet de résoudre, sous certaines hypothèses et de manière approchée, un MMDP présentant un caractère local. La résolution est cependant effectuée au niveau global, et le calcul des actions optimales suit une planification centralisée. Ces observations rendent cette méthode incompatible avec les systèmes multi-agents étudiés.

Contrairement à cette méthode qui distribue la solution du problème une fois qu'elle a été calculée, la méthode présentée dans la section suivante permet de distribuer la résolution du problème.

3.7.2 Une méthode de distribution des équations d'optimalité

Dans cette partie nous montrons qu'un MMDP factorisé possédant la propriété d'être *unichaîne* peut être résolu de manière distribuée.

3.7.2.1 Hypothèses

On considère un MMDP $(\mathcal{S}, \mathcal{A}, P, r)$ particulier possédant les propriétés suivantes :
– La loi d'évolution globale s'écrit sous la forme

$$p_{s,s'}(a) = \prod_{n=1}^{N} p^n_{s_n,s'_n}(a_n)$$

où $p^n_{s_n,s'_n}(a_n)$ est la probabilité que l'agent n passe de l'état s_n à s'_n en choisissant l'action

a_n [12]. On note $p^n(a_n) = \left(p^n_{s_n,s'_n}(a_n)\right)_{s_n,s'_n}$ la matrice de transition individuelle de l'agent n.

– Chaque MDP *local* $(\mathcal{S}_a, \mathcal{A}_a, p^n, r_n)$, où r_n est une fonction de (s_n, a_n), est *unichaîne*. Cela signifie que pour toute stratégie *locale*

$$\pi_n : \mathcal{S}_a \to \mathcal{A}_a$$

la chaîne de Markov $(s^t_n)_t$ correspondante possède une seule classe récurrente [123].

Intuitivement, le fait d'être unichaîne garantit que le comportement initial (et transitoire) du système n'a pas d'incidence sur sa dynamique à long terme. Une conséquence importante est

Proposition 4. *Soit $(\mathcal{S}, \mathcal{A}, p, r)$ un MDP unichaîne sur une espace d'états fini $\mathcal{S}$. Alors pour toute stratégie π, il existe une unique mesure stationnaire, c'est-à-dire une unique fonction $\rho : \mathcal{S} \to \mathbb{R}^+$ telle que*

$$\sum_{s' \in \mathcal{S}} \rho(s')\, p_{s,s'}\left(\pi(s)\right) = \rho(s), s \in \mathcal{S} \;\; et \;\; \sum_{s \in \mathcal{S}} \rho(s) = 1 \tag{3.20}$$

Une deuxième conséquence, importante pour les MDP, est :

Proposition 5. *Soit $(\mathcal{S}, \mathcal{A}, P, r)$ un MDP unichaîne, alors le gain moyen partant d'un état initial s donné*

$$G_\Pi(s) = \limsup_{T \to \infty} \mathbb{E}\left(\frac{1}{T}\sum_{t=0}^{T-1} r\big(s_t, \Pi(s_t)\big) \;\middle|\; s_0 = s\right).$$

est indépendant de s (et simplement noté G_Π).

Ces résultats sont démontrés dans [123].

Remarque 8.

– *La propriété d'être unichaîne pour un MDP ne dépend que de sa loi d'évolution, et est indépendante de sa fonction de récompense.*
– *Comme tous les MDP locaux $(\mathcal{S}_a, \mathcal{A}_a, p^n, r_n)$ sont unichaînes, le MDP global $(\mathcal{S}, \mathcal{A}, p, r)$ est unichaîne.*
– *Vérifier si un MDP donné est unichaîne est un problème NP-complet [143] dans le cas général. Il existe cependant des cas particuliers où cette vérification peut être réalisée en temps polynomial [57].*
– *Une condition suffisante simple pour qu'un MDP soit unichaîne est :*
« Si, en chaque état il existe une suite d'actions qui permet d'arriver à n'importe quel état (avec probabilité > 0), alors le MDP est unichaîne. »
– *De même, s'il existe un état accessible (avec probabilité > 0) depuis tous les autres états pour une action bien choisie, le MDP est automatiquement unichaîne. Il est donc possible de rendre un MDP unichaîne en introduisant artificiellement un tel état.*

12. Notez que cette condition est bien plus forte que la factorisation de la loi d'évolution.

3.7.2.2 Deux exemples d'application

Pour illustrer le type de problèmes qui entrent dans le cadre des hypothèses données dans la section 3.7.2.1, nous donnons deux exemples d'application.

Considérons un ensemble d'agents positionnés sur une grille rectangulaire finie $\mathcal{R}$, contenue dans $\mathbb{Z} \times \mathbb{Z}$, où plusieurs agents peuvent être situés sur une même cellule. À chaque instant, ces agents peuvent se déplacer sur une des cellules adjacentes, selon les quatre directions principales, ou rester sur place. L'objectif du système est que les agents essayent de s'éloigner le plus possible les uns des autres.

Ce problème entre dans le formalisme de la section 3.7.2.1. **L'espace d'états** de chaque agent est la grille des positions $\mathcal{S}_a = \mathcal{R}$, et **l'espace d'actions** de chaque agent l'ensemble des déplacements qu'il peut faire $\mathcal{S}_a = \{\leftarrow, \uparrow, \rightarrow, \downarrow, \emptyset\}$.

En vertu du quatrième point de la remarque 8, le MDP obtenu est unichaîne.

La récompense associée au problème est simplement la somme des distances euclidiennes entre les positions des agents :

$$\sum_{n,m} \sqrt{(x_n - x_m)^2 + (y_n - y_m)^2}$$

où (x_n, y_n) désigne la position de l'agent n.

L'algorithme de calcul d'une stratégie optimale présenté à la section 3.7.2.3 permet de calculer une stratégie de déplacement qui permet aux agents de maximiser ces distances mutuelles, en moyenne sur une longue période.

Un autre exemple d'application intéressant est donné par le même système d'agents localisés, avec les mêmes règles de transition, mais un objectif différent. On fixe une densité f sur l'ensemble des positions, c'est-à-dire une fonction telle que

$$f(x,y) \geq 0, \forall (x,y) \in \mathcal{R} \text{ et } \sum_{(x,y)\in\mathcal{R}} f(x,y) = 1.$$

L'objectif du système est que les agents se répartissent selon cette densité.

On définit **la récompense** associée à cet objectif comme une somme de récompenses locales

$$\sum_{(x,y)\in\mathcal{R}} \rho(x,y).$$

La récompense locale ρ évalue la réussite des agents situés au point (x,y), et est définie par

$$\rho(x,y) = \begin{cases} u(x,y)e^{1-\frac{u(x,y)}{f(x,y)}} & \text{si } f(x,y) > 0 \\ 0 & \text{sinon} \end{cases}$$

où $u(x,y)$ désigne la proportion d'agents situés en (x,y). La fonction ρ est représentée sur la figure 3.14, et vérifie les propriétés suivantes :

- elle récompense de manière croissante la présence d'agents, *tant qu'ils sont en nombre inférieur à la proportion ciblée*,
- elle pénalise le surencombrement, en décroissant au-delà de cette proportion.

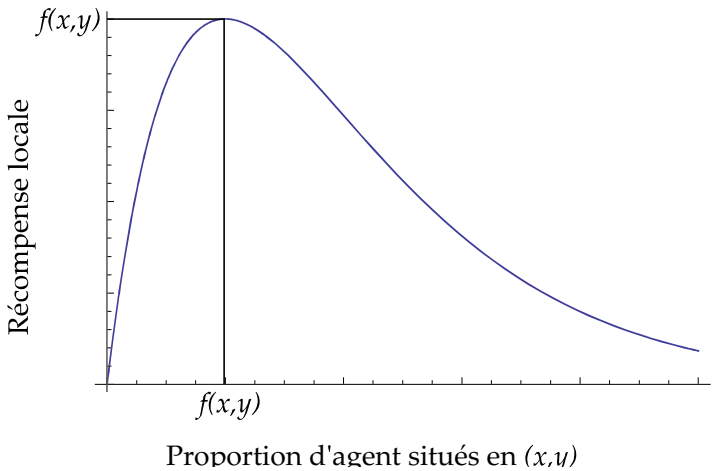


Figure 3.14 — Fonction de satisfaction locale des agents

L'algorithme présenté ci-dessous permet de déterminer les déplacements des agents qui maximisent la récompense (moyenne) à long terme. Par contre, en raison de sa complexité qui augmente fortement avec le nombre d'agents, cet algorithme n'est pas applicable pour de grands systèmes.

Dans le chapitre 5 nous reviendrons sur ce problème, et proposons une approximation continue qui permet calculer les stratégies de déplacement optimales pour de très grands systèmes.

3.7.2.3 Présentation de la méthode

La méthode de résolution présentée dans [36, 37] consiste à transformer le MDP global $(\mathcal{S}, \mathcal{A}, p, r)$ en N MDP locaux $(\mathcal{S}_a, \mathcal{A}_a, p^n, r_n)$, avec des récompenses r_n bien choisies, de sorte que la résolution collective et indépendante des MDP locaux aboutisse à un optimum global.

Considérons un agent n, et supposons toutes les stratégies locales des autres agents $\{\pi_m\}_{m \neq n}$ fixées. Sous l'hypothèse d'une perception complète du système, l'agent n détermine sa stratégie optimale en calculant le maximum suivant [13]

$$G_n^* = \max_{\pi_n \in \mathcal{A}_a^{\mathcal{S}}} \limsup_{T \to \infty} \mathbb{E} \left(\frac{1}{T} \sum_{t=0}^{T-1} r\big(s_t, (\pi_1(s_t), \ldots, \pi_n(s_t), \ldots, \pi_N(s_t))\big) \;\middle|\; s_0 = s \right). \tag{3.21}$$

Si tous les agents parviennent à réaliser ce maximum, l'ensemble de leurs stratégies constitue un équilibre de Nash (Cf. paragraphe 3.5.3). L'équation d'optimalité correspondant à ce problème est (Cf. théorème (8.4.3) dans [123]) :

$$G_n^* + V(s) = \max_{a_n \in \mathcal{A}_a} \left(r\left(s, (\pi_1(s_1), \ldots, a_n, \ldots, \pi_N(s_N)\right) + \sum_{s' \in \mathcal{S}} p^n_{s_n, s'_n}(a_n) \prod_{m \neq n} p_{s_m, s'_m}(\pi_m(s'_m))\, V(s') \right). \tag{3.22}$$

13. On rappelle que, par hypothèse, ce gain est indépendant de l'état initial s.

Soit, pour chaque $m \neq n$, ρ_m la mesure invariante [14] de l'agent m. L'idée de H. S. Chang [36, 37] consiste essentiellement à multiplier cette équation de par et d'autre par $\prod_{m \neq n} \rho_m(s_m)$ et à sommer sur tous les états possibles $s_m,\ m \neq n$. En utilisant l'invariance des mesures $\{\rho_m\}_m$ (Cf. (3.20)) on montre que cette opération aboutit à l'équation

$$G_n^* + V_n(s_n) = \max_{a_n \in \mathcal{A}} \left(r_n(s_n, a_n) + \sum_{s'_n \in \mathcal{S}_a} p^n_{s_n, s'_n}(a_n)\, V_n(s'_n) \right), \tag{3.23}$$

où

$$r_n(s_n, a_n) = \sum_{\substack{s_m \in \mathcal{S}_a \\ m \neq n}} \prod_{m \neq n} \rho_m(s_m) r\left(s, (\pi_1(s_1), \ldots, a_n, \ldots, \pi_N(s_N))\right) \tag{3.24}$$

et

$$V_n(s_n) = \sum_{\substack{s_m \in \mathcal{S}_a \\ m \neq n}} \prod_{m \neq n} \rho_m(s_m) V(s).$$

En d'autres termes, nous sommes amenés à résoudre le MDP local de l'agent n $(\mathcal{S}_a, \mathcal{A}_a, r_n, p^n)$.

Remarque 9. *Sans entrer dans le détail des calculs, le fait de multiplier par les différentes mesures invariantes permet de s'affranchir des probabilités de transition des autres agents $m \neq n$, et d'isoler l'équation d'optimalité du point de vue de l'agent n.*

On imagine sans difficultés un algorithme où les agents améliorent leur stratégies à tour de rôle, dans un ordre arbitraire choisi à l'avance. Cette démarche pose deux problèmes :

- Les agents doivent connaître l'ordre dans lequel ils agissent, pour attendre leur tour. Une telle convention n'existe pas nécessairement, et il n'est pas évident de l'établir de manière locale.
- L'exécution peut être très longue s'il y a un grand nombre d'agents.

Une autre possibilité est de permettre à tous les agents d'améliorer simultanément leurs stratégies. Cette implémentation a des chances de converger plus rapidement, mais risque de faire apparaître des phénomènes périodiques dans son déroulement.

3.7.2.4 Discussion

Nous avons présenté une méthode permettant de décentraliser la résolution d'un MDP multi-agent. Cette méthode présuppose deux hypothèses relativement restrictives :

- Les agents évoluent sans aucune interaction, hormis leur influence commune sur la récompense du système.
- Les MDP locaux sont unichaînes. Le comportement transitoire initial est sans importance sur le long terme.

Elles réduisent l'expressivité du MMDP, mais restent cependant acceptables pour certains types de problèmes comme ceux cités à la section 3.7.2.2.

En comparant la méthode de résolution à la précédente (section 3.7.1), on retrouve une qualité recherchée : cette méthode distribue la recherche des solutions optimales et non les

14. Par hypothèse, cette mesure est unique.

solutions. Le MDP global est décomposé en MDP locaux qui sont a priori plus faciles à résoudre.

La localisation du MDP a cependant un sérieux coût. Les agents doivent être capables, lorsque leur politique individuelle est fixée, de calculer leur distribution invariante. Cela exige la connaissance parfaite de la loi d'évolution des agents (puisque les probabilités de transition sont nécessaires pour le calcul des mesures invariantes), mais aussi la capacité de résoudre des problèmes linéaires.

De plus, le calcul de la récompense locale r_n est extrêmement coûteux, du point de vue calculatoire, si les agents sont nombreux (Cf. (3.24)). Une solution envisageable pour réduire son coût est de le baser sur un petit échantillon représentatif des agents, plutôt que leur totalité.

3.8 Conclusion

Dans ce chapitre nous avons présenté le formalisme des MDP et étudié sa pertinence pour représenter les SMA. Essentiellement les MDP permettent d'exprimer des problèmes où un système est confronté à des choix d'action, rémunérées ou pénalisées par une fonction de satisfaction. De ce fait, ils expriment de manière satisfaisante des problèmes de conception de systèmes multi-agents où on cherche à définir les comportements des agents en vue de réaliser une fonction adéquate au niveau du système.

Le formalisme des MDP souffre cependant de quelques limitations. D'une part, il exige que les agents aient des connaissances importantes sur l'état du système, mais aussi sur le modèle (loi d'évolution et récompense). Dans le contexte des systèmes multi-agents qui nous intéresse, ces informations ne sont généralement pas à la portée des agents :

- les agents peuvent ignorer une grand partie de l'état du système.
- le système peut être soumis à un bruit externe ou un environnement dynamique, inconnu des agents, ce qui rend sa loi d'évolution exacte inconnue.
- la loi d'évolution du système peut être hautement non linéaire vis-à-vis des actions des agents, et difficile à prévoir.

D'autre part, même lorsque les agents ont une connaissance complète de l'état du système et du modèle, il n'est pas clair qu'une stratégie optimale puisse être déterminée en temps raisonnable. De manière générale croît fortement avec la taille du système, et il n'existe à ce jour aucune théorie générale qui permet d'exploiter la structure spécifique du système de manière efficace. Des propositions ont été faites pour quelques cas très particuliers, mais leur applicabilité reste limitée.

À défaut de fournir des méthodes utilisables pour calculer des comportements optimaux, les MDP permettent d'inscrire les systèmes multi-agents dans une famille de problèmes d'optimisation bien définis. Cette classification explique sans doute l'intérêt croissant pour les MDP décentralisés au cours des dernières décennies, mais n'apporte pas de réponses utilisables en pratique pour construire des systèmes multi-agents.

4 Méthodes de champ moyen

4.1 Introduction

Dans le chapitre 3, nous avons étudié la pertinence des MDP pour l'étude des systèmes multi-agents. Une des conclusions importantes est que ces méthodes ne sont pas utilisables pour des grands systèmes, en raison de la forte croissance de la taille du système vis-à-vis du nombre d'agents.

Une méthode envisageable pour contourner cette difficulté croissante est de réaliser des approximations statistiques du système lorsque le nombre d'agents tends vers l'infini. Sous certaines hypothèses, qui seront précisées plus loin dans ce chapitre, le système converge vers une certaine limite appelée *champ moyen*.

Un avantage important de cette approche est que le champ moyen est indépendant du nombre d'agents (celui-ci étant supposé infini), et ce dernier n'influence donc pas la difficulté de son étude formelle. Ce point est un avantage énorme vis-à-vis des MDP.

Ce chapitre décrit précisément l'approche champ moyen qui a commencé au travail séminal de Kurtz [89] et qui a été repris par [93, 63] entre autres. L'approche générale consiste en deux phases : l'une est la description du système par un modèle stochastique avec des interactions à travers la densité, l'autre est l'approximation de ce modèle stochastique par un champ déterministe appelé *champ moyen*.

4.1.1 Présentation de l'approche

Les méthodes de champ moyen sont issues de la physique, et n'ont pour l'instant aucune formalisation mathématique globale. Les origines sont diverses : magnétisme, mécanique statistique, physique moléculaire, biologie, et au jour d'aujourd'hui elles ne cessent de trouver des applications surprenantes dans des domaines tels que l'optimisation des réseaux TCP [5], les phénomènes de propagation d'une rumeur [35], ou encore la théorie des jeux [92].

4.1.2 Systèmes de particules en interaction moyenne

La théorie des champs moyens repose sur l'idée statistique suivante : lorsqu'une particule subit l'interaction d'un grand nombre d'autres particules, alors nous pouvons appro-

cher l'effet de l'ensemble de ces interactions par la moyenne. Ainsi, chaque particule du système est soumise à un champ d'interaction moyen, résultant de l'ensemble des interactions du groupe. En définitive, un système de N particules avec des interactions mutuelles (éventuellement N^2) est identifié à un système de N particules interagissant avec une seule et même entité macroscopique. Cela sous-entend que le poids accordé à chaque interaction binaire est inversement proportionnel au nombre de particules constituant le système. Ce modèle est donc limité aux systèmes avec des interactions bi-particules relativement faibles.

Remarque 10. *Le terme **particule** dans cette description fait allusion aux constituants élémentaires du système. Pour les systèmes multi-agents il s'agira des agents.*

À titre d'exemple, considérons un ensemble de particules **identiques** distribuées dans l'espace en interaction gravitationnelle. Selon la loi de Newton, chaque paire de particules exerce une force mutuelle inversement proportionnelle à la distance qui les sépare au carré

$$\vec{F}_{m,n} = -\frac{K}{dist(m,n)^2}\,\vec{e}_{m,n}$$

où $\vec{e}_{m,n}$ est le vecteur unitaire orientée de la particule m vers la particule n. La force totale exercée sur une particule n donnée vaut

$$\vec{F}_{tot,n} = \sum_{m \neq n} \vec{F}_{m,n}$$

et ne dépend que de la distribution des particules dans l'espace.

Pour un exemple où les particules sont des individus, considérons une formation d'oiseaux migrateurs en vol. Selon le modèle de Cucker-Smale [42], chacun de ces oiseaux ajuste sa vitesse en y additionnant une moyenne pondérée de la vitesse relative de tous les autres oiseaux

$$v_n(t+1) = v_n(t) + \sum_{m \neq n} \alpha_{n,m} \cdot (v_n(t) - v_m(t)).$$

où les coefficients de pondération $\alpha_{m,n}$ dépendant de la distance séparant les individus n et m. L'influence totale sur un individu donné ne dépend donc que de

- la distribution des individus dans l'espace, et de
- la distribution des vitesses des individus dans l'espace.

Dans ces deux exemples les interactions des individus s'expriment à travers la densité, ce qui les rend adaptés à une approche de type champ moyen.

4.1.3 Approximation par un champ déterministe

Une fois décrit le système à l'aide d'un modèle stochastique avec interactions à travers la densité, il s'agira de faire tendre progressivement le nombre d'agents vers l'infini. Sous certaines hypothèses [1], on obtient un système limite déterministe appelé *champ moyen*. La loi d'évolution de ce système limite a l'avantage d'être régi par un (petit) nombre d'équations indépendant du nombre d'agents.

1. portant sur la loi d'évolution des agents, mais aussi sur leurs conditions initiales

Le résultat de l'approximation est analogue à la loi forte des grands nombres, à une différence près : les agents ne sont pas supposés indépendants à chaque instant. Lorsque les conditions initiales des agents sont indépendantes, on peut cependant prouver qu'à chaque instant, tout ensemble fini d'agents est asymptotiquement indépendant. Ce phénomène est connu sous le nom *propagation du chaos* [139].

Un point délicat est que le fait de faire augmenter le nombre d'agents du système peut demander un ajustement des paramètres internes. Faire varier les paramètres endogènes d'un système avec le nombre d'agents est une tâche relativement arbitraire, appelé *scaling*.

Un scaling est généralement réalisé pour des raisons de cohérence, comme par exemple augmenter la taille d'une zone pour pouvoir augmenter nombre d'agents qui s'y trouvent. Une autre motivation pour réaliser un scaling est de préserver l'intérêt du problème : "Comment doit-on lier les paramètres du système au nombre d'agents pour que la situation limite soit intéressante ?".

Un exemple historique de ce principe est celui de Boltzmann-Grad [91] : un gaz est assimilé à un ensemble de N sphères identiques de diamètre d, rebondissant de façon rigide les unes sur les autres. La limite de Bolzmann-Grad consiste à étudier le modèle limite $N \to \infty$ en supposant que $Nd^2 \to Cste$. D'autres exemples de scaling sont donnés dans les exemples de la section 4.3.

4.2 Modèle stochastique avec interactions à travers la densité

Dans cette partie, nous allons introduire le modèle markovien qui sera utilisé pour décrire les systèmes multi-agents étudiés. La description est faite aussi bien au niveau des agents qu'au niveau global. Les ingrédients essentiels sont

- les instants,
- les états,
- la ressource commune,
- la loi d'évolution.

qui peuvent être définies au niveau global (macroscopique) et au niveau des agents (microscopique).

Quelques discussions sur la nature de ces ingrédients seront menées. Nous terminons cette section par une vue d'ensemble succincte qui résume l'essentiel du formalisme introduit.

4.2.1 Formalisme

Dans ce premier paragraphe, nous définissions les différents éléments du modèle stochastique avec interactions à travers la densité. Ces définitions sont accompagnées de quelques commentaires sur leur nature et sur les choix qui ont été faits.

Instants Les différents instants auxquels le système et ses constituants (les agents) sont susceptibles d'évoluer sont supposés discrets[2], et fixés. On les notera $0, 1, 2, \ldots$ dans l'ordre chronologique, et on désignera par la lettre t un instant générique.

Il est également supposé que les agents évoluent de manière synchrone. Il faut donc que tous les agents aient la même notion du temps, et qu'ils partagent une horloge centralisée. Les instants sont définis de façon universelle, et perçus de façon identique à chaque niveau du système.

Agents et états Soit N le nombre d'agents constituant le système. Ce nombre est supposé constant au cours du temps, et il sera supposé que les agents du système sont toujours les mêmes (le système est fermé).

De plus, tout au long de cette partie les agents sont identiques : ils ont le même espace d'états, et les même règles de transition, et le système est obtenu comme N copies d'un même type d'agent. Ces agents sont numérotés par les entiers $\{1, \ldots, N\}$, et on note $s_n(t)$ l'état de l'agent n à l'instant t.

L'espace d'états commun aux agents, qui est invariant au cours du temps, est noté $\mathcal{S}$. Dans cette partie, son cardinal est supposé fini et noté $S := Card(\mathcal{S})$. Ainsi, à tout instant $t \geq 0$:

- Chaque agent $n \in \{1, \ldots, N\}$ a un état $s_n(t) \in \mathcal{S}$.
- Le système a un état $S^N(t) = (s_1(t), \ldots, s_N(t))$ dans $\mathcal{S}^N$. L'indice N est conservé dans la notation de l'état global, en vue de l'asymptotique $N \to \infty$. Pour des raisons de lourdeur nous omettrons volontairement de signaler cet indice sur les états des agents, et notons $s_n(t)$ l'état de l'agent n plutôt que $s_n^N(t)$.

Une première difficulté provient du fait que l'espace d'états du système $\mathcal{S}^N$ croît exponentiellement avec le nombre d'agents, ce qui présage une difficulté croissante pour le décrire. L'échangeabilité des agents permet de réduire cette taille, et d'étudier plutôt la densité d'occupation des agents en chaque état (ou, de façon équivalente, la mesure empirique du système). Le paragraphe suivant a pour but de définir ces notions.

Densité d'occupation Les agents du système sont identiques, et ne peuvent donc être distingués qu'à travers leurs états. Intervertir les agents ne change donc pas l'état global du système, ce qui permet de définir la relation d'équivalence :

$$(s_1, \ldots, s_N) \simeq (s'_1, \ldots, s'_N) \iff \exists \sigma \in \mathfrak{S}_N \,,\, \forall n \in \{1, \ldots, N\} \,,\, s_{\sigma(n)} = s'_n.$$

Deux états sont équivalents si l'un est obtenu à partir de l'autre par permutation des agents.

L'espace d'états naturel du système est alors $\mathcal{S}^N / \simeq$, c'est-à-dire l'ensemble des états de tous les agents *à permutation près*. Le lemme de factorisation permet de faire l'identification de ces états :

$$\mathcal{S}^N / \simeq = \mathcal{P}_N(S)$$

2. Ce choix nous éloigne déjà de la plupart des situations rencontrées en physique pour lesquelles le temps est continu.

où

$$\mathcal{P}_N(S) := \left\{ (v_i)_{i=1}^S \in \mathbb{N}^S \mid \sum_{i=1}^S v_i = N \right\}$$

désigne l'ensemble des partitions de l'entier N en somme de S entiers.

Cette identification revient à dire que l'état du système est complètement déterminé par le nombre d'agents en chaque état. A l'instant t, le nombre d'agents dans l'état i est noté :

$$N_i(t) := \sum_{i=1}^N \mathbb{1}_{s_n(t)=i}$$

et on pose :

$$M_i^N(t) := \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{s_n(t)=i}$$

La quantité $M_i^N(t)$ représente la proportion (ou *la densité*) d'agents dans l'état i à l'instant t. L'état du système est complètement décrit par le vecteur de densités d'occupation

$$M^N(t) = (M_i^N(t))_i$$

Ce vecteur est sans unité, invariant par permutation des agents, et stochastique : ses coefficients sont positifs et ont pour somme 1. Il prend ses valeurs dans l'ensemble $\frac{1}{N}\mathcal{P}_N(S)$. Pour chaque valeur de N, cet ensemble est inclus dans

$$\mathcal{P}(S) = \left\{ (v_i)_{i=1}^S \in (\mathbb{R}^+)^S \ / \ \sum_{i=1}^S v_i = 1 \right\},$$

l'ensemble des vecteurs stochastiques à S composantes.

L'ensemble $\mathcal{P}_N(S)$ a pour cardinal $\begin{pmatrix} N+1 \\ S-1 \end{pmatrix}$, qui tend toujours vers $+\infty$ lorsque $N \to +\infty$. En revanche, nous pouvons dire que $\frac{1}{N}\mathcal{P}_N(S) \to \mathcal{P}(S)$ (au sens de la convergence des ensembles de Hausdorff [110]). Cette convergence est un point important : l'ensemble des états du système devient infiniment grand, mais tend vers l'ensemble compact $\mathcal{P}(S)$.

Ressource du système Les agents ont connaissance du système via les densités d'occupation de chaque état (le vecteur $M^N(t)$) et une *ressource commune* notée $R^N(t)$ à l'instant t. Il s'agit d'une grandeur macroscopique, qui prend ses valeurs dans un espace vectoriel $\mathbb{R}^d$ d'une certaine dimension d fixe au cours du temps.

La variable ressource influence l'évolution des individus et, réciproquement, évolue en fonction de l'état du système. Dans les exemples qui suivent (Cf. partie 4.6.1.4) nous montrons qu'elle permet de conférer une certaine mémoire au système, en conservant une trace numérique des événements. La ressource peut également faire office de variable d'environnement (Cf. partie 6.5.2), dans le sens où elle incarne une entité macroscopique passive, avec laquelle le système interagit.

La ressource se renouvelle à chaque instant en fonction de son état précédent et de l'état courant du système, c'est-à-dire : il existe une fonction $g : \mathbb{R}^d \times \mathcal{P}_N(S) \to \mathbb{R}^d$ telle que :

$$R^N(t+1) = g\left(R^N(t), M^N(t+1)\right)$$

La fonction g sera supposée **continue**.

Remarque 11. *Lorsque le système évolue de façon aléatoire la ressource est aléatoire. Elle dépend cependant de manière déterministe de l'état du système.*

Loi d'évolution du système, et lois d'évolution marginales Les différents agents du système évoluent en prenant en compte l'état du système, et une ressource commune. L'état du système, décrit par le vecteur $S^N(t) = (s_1(t), \ldots, s_N(t))$ (ou encore la densité $M^N(t)$) ne peut donc être un processus stochastique markovien à proprement parler, car ses transitions dépendent de la ressource $R^N(t)$, qui peut avoir des dépendances vis-à-vis des états antérieurs du système. En revanche, nous supposons que le vecteur $(S^N(t), R^N(t))$ suit un processus aléatoire Markovien [3], d'espace d'états $\mathcal{S}^N \times \mathbb{R}^d$.

Nous admettons que la loi d'évolution du système se factorise comme un produit de lois d'évolution individuelles sous la forme :

$$\begin{aligned}\mathbb{P}\left(S^N(t+1) = s' \mid S^N(t) = s, R^N(t) = r\right) \\ = \prod_{n=1}^{N} \mathbb{P}\left(s_n(t+1) = s'_n \mid s_n(t) = s_n, M^N(t) = m, R^N(t) = r\right)\end{aligned}$$

où m est le vecteur de densités d'occupation associé à l'état s.

Écrire une telle factorisation revient à supposer que :
- la loi d'évolution est homogène en temps, et ne distingue pas les agents,
- les agents évoluent en n'observant que les densités des agents en chaque état et la ressource commune,
- les transitions des agents sont statistiquement indépendantes.

La matrice de transition de l'agent n, sachant que $M^N(t) = m$ et $R^N(t) = r$ est notée $P^N(m, r)$. Elle est formée des $S \times S$ coefficients

$$p^N_{i,j}(m, r) = \mathbb{P}\left(s_n(t+1) = j \mid s_n(t) = i, M^N(t) = m, R^N(t) = r\right)$$

Dans la suite, on notera

$$p^N_{i,j} : \begin{pmatrix} \mathcal{P}(S) \times \mathbb{R}^d & \to & \mathbb{R} \\ (m, r) & \mapsto & p^N_{i,j}(m, r) \end{pmatrix}$$

et

$$P^N : \begin{pmatrix} \mathcal{P}(S) \times \mathbb{R}^d & \to & \mathcal{M}_S(\mathbb{R}) \\ (m, r) & \mapsto & P^N(m, r) \end{pmatrix}$$

les fonctions sous-jacentes.

Dans le cas particulier où les transitions des agents ne dépendent que de l'état courant du système la matrice de transition est simplement notée $P^N(m)$. On dit alors que le système est *sans mémoire*.

3. Ce qui entraîne que le processus $(M^N(t), R^N(t))_t$ est Markovien

4.2.2 Bilan

Résumons les notations introduites. Pour un système de N agents, le modèle stochastique qui a été défini évolue comme un processus markovien $(S^N(t), R^N(t))_t$ à valeurs dans $\mathcal{S}^N \times \mathbb{R}^d$ tel que :

- les agents évoluent de manière synchrone à des instants discrets,
- le vecteur $M^N(t)$ contient les densités d'occupations des agents sur les états,
- chaque composante du vecteur $S^N(t)$ (chaque agent) évolue selon la même règle, qui dépend de la densité d'occupation $M^N(t)$ et le vecteur $R^N(t)$,
- la nouvelle ressource $R^N(t+1)$ est calculée à partir de sa valeur précédente $R^N(t)$ et la densité d'occupation courante $M^N(t+1)$.

L'évolution de ce modèle est illustrée par le schéma suivant :

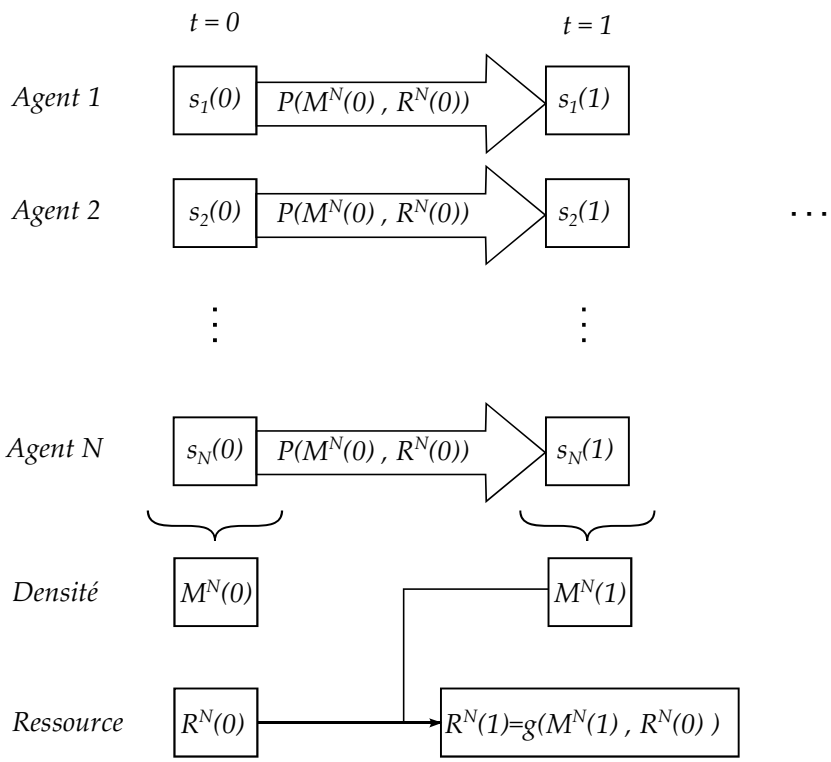


Figure 4.1 — Évolution du modèle stochastique avec interactions moyennes

4.3 Mise en oeuvre sur des exemples

Avant de donner le théorème de convergence de champ moyen, nous présentons deux situations concrètes qui sont modélisées à l'aide du formalisme introduit ci-dessus, afin de montrer l'applicabilité du modèle.

Pour chaque exemple, nous donnons la limite (le champ moyen) correspondant à la situation théorique où le nombre d'individus est infini. Cette limite est rigoureusement décrite dans la partie 4.4 qui suit.

4.3.1 Problème de trafic dans un ensemble de couloirs

Considérons la situation suivante : deux pièces sont reliées par un certain nombre de couloirs. Un groupe de personnes se trouve dans l'une des deux pièces et désire se rendre dans l'autre mais les couloirs ont une certaine capacité et ne permettent pas à toutes de passer simultanément.

Modélisation Cette situation peut être décrite à l'aide du modèle introduit à la partie 4.2.1. Les agents sont les N personnes se trouvant dans la pièce initiale. L'entier C désigne le nombre de couloirs.

Afin de prendre en compte la contrainte imposée par les capacités des couloirs, nous distinguons les états suivants :

- (1) : La personne se trouve dans la pièce initiale et ne demande à traverser aucun couloir. Initialement, tous les agents sont dans cet état.
- (D_c), $c \in \{1, \ldots, C\}$: La personne demande à traverser le couloir c.
- (T_c), $c \in \{1, \ldots, C\}$: La personne est parvenue à traverser le couloir c, et se dirige vers la seconde pièce.
- (2) : La personne se trouve dans la seconde pièce.

Ces états permettent de différencier les différentes phases d'une migration de la pièce 1 vers la pièce 2, en séparant l'intention et l'action effective de traverser (qui est soumise à la capacité des couloirs). Cette distinction s'appuie sur le principe d'*influence-réaction* [39], discuté à la section 2.1.3. Les agents formulent une intention (*influence*), en sollicitant un couloir, la *réaction* de l'environnement est le passage ou non selon la sollicitation totale.

Les états intermédiaires (T_c) permettent de mesurer le nombre de passages réussis. Cette information est utilisée ultérieurement (section 4.6.1) afin de représenter la réussite du système.

À un instant t donné on note, de manière transparente, $M_1^N(t)$, $M_{D_c}(t)$, $M_{T_c}(t)$ et $M_2^N(t)$ les proportions d'agents dans chacun des états. Les différentes transitions sont :

- $(1) \to (D_c)$: Une personne dans la pièce initiale demande à traverser le couloir c. Cette transition s'effectue avec une certaine probabilité a_c supposée fixe pour l'instant[4].
- Une personne qui demande à traverser un couloir c a deux transitions possibles :
  - $(D_c) \to (T_c)$ elle parvient à traverser avec une probabilité $p_c^N(M_{D_c}^N(t))$, qui dépend de la sollicitation totale du couloir c,
  - $(D_c) \to (1)$ en cas d'échec elle revient à l'état inactif (1).
- $(T_c) \to (2)$ une personne qui est parvenue à traverser se rend dans la seconde pièce.

Le graphe des transitions, sachant que $M^N(t) = m$, est représenté sur la figure 4.2.

4. Nous l'ajusterons ultérieurement (section 4.6.1) afin de maximiser le nombre de passages.

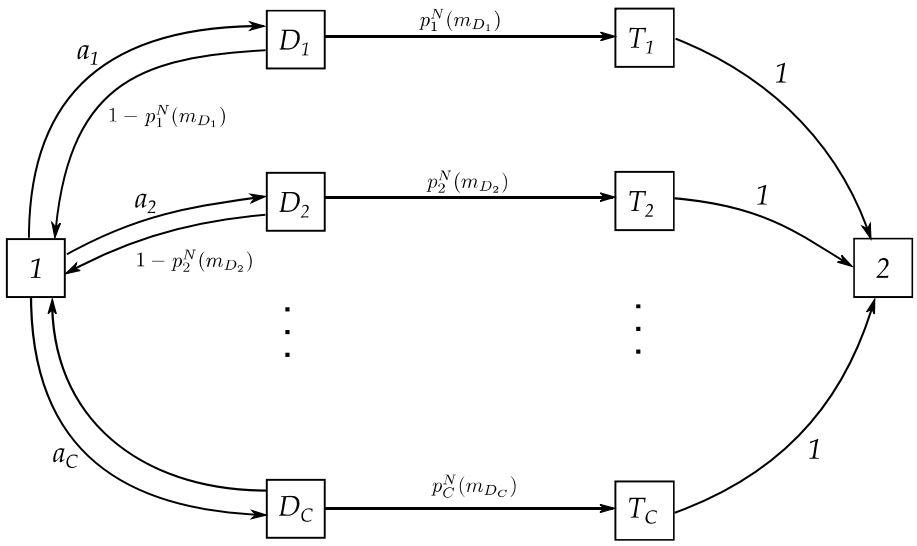


Figure 4.2 — Diagramme des transitions

L'évolution du système est sans mémoire. Si les états sont ordonnés selon $(1, D_1, \ldots, D_C, T_1, \ldots, T_C, 2)$, la matrice de transition d'un agent, sachant que $M^N(t) = m$, est

$$P^N(m) = \left(\begin{array}{c|ccc|cccc|c}
1 - \sum_{c=1}^{C} a_c & a_1 & \cdots & a_C & 0 & & & & 0 \\
\hline
1 - p_1^N(m_{D_1}) & 0 & & 0 & p_1^N(m_{D_1}) & 0 & & & \\
1 - p_2^N(m_{D_2}) & 0 & & & 0 & p_2^N(m_{D_2}) & & & \vdots \\
 & & & & & \ddots & & & \\
1 - p_C^N(m_{C+1}) & 0 & \ldots & & & 0 & & p_C^N(m_{C+1}) & 0 \\
\hline
0 & & \ldots & & & \ldots & & 0 & 1 \\
\vdots & & & & & & & & \\
0 & & \ldots & & & \ldots & & 0 & 1
\end{array}\right)$$

Remarque 12. *Un agent a la possibilité de rester dans l'état* (1) *d'un instant à l'autre (avec probabilité* $1 - \sum_{c=1}^{C} a_c$*). Une telle attitude peut être intéressante pour éviter l'encombrement dans les couloirs.*

Limite Champ Moyen En appliquant le théorème de convergence 5 (section 4.4.3) on montre que si la matrice de transition $P^N(m)$ converge vers une matrice $P(m)$ lorsque $N \to \infty$, le vecteur de densité peut être approché par le champ moyen μ défini itérativement par

$$\begin{cases} \mu(t+1) = \mu(t) \cdot P(\mu(t)) \\ \mu(0) = (1, 0, \ldots, 0) \end{cases}.$$

Ce champ moyen possède plusieurs avantages sur le système discret. D'une part, son évolution est régie par seulement $2C + 2$ équations, contre N lois d'évolution aléatoires pour le système aléatoire discret. Lorsque N est grand, le champ moyen offre donc une simplification intéressante pour l'étude formelle.

D'autre part, le champ moyen est déterministe, ce qui simplifie grandement la prédiction du système. En effet, à un instant futur, le champ moyen ne possède qu'un et un seul état. Pour le système aléatoire, plusieurs états futurs sont possibles, et pour les prédire il est nécessaire d'évaluer les probabilités de tous ces états.

4.3.2 Situation issue de la robotique

Cet exemple est issu de [79]. Un ensemble de robots identiques sont disposés dans une zone circulaire délimitée par une paroi infranchissable. La tâche des robots consiste à extraire un certain nombre de bâtons cylindriques partiellement enfoncés dans la surface de la zone (voir figure 4.3). Comme les robots ont les bras trop courts pour extraire les bâtons seuls, ils doivent coopérer avec un autre robot pour y parvenir. Une fois extrait le bâton, il est remis dans le trou et les robots poursuivent leur activité.

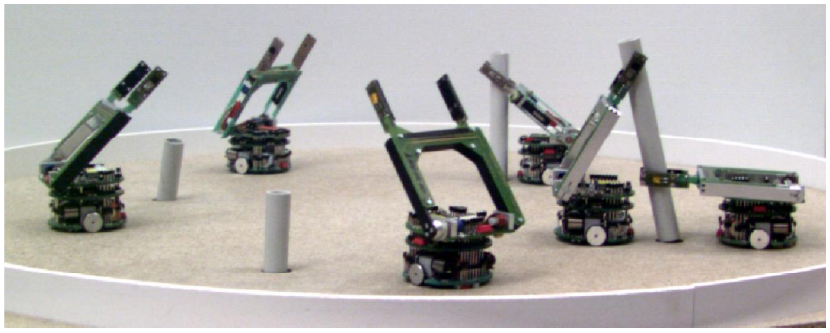


Figure 4.3 — Robots extracteurs de bâton (image issue de [79])

Modélisation Comme dans [93] nous représentons cette situation à l'aide du modèle introduit à la partie 4.2.1. Les N agents sont les robots, et ont deux états possibles :

- C : Chercher un bâton à tirer ou un autre robot à assister. Dans cet état, les robots se déplacent aléatoirement [5] dans la zone, en évitant les autres robots. Initialement, tous les robots sont dans cet état.
- A : Tenir un bâton et attendre de l'aide. Dans cette phase les robots sont inactifs.

De manière transparente, nous notons $M_C(t)$ et $M_A(t)$ la proportion de robots dans chacun de ces états.

Les transitions possibles sont :

- $C \to A$: Un robot à la recherche d'un bâton libre en trouve un et l'attrape.
  La probabilité de cette transition est :

$$p^N_{C,A}(M^N(t)) = (B - N \cdot M^N_A(t)) \cdot P$$

  où B désigne le nombre de bâtons, $B - N \cdot M_A(t)$ le nombre de bâtons libres et P la probabilité de rencontrer un bâton donné. Cette probabilité est représentée par la proportion d'aire occupée par un bâton relativement à l'aire de la zone.
- $A \to C$: Un robot qui tient un bâton peut revenir à l'état C de deux façons :

5. Plus de détails sont donnés dans [79].

- il est aidé par un autre robot avec probabilité

$$N \cdot M_C^N(t) \cdot P \cdot \nu$$

  où le facteur ν représente le fait que le robot assistant doit se présenter sous un angle convenable pour une extraction à deux.
- il abandonne son bâton et revient à l'état C. Cette transition s'effectue avec une certaine probabilité a. Dans [93, 79] cette probabilité est notée $1/T$, et interprétée comme l'inverse d'un temps d'attente T.

Par conséquent, les probabilités de transition valent :

$$p_{A,C}^N(M^N(t)) = N \cdot M_C^N(t) \cdot P \cdot \nu + a$$

Scaling Le taux d'occupation de l'état A est majoré par le nombre de bâtons par robot : $M_A^N(t) \leq \frac{B}{N}$. Faire grandir le nombre de robots en laissant constant le nombre de bâtons réduit donc fatalement l'état A à être inoccupé.

Nous allons donc supposer que le ratio $\frac{B}{N}$ converge vers une certaine constante $b > 0$. En d'autres termes, le nombre de bâtons est asymptotiquement proportionnel au nombre de robots.

De la même façon, augmenter le nombre de robots (qui sont de taille constante) sans augmenter la taille de la zone n'est pas raisonnable pour des contraintes évidentes de volume. Nous supposons donc l'aire de la zone asymptotiquement proportionnelle au nombre de robots, ce qui entraîne que :

$$P = \frac{\text{Taille(Robot)}}{\text{Taille(Zone)}} \simeq \frac{p}{N}$$

pour une certaine constante p.

Limite Champ Moyen En tenant compte des hypothèses de scaling ci-dessus, les probabilités de transition de chaque agent tendent vers :

$$p_{C,A} = (b - m_A(t)) \cdot p \quad \text{et} \quad p_{A,C} = m_C(t)\nu p + a$$

lorsque le nombre d'agents tend vers l'infini, et le diagramme des transitions asymptotique est le suivant :

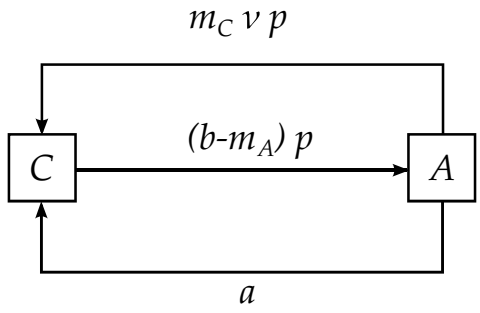


Figure 4.4 — Diagramme des transitions asymptotique

Comme $0 \leq m_A(0) \leq b$, la valeur initiale de $p_{C,A}$ définit bien une probabilité. Pour éviter que cette probabilité de la transition ne dépasse 1 par la suite, nous ajoutons la condition

$$a \leq 1 - \nu p,$$

qui garantit que $p_{A,C} = m_C(t)\nu p + a \in [0,1]$.

L'évolution du système est sans mémoire, et la matrice de transition marginale d'un agent tend vers

$$P(m) = \begin{pmatrix} 1-(b-m_A)\cdot p & (b-m_A)\cdot p \\ m_C\nu p + a & 1 - m_C\nu p - a \end{pmatrix}$$

En appliquant le théorème 5 (section 4.4.3), on montre que le vecteur de densité d'occupation $M^N(t)$ converge, lorsque N tend vers l'infini, vers le champ moyen μ défini itérativement par :

$$\begin{cases} \mu(t+1) = \mu(t)\cdot P(\mu(t)) \\ \mu(0) = (1,0) \end{cases}.$$

Comme pour l'exemple précédent, ce champ moyen possède des avantages considérables par rapport au système aléatoire. D'une part, son évolution est décrite par deux équations seulement, contre N probabilités de transition pour le système aléatoire. D'autre part, le champ moyen est déterministe, ce qui offre un grand avantage pour la prévision de l'état futur du système.

4.4 La limite champ moyen

Dans cette partie, nous présentons le théorème de convergence vers le champ moyen ainsi que quelques résultats annexes.

4.4.1 Le théorème de convergence vers le champ moyen

Avec les notations introduites précédemment, on peut énoncer le résultat clé de convergence :

Théorème 5. *On considère le modèle Markovien décrit à la partie 4.2.1. On suppose que, lorsque $N \to \infty$:*

(i) Les conditions $M^N(0)$ et $R^N(0)$ convergent presque sûrement vers des vecteurs $\mu(0)$ et $\rho(0)$.

(ii) La suite de fonctions $(P^N)_N$ converge uniformément vers une fonction continue

$$P : \mathcal{P}(S) \times \mathbb{R}^d \to \mathcal{M}_S(\mathbb{R}).$$

Alors pour tout $t \geq 0$, $M^N(t) \xrightarrow{PS} \mu(t)$ et $R^N(t) \xrightarrow{PS} \rho(t)$, où les suites ρ et μ sont définies récursivement par

$$\begin{cases} \mu(t+1) = \mu(t)\cdot P(\mu(t),\rho(t)) \\ \rho(t+1) = g(\rho(t),\mu(t+1)) \end{cases} \tag{4.1}$$

Définition 1. *Les équations du système* (4.1) *sont appelées* équations champ moyen, *et la suite* $(\mu(t))_t$ *est appelée* champ moyen.

Démonstration

Le théorème 5 est démontré dans [93]. □

Le théorème 5 montre donc que, si
- les états initiaux des agents convergent,
- leurs probabilités de transition convergent,

lorsque le nombre d'agents tend vers l'infini, alors la densité des agents tend vers une quantité déterministe appelée *champ moyen*.

Les avantages de ce champ moyen ont été constatés sur les exemples. D'une part, le nombre d'équations qui régissent ce champ moyen est égal au nombre d'états, et donc invariant devant le nombre d'agents. Ce fait permet de contourner l'explosion combinatoire lorsque le nombre d'agents tend vers l'infini.

D'autre part, le caractère déterministe du champ moyen offre un avantage pour la prévision. À un état initial donné correspond un et un seul état futur. Le système stochastique ne possède pas cet avantage, et ses états futurs ne peuvent être prédits qu'à des fluctuations aléatoires près.

Remarque 13.
- *L'hypothèse (i) est en particulier vraie si pour tout N, $M^N(0) = \mu(0)$ et $R^N(0) = \rho(0)$, c'est-à-dire lorsque les densités initiales et la ressource initiale sont indépendantes du nombre d'agents.*
- *L'hypothèse (ii) équivaut à la convergence uniforme de chaque probabilité de transition : Pour tout $(i,j) \in \mathcal{S}^2$, la suite $(p^N_{i,j})_{i,j}$ converge uniformément vers une fonction continue*

$$p_{i,j} : \mathcal{P}(S) \times \mathbb{R}^d \to \mathbb{R},$$

et la matrice de transition limite est définie par

$$P(m,r) := \big(p_{i,j}(m,r)\big)_{i,j}$$

Une question pertinente est de savoir à quelle vitesse le système converge vers ce champ moyen. La réponse à cette question permettrait de connaitre la qualité de l'approximation champ moyen pour un nombre d'agents donné, ou encore d'évaluer le nombre d'agents nécessaire pour avoir une approximation à un seuil de précision donné. Dans [62] on trouve une réponse partielle à cette question :

Théorème 6. *Si, en plus des hypothèses du théorème 5, on suppose les fonctions g et P lipschitziennes, alors pour tout $t \geq 0$, il existe des constantes α_t et β_t telles que*

$$\mathbb{E}\left(\left\|(M^N(t), R^N(t)) - (\mu(t), \rho(t))\right\|\right) \leq \frac{\alpha_t}{\sqrt{N}} + \beta_t\, \mathbb{E}\left(\left\|(M^N(0), R^N(0)) - (\mu(0), \rho(0))\right\|\right)$$

Par conséquent la vitesse de convergence vers le champ moyen à un instant t fixé est de l'ordre de $\frac{1}{\sqrt{N}}$, soit inversement proportionnelle à la racine carrée du nombre d'agents [6]. Le théorème 6 donne cependant peu de guides sur le nombre d'agents nécessaire pour valider l'approximation par le champ moyen, puisque les coefficients α_t,β_t dépendent a priori de l'instant auquel le système se trouve.

4.4.2 Évolution d'un individu isolé

Le théorème d'approximation 5 donne un résultat sur l'évolution macroscopique de la densité du système. Nous donnons ici deux conséquences de ce résultat sur l'évolution des individus, sans trop de formalisme.

Le premier résultat est issu de [93] : dans le système décrit dans la partie 4.2.1, prenons une particule n. On suppose que son état initial $s_n(0)$ ne dépend pas du nombre d'agents N. Considérons les deux lois d'évolution suivantes :

- à chaque instant, la particule prend un nouvel état s'_n selon $p_{s_n,s'_n}(M^N(t), R^N(t))$.
- à chaque instant, la particule prend un nouvel état s'_n selon $p_{s_n,s'_n}(\mu(t), \rho(t))$.

Il s'avère que lorsque N est grand, ces deux lois d'évolution mènent à la même loi, à condition que les graines utilisées pour les tirages aléatoires soient identiques.

Un autre résultat intéressant, rencontré dans [62, 93, 11] est la *propagation du chaos* [139] :

Si les particules ont des états initiaux $s_1(0), \ldots, s_N(0)$ indépendants, alors tout ensemble fini d'entre elles $\{i_1, i_2, \ldots, i_k\}$ est asymptotiquement indépendant :

$$\lim_{N\to\infty} \mathbb{P}\left(s_{i_1}(t) = z_1, \ldots, s_{i_k}(t) = z_k\right) = \mu_{z_1}(t) \cdot \ldots \cdot \mu_{z_k}(t)$$

En d'autres termes, les lois de probabilité des individus tendent vers des lois indépendantes et identiques. Cette loi commune est donnée par le champ moyen.

Ces deux résultats permettent d'interpréter le champ moyen au niveau des agents. D'une part, l'évolution d'un agent en interaction avec le champ moyen correspond bien à celle d'un agent en interaction "moyenne" avec le système.

D'autre part, le champ moyen n'approche pas seulement la distribution des agents, mais aussi à la loi de probabilités de chaque agent isolé.

4.4.3 Convergence du champ moyen

La démonstration du théorème 5 proposée dans [93] est réalisée par récurrence sur t. Elle met en évidence une limitation du résultat : il ne sera valable que sur des durées finies. Pour pouvoir rapprocher les comportements du système stochastique et celui du champ moyen à très long terme, il faut que les limites (dans la mesure où elles ont du sens)

$$\lim_{t\to\infty} \lim_{N\to\infty} M^N(t)$$

et

$$\lim_{N\to\infty} \lim_{t\to\infty} M^N(t)$$

6. Cette vitesse est relativement faible, des résultats plus forts (moyennant plus d'hypothèses) peuvent être trouvés dans [62].

soient égales.

Un contre-exemple Considérons un système de N agents répartis sur deux états $\{1,2\}$, selon des densités respectives $M_1^N(t)$ et $M_2^N(t)$, tel que la matrice de transition de chaque agent sachant que $M^N(t) = m$ soit

$$P(M) = \begin{pmatrix} m_2 & m_1 \\ m_2 & m_1 \end{pmatrix}$$

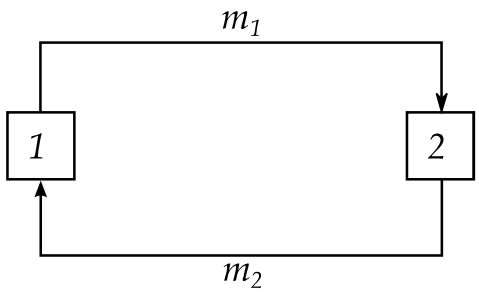


Figure 4.5 — Diagramme des transitions

En moyenne, à chaque instant une proportion de $M_1^N(t) \cdot M_2^N(t)$ d'individus effectue la transition $1 \to 2$, et un proportion identique effectue la transition $2 \to 1$. En moyenne les proportions d'individus en chaque état restent donc constants.

Remarque 14. *Pour une une preuve formelle, il suffit d'écrire l'équation champ moyen :*

$$\mu(t+1) = \mu(t) \cdot \begin{pmatrix} \mu_2(t) & \mu_1(t) \\ \mu_2(t) & \mu_1(t) \end{pmatrix} = \mu(t)$$

qui montre que le champ moyen est constant.

Or, pour le système stochastique, il est possible (avec probabilité non nulle) qu'un des deux états se vide complètement à un instant donné. Ces états ($M_1^N(t) = 0$ ou $M_2^N(t) = 0$) sont des attracteurs stables : une fois qu'ils sont atteints, le système n'évolue plus.

Remarque 15. *Pour un argument plus formel, on peut montrer que la suite $(M_1^N(t))_t$ (à N fixé) est une martingale bornée, et converge donc vers une variable aléatoire L^N sur $\{0,1\}$ [113].*

Dire que les limites $N \to \infty$ et $t \to \infty$ permutent revient à dire que la limite du système stochastique, qui est une loi de probabilités sur $\{(0,1),(1,0)\}$ est égale au champ moyen qui est une constante $\mu(0)$. Cette égalité est exclue pour peu que $\mu(0) \notin \{(0,1),(1,0)\}$.

Des résultats positifs Malgré le contre-exemple que nous venons de présenter, il existe un résultat assez général permettant de relier les mesures invariantes du système stochastique et le régime stationnaire du champ moyen [10]. Nous l'expliquons de manière heuristique :

Soit (L^N) une suite de distribution invariantes pour les chaînes de Markov $(M^N(t))_t$. Si (L^N) converge, alors la limite est nécessairement une position d'équilibre du champ moyen.

En renforçant les hypothèses, on obtient un résultat plus précis [62, 10] :

Si le champ moyen possède une unique position d'équilibre μ_* vers laquelle toutes les trajectoires convergent, alors la suite L_N converge nécessairement vers μ_*.

En d'autres termes : si le champ moyen possède une unique position d'équilibre qui est un attracteur global, alors cet état est une approximation de l'équilibre (statistique) du système stochastique.

Remarque 16. *L'exemple n'est pas en conflit avec ces résultats, puisque tous les points du segment* $\{\lambda(0,1) + (1-\lambda)(1,0)\,,\ \lambda \in [0,1]\}$ *sont des positions d'équilibre pour le champ moyen.*

4.5 Méthodes de champ moyen pour la résolution d'un MDP

Une application intéressante des méthodes de champ moyen est la résolution de certains MDP. Nous avons constaté que le formalisme des MDP permet de représenter des systèmes d'individus avec un objectif, ce qui est important pour la représentation des SMA, mais la difficulté de résolution croît de manière exponentielle avec le nombre d'individus. Cette difficulté a été appelé *la malédiction de la dimension*[7] par Bellmann [9].

À l'opposé des méthodes proposées dans les parties 3.7.1, 3.7.2 qui essayent d'exploiter le caractère local des interactions, [63] et [19] proposent de résoudre les MDP à l'aide d'une approximation de type champ moyen. Pour cela il faut étendre le modèle formel de la partie 4.2.1 pour prendre en compte les actions, c'est-à-dire les paramètres de contrôle du système, ainsi qu'une récompense qui évalue l'accomplissement du système dans sa tâche. Le théorème 5 montre que ce modèle converge vers un champ moyen contrôlé. De plus, si on associe un MDP à ce modèle stochastique, [63] montre que l'équation d'optimalité associée converge vers une équation d'optimalité déterministe associée au champ moyen.

4.5.1 Modèle stochastique contrôlé

Reprenons le modèle formel de la partie 4.2.1 pour y ajouter des actions. On considère un système de N agents identiques évoluant en temps discret $t = 0, 1, \ldots$.

États Chaque agent possède un état dans un espace fini $\mathcal{S}$, de cardinal $S = Card(\mathcal{S})$. On note $s_n(t)$ l'état de l'agent n à l'instant t, et $S^N(t) = (s_1(t), \ldots, s_N(t))$.

Densité d'occupation En raison de l'échangeabilité des agents, l'état du système peut être décrit par le nombre d'individus en chaque état. On note

$$N_i(t) := \sum_{i=1}^{N} \mathbb{1}_{s_n(t)=i}$$

7. En Anglais : *curse of dimensionality*

le nombre d'individus en l'état i à l'instant t et

$$M_i^N(t) := \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{s_n(t)=i}$$

la proportion d'individus dans l'état i.

Actions À chaque instant t, une action collective $A(t)$ est choisie [8]. Cette action appartient à un espace d'actions $\mathcal{A}$ supposé **compact**.

Ressource Les agents évoluent en présence d'une ressource notée $R^N(t)$ à l'instant t. Cette ressource prend ses valeurs dans un espace vectoriel réel $\mathcal{R}^d$ et se renouvelle avec le vecteur de densités d'occupation

$$R^N(t+1) = g\left(R^N(t), M^N(t+1)\right)$$

à l'aide d'une fonction g. Cette fonction sera supposée **continue**.

Loi d'évolution Comme dans la partie 4.2.1 on suppose les transitions des agents aléatoires, indépendantes et de même loi. Cette loi dépend de la densité d'occupation, de la ressource *et de l'action collective*. Chaque agent évolue selon

$$p_{i,j}^N(m,r,a) = \mathbb{P}\left(s_n(t+1) = j \mid s_n(t) = i, M^N(t) = m, R^N(t) = r, A(t) = a\right)$$

qui est la probabilité de passer de l'état i à l'état j lorsque la densité d'occupation vaut m, la ressource vaut r, et l'action collective est a. Le système évolue selon

$$\mathbb{P}\left(S^N(t+1) = s' \mid S^N(t) = s, R^N(t) = r, A(t) = a\right) = \prod_{n=1}^{N} p_{i,j}^N(m,r,a).$$

Par la suite, on notera $P^N(m,r,a) = \left(p_{i,j}^N(m,r,a)\right)_{i,j}$ la matrice de transition d'un agent, et

$$p_{i,j}^N : \begin{pmatrix} \mathcal{P}(S) \times \mathbb{R}^d & \to & \mathbb{R} \\ (m,r,a) & \mapsto & p_{i,j}^N(m,r,a) \end{pmatrix}$$

et

$$P^N : \begin{pmatrix} \mathcal{P}(S) \times \mathbb{R}^d & \to & \mathcal{M}_S(\mathbb{R}) \\ (m,r,a) & \mapsto & P^N(m,r,a) \end{pmatrix}$$

les fonctions sous-jacentes.

Récompense La réussite du système est évaluée par une fonction $r(M^N(t), R^N(t))$ de récompense. Cette fonction est supposée **continue** et **bornée**. Notons que, contrairement aux MDP, la récompense est indépendante de l'action du système.

8. Cette notion d'*action* est dépourvue de sémantique. Il s'agit simplement d'un paramètre de contrôle.

4.5.2 Limite champ moyen

Nous venons de présenter un modèle stochastique similaire à celui introduit à la partie 4.2.1, en ajoutant deux ingrédients nouveaux :

- Un ensemble d'**actions**, permettant de contrôler l'évolution du système.
- Une **fonction de récompense**, permettant d'exprimer la réussite du système vis-à-vis d'un objectif.

Le fait d'ajouter des actions n'invalide pas l'approximation par le champ moyen. En effet, le théorème 5 est démontré pour chaque t fixé, et reste donc valable pour des processus qui ne sont pas homogènes en temps.

Théorème 7. *On considère le modèle Markovien décrit à la partie 4.5.1, et on suppose que*

(i) Les conditions initiales $M^N(0), R^N(0)$ convergent presque sûrement vers des vecteurs $\rho(0)$ et $\mu(0)$.

(ii) La suite de fonctions (P^N) converge uniformément vers une fonction continue

$$P : \mathcal{P}(S) \times \mathbb{R}^d \times \mathcal{A} \to \mathcal{M}_S(\mathbb{R}).$$

Alors pour tout $t \geq 0$, $M^N(t) \xrightarrow{PS} \mu(t)$ et $R^N(t) \xrightarrow{PS} \rho(t)$, où les suites ρ et μ sont définies récursivement par

$$\begin{cases} \mu(t+1) = \mu(t) \cdot P(\mu(t), \rho(t), A(t)) \\ \rho(t+1) = g(\rho(t), \mu(t+1)) \end{cases}$$

À l'aide de ce théorème, on peut approcher le système stochastique par le champ moyen (déterministe). Si le système stochastique correspond à un MDP, ce résultat permet de le rapprocher d'un MDP déterministe, dont la résolution en généralement plus aisée. Les théorèmes des sections suivantes, issus de [63, 62], énoncent précisément ces faits.

4.5.3 Application aux MDP sur un horizon fini

Une conséquence immédiate du théorème 7 est que, pour une suite d'actions $A(0), \ldots, A(T)$ donnée, le gain moyen sur une période finie $[0, T]$

$$G^N_{A(0),\ldots,A(T)} = \mathbb{E}\left(\sum_{t=0}^{T} r(M^N(t), R^N(t))\right)$$

converge vers le gain du champ moyen

$$\Gamma_{A(0),\ldots,A(T)} = \sum_{t=0}^{T} r(\mu(t), \rho(t)).$$

En effet d'après le théorème 7 chaque terme du gain $G_{A(0),\ldots,A(T)}$ converge vers le terme correspondant dans $\Gamma_{A(0),\ldots,A(T)}$.

Une autre conséquence, moins immédiate, est que le gain optimal du système stochastique

$$G^{*N} = \max_{A(0),\ldots,A(T)} G_{A(0),\ldots,A(T)}$$

converge presque sûrement vers le gain optimal du champ moyen

$$\Gamma^* = \max_{A(0),\ldots,A(T)} \Gamma_{A(0),\ldots,A(T)} \tag{4.2}$$

Avec la propriété intéressante suivante : soit $A^*(0), \ldots, A^*(T)$ une suite d'actions optimales pour le champ moyen (c'est-à-dire une suite réalisant le maximum (4.2)). Alors cette suite est asymptotiquement optimale pour le système stochastique.

Théorème 8. *Sous les hypothèses du théorème 7, et avec les notations introduites ci-dessus, les limites suivantes coïncident*

$$\lim_{N\to\infty} G^{*N} = \Gamma^* = \lim_{N\to\infty} G^N_{A^*(0),\ldots,A^*(T)}$$

Le théorème 8 et les observations qui le précèdent permettent d'exploiter le MDP associé au champ moyen :

- D'une part, ce MDP a un espace d'états compact et se résout à l'aide de méthodes d'analyse continue. Ces méthodes permettent en partie de contourner la difficulté exponentielle associée aux MDP discrets (la *malédiction de la dimension*). De plus, le MDP associé au champ moyen est déterministe. L'équation d'optimalité ne fera donc pas intervenir d'espérance sur l'état suivant (Cf. par exemple (3.11)), qui pouvait donner lieu à une grande complexité algorithmique.
- D'autre part, la résolution du MDP associé au champ moyen fournit deux réponses sur la solution du MDP stochastique. Premièrement, il donne le gain optimal du système stochastique, lorsque le nombre d'agents tend vers l'infini. Ensuite les actions optimales du champ moyen sont asymptotiquement optimales pour le système stochastique, et peuvent donc constituer une stratégie intéressante lorsque le système stochastique est grand.

4.5.4 Application au cas dévalué

La deuxième application aux MDP concerne le cas dévalué. On se limite aux stratégies stationnaires, c'est-à-dire les stratégies de la forme $(\Pi, \Pi, \ldots)$. À chaque instant t, l'action choisie est $A(t) = \Pi(M^N(t))$. Le gain dévalué associé à la politique Π s'écrit :

$$G^N_\Pi = \mathbb{E}\left(\sum_{t=0}^{\infty} \delta^t\, r(M^N(t), R^N(t))\right)$$

où $\delta \in]0,1[$ est le facteur de dévaluation. Le gain optimal du système stochastique est noté :

$$G^{*N} = \sup_{\Pi\in\mathcal{A}^{\mathcal{P}(\mathcal{S})}} G^N_\Pi.$$

Comme la politique Π n'est pas nécessairement continue, rien ne garantit que le processus $(M^N(t), R^N(t))_t$ associé à la politique Π vérifie les hypothèses du théorème 5 et qu'il converge vers un champ moyen.

On peut cependant écrire :

$$G_{\Pi}^{N} = \lim_{T\to\infty} \mathbb{E}\left(\sum_{t=0}^{T} \delta^t \, r(M^N(t), R^N(t))\right)$$

et approcher le cas dévalué infini en appliquant successivement le théorème 8 pour les MDP dévalués finis. Dans [63, 62], cet argument est utilisé pour montrer que la value du MDP dévalué peut être approché tend vers la solution Γ^* de l'équation d'optimalité stochastique

$$\Gamma^*(\mu,\rho) = r(\mu,\rho) + \delta \sup_{a\in\mathcal{A}} \Gamma^*\Big(F(\mu,\rho,a)\Big). \tag{4.3}$$

et où la fonction F correspond à l'itération déterministe :

$$F(\mu,\rho,a) = \Big(\mu\cdot P\left(\mu,\rho,a\right), g\Big(\mu\cdot P\left(\mu,\rho,a\right),\rho\Big)\Big).$$

Ce résultat permet d'exploiter l'approximation champ moyen pour résoudre un MDP dévalué. D'une part, il donne une valeur approchée du gain optimal du système stochastique, et des actions asymptotiquement optimales. D'autre part, l'équation d'optimalité (4.3) est déterministe, et se résout à l'aide de méthodes d'analyse continue. L'approche permet donc de contourner en partie la malédiction de la dimension.

4.5.5 Commentaire

Les auteurs de [63] précisent un fait important dans leurs article. Pour tout MDP il existe une stratégie optimale qui est sans mémoire. Cette affirmation est illustrée par le fait que les méthodes de résolution proposées dans les sections 4.5.3 et 4.5.4 fournissent des stratégies optimales sans mémoire.

Comme nous l'avons mentionné dans la section 3.4.5, ces stratégies optimales font généralement intervenir la loi d'évolution du système, ainsi que la fonction de récompense. De ce fait, elles ne sont pas pertinentes pour les systèmes multi-agents étudiés dans cette thèse, dont les agents ignorent généralement la loi d'évolution macroscopique du système ainsi que la fonction de récompense globale.

Il est toutefois possible de mettre les stratégies optimales à profit, sans avoir la totalité des informations qu'elles utilisent à disposition. Les informations manquantes nécessaires pour utiliser la stratégie optimale peuvent par exemple être obtenues en devinant l'état du système (Cf. section 4.6.1.3). Une autre méthode envisageable pour combler le manque d'informations est d'introduire des mécanismes d'apprentissage ou d'adaptation dans les agents. Ce fait est illustré dans l'exemple présenté dans la section 4.6.1.

4.6 Retour sur les exemples

Dans cette partie nous étendons les exemples de la partie 4.3 à des MDP, afin d'illustrer la mise en oeuvre les résultats théoriques de la section 4.5.

4.6.1 Trafic unilatéral dans deux couloirs

Reprenons l'exemple de la partie 4.3.1 dans le cas particulier où il y a deux couloirs. Un groupe de N personnes se trouve dans une pièce, et souhaite se rendre dans une seconde pièce. Pour ce faire, elles peuvent emprunter deux couloirs dont les capacités de passage sont différentes.

4.6.1.1 Notations et scaling

Rappelons rapidement les notations : les individus ont quatre états possibles
- (1) et (2) : présence dans la première ou la seconde pièce.
- (D_1) et (D_2) : l'individu sollicite un des couloirs 1 ou 2.
- (T_1) et (T_2) : l'individu est parvenu à traverser un des couloirs 1 ou 2, et se rend dans la pièce 2.

Les proportions d'individus dans chacun de ces états sont notées de manière transparente $M_1^N(t)$, $M_2^N(t)$, $M_{D_1}^N(t)$, $M_{D_2}^N(t)$, $M_{T_1}^N(t)$ et $M_{T_2}^N(t)$.

À chaque instant, les personnes dans la pièce initiale peuvent essayer de s'engager dans un des deux couloirs avec des probabilités respectives a_1 et a_2. Ces probabilités feront office de paramètres de contrôle du système. Elles paramètrent les stratégies aléatoires individuelles sur les actions $\{\text{Solliciter le couloir } 1, \text{Solliciter le couloir } 2\}$.

Les capacités des couloirs sont supposées exponentiellement décroissantes : si k personnes demandent à traverser le couloir, alors chacune y parvient avec une probabilité de e^{-Ak}. Cette probabilité décroît exponentiellement avec le coefficient A, qui représente donc la résistance du couloir. Les probabilités de traverser, sachant que $M^N(t) = m$, sont donc :

$$p_{D_1,T_1}^N(m) = e^{-A_1 N m_{D_1}} \text{ et } p_{D_2,T_2}^N(m) = e^{-A_2 N m_{D_2}}$$

où α_1 et α_2 désignent les résistances respectives des couloirs.

Il est clair que si A_1 et A_2 sont constantes devant le nombre d'individus, ces probabilités tendent vers 0 lorsque $N \to \infty$. Pour cette raison, nous faisons l'hypothèse de scaling suivante : asymptotiquement, les résistances A_1 et A_2 sont inversement proportionnelles à N, c'est-à-dire :

$$A_1 \simeq \frac{\alpha_1}{N} \text{ et } A_2 \simeq \frac{\alpha_2}{N}$$

Sous cette hypothèse, les probabilités asymptotiques de passage d'une pièce à l'autre, lorsque $N \to \infty$, sont :

$$p_{D_1,T_1}(m) = e^{-\alpha_1 m_{D_1}} \text{ et } p_{D_2,T_2}(m) = e^{-\alpha_2 m_{D_2}}.$$

Le diagramme des transitions asymptotique est le suivant :

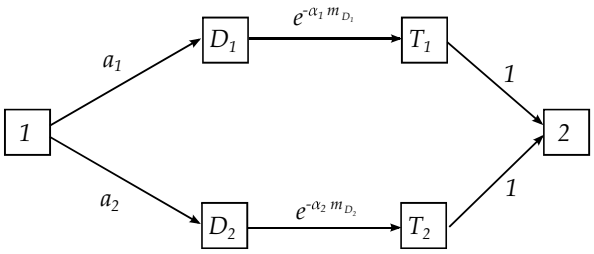


Figure 4.6 — Diagramme des transitions asymptotique

En ordonnant les états selon $(1, D_1, D_2, T_1, T_2, 2)$, la matrice de transition asymptotique s'écrit :

$$P(m,(a_1,a_2)) = \left(\begin{array}{c|cc|cc|c} 1-a_1-a_2 & a_1 & a_2 & 0 & 0 & 0 \\ \hline 1-e^{-\alpha_1 m_2} & 0 & 0 & e^{-\alpha_1 m_2} & 0 & 0 \\ 1-e^{-\alpha_2 m_3} & 0 & 0 & 0 & e^{-\alpha_2 m_3} & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right)$$

et le champ moyen évolue selon la règle de récurrence :

$$\begin{cases} \mu(t+1) = \mu(t) \cdot P\left(\mu(t), (a_1(t), a_2(t))\right) \\ \mu(0) = (1,0,0,0,0,0) \end{cases}.$$

Les paramètres $a_1(t), a_2(t)$ sont nos paramètres de contrôle. Dans la suite, nous les ajustons afin de maximiser un certaine récompense définie dans la partie suivante.

4.6.1.2 Récompense et gain

Comme l'objectif est que les personnes traversent les couloirs, nous définissons la récompense instantanée comme la proportion de personnes qui y parvient à chaque instant

$$r(m) = m_{T_1} + m_{T_2}.$$

Considérons une durée T, et le gain cumulé sur la période $[0, T]$. Pour le système discret, il s'écrit

$$G_{a_1(0),a_2(0),\ldots} = \mathbb{E}\left(\sum_{t=0}^{T} r(M^N(t))\right)$$

et pour le champ moyen il vaut

$$\Gamma_{a_1(0),a_2(0),\ldots} = \sum_{t=0}^{T} r(\mu(t)).$$

Remarque 17. *Le gain du système stochastique correspond à la proportion de personnes dans les états T_1, T_2 et 2 à l'instant T*

$$G_{a_1(0),a_2(0),\ldots} = \mathbb{E}\left(M^N_{T_1}(T) + M^N_{T_2}(T) + M^N_2(T)\right)$$

et de manière similaire, pour le champ moyen

$$\Gamma_{a_1(0),a_2(0),\ldots} = \mu_{T_1}(T) + \mu_{T_2}(T) + \mu_2(T).$$

En effet, les récompenses cumulées correspondent au nombre total de personnes qui ont traversé le couloir.

4.6.1.3 Optimisation du champ moyen

Nous proposons de calculer le gain optimal du champ moyen, ainsi qu'une suite d'actions optimales pour le champ moyen, à l'aide d'une méthode rétrogradée (Cf. parties 3.4.1 et 4.5.3). On définit la value à l'instant final T par $V_T(\mu) = 0$. Cette value correspond au meilleur gain que l'on peut obtenir à partir de l'instant T et est nul car il n'y a plus d'instant ultérieur. La value à l'instant $T-1$ est définie par $V_{T-1}(\mu) = \mu_{T_1} + \mu_{T_2}$ et la value en $T-2$ vaut :

$$\begin{aligned} V_{T-2}(\mu) &= \mu_{T_1} + \mu_{T_2} + \max_{(a_1,a_2)} V_{T-1}\Big(\mu \cdot P(\mu,(a_1,a_2))\Big) \\ &= \mu_{T_1} + \mu_{T_2} + \mu_{D_1} e^{-\alpha_1 \mu_{D_1}} + \mu_{D_2} e^{-\alpha_1 \mu_{D_2}}. \end{aligned}$$

Rappelons le sens de la value $V_{T-2}(\mu)$ il s'agit du meilleur gain que l'on peut obtenir en partant de l'état μ à l'instant $T-2$.

Pour l'instant les paramètres de contrôle ne sont pas intervenus, étant donné la latence de deux transitions entre l'action (un choix de $a_1(t)$, $a_2(t)$) et la récompense résultante. Ensuite,

$$\begin{aligned} V_{T-3}(\mu) &= \mu_{T_1} + \mu_{T_2} + \max_{(a_1,a_2)} V_{T-2}\Big(\mu \cdot P(\mu,(a_1,a_2))\Big) \\ &= \mu_{T_1} + \mu_{T_2} + \max_{(a_1,a_2)} \Big(\mu_{D_1} e^{-\alpha_1 \mu_{D_1}} + \mu_{D_2} e^{-\alpha_1 \mu_{D_2}} + \mu_1 a_1 e^{-\alpha_1 a_1 \mu 1} + \mu_1 a_2 e^{-\alpha_2 a_2 \mu 1}\Big) \\ &= \mu_{T_1} + \mu_{T_2} + \mu_{D_1} e^{-\alpha_1 \mu_{D_1}} + \mu_{D_2} e^{-\alpha_1 \mu_{D_2}} + \max_{(a_1,a_2)} \Big(\mu_1 a_1 e^{-\alpha_1 a_1 \mu 1} + \mu_1 a_2 e^{-\alpha_2 a_2 \mu 1}\Big). \end{aligned} \tag{4.4}$$

L'expression à maximiser dans (4.4) est dérivable et on peut procéder par un calcul de dérivées partielles. On montre que le couple (a_1, a_2) réalisant le maximum, sous les contraintes $0 \leq a_1 \leq 1$, $0 \leq a_2 \leq 1$ et $a_1 + a_2 \leq 1$, est

$$\begin{cases} \left(\frac{1}{\alpha_1 \mu_1}, \frac{1}{\alpha_2 \mu_1}\right) & \text{si } \mu_1 \geq \frac{1}{\alpha_1} + \frac{1}{\alpha_2} \\ \left(\frac{\alpha_2}{\alpha_1+\alpha_2}, \frac{\alpha_1}{\alpha_1+\alpha_2}\right) & \text{sinon} \end{cases}$$

et que

$$V_{T-3}(\mu) = \mu_{T_1} + \mu_{T_2} + \mu_{D_1} e^{-\alpha_1 \mu_{D_1}} + \mu_{D_2} e^{-\alpha_2 \mu_{D_2}} + \begin{cases} \left(\frac{1}{\alpha_1} + \frac{1}{\alpha_2}\right) e^{-1} & \text{si } \mu_1 \geq \frac{1}{\alpha_1} + \frac{1}{\alpha_2} \\ \mu_1 e^{-\frac{\alpha_1 \alpha_2}{\alpha_1+\alpha_2} \mu_1} & \text{sinon} \end{cases}.$$

Il est possible d'interpréter ce résultat simplement. Si la proportion de personnes restantes est supérieure à $\frac{1}{\alpha_1} + \frac{1}{\alpha_2}$, alors il faut s'engager dans chacun des couloirs avec les probabilités suivantes

$$a_1(\mu) = \frac{1}{\alpha_1 \mu_1},\ a_2(\mu) = \frac{1}{\alpha_2 \mu_1}$$

afin d'avoir le taux de passage optimal $\mu_1 e^{-\frac{\alpha_1\alpha_2}{\alpha_1+\alpha_2}\mu_1}$ à l'instant suivant.

Si la proportion de personnes restantes est inférieure à $\frac{1}{\alpha_1}+\frac{1}{\alpha_2}$, c'est-à-dire s'il ne reste que peu de personnes, alors il faut s'engager avec probabilité 1. Cette probabilité est répartie sur les deux couloirs selon les résistances relatives $\frac{\alpha_2}{\alpha_1+\alpha_2}, \frac{\alpha_1}{\alpha_1+\alpha_2}$ et assure un taux de passage de $\mu_1 e^{-\frac{\alpha_1\alpha_2}{\alpha_1+\alpha_2}\mu_1}$ à l'instant suivant.

Ces constats permettent de réaliser les calculs de value suivants $V_{T-4}, V_{T-5}, \ldots$ rapidement. Le premier de ces calculs est :

$$V_{T-4}(\mu) = \mu_{T_1} + \mu_{T_2} + \sup_{(a_1,a_2)} V_{T-3}\Big(\mu \cdot P(\mu,(a_1,a_2))\Big).$$

La fonction à maximiser est dérivable par morceaux, et on peut procéder à un calcul de dérivées partielles comme cela a été fait pour V_{T-3}. On montre que :

$$\begin{cases} \left(\frac{1}{\alpha_1\mu_1}, \frac{1}{\alpha_2\mu_1}\right) & \text{si } \mu_1 \geq \frac{1}{\alpha_1}+\frac{1}{\alpha_2} \\ \left(\frac{\alpha_2}{\alpha_1+\alpha_2}, \frac{\alpha_1}{\alpha_1+\alpha_2}\right) & \text{sinon} \end{cases}.$$

réalise le maximum, et on retrouve la stratégie optimale à $T-3$. En poursuivant ces calculs à l'aide d'un raisonnement par récurrence, on montre qu'elle est optimale à chaque instant.

Il existe donc une stratégie optimale qui est stationnaire, et qui ne dépend que du nombre de personnes restantes dans la première pièce. Lorsqu'il reste une proportion μ_1 de personnes dans la pièce initiale, la stratégie optimale consiste à solliciter les couloirs 1 et 2 avec des probabilités (a_1, a_2) égales à :

$$\begin{cases} \left(\frac{1}{\alpha_1\mu_1}, \frac{1}{\alpha_2\mu_1}\right) & \text{si } \mu_1 \geq \frac{1}{\alpha_1}+\frac{1}{\alpha_2} \\ \left(\frac{\alpha_2}{\alpha_1+\alpha_2}, \frac{\alpha_1}{\alpha_1+\alpha_2}\right) & \text{sinon} \end{cases}.$$

Avec cette stratégie on assure que la proportion de personnes qui traverse à l'instant suivant est maximale, égal à

$$\begin{cases} \left(\frac{1}{\alpha_1}+\frac{1}{\alpha_2}\right) e^{-1} & \text{si } \mu_1 \geq \frac{1}{\alpha_1}+\frac{1}{\alpha_2} \\ \mu_1 e^{-\frac{\alpha_1\alpha_2}{\alpha_1+\alpha_2}\mu_1} & \text{sinon} \end{cases}. \tag{4.5}$$

Notons que la stratégie optimale dépend essentiellement de la proportion (et donc du nombre) de personnes restantes, et fait apparaître un *seuil* : tant que la proportion de personnes restantes dépasse $\frac{1}{\alpha_1}+\frac{1}{\alpha_2}$, elles doivent s'engager avec un probabilité inversement proportionnelle au nombre de personnes restantes.

En revanche, lorsque le nombre de personnes arrive en dessous de ce seuil, elles s'engagent avec probabilité 1 à chaque instant. Cette probabilité est divisée sur les deux couloirs par les résistances relatives $\frac{\alpha_2}{\alpha_1+\alpha_2}$ et $\frac{\alpha_1}{\alpha_1+\alpha_2}$.

Remarque 18. *De manière naturelle, la moyenne harmonique des résistances α_1 et α_2 s'est introduit dans les calculs. En effet, la seconde ligne de* (4.5) *correspond à la probabilité de passage dans un seul couloir dont la résistance est $\frac{\alpha_1\alpha_2}{\alpha_1+\alpha_2} = \frac{1}{\frac{1}{\alpha_1}+\frac{1}{\alpha_2}}$ soit la moyenne harmonique des résistances des deux couloirs. On retrouve la formule classique des résistances électriques montées en parallèle.*

Avec la stratégie optimale trouvée ci-dessus, le gain croît de manière linéaire tant que la proportion de personnes dans la première pièce est supérieure à $\frac{1}{\alpha_1} + \frac{1}{\alpha_2}$. Ensuite cette croissance se fait avec un taux de $\mu_1 e^{-\frac{\alpha_1\alpha_2}{\alpha_1+\alpha_2}\mu_1}$ par unité de temps.

Dans la section suivante, nous effectuons des simulation numériques pour évaluer cette performance et confirmer les résultats théoriques présentés à la section 4.5.3. De plus, nous proposons de comparer la stratégie optimale à deux stratégies heuristiques dont le choix sera justifié.

4.6.1.4 Analyse numérique et comparaison de la performance

Dans cette partie nous évaluons la performance de la stratégie optimale à l'aide de simulations numériques. Cette stratégie fait intervenir de manière explicite les capacités α_1 et α_2 des couloirs. Conformément à la discussion en section 4.5.5, ces informations ne sont pas nécessairement connues par les agents.

Nous allons donc proposer deux stratégies heuristiques, notées **(MC)** et **(BO)**, qui ne font pas appel aux valeurs de α_1 et α_2. L'idée de ces stratégies est d'introduire une variable mémoire qui permet de décider de l'attitude que les agents doivent adopter. Les deux stratégies heuristiques sont exposées dans les deux paragraphes suivants, et leur performance est ensuite comparée à celle de la stratégie optimale.

Il est important de noter que ces stratégies heuristiques sont particulières, car elles s'inscrivent dans la cadre formel de ce chapitre, défini à la section 4.2.1, et parce qu'elles vérifient toutes les deux les conditions d'application du théorème 5. De ce fait, nous pouvons les comparer de manière équitable à la stratégie optimale établie à la section 4.6.1.3.

Une question intéressante, mais très délicate, est de savoir comment ces stratégies peuvent être comparées à des stratégies heuristiques qui ne vérifient pas les conditions du théorème 5 (c'est-à-dire des comportements d'agents qui ne passent pas à la limite champ moyen). La réflexion autour de cette question n'a pas été approfondie dans cette thèse, mais elle figure parmi les perspectives de ce travail.

(MC) : Mémorisation des collisions La première stratégie heuristique consiste à mémoriser le nombre de collisions qui ont eu lieu dans chaque couloir, et à déterminer la probabilité de chaque couloir en fonction de cette information.

Afin d'inscrire cette stratégie dans le modèle proposé, il faut introduire de la mémoire dans le système à l'aide de la variable ressource. Les deux premières composantes de la ressource $R_1(t), R_2(t)$ représentent le nombre de collisions (moyen par agent) cumulé au cours du temps dans chacun des couloirs. Pour le premier couloir, ce nombre est calculé par

$$R_1(t) = R_1(t-1) + M_{D_1}(t-1) - M_{T_1}(t).$$

En termes simples, on ajoute un terme (sollicitation) - (passage) au nombre de collisions cumulées. Afin de conserver la valeur $M_{D_1}(t-1)$ de l'instant précédent, nous l'inscrivons dans la troisième composante de la ressource $R_3(t) = M_{D_1}(t)$. La ressource complète est

donc un vecteur $R(t) = (R_1(t), R_2(t), R_3(t), R_4(t))$ mise à jour par la règle :

$$\begin{cases} R_1(t) = R_1(t-1) + R_3(t-1) - M_{T_1}(t) \\ R_2(t) = R_2(t-1) + R_4(t-1) - M_{T_2}(t) \\ R_3(t) = M_{D_1}(t) \\ R_4(t) = M_{D_2}(t) \end{cases}$$

et initialisée à $R(0) = (1, 1, 0, 0)$. Cette initialisation permet de commencer avec une probabilité identique pour les trois choix $\{\text{Solliciter Couloir 1}, \text{Solliciter Couloir 2}, \text{Ne rien faire}\}$.

À chaque instant les personnes choisissent les couloirs, en privilégiant celui où le nombre de collisions est le plus faible, avec les probabilités suivantes :

$$\begin{cases} a_1(t) = \frac{R_2(t)}{1+R_1(t)+R_2(t)} \\ a_2(t) = \frac{R_1(t)}{1+R_1(t)+R_2(t)} \end{cases}.$$

Avec cette règle, la probabilité totale de s'engager vaut $a_1(t) + a_2(t) = \frac{R_1(t)+R_2(t)}{1+R_1(t)+R_2(t)}$ et tend vers 1 lorsque le nombre de collisions est grand. Pour cette stratégie, les agents deviennent donc de plus ne plus agressifs au cours de leurs échecs successifs.

(BO) : Back-off La deuxième stratégie proposée consiste à fixer des probabilités a_1 et a_2 initiales, et à les ajuster au cours du temps en fonction du nombre de collisions observées dans chacun des couloirs.

Pour les mêmes raisons que dans la stratégie précédente, nous utilisons la variable ressource afin de mesurer la nombre de collisions qui ont lieu dans chaque couloir. Pour appliquer cette stratégie, nous utilisons seulement le nombre de collisions instantané (et non cumulé). Les quatre premières composantes de la ressource sont donc mises à jour selon

$$\begin{cases} R_1(t) = R_3(t-1) - M_{T_1}(t) \\ R_2(t) = R_4(t-1) - M_{T_2}(t) \\ R_3(t) = M_{D_1}(t) \\ R_4(t) = M_{D_2}(t) \end{cases}.$$

et initialisées à $(1, 1, 0, 0)$. Les composantes $R_1(t)$ et $R_2(t)$ représentent respectivement le nombre de collisions instantané entre $t-1$ et t dans le couloir 1 et 2.

De plus, il est nécessaire de conserver les valeurs courantes des probabilités a_1 , a_2. Nous les inscrivons dans les composantes 5 et 6 de la ressource, c'est-à-dire $a_1(t) = R_5(t)$ et $a_2(t) = R_6(t)$. Initialisées à $R_5(0) = R_6(0) = \frac{1}{2}$, ces probabilités se mettent à jour selon

$$\begin{cases} a_1(t) = a_1(t-1)\ (1 - R_1(t-1)) \\ a_2(t) = a_2(t-1)\ (1 - R_2(t-1)) \end{cases} \text{soit} \begin{cases} R_5(t) = R_5(t-1)\ (1 - R_1(t-1)) \\ R_6(t) = R_6(t-1)\ (1 - R_2(t-1)) \end{cases}.$$

L'idée est que si un couloir est obstrué, les agents vont préférer l'autre couloir. Cette préférence se précise de manière exponentielle lorsque l'obstruction perdure.

Nous insistons sur le fait qu'aucune des stratégies proposées dans cette section ne fait appel aux résistances α_1 et α_2 des couloirs. Ce manque d'information a priori est compensé par la mémoire (collective), à travers la ressource R. La mémoire permet aux stratégies d'améliorer leurs performances au cours du temps, la première par un processus d'apprentissage et la seconde par un processus d'ajustement.

Simulations, comparaison des stratégies Les deux méthodes heuristiques proposées entrent dans le modèle stochastique (avec ressource) décrit à la partie 4.2.1. Nous avons simulé le modèle stochastique, ainsi que le champ moyen correspondant aux stratégies (BO), (MC) et la stratégie optimale calculée à la partie 4.6.1.3.

Pour les simulations, nous avons fixé :
- le nombre d'agents à $N = 100$,
- les résistances des couloirs à $\alpha_1 = 10$ et $\alpha_2 = 50$,
- la durée des simulations à $t = 100$.

La figure 4.7 montre l'évolution de la proportion de personnes restantes (états 1, D_1 et D_2) pour les trois stratégies [9].

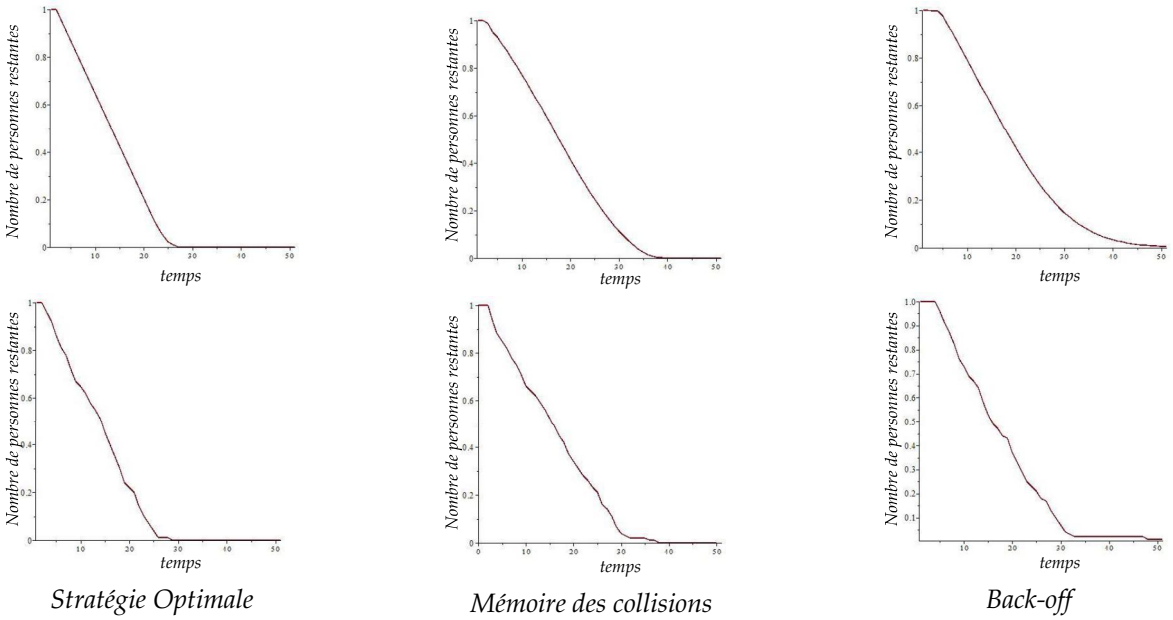


Figure 4.7 — Évolution du système avec les trois stratégies. Le champ moyen est représenté au-dessus, et le système stochastique en dessous

Les trois stratégies permettent de vider la pièce initiale. La stratégie optimale est un peu plus rapide que les deux autres, mais elle dépend des capacités des couloirs, et ne sera applicable que pour des valeurs de α_1 et α_2 fixes et connues.

La stratégie **(MO)** consiste à mémoriser les collisions pour en déduire son attitude. Ces collisions mémorisées ont une interprétation intéressante, qui sera réutilisée dans le chapitre 6 : il s'agit de «phéromones négatives» avec un effet répulsif sur les agents. De manière générale, on peut les voir comme des marqueurs environnementaux permettant aux agents de communiquer leur expérience.

Finalement, la stratégie Back-off **(BO)** a également une performance intéressante, et ne demande aucune connaissance sur les coefficients α_1 et α_2. Elle exige cependant de stocker les probabilités courantes a_1 et a_2, qui peut être problématique si ces nombres n'ont pas de représentation compacte. Dans un contexte général ce fait n'est pas garanti.

9. Ce nombre est simplement le complément à 1 du gain.

Le système étudié dans cette partie possède un fort caractère monotone : le nombre de personnes dans la seconde pièce ne fait qu'augmenter (au sens large) au cours du temps. Le système tend inexorablement vers la position d'équilibre où toutes les personnes se trouvent dans la seconde pièce, et présente donc un l'intérêt limité. Nous étudions donc un exemple un peu plus varié ci-dessous, dont la position d'équilibre n'est pas triviale.

4.6.2 Attitude optimale pour les robots extracteurs de bâtons

Reprenons l'exemple issu de la robotique (partie 4.3.2). Un collectif de N robots se déplacent dans une zone circulaire, en présence d'un certain nombre de bâtons qu'ils souhaitent extraire du sol. L'extraction se fait par un processus de coopération et nécessite la rencontre et la coordination de deux robots.

4.6.2.1 Notations, récompense et gain

Les robots se déplacent selon des marches aléatoires, et leurs positions sont supposées uniformes sur la zone. En simplifiant les transitions (détaillées dans la partie 4.3.2) :

- Un robot rencontre un bâton libre (resp. occupé) avec une probabilité proportionnelle au nombre de bâtons libres (resp. occupés).
- Une extraction a lieu lorsqu'un robot rencontre un bâton occupé.
- Un robot qui tient un bâton peut le relâcher à chaque instant avec une probabilité a.

Le diagramme des transitions est

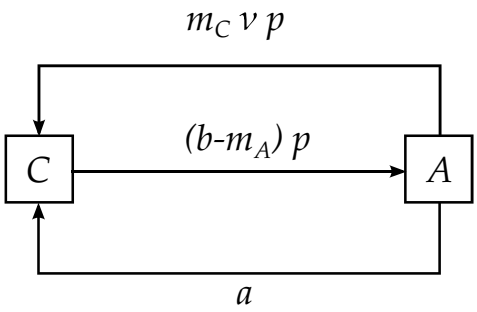


Figure 4.8 — Diagramme des transitions

avec

- les états C et A désignent respectivement l'état où le robot explore et l'état où il tient un bâton,
- le nombre p désigne la probabilité de rencontrer un bâton libre,
- le nombre ν désigne l'angle sous lequel les robots doivent se rencontrer pour pouvoir coopérer et extraire un bâton,
- les nombres m_C et m_A désignent les densités de robots en chaque état,
- le nombre a représente la probabilité avec laquelle un robot dans l'état A abandonne son bâton.

La matrice de transition est

$$P(m,a) = \begin{pmatrix} 1-(b-m_A)\cdot p & (b-m_A)\cdot p \\ m_C\nu p + a & 1 - m_C\nu p - a \end{pmatrix}.$$

Contrairement à la partie 4.3.2, le paramètre a n'est plus supposé fixe. Dans cette section il fait office de paramètre de contrôle. Afin d'assurer que toutes les probabilités de transition soient comprises entre 0 et 1 nous imposons la contrainte :

$$a \leq 1 - m_C\nu p$$

qui garantit que $m_C\nu p + a \leq 1$, et que la probabilité de transition de A vers C ne dépasse pas 1.

Nous proposons de déterminer la probabilité d'abandon optimale a, qui maximise le nombre de bâtons extraits sur une période finie $[0, T]$. La récompense instantanée est définie par le nombre moyen d'extractions par unité de temps

$$r(m_A, m_C) = m_A m_C \nu p.$$

Le gain du système stochastique est donné par le nombre moyen d'extractions effectuées sur la période $[0, T]$

$$G_{a(0),a(1),\ldots} = \mathbb{E}\left(\sum_{t=0}^{T} r(M^N(t))\right)$$

et pour le champ moyen, ce gain est donné par

$$\Gamma_{a(0),a(1),\ldots} = \sum_{t=0}^{T} r(\mu(t)).$$

L'objectif de la partie suivante est de déterminer les valeurs de $a(0), a(1), \ldots, a(T)$ qui maximisent ces gains.

4.6.2.2 Optimisation du champ moyen

Comme dans la section 4.6.1.3, nous allons utiliser une méthode rétrogradée pour déterminer la stratégie optimale du champ moyen. On définit donc les values[10] suivantes $V_T(\mu) = 0$, $V_{T-1}(\mu) = \mu_C\mu_A\nu p$, puis

$$V_{T-2}(\mu) = \mu_C\mu_A\nu p + \sup_a V_{T-1}(\mu \cdot P(\mu, a))$$

avec :

$$\begin{aligned} V_{T-1}(\mu \cdot P(\mu,a)) = {} & (\mu_C(1-(b-\mu_A)p) + \mu_A(\mu_C\nu p + a)) \\ & \times (1 - (\mu_C(1-(b-\mu_A)p) + \mu_A(\mu_C\nu p + a)))\,\nu p. \end{aligned}$$

10. Nous rappelons que la value $V_t(\mu)$ représente le meilleur gain que l'on peut obtenir en partant de l'état μ à l'instant t.

Nous calculons le maximum dans l'expression de $V_{T-2}(\mu)$ grâce à une simplification de l'expression de V_{T-1}. À l'aide de l'égalité $X \cdot (1-X) = \frac{1}{4} - (X - \frac{1}{2})^2$ on montre que :

$$V_{T-1}(\mu \cdot P(\mu, a)) = \frac{1}{4}\nu p - \nu p \left(\mu_C(1-(b-\mu_A)p) + \mu_A(\mu_C \nu p + a) - \frac{1}{2}\right)^2$$

et cette expression est maximale lorsque $X = \frac{1}{2}$, soit :

$$\mu_C(1-(b-\mu_A)p) + \mu_A\left(\mu_C \nu p + a\right) = \frac{1}{2}.$$

La valeur optimale de a, respectant l'encadrement $a \in [0, 1-\mu_C \nu p]$, est donc

$$a_{T-2}(\mu) = \begin{cases} 0 & \text{si } \frac{\frac{1}{2}-\mu_C(1-(b-\mu_A)p)-\mu_C\mu_A\nu p}{\mu_A} < 0 \\ 1-\mu_C\nu p & \text{si } \frac{\frac{1}{2}-\mu_C(1-(b-\mu_A)p)-\mu_C\mu_A\nu p}{\mu_A} > 1-\mu_C\nu p \\ \frac{\frac{1}{2}-\mu_C(1-(b-\mu_A)p)-\mu_C\mu_A\nu p}{\mu_A} & \text{sinon} \end{cases} . \tag{4.6}$$

et la value à l'instant $T-2$ vaut

$$V_{T-2}(\mu) = \mu_C\mu_A\nu p + V_{T-1}\left(\mu \cdot P\left(\mu, a_{T-2}(\mu)\right)\right)$$

Remarque 19.

- *Le cas $\mu_A = 0$, problématique en apparence, ne pose en réalité pas de problème : lorsque l'état 2 est vide il n'est pas nécessaire de définir la probabilité a, puisqu'aucun robot ne peut effectuer la transition $A \to C$.*
- *La distinction de cas dans (4.6) s'explique simplement. La troncature vis-a-vis des valeurs extrémales 0 et $1-\mu_C\nu p$ correspond aux deux cas de figure :*
  - *l'état 2 est trop peu occupé, et les agents doivent s'efforcer d'y rester.*
  - *l'état 2 est trop occupé, et les agents doivent s'efforcer de le quitter.*

Le calcul de V_{T-3} fait intervenir une discussion sur l'effet de a sur l'état du système à $T-2$. Cet état détermine la valeur de a_{T-2} selon les cas de figure dans (4.6). Bien que cette discussion soit intéressante, elle est délicate et risque d'obscurcir le but de cet exemple qui est d'illustrer la méthode d'optimisation.

Nous réduisons donc le degré de généralité de cet exemple en supposant que :

$$bp \leq \frac{1}{2} \tag{4.7}$$

qui assure que

$$\frac{\frac{1}{2}-\mu_C(1-(b-\mu_A)p)-\mu_C\mu_A\nu p}{\mu_A} \leq 1-\mu_C\nu p.$$

La condition (4.7) contraint le produit bp à rester inférieur à $\frac{1}{2}$, ce qui entraîne en particulier que lorsque **tous** les robots sont en phase d'exploration, au plus la moitié d'entre eux rencontrent des bâtons.

Par conséquent la deuxième option de (4.6) est exclue, et

$$a_{T-2}(\mu) = \max\left(\frac{\frac{1}{2}-\mu_C(1-(b-\mu_A)p)-\mu_C\mu_A\nu p}{\mu_A}, 0\right).$$

En étudiant le signe de $\frac{\frac{1}{2}-\mu_C(1-(b-\mu_A)p)-\mu_C\mu_A\nu p}{\mu_A}$, on montre qu'il existe un *seuil* μ^* tel que

$$a_{T-2}(\mu) = \begin{cases} 0 & \text{si } \mu^* \leq \mu_C \leq 1 \\ \frac{\frac{1}{2}-\mu_C(1-(b-\mu_A)p)-\mu_C\mu_A\nu p}{\mu_A} & \text{si } \mu_C \leq \mu^* \end{cases}.$$

D'autre part, ce seuil permet d'interpréter précisément la stratégie optimale à $T-2$:

- Si le taux d'occupation de l'état C est plus grand que μ^*, alors il faut poser $a = 0$.
  Dans ce cas, le taux d'occupation μ_C diminue en passant à
  $$\mu_C(1-(b-\mu_A)p) + \mu_C\mu_A\nu p$$
  à l'instant $T-1$.
- Sinon, il faut poser $a = \frac{\frac{1}{2}-\mu_C(1-(b-\mu_A)p)-\mu_C\mu_A\nu p}{\mu_A}$.
  Dans ce cas, les taux d'occupation passent à $\mu_C = \mu_A = \frac{1}{2}$ à l'instant $T-1$.

Ces observations permettent de réaliser rapidement tous les calculs de value suivants. Montrons comment le premier est fait :

$$V_{T-3}(\mu) = \mu_C\mu_A\nu p + \sup_a V_{T-2}\left(\mu \cdot P(\mu, a)\right)$$

Si le taux d'occupation de l'état C est inférieur à μ^*, alors $a^* = \frac{\frac{1}{2}-\mu_C(1-(b-\mu_A)p)-\mu_C\mu_A\nu p}{\mu_A}$ réalise le maximum.

Dans le cas contraire, la valeur optimale est $a^* = 0$, et

- si le taux d'occupation chute en dessous du seuil μ^*, le système passe à $\mu_C = \mu_A = \frac{1}{2}$ dans tous les états suivants.
- sinon, on choisit à nouveau $a = 0$ à l'instant $T-2$.

En répétant ce raisonnement pour les values à $T-4, T-5, \ldots$, on montre que la stratégie consistant à choisir

$$a(\mu) = \begin{cases} 0 & \text{si } \mu^* \leq \mu_C \leq 1 \\ \frac{\frac{1}{2}-\mu_C(1-(b-\mu_A)p)-\mu_C\mu_A\nu p}{\mu_A} & \text{si } \mu_C \leq \mu^* \end{cases}.$$

est optimale à chaque instant. La dynamique optimale du système[11] consiste en deux phases :

- Tant que le taux d'occupation μ_C est supérieur au seul μ^*, choisir $a = 0$.
  Ce choix a pour effet de diminuer le taux d'occupation de l'état C.
- Dès que $\mu_C \leq \mu^*$, choisir $a = \frac{\frac{1}{2}-\mu_C(1-(b-\mu_A)p)-\mu_C\mu_A\nu p}{\mu_A}$.
  Avec ce choix, l'état du système passe à $\mu_C = \mu_A = \frac{1}{2}$ pour tous les instant suivants.

La dynamique optimale de ce système possède un caractère monotone et un phénomène de seuil. Contrairement à la partie précédente, le système n'est monotone que sous la stratégie optimale. De plus, contrairement à l'exemple précédent, le seuil dans cet exemple a l'effet inverse : une fois que le système passe en dessous du seuil μ^*, le système arrive à des gains instantanés maximaux et une performance optimale.

Dans la partie suivante nous confirmons les affirmations ci-dessus à l'aide de simulations numériques.

11. i.e. la dynamique sous la stratégie optimale

4.6.2.3 Analyse et comparaison de la performance

Pour évaluer la performance de la stratégie optimale trouvée dans la partie précédente, nous faisons des simulations. Une étude isolée de la dynamique optimale du système, et de sa récompense optimale n'a que peu de sens. Pour cette raison nous proposons de la comparer à la performance d'une autre stratégie établie très simplement. Cette stratégie est présentée dans le paragraphe suivant.

Stratégie constante et optimisation à l'équilibre Une méthode simple pour obtenir une valeur intéressante de la probabilité de lâcher un bâton a est de la supposer constante, et la rendre optimale à l'équilibre du système [12].

Pour une valeur constante de a, le champ moyen évolue comme :

$$\mu(t+1) = \mu(t) \cdot \begin{pmatrix} 1-(b-\mu_2(t))\cdot p & (b-\mu_2(t))\cdot p \\ \mu_1(t)\nu p + a & 1-\mu_1(t)\nu p - a \end{pmatrix}.$$

Étudions ses positions d'équilibre. Elles sont données par l'équation

$$\mu = \mu \cdot \begin{pmatrix} 1-(b-\mu_2)\cdot p & (b-\mu_2)\cdot p \\ \mu_1\nu p + a & 1-\mu_1\nu p - a \end{pmatrix}.$$

En particulier, la densité d'agents en l'état C à l'équilibre est donnée par

$$\mu_1 = \mu_1(1-(b-\mu_2)p) + \mu_2(\mu_1\nu p + a), \tag{4.8}$$

L'approche que nous proposons est de rendre la récompense instantanée $r(\mu) = \mu_1\mu_2\nu p$ maximale à l'équilibre. Lorsque $b \geq \frac{1}{2}$, cette récompense est maximale pour $\mu_1 = \mu_2 = \frac{1}{2}$.

Admettons que cette répartition soit atteinte à l'équilibre. L'équation (4.8) devient

$$\frac{1}{2} = \frac{1}{2}\left(1-(b-\frac{1}{2})p\right) + \frac{1}{2}\left(\frac{1}{2}\nu p + a\right).$$

et a pour solution

$$a = \left(b - \frac{1}{2}\right)p - \frac{1}{2}\nu p. \tag{4.9}$$

D'une part, nous vérifions que

- $b \geq \frac{1}{2}$ entraîne que $a \geq 0$,
- La condition $bp \leq \frac{1}{2}$ dans l'hypothèse (4.7) entraîne que $a \leq 1$,

et par conséquent a définit correctement une probabilité.

D'autre part, il faut montrer qu'avec ce choix de a le champ moyen μ converge effectivement vers la répartition optimale $(\frac{1}{2}, \frac{1}{2})$. Pour cela, nous réécrivons la règle d'évolution de la première composante de μ :

$$\begin{aligned}\mu_1(t+1) &= \mu_1(t)(1-(b-\mu_2(t))p) + \mu_2(t)(\mu_1(t)\nu p + a) \\ &= \frac{1}{2} - \left(\mu_1(t) - \frac{1}{2}\right)\cdot\left(p(1+\nu)\mu_1(t) + p(2b-\nu-1) - 1\right)\end{aligned}$$

12. Cette approche sera à nouveau utilisée à la partie 5.

en utilisant le fait que $\mu_2(t) = 1 - \mu_1(t)$. Par conséquent :

$$\left|\mu_1(t+1) - \frac{1}{2}\right| = \left|\mu_1(t) - \frac{1}{2}\right| \cdot \left|p(1+\nu)\mu_1(t) + p(2b-\nu-1) - 1\right|.$$

Or

$$p(1+\nu)\mu_1(t) + p(2b-\nu-1) - 1 = 2bp - (1+\nu)(1-\mu_1(t))p - 1.$$

Grâce à la contrainte $bp \leq \frac{1}{2}$ et le fait que $1 - \mu_1(t) \leq b$, on obtient que

$$0 \leq 2bp - (1+\nu)(1-\mu_1(t))p < 1.$$

dés que $\nu < 1$ ou $p < 1$. Par conséquent,

$$\left|p(1+\nu)\mu_1(t) + p(2b-\nu-1) - 1\right| \leq k$$

pour une certaine constante $k < 1$, et

$$\left|\mu_1(t+1) - \frac{1}{2}\right| \leq k \cdot \left|\mu_1(t) - \frac{1}{2}\right|.$$

Cette dernière inégalité montre que le champ moyen converge exponentiellement vite vers $(\frac{1}{2}, \frac{1}{2})$, et ce quelle que soit sa valeur initiale.

Le champ moyen admet donc une unique position d'équilibre vers laquelle toutes les trajectoires convergent. Selon le résultat énoncé à la section 4.4.3 le système stochastique a donc également tendance à se stabiliser selon cet équilibre, si le nombre d'agents N est suffisamment grand.

En résumé, la valeur de a donnée par l'équation (4.9) mène le système stochastique vers la répartition optimale $(\frac{1}{2}, \frac{1}{2})$ à l'équilibre, si le nombre d'agents est suffisamment grand.

Dans le paragraphe suivant, nous confirmons ces affirmations à l'aide de simulations numériques. De plus, nous y simulons la dynamique optimale calculée à la section 4.6.2.2.

Simulations, comparaison des stratégies Comme pour l'exemple précédent (section 4.6.1.4), nous proposons d'évaluer la performance de la stratégie optimale par des simulations numériques. Elle sera comparée avec la stratégie constante déterminée au paragraphe ci-dessus.

Pour les simulations, nous avons fixé
- le nombre de robots à $N = 100$,
- les paramètres $p = 0,05$, $b = 0,8$ et $\nu = 0,1$. Il y a donc
  - peu de robots par rapport à la taille de la zone,
  - beaucoup de bâtons (relativement au nombre de robots),
  - 10% de chances qu'une rencontre de robots mène à une extraction de bâtons,
- la durée des simulations à $t = 80$ itérations.

Ces paramètres ont été choisis arbitrairement, et vérifient la contrainte (4.7). Ils permettent d'utiliser la probabilité constante optimale

$$a = \left(b - \frac{1}{2}\right)p - \frac{1}{2}\nu p = 0,0125.$$

La figure 4.9 montre l'évolution du taux de robots chercheurs pour les différentes stratégies proposées.

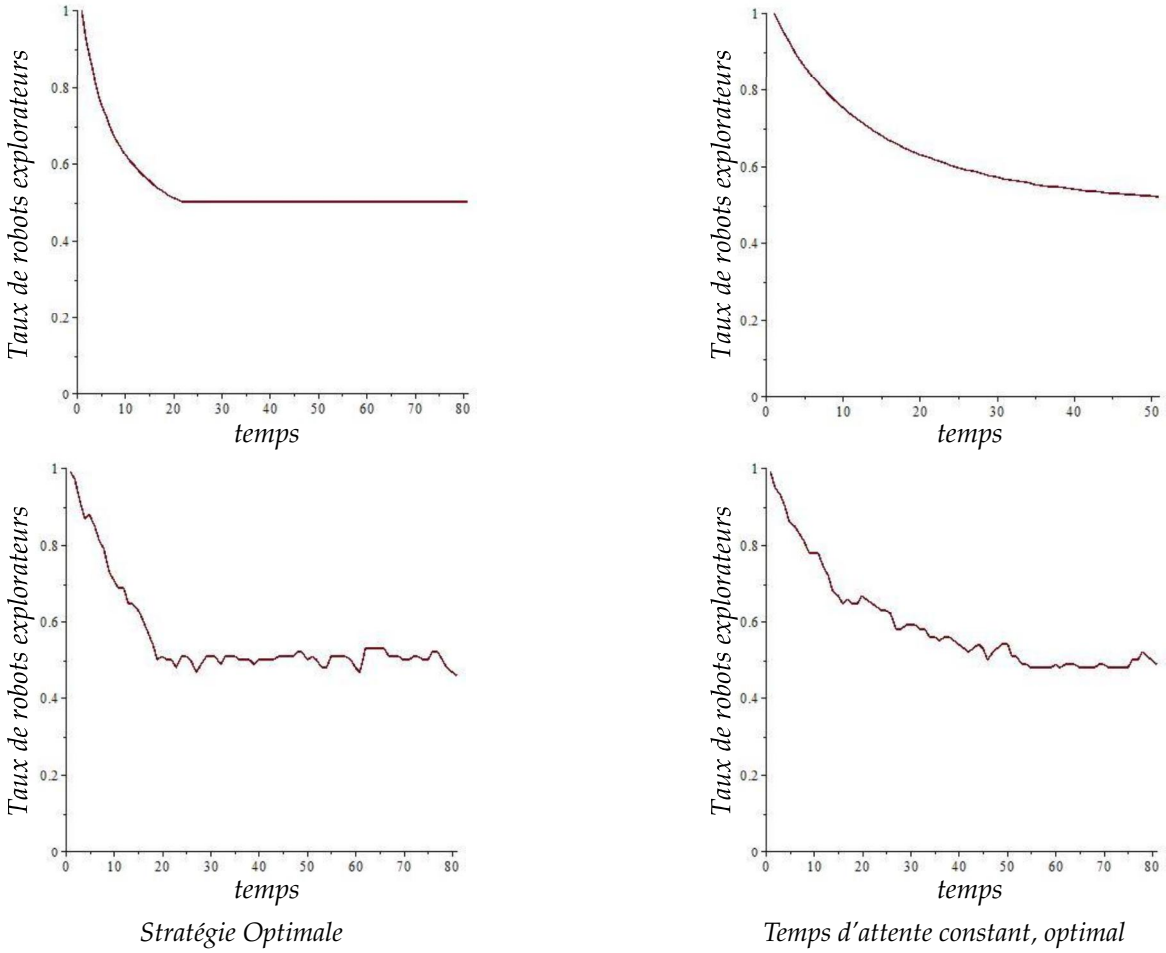


Figure 4.9 — Évolution du système avec les deux stratégies. Le champ moyen est représenté au-dessus, et le système stochastique en dessous.

On observe la dynamique escomptée : les deux systèmes convergent vers la répartition optimale $\mu_C = \mu_A = \frac{1}{2}$. La figure 4.10 montre l'évolution de la probabilité a au cours du temps.

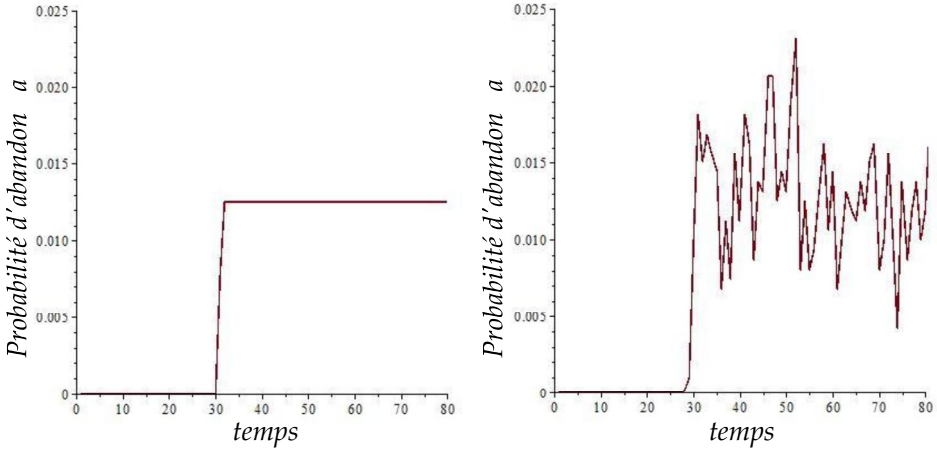


Figure 4.10 — Évolution du la probabilité d'abandon a pour le champ moyen (à gauche) et pour le système stochastique (à droite)

La courbe du champ moyen illustre ce qui a été avancé. Pour $t \leq 30$ la probabilité d'abandon a reste constante, égale à 0. Durant cette période le taux d'occupation de l'état C diminue jusqu'à atteindre $\frac{1}{2}$. Ensuite la probabilité d'abandon monte brutalement, afin de préserver les proportions optimales $\mu_C = \mu_A = \frac{1}{2}$, et se stabilise instantanément à $a = 0,0125$. Ce dernier point confirme au passage la valeur optimale constante de a.

La courbe du système stochastique est vaguement similaire, mais le résultat d'approximation paraît beaucoup plus contrasté. Une explication possible est que la stratégie optimale a est continue, mais qu'elle n'est pas plus régulière que cela avec un point de non-dérivabilité au niveau du seuil μ^*.

4.7 Conclusion

Dans ce chapitre, nous avons présenté une méthode d'approximation de type champ moyen. Cette méthode permet, sous certaines conditions, d'approcher le système par un champ déterministe. En particulier, elle exige que les agents aient une interaction «moyenne» avec les autres agents, qui converge vers une limite continue lorsque le nombre d'individus tend vers l'infini.

Approcher un système multi-agent par un champ moyen exige un nombre d'agents N important devant les autres grandeurs caractéristiques du système, comme le nombre d'états ou la taille de la mémoire. Dans [11] les auteurs proposent une extension qui permet d'échelonner le temps avec le nombre d'agents, et d'aboutir à des équations différentielles. Cette approche a deux avantages : d'une part, elle permet de modéliser les systèmes asynchrones. D'autre part, l'étude formelle d'une équation différentielle se fait à l'aide d'outils d'analyse mathématique (continue). Ces outils ne sont pas affectés par la taille du système, et peuvent s'avérer plus simples à utiliser et plus performants que des méthodes discrètes pour de grands systèmes. Nous présenterons une autre extension de ce type dans le chapitre 5, en utilisant les méthodes de champ moyen dans un contexte où les agents sont disposés dans un espace géométrique.

Le modèle stochastique utilisé pour représenter le système multi-agent est pertinent pour les systèmes multi-agents étudiés dans cette thèse. Le fait de considérer des agents identiques peut être contourné en introduisant artificiellement des états. Par ailleurs, le fait de ne considérer que les espaces d'états fini n'est pas obligatoire. Dans [20] des résultats similaires dans le cas d'un espace d'états infini sont exposés. Pour finir, le fait que la loi d'évolution dépende de la densité d'occupation des agents n'entraîne absolument pas que ces agents aient une connaissance complète de l'état du système. En effet, il est possible que les transitions des agents ne dépendent que de la densité en quelques états «voisins», ce qui est bien conforme à la localité des interactions des systèmes auquel nous nous intéressons.

La représentation d'un système multi-agent par le modèle stochastique introduit à la section 4.2.1 peut être compactifiée en omettant certaines caractéristiques du système. Dans ce cas, les paramètres qui demeurent dans le modèle réduit sont difficiles à interpréter.

Prenons l'exemple des robots tireurs de bâtons (section 4.3.2), où les positions exactes des robots ne sont pas prises en compte. Le modèle réduit est obtenu à l'aide de considérations géométriques, et présuppose une répartition uniforme des robots. Les paramètres p et ν ont des interprétations simples, mais il est difficile de savoir comment ces paramètres sont liés aux paramètres internes des robots.

Pour cette raison une perspective intéressante est d'inférer un modèle stochastique avec interactions à travers la moyenne (section 4.2.1) à partir de mesures statistiques sur un système donné. Pour ce faire on peut utiliser des méthodes d'inférence statistique comme celles proposées dans [51], qui permettent de trouver le modèle stochastique le plus proche d'une série de simulations.

Une fois obtenu un modèle Markovien avec interactions à travers la moyenne, il est possible de procéder à une approximation champ moyen du modèle Markovien. Ainsi, par le biais d'une modélisation statistique suivie d'une approximation champ moyen, on peut étudier le système multi-agent considéré.

Troisième partie

Étude de systèmes multi-agents situés réactifs à l'aide d'équations à réaction-diffusion-advection

5 Utilisation de l'équation de Fokker-Planck pour la paramétrisation d'un système multi-agent situé réactif

Dans cette section nous étudions un système simple, formé d'agents avec une perception locale et partielle de l'environnement. L'idée est de passer d'une loi d'évolution microscopique comprenant déplacements et interactions locales à une loi d'évolution macroscopique. Cela conduit à une étude asymptotique lorsque certaines grandeurs caractéristiques du système tendent vers des valeurs limites. Le modèle obtenu permet non seulement de caractériser la dynamique du système de manière précise, mais il permet aussi d'ajuster les comportements des agents afin de maximiser une certaine fonction de satisfaction.

Dans un premier temps (section 5.1) nous présentons le problème ainsi que le modèle formel utilisé pour l'étudier. Ensuite, dans la deuxième partie (section 5.2) nous dérivons une limite continue de ce modèle appelée équation de Fokker-Planck. La troisième partie (section 5.3) est consacrée à l'étude de cette équation dans un cas particulier pertinent dans le cadre du problème étudié. Grâce à cette étude, nous déterminons deux stratégies optimales dans la quatrième partie (section 5.4). Les approximations et l'optimalité des stratégies sont justifiées par des résultats théoriques, et nous proposons une validation numérique dans la section 5.5. Ensuite, nous proposons une rapide extension au cas bidimensionnel dans la section 5.6. Ce chapitre se termine par un bilan général, et une ouverture vers d'éventuelles perspectives à la section 5.7.

5.1 Position du problème et modèle utilisé

Nous étudions un système d'agents identiques qui se déplacent sur un réseau discret. L'objectif que nous fixons pour les agents est qu'ils se répartissent selon une certaine fonction de répartition f choisie à l'avance. Les déplacements individuels des agents sont aléatoires, mais orientés par la perception de leur environnement et par les interactions locales.

Il s'agit d'un problème général, rencontré, par exemple, en entomologie, lorsqu'une co-

lonie d'insectes essaye de récolter de la nourriture disposée dans l'espace [1], ou encore en robotique, lors du déploiement d'un collectif de robots secouristes dans une situation critique où des victimes humaines sont réparties dans l'espace.

Le but de notre approche est donc de définir et paramétrer les comportements des agents afin qu'ils se répartissent selon une distribution souhaitée (la nourriture pour les fourmis, ou les victimes pour les robots de secours).

5.1.1 Mouvement collectif des agents

Le modèle formel utilisé pour décrire la situation est discret. Le temps, ainsi que l'ensemble des positions des agents sont des ensembles discrets. Initialement les agents sont disposés aléatoirement et de manière indépendante selon des lois de probabilités identiques. Les déplacements des agents sont modélisés par des marches aléatoires discrètes, sans mémoire.

Le caractère aléatoire des déplacements a plusieurs motivations que nous avons déjà discutées. Nous travaillons sous l'hypothèse d'un système d'agents identiques. Si ces agents disposent des mêmes (par exemple, si tous disposent de la connaissance complète de l'état du système), et suivent une règle de décision déterministe pour décider de leurs déplacements, alors les comportements collectifs obtenus sont grégaires : tous les agents situés au même endroit se déplacent de manière identique.

De tels comportements collectifs sont clairement sous-optimaux pour l'objectif du système. De plus l'exploration de l'espace par les agents correspond naturellement à un processus de diffusion. De tels processus sont essentiellement modélisés par des marches aléatoires.

Pour commencer simplement, nous supposons que ces agents sont en mouvement sur le cercle $\mathcal{C} = \mathbb{R}/2\pi\mathbb{Z}$. Cet espace est divisé de manière régulière en n_p positions

$$\mathcal{C}_\delta = \left\{ \frac{2k\pi}{n_p} \ / \ k \in \{0, \ldots, n_p - 1\} \right\} = \left\{ k\delta \ / \ k \in \{0, \ldots, n_p - 1\} \right\}$$

en notant $\delta = \frac{2\pi}{n_p}$ le pas spatial. Nous le ferons tendre vers 0 ultérieurement, de sorte que $\mathcal{C}_\delta \to \mathcal{C}$ (au sens de la convergence des ensembles de Hausdorff [110]).

1. Sans se préoccuper du rapatriement

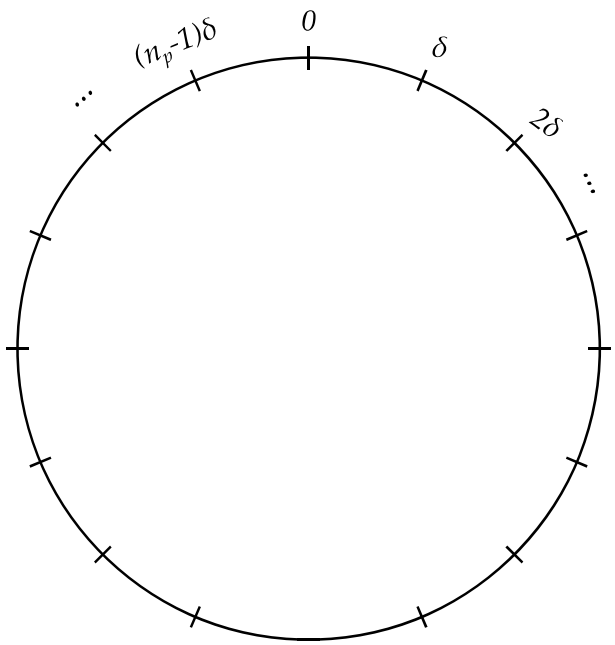


Figure 5.1 — Espace des positions des agents

La topologie périodique et compacte du domaine considéré offre quelques avantages théoriques pour la démonstration des théorèmes de ce chapitre. Néanmoins, l'approche générale présentée est aisément prolongeable aux domaines non-périodiques ou non bornés, ou encore des domaines multidimensionnels.

Comme les agents sont supposés identiques, l'état global du système est complètement décrit par le nombre d'agents en chaque position x de l'espace des positions $\mathcal{C}_\delta$. Notons $(X_n(t))_{n=1}^N$ les positions des agents à l'instant t, et $u^N(x,t)$ la proportion d'agents située au point x à l'instant t :

$$u^N(x,t) = \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{X_n(t)=x} \quad \text{où} \quad \mathbb{1}_{X_n(t)=x} = \begin{cases} 1 & \text{si } X_n(t) = x \\ 0 & \text{sinon} \end{cases} .$$

Le vecteur de *densité d'occupation* est donné par

$$u^N(t) = \Big(u^N(x,t)\Big)_{x\in\mathcal{C}_\delta} .$$

Durant un intervalle de temps τ, chaque agent pourra se déplacer de $-\delta$ ou $+\delta$ à partir de sa position courante ou rester en place. Ce déplacement (ou non-déplacement) est effectué de manière aléatoire, selon une loi de probabilités qui dépend de la positon courante de l'agent, ainsi que les positions de tous les autres agents.

Comme les agents sont incapables de se distinguer, ils ne peuvent observer que les densités d'occupation en chaque position. On notera $p^N_{\delta,\tau}(x,u)$ (resp. $q^N_{\delta,\tau}(x,u)$) la probabilité qu'un agent donné se déplace de $-\delta$ (resp. $+\delta$) à δ, τ et N fixés, lorsque l'ensemble des agents est réparti selon le vecteur de densité d'occupation u (voir figure 5.2).

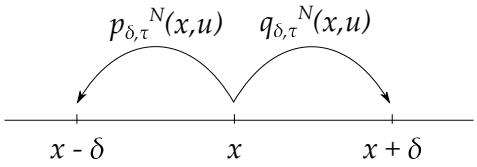


Figure 5.2 — Probabilités des déplacements de chaque agent

À ce stade nous avons défini un modèle formel générique qui décrit les déplacements des agents. Notons que ce modèle n'impose pas la dépendance globale des agents. Il est possible par exemple que la probabilité $p_{\delta,\tau}(x,u)$ ne dépende en réalité que de la densité au point x.

Les probabilités des déplacements $p^N_{\delta,\tau}$ et $q^N_{\delta,\tau}$ n'ont pas encore été choisies. Elles font office de paramètres de contrôle du système et devront être choisies conformément à l'objectif que nous définissons dans la partie suivante.

5.1.2 Fonction de satisfaction

Pour intégrer la finalité du système dans le modèle, nous lui associons une fonction de satisfaction qui reflète la réussite du système dans l'objectif que nous lui avons fixé (organisation des agents selon la fonction f). Nous imposons deux contraintes sur sa forme :

- elle doit croître avec la densité d'agents tant que celle-ci reste inférieure à la valeur de f,
- elle doit pénaliser la surexploitation, en décroissant au-delà de cette valeur.

Soit f une fonction 2π-périodique représentant la répartition ciblée. Cette fonction est supposée positive, normalisée de sorte que $\int_C f(x)dx = 1$, et sans point d'annulation. La fonction

$$r : (x,u) \mapsto u(x)\, e^{1-\frac{u(x)}{f(x)}}$$

est une fonction de satisfaction qui respecte les contraintes énoncées ci-dessus : pour une position x fixée elle croît tant que $u(x) \leq f(x)$ et décroît exponentiellement pour $u(x) > f(x)$ (voir figure 5.3).

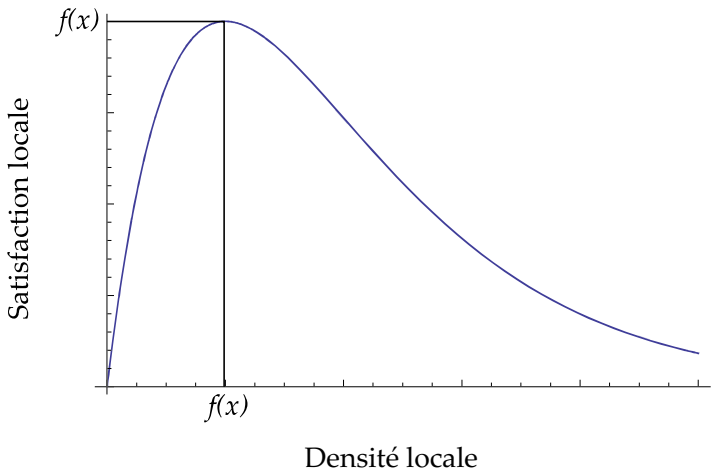


Figure 5.3 — Fonction de satisfaction des agents

La fonction r représente localement la réussite du système, dans chaque unité de longueur $[x, x+\delta[$. La récompense globale du système à un instant t est définie par

$$R_\delta(t) = \sum_{x \in \mathcal{C}_\delta} r(x, u(x,t))\, \delta.$$

L'objectif est que les agents se répartissent selon la distribution f. Il s'agit donc de déterminer un comportement d'agent (donc des fonctions $p^N_{\delta,\tau}(x,u)$ et $q^N_{\delta,\tau}(x,u)$) qui maximise la récompense R.

Nous nous limitons au gain instantané à long terme, c'est-à-dire $R(T)$, avec T grand

$$R(T) = \max_{p_{\delta,\tau}, q_{\delta,\tau}} \mathbb{E}\left(R_\delta(T)\right). \tag{5.1}$$

Nous aurions également pu considérer la maximisation des récompenses cumulées au cours du temps $\mathbb{E}\left(\sum_{t=1}^T R_\delta(t)\right)$ auquel cas le problème s'inscrit dans les processus décisionnels markoviens (Cf. chapitre 3).

Il est facile de simuler ce modèle pour un couple $(p_{\delta,\tau}, q_{\delta,\tau})$ donné, et d'en évaluer la performance. L'étude formelle est cependant plus délicate. En effet, si le système a un grand nombre d'agents et de positions possibles, la chaîne de Markov $\left((X_n(t))_{n=1}^N\right)_t$ (ou encore $(u^N(t))_t$) décrivant l'évolution du système devient très difficile à étudier. Cette difficulté est due à la forte croissance de sa matrice de transition, lorsque le nombre d'agents augmente.

Pour palier cette difficulté, nous réalisons une approximation continue. Cette approximation est réalisée en deux temps, et sera exposée dans la partie suivante.

5.2 Approximation du modèle discret

Dans cette partie, nous approchons le modèle discret aléatoire présenté ci-dessus par un modèle continu et déterministe. Avant de procéder à la dérivation de ce modèle continu, nous précisons les notations introduites.

Le modèle aléatoire proposé possède trois grandeurs caractéristiques :
- le nombre d'agents N,
- le pas d'espace δ,
- le pas de temps τ,

et la variable d'état du système sera notée $u^N_{\delta,\tau}(t)$ par la suite.

La dérivation du modèle continu est réalisée en deux temps. Tout d'abord, nous effectuons une limite de type champ moyen ($N \to \infty$), en procédant comme dans le chapitre 4. Ensuite, nous effectuons une limite spatiotemporelle, en faisant tendre $\tau, \delta \to 0$ d'une manière bien choisie. Ces deux opérations aboutissent à une équation aux dérivées partielles appelée *équation de Fokker-Planck*.

5.2.1 Limite champ moyen

Pour des valeurs de δ et τ fixées, le système aléatoire discret entre parfaitement dans le formalisme de la partie 4. La loi d'évolution de chaque individu peut s'écrire

$$\begin{cases} \mathbb{P}(X(t+\tau) = x-\delta \mid X(t) = x\,,\, u^N(t) = u) = p^N_{\delta,\tau}(x,u) \\ \mathbb{P}(X(t+\tau) = x \mid X(t) = x\,,\, u^N(t) = u) = 1 - p^N_{\delta,\tau}(x,u) - q^N_{\delta,\tau}(x,u) \\ \mathbb{P}(X(t+\tau) = x+\delta \mid X(t) = x\,,\, u^N(t) = u) = q^N_{\delta,\tau}(x,u) \end{cases}$$

ou encore

$$\begin{cases} \mathbb{P}(X(t+\tau) = x \mid X(t) = x-\delta\,,\, u^N(t) = u) = q^N_{\delta,\tau}(x-\delta,u) \\ \mathbb{P}(X(t+\tau) = x \mid X(t) = x\,,\, u^N(t) = u) = 1 - p^N_{\delta,\tau}(x,u) - q^N_{\delta,\tau}(x,u) \\ \mathbb{P}(X(t+\tau) = x \mid X(t) = x+\delta\,,\, u^N(t) = u) = p^N_{\delta,\tau}(x+\delta,u) \end{cases}.$$

Cette deuxième forme peut s'interpréter comme un bilan au point x à l'instant $t+\tau$. Un agent situé en x peut être

- arrivé par la gauche, avec probabilité $q^N_{\delta,\tau}(x-\delta,u)$,
- resté sur place, avec probabilité $1 - p^N_{\delta,\tau}(x,u) - q^N_{\delta,\tau}(x,u)$,
- être arrivé par la droite, avec probabilité $p^N_{\delta,\tau}(x+\delta,u)$.

Ces transitions aléatoires ne dépendent que de la densité d'individus en chaque position, et nous pouvons donc appliquer le théorème 5.

Proposition 6. *On suppose que pour chaque position x,*

(i) Les densités initiales $u^N_{\delta,\tau}(x,0)$ convergent presque sûrement vers des quantités $u^0_{\delta,\tau}(x)$.

(ii) Les probabilités des déplacements $p^N_{\delta,\tau}(x,u)$ et $q^N_{\delta,\tau}(x,u)$ convergent uniformément (en u) vers des fonctions continues $u \mapsto p_{\delta,\tau}(x,u)$ et $u \mapsto q_{\delta,\tau}(x,u)$

lorsque $N \to \infty$.

Alors pour chaque instant t, et pour chaque position x, la densité $u^N_{\delta,\tau}(x,t)$ converge presque sûrement vers $u_{\delta,\tau}(x,t)$ défini récursivement par

$$u_{\delta,\tau}(x,0) = u^0_{\delta,\tau}(x)$$

et

$$\begin{aligned} u_{\delta,\tau}(x,t+\tau) = u_{\delta,\tau}(x,t)(1 - p_{\delta,\tau}(x,u_{\delta,\tau}(t)) - q_{\delta,\tau}(x,u_{\delta,\tau}(t))) \\ + u_{\delta,\tau}(x-\delta,t)q_{\delta,\tau}(x-\delta,u_{\delta,\tau}(t)) \\ + u_{\delta,\tau}(x+\delta,t)p_{\delta,\tau}(x+\delta,u_{\delta,\tau}(t)). \end{aligned} \tag{5.2}$$

Ce résultat peut être compris de manière intuitive. L'hypothèse (i) signifie que la distribution initiale (aléatoire) des agents converge vers une certaine distribution lorsque le nombre d'agents tend vers l'infini. L'hypothèse (ii) signifie que les probabilités de déplacement des agents convergent lorsque le nombre d'agents tends vers l'infini. En termes simples, il est exigé que l'état initial des agents et leurs probabilités de déplacement convergent lorsque le nombre d'agents tend vers l'infini.

Sous ces hypothèses, les densités d'occupations qui décrivent l'état du système converge vers une quantité déterministe lorsque le nombre d'agents tend vers l'infini. Cette quantité

$u_{\delta,\tau}$ est appelée *champ moyen* et représente la densité du système dans le contexte théorique où le nombre d'agents est infini. Il évolue de manière déterministe et est régi par les n_p relations de récurrence définies en (5.2). Chacune de ces équations peut être interprétée comme un bilan, qui décrit le nombre d'agents restés sur place, arrivés par la gauche et arrivés par la droite (par les 3 lignes respectives de l'équation (5.2)).

Si le nombre de positions n_p est grand, il peut être difficile d'étudier le champ moyen. Pour cette raison nous en faisons une approximation continue dans la section 5.2.2.

5.2.2 Limite spatiotemporelle

Un modèle continu est obtenu en faisant tendre les intervalles de temps τ et d'espace δ vers 0. Pour cela considérons une solution régulière u de (5.2), au moins dérivable vis-à-vis de la variable t et deux fois dérivable vis-à-vis de la variable x.

Un développement limité à l'ordre 2 en δ des membres de droite de (5.2) donne :

$$\begin{aligned} u(x-\delta,t)q_{\delta,\tau}(x-\delta,u) = u(x,t)q_{\delta,\tau}(x,u) \\ - \delta\frac{\partial}{\partial x}\left(u(x,t)q_{\delta,\tau}(x,u)\right) \\ + \frac{\delta^2}{2}\frac{\partial^2}{\partial x^2}\left(u(x,t)q_{\delta,\tau}(x,u)\right) + o(\delta^2) \end{aligned}$$

$$\begin{aligned} u(x+\delta,t)p_{\delta,\tau}(x+\delta,u) = u(x,t)p_{\delta,\tau}(x,u) \\ + \delta\frac{\partial}{\partial x}\left(u(x,t)p_{\delta,\tau}(x,u)\right) \\ + \frac{\delta^2}{2}\frac{\partial^2}{\partial x^2}\left(u(x,t)p_{\delta,\tau}(x,u)\right) + o(\delta^2) \end{aligned}$$

En développant le membre de gauche de (5.2) à l'ordre 1 en τ, et en réduisant l'ensemble on obtient

$$\begin{aligned} \tau\frac{\partial}{\partial t}u(x,t) + o(\tau) = \delta\frac{\partial}{\partial x}\left((p_{\delta,\tau}(x,u) - q_{\delta,\tau}(x,u))u(x,t)\right) \\ + \frac{\delta^2}{2}\frac{\partial^2}{\partial x^2}\left((p_{\delta,\tau}(x,u) + q_{\delta,\tau}(x,u))u(x,t)\right) + o(\delta^2) \end{aligned}$$

Ensuite, on divise chaque membre par τ et on fait tendre $\delta, \tau \to 0$, de sorte que $\frac{\delta^2}{\tau}$ reste borné [2], en supposant que les limites suivantes existent, uniformément en u :

$$c(x,u) = \lim_{\delta,\tau}\left(\frac{q_{\delta,\tau}(x,u) - p_{\delta,\tau}(x,u)}{\tau}\delta\right) \tag{5.3}$$

$$d(x,u) = \lim_{\delta,\tau}\left(\frac{p_{\delta,\tau}(x,u) + q_{\delta,\tau}(x,u)}{2\tau}\delta^2\right) \tag{5.4}$$

On obtient alors l'*équation de Fokker-Planck*

$$\frac{\partial u}{\partial t}(x,t) = -\frac{\partial}{\partial x}\left(c(x,u)\,u(x,t)\right) + \frac{\partial^2}{\partial x^2}\left(d(x,u)\,u(x,t)\right) \tag{5.5}$$

2. Plus précisément $a < \frac{\delta^2}{\tau} < b$

qui décrit l'évolution de la densité d'individus en chaque position.

Le coefficient c est appelé *coefficient de convection* et peut être interprété comme la vitesse moyenne des agents situés en x. Le coefficient d est appelé *coefficient de diffusion* et indique la tendance locale des agents à quitter un lieu donné. De par notre approche microscopique $d \geq 0$.

La dynamique du modèle continu que nous avons obtenu est entièrement décrite par les deux coefficients c et d, et le retour aux probabilités des déplacements peut s'effectuer par les relations

$$p_{\delta,\tau}(x,u) \simeq \frac{\tau}{2\delta^2}\left(2d(x,u) - \delta c(x,u)\right) \tag{5.6}$$

$$q_{\delta,\tau}(x,u) \simeq \frac{\tau}{2\delta^2}\left(2d(x,u) + \delta c(x,u)\right). \tag{5.7}$$

Les relations (5.3) et (5.4) peuvent être interprétées comme un ensemble de contraintes asymptotiques sur les grandeurs caractéristiques du système (N, δ et τ) et ses paramètres endogènes ($p_{\delta,\tau}$ et $q_{\delta,\tau}$), nécessaires à la validité du modèle continu (5.5). Ces relations constituent les *conditions de scaling* (ou *conditions de passage à l'échelle*) dans ce contexte.

Plusieurs paramètres $p_{\delta,\tau}$ et $q_{\delta,\tau}$ peuvent donner lieu aux mêmes coefficients c et d. Les systèmes discrets aléatoires correspondants sont représentés par la même équation de Fokker-Planck. Un de ces systèmes discrets est obtenu à l'aide des équations de retour (5.6), (5.7).

Il est intéressant de mentionner que si les déplacements des agents ne sont pas limités à $\pm\delta$, alors l'équation maîtresse (5.2) prend la forme générale suivante :

$$u(x,t+\tau) = \int_y u(y,t)W(y,x) - u(x,y)\int_z W(x,z)dz$$

où $W(y,z)$ est la probabilité de transition de y vers z. La dérivation de l'équation de Fokker-Planck, présentée dans cette section, s'applique également à cette équation [127, 150].

5.2.3 Commentaires

Au moyen d'une approximation en deux étapes (section 5.2), nous avons dérivé une équation d'évolution de type Fokker-Planck à partir du modèle aléatoire initial. Ici, nous faisons une brève digression sur cette équation, en trois remarques. La première concerne la forme de cette équation, qui permet de la rapprocher la thermodynamique, mécanique statistique et la dynamique des populations. La deuxième concerne le choix du domaine circulaire, qui est important pour la bonne définition de l'équation. La troisième concerne la méthode de dérivation en deux étapes qui a été utilisée.

5.2.3.1 Le flux macroscopique

L'équation de Fokker-Planck (5.5) peut s'écrire

$$\frac{\partial u}{\partial t}(x,t) = -\frac{\partial}{\partial x}\left(c(x,u(t))u(x,t) + \frac{\partial}{\partial x}\left(d(x,u(t))u(x,t)\right)\right) = \frac{\partial}{\partial x}J_u(x,t) \tag{5.8}$$

en notant

$$J_u(x,t) = c(x,u(t))u(x,t) + \frac{\partial}{\partial x}\left(d(x,u(t))u(x,t)\right)$$

le flux macroscopique.

Ce flux J_u admet une interprétation physique intéressante. Pour tout sous-intervalle $[a,b]$ du domaine considéré, on observe que

$$\frac{d}{dt}\int_a^b u(x,t)dx = J_u(b,t) - J_u(a,t)$$

Ainsi les variations d'effectif dans un intervalle sont données par les échanges au bords, et quantifiées par la valeur du flux J_u.

Une forme particulière de l'équation de Fokker-Planck est l'équation de la chaleur

$$\frac{\partial u}{\partial t} = d\,\frac{\partial^2 u}{\partial x^2} \tag{5.9}$$

proposée en 1822 par Fourier [60] pour modéliser la diffusion de la chaleur dans un matériau homogène. Il est intéressant de noter qu'en 1855 Fick [59] a proposé la même équation pour modéliser un flux de particules à travers un milieu homogène. Essentiellement, sa démarche peut être comprise en partant de (5.8) et en y intégrant le principe suivant :

« Le flux de diffusion est proportionnel au gradient de concentration. »

qui porte le nom *Loi de Fick*. Dans notre contexte elle se traduit par $J_u(x,t) = d\,\frac{\partial u}{\partial x}(x,t)$ et aboutit immédiatement à l'équation de la chaleur (5.9). L'utilisation de cette équation est donc pertinente en thermodynamique, mécanique statistique et dynamique des populations.

5.2.3.2 Les conditions au bord

Dans cet exemple, le domaine sur lequel les agents se déplacent est le cercle $\mathcal{C} = \mathbb{R}/2\pi\mathbb{Z}$. Il est important de noter que l'équation d'évolution (5.5) n'est valable qu'à l'intérieur de ce domaine. Pour décrire la situation de manière complète il faut joindre des *conditions au bord* à l'équation (5.5), qui précisent explicitement le caractère périodique du domaine.

Les fonctions c et d sont supposées périodiques en x. La périodicité du domaine peut être intégrée à l'équation (5.5) en imposant la condition de périodicité sur le flux

$$J_u(x+2\pi,t) = J_u(x,t),\ \forall x,t$$

qui est trivialement vérifiée lorsque la fonction u est 2π-périodique en x.

Comme nous l'avons indiqué, le choix du domaine périodique offre des avantages théoriques pour les démonstrations formelles qui suivent, et impose des conditions au bord périodiques. On peut envisager d'autres situations, comme la condition de Neumann : les agents se déplacent sur l'intervalle $[0,2\pi]$, et ne peuvent quitter cet espace. Dans ce cas, il faut joindre la condition de Neumann

$$J_u(0,t) = 0 \text{ et } J_u(2\pi,t) = 0,\ \forall t \geq 0$$

qui indique que le flux sortant est nul au bord du domaine. L'équation d'évolution (5.5) est préservée à l'intérieur du domaine.

Les conditions au bord correspondent donc au comportement du système aux limites de l'espace dans lequel les agents se déplacent. Elles ont une forte influence sur l'évolution du système, et sont nécessaires pour une description complète.

En résumé, pour rendre l'équation de Fokker-Planck consistante, il faut y joindre la condition au bord périodique, et une condition initiale. Le modèle continu complet est donc donné par

$$\frac{\partial u}{\partial t}(x,t) = -\frac{\partial}{\partial x}\left(c(x,u)\,u(x,t)\right) + \frac{\partial^2}{\partial x^2}\left(d(x,u)\,u(x,t)\right)$$

muni de la condition initiale

$$u(x,0) = u_0(x)$$

et la condition au bord

$$J_u(x+2\pi,t) = J_u(x,t).$$

5.2.3.3 À propos des raisonnements asymptotiques

L'équation de Fokker-Planck a été obtenue via deux limites successives : $N \to \infty$ puis $\delta, \tau \to 0$. Dans un premier temps, nous avons supposé que le nombre d'agent est grand, ensuite nous avons supposé les pas de temps et espace petits, avec des contraintes spécifiques sur leurs tailles respectives.

Effectuer des limites à tour de rôle présente toujours des risques de non-commutativité, à savoir : qu'en est-il si l'on effectue les trois limites sur N, δ, τ dans un ordre différent ? Il est possible d'obtenir un résultat différent, ou même un résultat complètement incohérent, en permutant ces limites.

La figure 5.4 illustre ces propos. Étant donné un système de tailles caractéristiques (N, δ, τ), il s'agit d'étudier sa position par rapport au point limite $(\infty, 0, 0)$.

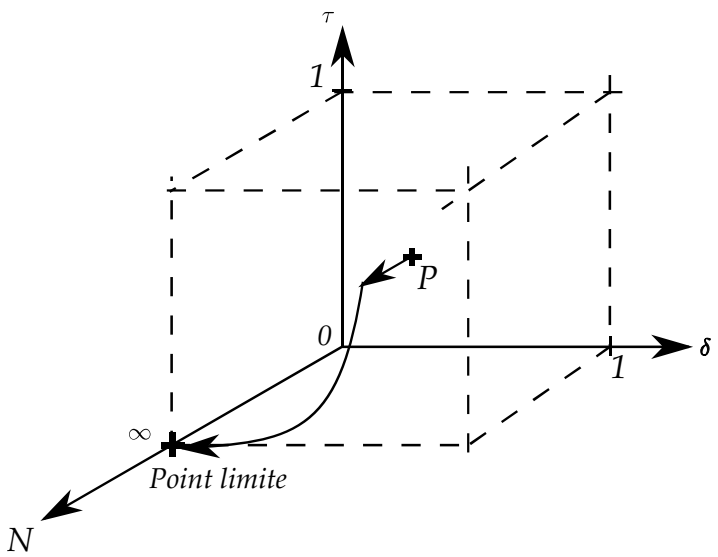


Figure 5.4 — Représentation géométrique des limites effectuées

L'approche que nous utilisons est de projeter le point P sur le plan $N = \infty$, puis à tendre vers le point limite $(\infty, 0, 0)$ de sorte que $\frac{\delta^2}{\tau}$ reste à peu près « constant ». Ce raisonnement est matérialisé par le chemin en traits pleins sur la figure 5.4.

Pour un système discret, avec des valeurs de N, δ et τ données, nous n'avons pas un réel critère qui permet d'évaluer la pertinence de l'approximation caricaturale $N = \infty$, $\delta = \tau = 0$. Une étude possible de cette question, que nous n'avons pas abordée, est de réaliser un grand nombre de simulations avec différentes valeurs de N, δ, τ, et de comparer systématiquement le système discret, le champ moyen et le système continu.

On peut donner une réponse satisfaisante sur la possibilité de permuter les limites $N \to \infty$ et $\delta, \tau \to 0$. En effet, dans [126] Chap. 5, il est montré qu'une famille de marches aléatoires couplées converge vers un *mouvement Brownien* si $\delta, \tau \to 0$ avec $\frac{\tau}{\delta^2} \simeq cste$. Ce résultat correspond à l'équivalent *continu* (en espace et en temps) d'une famille de marches aléatoires discrètes.

En supposant que ces mouvements aléatoires continus sont couplés à travers la densité (continue), on peut effectuer une *limite champ moyen* [18] et montrer que, lorsque $N \to \infty$, la loi d'évolution tend vers une équation de Fokker-Planck. Plus de détails sur cette démarche peuvent être trouvés dans [67].

5.2.4 Formulation continue du problème étudié

À l'aide des équations (5.3) et (5.4) nous mettons en relation les comportements microscopiques (représentés par les fonctions $p^N_{\delta,\tau}$ et $q^N_{\delta,\tau}$) et les paramètres macroscopiques c et d, dans une contexte théorique où $N = \infty$ et $\delta = \tau = 0$.

Grâce à cette correspondance, le choix optimal des déplacements des agents (5.1) est remplacé par un problème continu portant sur c et d :

$$\max_{c,d} \int_{\mathcal{C}} r(x, u(x, T))dx \tag{5.10}$$

sous la contrainte donnée par l'équation (5.5), munie d'une condition initiale u_0 et de la condition au bord indiquée ci-dessus.

Une couple de solutions (c, d) à ce problème ne fournit pas une unique solution au problème initial. En effet, plusieurs probabilités de déplacement $p_{\delta,\tau}, q_{\delta,\tau}$ peuvent vérifier les conditions de passage à la limite (5.3), (5.4) avec cette solution (c, d). Les (5.6), (5.7) donnent une de ces solutions.

Il n'est pas évident que la formulation continue (5.10) ait réellement simplifié le problème discret exposé à la section 5.1.2. L'équation de Fokker-Planck (5.5) peut être difficile à étudier selon la forme des coefficients c et d. Il se peut également que les solutions de cette équation soient irrégulières (même si les coefficients c et d sont réguliers). Ces *solutions faibles* sont des objets purement mathématiques sans interprétation concrète [3].

3. Elles suscitent toutefois un intérêt mathématique. Dans le cadre de l'équation de Fokker-Planck, on peut consulter l'article [94] et l'ouvrage [101] pour de nombreux théorèmes d'existence et de caractérisation des solutions faibles.

Il existe cependant des cas particuliers où l'on dispose de suffisamment de connaissances sur cette équation pour résoudre le problème. La partie suivante est consacrée à un de ces cas particuliers.

5.3 Propriétés de l'équation de Fokker-Planck linéaire

Dans cette partie nous présentons le cas particulier, bien maîtrisé, de l'équation de Fokker-Planck où les coefficients c et d ne dépendent que de x (et pas de la densité u). Ce cas de figure est rencontré si les agents ne tiennent compte que de leur position courante au cours de leurs déplacements, et ignorent leurs congénères.

L'équation de Fokker-Planck correspondante s'écrit

$$\frac{\partial u}{\partial t}(x,t) = -\frac{\partial}{\partial x}\left(c(x)\,u(x,t) + \frac{\partial}{\partial x}\left(d(x)\,u(x,t)\right)\right) \tag{5.11}$$

et est toujours accompagnée d'une condition initiale $u(x,0) = u_0(x)$ et la condition au bord périodique $J_u(x+2\pi,t) = J_u(x,t)$.

L'équation de Fokker-Planck (5.11) est linéaire au sens mathématique, et dans cette partie nous montrons que ce cas particulier est suffisamment bien maîtrisé pour résoudre le problème (5.10).

Dans un premier temps nous vérifions que cette équation possède une solution triviale, qui est accessoirement la solution stationnaire (c'est-à-dire invariante au cours du temps). Ensuite, nous vérifions que l'équation (5.5) constitue un «bon» modèle, dans le sens où elle ne contredit pas immédiatement le sens physique qu'on lui donne. Pour cela, nous montrons qu'elle admet des solutions régulières, et que ces solutions correspondent bien à des densités. Ensuite, nous verrons que sous certaines hypothèses sur les coefficients c et d toutes ces solutions convergent vers la solution stationnaire, et que le système a donc tendance à se stabiliser.

5.3.1 Régime stationnaire de l'équation d'évolution

Une des première questions sur l'équation (5.11) que nous étudions est l'existence de *solutions stationnaires*. Il s'agit des solutions qui ne dépendent pas du temps, et correspondent aux configurations dans lesquelles le système n'évolue pas. Dans cette section, nous montrons qu'une telle position d'équilibre existe, et qu'on peut la calculer de manière explicite.

L'étude de ces états d'équilibre est importante pour les systèmes multi-agents. En effet, il est souhaité que le système ne termine son évolution que lorsqu'il atteint un état satisfaisant pour le concepteur.

L'état stationnaire du système est donné par l'*équation de Fokker-Planck stationnaire* (d'inconnue u_∞) :

$$-\frac{d}{dx}\left(c(x)u_\infty(x)\right) + \frac{d^2}{dx^2}\left(d(x)u_\infty(x)\right) = 0 \tag{5.12}$$

Proposition 7. *On suppose que les fonctions c et d sont continues, et que d est strictement positive.*

(i) Les solutions de l'équation stationnaire (5.12) *sont de la forme*

$$u_\infty(x) = \frac{e^{\int_0^x \frac{c(z)}{d(z)}dz}}{d(x)} \left(C_1 \int_{y=0}^{x} e^{-\int_0^y \frac{c(z)}{d(z)}dz} dy + C_2 \right) \tag{5.13}$$

où C_1 et C_2 sont des constantes réelles.

(ii) L'unique solution de l'équation stationnaire (5.12) *vérifiant $\int_{\mathcal{C}} u_\infty(x)dx = 1$ et la condition de périodicité $u_\infty(0) = u_\infty(2\pi)$ est strictement positive.*

En termes simples, dans le contexte de notre problème, il existe une unique distribution d'agents qui est invariante par la règle d'évolution du système. Cela ne signifie pas que tous les agents sont immobiles. En effet, les agents peuvent continuer leurs déplacements en conservant la même distribution globale.

Démonstration

Pour l'assertion (i), nous intégrons (5.12) par rapport à la variable d'espace x

$$-c(x)u_\infty(x) + \frac{d}{dx}\left(d(x)u_\infty(x)\right) = C_1.$$

Cette égalité traduit que le flux J_{u_∞} est constant à l'équilibre, égal à la constante d'intégration C_1. En remarquant que le terme de gauche peut s'écrire

$$e^{\int_0^x \frac{c(z)}{d(z)}dz} \frac{d}{dx}\left(e^{-\int_0^x \frac{c(z)}{d(z)}dz} d(x)u_\infty(x) \right)$$

(puisque $d(x) > 0$), il vient

$$\frac{d}{dx}\left(e^{-\int_0^x \frac{c(z)}{d(z)}dz} d(x)u_\infty(x) \right) = C_1\, e^{-\int_0^x \frac{c(z)}{d(z)}dz}$$

en intégrant cette égalité on obtient la forme voulue

$$u_\infty(x) = \frac{e^{\int_0^x \frac{c(z)}{d(z)}dz}}{d(x)} \left(C_1 \int_{y=0}^{x} e^{-\int_0^y \frac{c(z)}{d(z)}dz} dy + C_2 \right),\ C_2 \in \mathbb{R}.$$

Pour l'assertion (ii), nous écrivons les conditions $\begin{cases} u_\infty(0) - u_\infty(2\pi) = 0 \\ \int_{\mathcal{C}} u_\infty(x)dx = 1 \end{cases}$ sous forme matricielle

$$A \cdot \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{5.14}$$

où A est la matrice 2×2 définie par

$$A_{1,1} = -\frac{e^{\int_0^{2\pi} \frac{c(z)}{d(z)}dz}}{d(2\pi)} \int_{y=0}^{2\pi} e^{-\int_0^y \frac{c(z)}{d(z)}dz} dy \qquad A_{1,2} = \frac{1}{d(0)} - \frac{e^{\int_0^{2\pi} \frac{c(z)}{d(z)}dz}}{d(2\pi)}$$
$$A_{2,1} = \int_0^{2\pi} \frac{e^{\int_0^x \frac{c(z)}{d(z)}dz}}{d(x)} \int_{y=0}^{x} e^{-\int_0^y \frac{c(z)}{d(z)}dz} dy\, dx \quad A_{2,2} = \int_0^{2\pi} \frac{e^{\int_0^x \frac{c(z)}{d(z)}dz}}{d(x)} dx$$

La matrice A a pour déterminant $\det(A) = A_{1,1}A_{2,2} - A_{2,1}A_{1,2}$, que l'on peut réécrire

$$\det(A) = -\frac{e^{\int_0^{2\pi} \frac{c(z)}{d(z)}dz}}{d(2\pi)} \int_{x=0}^{2\pi} \int_{y=x}^{2\pi} e^{-\int_0^y \frac{c(z)}{d(z)}dz} \frac{e^{\int_0^x \frac{c(z)}{d(z)}dz}}{d(x)} dydx - \frac{1}{d(0)} \int_0^{2\pi} \frac{e^{\int_0^x \frac{c(z)}{d(z)}dz}}{d(x)} \int_{y=0}^{x} e^{-\int_0^y \frac{c(z)}{d(z)}dz} dydx.$$

Sous cette forme il est clair que $\det(A) < 0$. Par conséquent, il existe un unique couple (C_1, C_2) solution du système (5.14), et donc une unique solution stationnaire vérifiant les conditions $\begin{cases} u_\infty(0) - u_\infty(2\pi) = 0 \\ \int_{\mathcal{C}} u_\infty(x)dx = 1 \end{cases}$. Le couple solution est donné par

$$\begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = A^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{\det(A)} \begin{pmatrix} -A_{1,2} \\ A_{1,1} \end{pmatrix} \tag{5.15}$$

et la solution stationnaire correspondante est donnée par

$$u_\infty(x) = \frac{e^{\int_0^x \frac{c(z)}{d(z)}dz}}{d(x)} \begin{pmatrix} C_1 & C_2 \end{pmatrix} \begin{pmatrix} \int_{y=0}^x e^{-\int_0^y \frac{c(z)}{d(z)}dz} dy \\ 1 \end{pmatrix} = \frac{e^{\int_0^x \frac{c(z)}{d(z)}dz}}{d(x)\det(A)} \begin{pmatrix} -A_{1,2} & A_{1,1} \end{pmatrix} \begin{pmatrix} \int_{y=0}^x e^{-\int_0^y \frac{c(z)}{d(z)}dz} dy \\ 1 \end{pmatrix}.$$

Par conséquent

$$u_\infty(x) = \frac{e^{\int_0^x \frac{c(z)}{d(z)}dz}}{d(x)\det(A)} \left(-\int_0^{2\pi} \frac{e^{\int_0^x \frac{c(z)}{d(z)}dz}}{d(x)} \int_{y=0}^x e^{-\int_0^y \frac{c(z)}{d(z)}dz} dydx \int_{y=0}^x e^{-\int_0^y \frac{c(z)}{d(z)}dz} dy - \frac{e^{\int_0^{2\pi} \frac{c(z)}{d(z)}dz}}{d(2\pi)} \int_{y=0}^{2\pi} e^{-\int_0^y \frac{c(z)}{d(z)}dz} dy \right)$$

et cette expression est strictement positive en vertu des facteurs exponentiels. □

On obtient également une propriété du flux à l'équilibre :

Corollaire 1. *Le flux à l'équilibre est nul si et seulement si*

$$\int_0^{2\pi} \frac{c(z)}{d(z)} dz = 0 \tag{5.16}$$

Démonstration

Dans la démonstration de la proposition 7 le flux à l'équilibre est noté C_1. Ce dernier est nul si et seulement si le coefficient $A_{1,2}$ est nul (Cf. équation (5.15)). Par définition $A_{1,2} = 0$ équivaut à $\int_0^{2\pi} \frac{c(z)}{d(z)} dz = 0$. □

Il est possible d'interpréter le fait que le flux à l'équilibre soit non nul. Concrètement cela signifie que les agents se déplacent sans que la densité de répartition globale ne change, et que la vitesse moyenne macroscopique est non nulle. Dans le cas du cercle, le signe du flux constant indique un sens de circulation des agents à l'équilibre qui préserve l'état global du système (la densité). Une telle situation ne correspond clairement pas à notre objectif si les déplacements ont un coût.

Dans le cadre de notre problème, l'équation Fokker-Planck admet une unique solution stationnaire, c'est-à-dire une unique distribution invariante au cours du temps. Si le système est initialement dans cet état, alors il y reste.

5.3.2 L'équation de Fokker-Planck linéaire est bien posée

Dans cette section nous vérifions que l'équation de Fokker-Planck (5.11) admet des solutions régulières, et que ces solutions correspondent à des densités (positives, d'intégrale 1).

Théorème 9. *On suppose que*

- *la condition initiale u_0 et les coefficients c et d sont indéfiniment dérivables.*
- *le coefficient de diffusion d est borné inférieurement*

$$d(x) > \epsilon > 0, \ \forall x \in \mathcal{C}$$

Alors l'équation de Fokker-Planck linéaire (5.11) *admet une unique solution périodique u indéfiniment dérivable.*

Démonstration

La démonstration de ce résultat peut être trouvée dans [53] (Section 7.1.3), [61] (Chap. 9, section 6) ou dans les nombreuses références citées dans [26] (Commentaires du chapitre X). □

Ce résultat montre que l'équation de Fokker-Planck linéaire admet des solutions régulière si les coefficients sont réguliers [4]. En particulier, les solutions n'ont pas de discontinuité et sont bornées.

La condition $d(x) > \epsilon > 0$, parfois appelée *condition de parabolicité*, possède une interprétation simple. Elle signifie que la solution est régulière si la diffusion est suffisamment forte, i.e. si les agents sont suffisamment mobiles. Si tel est le cas on peut exclure des phénomènes de concentration, où la totalité des agents se trouve en une seule position. La densité correspondante (une fonction de Dirac) n'est clairement pas régulière. De manière générale, proposer l'équation de Fokker-Planck comme modèle est pertinent si le système est suffisamment « agité ».

Nous allons voir que les solutions de l'équation de Fokker-Planck linéaire (5.11) correspondent à des densités : elles sont positives et leur intégrale sur le cercle complet $\mathcal{C}$ vaut 1. Avant d'énoncer ce résultat, nous introduisons un lemme très utile qui sera utilisé à de multiples reprises.

Lemme 1. *Soit u une solution régulière de l'équation de Fokker-Planck linéaire* (5.11)*, u_∞ la solution stationnaire de la proposition 7, et h une fonction de classe $\mathcal{C}^2$.*

Alors la fonction H définie par

$$H(t) = \int_{\mathcal{C}} u_\infty \, h\left(\frac{u}{u_\infty}\right)$$

vérifie

$$\dot{H}(t) = -\int_{\mathcal{C}} d\, u_\infty \left[\frac{\partial}{\partial x}\left(\frac{u}{u_\infty}\right)\right]^2 h''\left(\frac{u}{u_\infty}\right).$$

Démonstration

En dérivant la fonction H sous le signe intégral, il vient

$$\dot{H}(t) = \int_{\mathcal{C}} \frac{\partial u}{\partial t} h'\left(\frac{u}{u_\infty}\right) = \int_{\mathcal{C}} \frac{\partial}{\partial x}\left[-cu + \frac{\partial}{\partial x}(du)\right] h'\left(\frac{u}{u_\infty}\right).$$

Nous vérifions que

$$-cu + \frac{\partial}{\partial x}(du) = \frac{u}{u_\infty}\left(-cu_\infty + \frac{\partial}{\partial x}(du_\infty)\right) + du_\infty \frac{\partial}{\partial x}\left(\frac{u}{u_\infty}\right).$$

En remplaçant cette expression dans $\dot{H}$, on obtient

$$\begin{aligned}\dot{H}(t) &= \int_{\mathcal{C}} \frac{\partial}{\partial x}\left[\frac{u}{u_\infty}\left(-cu_\infty + \frac{\partial}{\partial x}(du_\infty)\right) + du_\infty \frac{\partial}{\partial x}\left(\frac{u}{u_\infty}\right)\right] h'\left(\frac{u}{u_\infty}\right) \\ &= \int_{\mathcal{C}} \frac{\partial}{\partial x}\left[\frac{u}{u_\infty}\left(-cu_\infty + \frac{\partial}{\partial x}(du_\infty)\right)\right] h'\left(\frac{u}{u_\infty}\right) \\ &\qquad + \int_{\mathcal{C}} \frac{\partial}{\partial x}\left[du_\infty \frac{\partial}{\partial x}\left(\frac{u}{u_\infty}\right)\right] h'\left(\frac{u}{u_\infty}\right).\end{aligned}$$

4. En réalité les solutions sont au moins aussi régulières que les coefficients [53], et on a un résultat similaire si les coefficients sont seulement de classe $\mathcal{C}^k, k \geq 1$.

Nous montrons que la première intégrale

$$I_1 = \int_{\mathcal{C}} \frac{\partial}{\partial x}\left[\frac{u}{u_\infty}\left(-cu_\infty + \frac{\partial}{\partial x}(du_\infty)\right)\right] h'\left(\frac{u}{u_\infty}\right)$$

est nulle. Comme u_∞ est une solution stationnaire $C_1 = -cu_\infty + \frac{\partial}{\partial x}(du_\infty)$ est constant (Cf. proposition 7). Par conséquent

$$I_1 = C_1 \int_{\mathcal{C}} \left[\frac{\partial}{\partial x}\left(\frac{u}{u_\infty}\right)\right] h'\left(\frac{u}{u_\infty}\right) = C_1 \int_{\mathcal{C}} \frac{\partial}{\partial x}\left[h\left(\frac{u}{u_\infty}\right)\right] = 0$$

comme toutes les fonction en jeu sont périodiques. Il reste donc

$$\dot{H}(t) = \int_{\mathcal{C}} \frac{\partial}{\partial x}\left[du_\infty \frac{\partial}{\partial x}\left(\frac{u}{u_\infty}\right)\right] h'\left(\frac{u}{u_\infty}\right) = -\int_{\mathcal{C}} du_\infty \left[\frac{\partial}{\partial x}\left(\frac{u}{u_\infty}\right)\right]^2 h''\left(\frac{u}{u_\infty}\right)$$

en intégrant par parties, avec les termes au bord nuls en raison de la périodicité des fonctions en jeu. □

Nous pouvons à présent énoncer le résultat qui assure qu'en partant initialement d'une fonction de densité (positive, d'intégrale 1), la solution de l'équation (5.11) reste une fonction de densité :

Proposition 8. *En plus des hypothèses du théorème 9, on suppose que la condition initiale u_0 est positive et vérifie $\int_{\mathcal{C}} u_0(x)dx = 1$. Alors toute solution de l'équation de Fokker-Planck vérifie*

$$u(x,t) \geq 0,\ \forall x, t$$

et

$$\forall t \geq 0, \int_{\mathcal{C}} u(x,t) = 1$$

Démonstration

On introduit la fonction

$$H_0(t) = \int_{\mathcal{C}} u_\infty(x)\, h\left(\frac{u(x,t)}{u_\infty(x)}\right)$$

où u_∞ est la solution stationnaire de la proposition 7, et h la fonction convexe de classe $\mathcal{C}^2$ définie par :

$$h : x \mapsto \begin{cases} -x^3 & \text{si } x \leq 0 \\ 0 & \text{si } x > 0 \end{cases}.$$

En appliquant le lemme 1, on obtient

$$\dot{H}_0(t) = -\int_{\mathcal{C}} du_\infty \left[\frac{\partial}{\partial x}\left(\frac{u}{u_\infty}\right)\right]^2 h''\left(\frac{u}{u_\infty}\right) \leq 0,$$

puisque $h'' \geq 0$.

La fonction H_0 est donc positive décroissante. Or $H_0(0) = 0$ car u_0 est positive, donc la fonction H_0 est nulle. Comme la fonction u_∞ ne s'annule pas, cela entraîne que $h\left(\frac{u}{u_\infty}\right) = 0$ pour tout x et tout t. On conclut que $u(x,t)$ est de même signe que $u_0(x)$, c'est-à-dire positive.

Le théorème de dérivation des intégrales à paramètre donne

$$\frac{d}{dt}\int_{\mathcal{C}} u = \int_{\mathcal{C}} \frac{\partial u}{\partial t} = \int_{\mathcal{C}} \frac{\partial}{\partial x}\left(-cu + \frac{\partial}{\partial x}(du)\right) = 0$$

grâce à la condition au bord périodique. Par conséquent $\int_{\mathcal{C}} u(x,t)dx = \int_{\mathcal{C}} u_0(x)dx = 1, \forall t \geq 0$. □

Remarque 20. *Par un raisonnement analogue, on peut montrer que si la condition initiale u_0 est strictement positive, alors la solution de l'équation de Fokker-Planck est strictement positive.*

Pour résumer les résultats de cette section, l'équation de Fokker-Planck linéaire est « bien posée » : elle admet des solutions régulières qui correspondent à des densités.

5.3.3 Convergence vers le régime stationnaire

Dans ce paragraphe nous étudions la dynamique des solutions non-stationnaires de l'équation de Fokker-Planck. Nous montrons que ces solutions convergent vers la solution stationnaire à vitesse exponentielle.

Cette convergence est importante pour les systèmes multi-agents de deux points de vue. D'une part elle permet de conclure que le système a tendance à se stabiliser selon la distribution stationnaire, ce qui peut être lié à l'auto-organisation du système.

D'autre part, les bornes de convergence dans les théorèmes 10 et 11 quantifient la vitesse de convergence. Cette quantification fournit des réponses sur la robustesse du système vis-à-vis des perturbations. En effet, si l'on perturbe un système qui a atteint sa distribution stationnaire, et nous pouvons quantifier le temps qu'il faut pour que le système y revienne.

La tendance à la stabilisation peut être démontrée à l'aide de plusieurs arguments que nous détaillons dans cette partie. Comme les résultats annoncés sont calculatoires, nous commençons par proposer une légère simplification de l'équation de Fokker-Planck linéaire (5.11).

Proposition 9. *Par le changement de variable $y = \phi(x) = \int_0^x \frac{dz}{\sqrt{d(z)}}$, l'équation de Fokker-Planck est équivalente à*

$$\frac{\partial v}{\partial t} = -\frac{\partial}{\partial y}(c_1 v) + \frac{\partial^2 v}{\partial y^2}$$

où

$$v(y,t) = \sqrt{d(\phi^{-1}(y))}\, u(\phi^{-1}(y), t)$$

et

$$c_1(y) = \frac{1}{\sqrt{d(\phi^{-1}(y)}} \left(c(\phi^{-1}(y)) - \frac{1}{2} d'(\phi^{-1}(y)) \right)$$

Notons que la fonction $(y,t) \mapsto v(y,t)$ est périodique en y, de période $\int_0^{2\pi} \frac{dz}{\sqrt{d(z)}}$.

Démonstration

On note $\tilde{c}(y) = c(\phi^{-1}(y))$ et $\tilde{d}(y) = d(\phi^{-1}(y))$. Des calculs montrent (moyennant une notation abusive confondant les fonction de x et de y) que

$$\frac{\partial u}{\partial t} = \frac{1}{\sqrt{\tilde{d}}} \frac{\partial v}{\partial t},$$

$$\frac{\partial}{\partial x}(cu) = \frac{1}{\sqrt{\tilde{d}}} \frac{\partial}{\partial y} \left(\frac{\tilde{c}}{\sqrt{\tilde{d}}} v \right),$$

puis

$$\frac{\partial}{\partial x}(du) = \frac{1}{\sqrt{\tilde{d}}} \frac{\partial}{\partial y} \left(\sqrt{\tilde{d}} v \right)$$

et

$$\frac{\partial^2}{\partial x^2}(du) = \frac{1}{\sqrt{\tilde{d}}}\frac{\partial}{\partial y}\left(\frac{1}{\sqrt{\tilde{d}}}\frac{\partial}{\partial y}\left(\sqrt{\tilde{d}}v\right)\right) = \frac{1}{\sqrt{\tilde{d}}}\frac{\partial}{\partial y}\left(\frac{d'}{2\sqrt{\tilde{d}}}v + \frac{\partial v}{\partial y}\right).$$

Il reste seulement à vérifier que l'équation $\frac{\partial u}{\partial t} = -\frac{\partial}{\partial x}(cu) + \frac{\partial^2}{\partial x^2}(du)$ devient $\frac{\partial v}{\partial t} = -\frac{\partial}{\partial y}(c_1 v) + \frac{\partial^2 v}{\partial y^2}$ avec les notations introduites dans la proposition. □

Nous supposons donc que $d = 1$ dans la suite de ce chapitre, ce qui n'enlève rien à la généralité des résultats énoncés.

5.3.3.1 Un argument d'analyse linéaire

Avec la simplification motivée par la proposition précédente, l'équation de Fokker-Planck linéaire étudiée est

$$\frac{\partial u}{\partial t} = -\frac{\partial}{\partial x}(cu) + \frac{\partial^2 u}{\partial x^2} = \mathcal{L}u$$

où $\mathcal{L}$ désigne l'opérateur différentiel de Fokker-Planck. La convergence vers le régime stationnaire peut alors se démontrer à l'aide d'une étude spectrale de cet opérateur différentiel.

Proposition 10. *On suppose que $\int_{\mathcal{C}} c(x)dx = 0$. Alors l'opérateur de Fokker-Planck $\mathcal{L}$*

- *est auto-adjoint négatif dans l'espace $L^2(\mu)$, où μ est une mesure bien choisie.*
- *a un unique vecteur propre positif associé à la valeur propre 0 (i.e : une unique mesure stationnaire).*

De plus, toute solution de l'équation de Fokker-Planck converge exponentiellement vite vers la solution stationnaire dans l'espace $L^2(\mu)$.

Cette proposition fournit une réponse satisfaisante sur la convergence du système. En effet, l'écart entre la distribution des agents à l'instant t et la distribution stationnaire tend vers 0 à vitesse exponentielle.

Démonstration

On définit la fonction (périodique) $\mu(x) = e^{-\int_{\eta=0}^{x} c(\eta)d\eta}$ et le produit scalaire

$$< u, v >_{L^2(\mu)} = \int_0^{2\pi} u(x)\, v(x)\mu(x)dx.$$

On constate que

$$\mathcal{L}u = \frac{\partial}{\partial x}\left(e^{\int_{\eta=0}^{x} c(\eta)d\eta}\frac{\partial}{\partial x}\left(e^{-\int_{\eta=0}^{x} c(\eta)d\eta}\, u\right)\right).$$

Par conséquent

$$< \mathcal{L}u, v >_{L^2(\mu)} = \int_0^{2\pi} \frac{\partial}{\partial x}\left(e^{\int_{\eta=0}^{x} c(\eta)d\eta}\frac{\partial}{\partial x}\left(e^{-\int_{\eta=0}^{x} c(\eta)d\eta}\, u\right)\right) v\, \mu\, dx$$

Une intégration par parties, avec un terme au bord nul grâce à la périodicité des fonctions en jeu, laisse

$$< \mathcal{L}u, v >_{L^2(\mu)} = -\int_0^{2\pi} \frac{\partial}{\partial x}\left(e^{-\int_{\eta=0}^{x} c(\eta)d\eta}\, u\right) \frac{\partial}{\partial x}\left(e^{-\int_{\eta=0}^{x} c(\eta)d\eta}\, v\right) e^{\int_{\eta=0}^{x} c(\eta)d\eta} dx.$$

Cette expression est clairement symétrique en (u, v). De plus, elle montre que

$$< \mathcal{L}u, u >_{L^2(\mu)} = -\int_0^{2\pi} \left[\frac{\partial}{\partial x}\left(e^{-\int_{\eta=0}^{x} c(\eta)d\eta}\, u\right)\right]^2 e^{\int_{\eta=0}^{x} c(\eta)d\eta} dx. \tag{5.17}$$

Ainsi, l'opérateur $\mathcal{L}$ est auto-adjoint négatif dans l'espace $L^2(\mu)$. L'existence d'une unique mesure stationnaire et d'un gap spectral est établie par le théorème de Krein-Rutman [32]. Un argument d'analyse fonctionnelle [107], basé sur la théorie des semi-groupes, montre alors que toute solution converge vers la solution stationnaire en norme $L^2(\mu)$. □

L'équation (5.17) montre au passage que la solution stationnaire est un multiple de $e^{\int_{\eta=0}^{x} c(\eta)d\eta}$, ce qui confirme le résultat énoncé dans la proposition 7 dans le cas où $d = 1$ et $\int_{\mathcal{C}} c(x)dx = 0$.

L'approche spectrale présentée ici est limitée aux équations de Fokker-Planck linéaires. Nous avons l'intention de sortir de ce cadre par la suite. De plus, même lorsque l'équation est linéaire, la localisation du spectre peut être une question délicate. Pour ces raisons, nous présentons un argument différent basé sur la notion d'*entropie*.

5.3.3.2 L'entropie quadratique

Soit u une fonction continue, et u_∞ la solution stationnaire de l'équation de Fokker-Planck (5.5). On suppose que la fonction

$$H_1(t) = \int_{x\in\mathcal{C}} \left(\frac{u(x,t)}{u_\infty(x)} - 1\right)^2 u_\infty(x)dx$$

est bien définie et suffisamment régulière pour permettre les calculs qui vont suivre.

Proposition 11. *Si u est une fonction continue et positive telle que $\int_{\mathcal{C}} u(x,t)dx = 1,\ \forall t \geq 0$, alors la fonction H_1 a les propriétés suivantes :*

(i) $H_1(t) = 0 \Leftrightarrow u(\cdot,t) = u_\infty$

(ii) la quantité $H_1(t)$ contrôle l'écart de u à la solution stationnaire en norme L^2 et en norme L^1 :

$$\|u - u_\infty\|_{L^2}^2 \leq \|u_\infty\|_{L^\infty} H_1(t),\ t \geq 0$$

et

$$\|u - u_\infty\|_{L^1}^2 \leq 2\pi\|u_\infty\|_{L^\infty} H_1(t),\ t \geq 0.$$

Démonstration

L'assertion (i) est évidente car la solution stationnaire u_∞ est non-nulle presque partout (Cf. 7). Pour l'assertion (ii), il suffit de réécrire

$$H_1(t) = \int_{\mathcal{C}} (u - u_\infty)^2 \frac{1}{u_\infty}.$$

Par hypothèse, la fonction u_∞ est bornée, donc

$$H_1(t) \geq \frac{1}{\|u_\infty\|_{L^\infty}} \|u - u_\infty\|_{L^2}^2.$$

L'inégalité de Cauchy-Schwartz donne l'inclusion classique $L^1 \subset L^2$:

$$\|u - u_\infty\|_{L^1} = \int_{\mathcal{C}} 1 \times |u - u_\infty| \leq \sqrt{\int_{\mathcal{C}} 1\, dx} \sqrt{\int_{\mathcal{C}} (u - u_\infty)^2 dx}$$

soit

$$\|u - u_\infty\|_{L^1} \leq \sqrt{2\pi}\|u - u_\infty\|_{L^2}$$

□

La proposition 11 montre que la fonction H_1 mesure l'écart du système vis-à-vis de sa position d'équilibre. En effet, la fonction H_1 est nulle si et seulement si $u = u_\infty$, et domine l'écart entre u et u_∞ en norme L^1 et en norme L^2.

Le résultat important qui motive l'introduction de la fonction H_1 est le suivant :

Théorème 10. *Si u est une solution régulière de l'équation de Fokker-Planck linéaire* (5.11) *telle que* $\forall t \geq 0, \int_{\mathcal{C}} u(x,t)dx = 1$ *, alors la fonction H_1 décroît exponentiellement vite vers* 0 *:*

$$H_1(t) \leq H_1(0)e^{-\frac{2}{C_1}t},\ t \geq 0$$

avec $C_1 = 2\pi \| \frac{1}{u_\infty} \|_{L^\infty}$

Pour démontrer ce résultat, nous introduisons un lemme qui sera démontré à l'issue de la démonstration du théorème 10.

Lemme 2 (Inégalité de Poincaré-Wirtinger)**.** *Soit ρ est une mesure de probabilités sur le cercle $\mathcal{C}$ telle que $\frac{1}{\rho}$ est bornée. Alors pour toute fonction f dérivable*

$$\int_{\mathcal{C}} (f - \int_{\mathcal{C}} f(z)\rho(z)dz)^2 \rho \leq C_1 \int_{\mathcal{C}} \left(\frac{df}{dx}\right)^2 \rho$$

avec $C_1 = 2\pi \| \frac{1}{\rho} \|_{L^\infty}$

Démonstration

Nous appliquons directement le lemme 1 avec $h(z) = (z-1)^2$:

$$\dot{H}_1(t) = -2 \int_{\mathcal{C}} u_\infty \left[\frac{\partial}{\partial x} \left(\frac{u}{u_\infty} \right) \right]^2 \tag{5.18}$$

puisque $h''(z) = 2$. À présent, nous utilisons l'inégalité de Poincaré-Wirtinger (lemme 2), en substituant $f = \frac{u}{u_\infty}$ et $\rho = u_\infty$:

$$\int_{\mathcal{C}} \left(\frac{u}{u_\infty} - 1 \right)^2 u_\infty \leq C_1 \int_{\mathcal{C}} \left[\frac{\partial}{\partial x} \left(\frac{u}{u_\infty} \right) \right]^2 u_\infty.$$

On identifie clairement $H_1(t)$ dans le membre de gauche et, grâce à (5.18), une expression proportionnelle à $\dot{H}_1$ dans le membre de droite. Ces identifications mènent à l'inégalité

$$H_1(t) \leq -\frac{C_1}{2} \dot{H}_1(t)$$

soit

$$\dot{H}_1(t) \leq -\frac{2}{C_1} H_1(t).$$

Le lemme de Gronwall permet de conclure que

$$H_1(t) \leq H_1(0)e^{-\frac{2}{C_1}t},$$

ce qui prouve le résultat souhaité. □

Démonstration

(Inégalité de Poincaré-Wirtinger) On part du constat

$$(f(x) - f(y))\rho(y) = \int_{z=y}^{z=x} f'(z)dz\rho(y)$$

qu'on intègre entre $y = 0$ et $y = 2\pi$, en utilisant $\int_{\mathcal{C}} \rho = 1$:

$$f(x) - \int_{y=0}^{2\pi} f(y)\rho(y)dy = \int_{y=0}^{2\pi} \left(\int_{z=y}^{x} f'(z)dz \right) \rho(y)dy.$$

Ainsi

$$\left[f(x) - \int_{y=0}^{2\pi} f(y)\rho(y)dy\right]^2 = \left[\int_{y=0}^{2\pi}\left(\int_{z=y}^{x} f'(z)dz\right)\rho(y)dy\right]^2.$$

Le théorème de Cauchy-Schwartz, relativement à la mesure d'intégration ρ, appliqué au membre de droite laisse

$$\left[f(x) - \int_{y=0}^{2\pi} f(y)\rho(y)dy\right]^2 \leq \int_{y=0}^{2\pi}\left(\int_{z=y}^{x} f'(z)dz\right)^2 \rho(y)dy.$$

En appliquant à nouveau le théorème de Cauchy-Schwartz à l'intégrale portant sur la variable z, le membre de droite est majoré par

$$\int_{y=0}^{2\pi}\left|\int_{z=y}^{x} f'(z)^2 dz\right| \cdot \left|\int_{z=y}^{x} dz\right| \rho(y)dy = \int_{y=0}^{2\pi}\left|\int_{z=y}^{x} f'(z)^2 dz\right| \, |y-x|\rho(y)dy$$

On en déduit que

$$\left(f(x) - \int_{y=0}^{2\pi} f(y)\rho(y)dy\right)^2 \leq \int_{y=0}^{2\pi}\int_{z=0}^{2\pi} f'(z)^2 dz \, |y-x|\rho(y)dy.$$

Le facteur $|y-x|$ est borné par la circonférence du cercle, d'où

$$\left(f(x) - \int_{y=0}^{2\pi} f(y)\rho(y)dy\right)^2 \leq 2\pi \int_{z=0}^{2\pi} f'(z)^2 dz,$$

puis en introduisant artificiellement la mesure de probabilités ρ dans l'intégrale de droite

$$\left(f(x) - \int_{y=0}^{2\pi} f(y)\rho(y)dy\right)^2 \leq 2\pi \|\frac{1}{\rho}\|_{L^\infty} \int_{z=0}^{2\pi} f'(z)^2 \rho(z)dz.$$

On note que le membre de droite ne dépend pas de x. En multipliant de part et d'autre par $\rho(x)$ et en intégrant pour $x \in \mathcal{C}$, on obtient immédiatement le résultat souhaité

$$\int_{x=0}^{2\pi}\left(f(x) - \int_{y=0}^{2\pi} f(y)\rho(y)dy\right)^2 dx \leq 2\pi \|\frac{1}{\rho}\|_{L^\infty} \int_{z=0}^{2\pi} f'(z)^2 \rho(z)dz.$$

□

En combinant la proposition 11 et le théorème 10, nous montrons que les solutions de l'équation de Fokker-Planck linéaire (5.11) vérifient les majorations

$$\|u - u_\infty\|_{L^2} \leq Ae^{-Bt} \quad \text{et} \quad \|u - u_\infty\|_{L^1} \leq \sqrt{2\pi}Ae^{-Bt}$$

avec $A = \sqrt{2\|u_\infty\|_{L^\infty} H_1(0)}$ et $B = \frac{1}{C_1}$. On conclut que ces solutions convergent exponentiellement vite vers la solution stationnaire en norme L^1 et en norme L^2. L'argument clé est une fonction H_1 qui caractérise la convergence des solutions, et qui tend exponentiellement vite vers 0.

5.3.3.3 L'entropie relative de Kullback

Un autre candidat intéressant est la fonction

$$H_2(t) = \int_{z\in\mathcal{C}} u(z,t) \ln\left(\frac{u(z,t)}{u_\infty(z)}\right) dz$$

appelée entropie relative de Kullback [3] de u par rapport à u_∞. Cette fonction n'est définie que pour les solutions strictement positives [5].

5. Nous rappelons que pour une condition initiale strictement positive, la solution est strictement positive.

Proposition 12. *Soit u une fonction continue et positive telle que $\int_{\mathcal{C}} u(x)dx = 1$. Alors la fonction H_2 a les propriétés suivantes :*

(i) La quantité H_2 est positive, et $H_2(t) = 0 \Leftrightarrow u(\cdot, t) = u_\infty$.

(ii) ***Inégalité de Csiszar-Kullback :*** *H_2 contrôle l'écart à la solution stationnaire en norme L^1*

$$\|u - u_\infty\|_{L^1} \leq \sqrt{2H_2(t)}$$

Démonstration

Pour montrer (i), on constate que

$$H_2(t) = \int_{\mathcal{C}} u \ln\left(\frac{u}{u_\infty}\right) - u + u_\infty = \int_{\mathcal{C}} u_\infty \left(\frac{u}{u_\infty} \ln\left(\frac{u}{u_\infty}\right) - \frac{u}{u_\infty} + 1\right)$$

grâce à la normalisation des fonctions u et u_∞. On utilise ensuite l'inégalité réelle

$$x \ln(x) - x + 1 \geq 0,\ x > 0$$

qui ne réalise l'égalité que lorsque $x = 1$. Par conséquent

$$H_2(t) \geq 0 \text{ et } H_2(t) = 0 \Leftrightarrow u(\cdot, t) = u_\infty$$

La démonstration rapide de l'assertion (ii), due à Pinsker[6] [122], consiste a réécrire

$$\|u - u_\infty\|_{L^1} = \int_{\mathcal{C}} \left|\frac{u(x,t)}{u_\infty(x)} - 1\right| u_\infty(x)dx \tag{5.19}$$

et à utiliser l'inégalité

$$3(h-1)^2 \leq (4+2h)(h\ln(h) - h + 1),\ h > 0$$

qui donne

$$|h-1| \leq \frac{1}{\sqrt{3}}\sqrt{4+2h}\sqrt{h\ln(h) - h + 1},\ h > 0.$$

En substituant $h = \frac{u(x,t)}{u_\infty(x)}$ dans (5.19) et en appliquant l'inégalité de Cauchy-Schwartz relativement à la mesure d'intégration $u_\infty(x)$

$$\|u - u_\infty\|_{L^1} = \int_{\mathcal{C}} |h-1| u_\infty(x)dx \leq \frac{1}{\sqrt{3}}\sqrt{\int_{\mathcal{C}} (4+2h)u_\infty(x)dx}\sqrt{\int_{\mathcal{C}} h(\ln(h) - h + 1)u_\infty(x)dx}.$$

D'une part

$$\int_{\mathcal{C}} (4+2h)u_\infty(x)dx = \int_{\mathcal{C}} 4\, u_\infty(x)dx + \int_{\mathcal{C}} 2\, u(x)dx = 6$$

et d'autre part

$$\int_{\mathcal{C}} h(\ln(h) - h + 1)u_\infty(x)dx = H_2(t).$$

Ces égalités permettent de conclure que

$$\|u - u_\infty\|_{L^1} \leq \frac{1}{\sqrt{3}}\sqrt{6}\sqrt{H_2(t)} \leq \sqrt{2H_2(t)}.$$

□

La proposition 12 montre que la fonction H_2 mesure l'écart du système par rapport à sa position d'équilibre, tout comme la fonction H_1. Contrairement à H_1, la fonction H_2 ne domine la distance de u à u_∞ qu'en norme L^1.

La fonction H_2 possède également la propriété importante de tendre vers 0, ce qui prouve la convergence du système vers sa position d'équilibre :

Théorème 11. *Si u est une solution régulière de l'équation de Fokker-Planck linéaire* (5.11) *telle que $\int_{\mathcal{C}} u(x,t)dx = 1,\ \forall t \geq 0$, alors la fonction H_2 décroît exponentiellement vite vers* 0 *:*

$$H_2(t) \leq H_2(0)e^{-\frac{4}{C_2}t}$$

pour une certaine constante C_2

Pour démontrer ce théorème, nous utilisons le résultat suivant :

Lemme 3 (Inégalité de Sobolev logarithmique)**.** *Soit ρ une mesure de probabilité sur le cercle $\mathcal{C}$, alors pour toute fonction f régulière*

$$\int_{\mathcal{C}} f^2 \ln\left(\frac{f^2}{\int_{\mathcal{C}} f^2(z)\rho(z)dz}\right)\rho \leq C_2 \int_{\mathcal{C}} (f')^2\rho$$

où C_2 est une constante positive.

Ce lemme sera admis. Pour sa démonstration, nous renvoyons le lecteur vers [107, 3, 52, 72, 95] ou encore [47].

Démonstration

Nous constatons que

$$H_2(t) = \int_{\mathcal{C}} u \ln\left(\frac{u}{u_\infty}\right) - u + u_\infty = \int_{\mathcal{C}} u_\infty \left(\frac{u}{u_\infty}\ln\left(\frac{u}{u_\infty}\right) - \frac{u}{u_\infty} + 1\right)$$

grâce à la normalisation des fonctions u et u_∞. Par conséquent

$$H_2(t) = \int_{\mathcal{C}} u_\infty\, h\left(\frac{u}{u_\infty}\right)$$

avec $h(z) = z\ln(z) - z + 1$. Nous appliquons directement le lemme 1 :

$$\dot{H}_2(t) = -\int_{\mathcal{C}} u_\infty \frac{\left[\frac{\partial}{\partial x}\left(\frac{u}{u_\infty}\right)\right]^2}{\frac{u}{u_\infty}}$$

puisque $h''(z) = \frac{1}{z}$. Cette égalité se réécrit :

$$\dot{H}_2(t) = -4\int_{\mathcal{C}} u_\infty \left(\frac{\partial}{\partial x}\left(\sqrt{\frac{u}{u_\infty}}\right)\right)^2 \tag{5.20}$$

Nous utilisons à présent l'inégalité de Sobolev logarithmique (lemme 3), en substituant $f^2 = \frac{u}{u_\infty}$ et $\rho = u_\infty$. Tout d'abord :

$$\int_{\mathcal{C}} f^2(z)\rho(z)dz = \int_{\mathcal{C}} u(z)dz = 1.$$

Ensuite, l'inégalité de Sobolev logarithmique donne

$$\int_{\mathcal{C}} u \ln\left(\frac{u}{u_\infty}\right) \leq C_2 \int_{\mathcal{C}} \left(\frac{\partial}{\partial x}\left(\sqrt{\frac{u}{u_\infty}}\right)\right)^2 u_\infty.$$

On identifie clairement la fonction H_2 dans le membre de gauche et, grâce à l'égalité (5.20), une expression proportionnelle à $\dot{H}_2(t)$ dans le membre de droite

$$H_2(t) \leq -\frac{C_2}{4}\dot{H}_2(t).$$

Ainsi

$$\dot{H}_2(t) \leq -\frac{4}{C_2}H_2(t)$$

et le lemme de Gronwall permet de conclure que

$$H_2(t) \leq H_2(0)e^{-\frac{4}{C_2}t},$$

ce qui montre le résultat souhaité. □

En combinant la proposition 12 et le théorème 11, nous montrons que les solutions de l'équation de Fokker-Planck linéaire (5.11) vérifient la majoration

$$\|u - u_\infty\|_{L^1} \leq Ae^{-Bt}$$

pour les deux constantes positives $A = \sqrt{2H_2(0)}$ et $B = \frac{2}{C_2}$. On conclut que ces solutions convergent vers la solution stationnaire à vitesse exponentielle.

Le fait d'introduire la fonction H_2 peut sembler superflu, car la fonction H_1 permet déjà de démontrer la convergence exponentielle des solutions. L'intérêt de la fonction H_2 apparaît lorsqu'on veut généraliser le raisonnement. Si l'espace d'états considéré n'est plus de mesure finie, l'inclusion $L^1 \subset L^2$ n'est plus valable, et H_1 offre seulement un contrôle sur la norme L^2. Or l'hypothèse de conservation d'effectif $\int u = 1$, inhérente aux système fermés, se traduit naturellement au niveau de la norme L^1.

De plus, nous verrons que dans un exemple non-linéaire traité plus loin (section 5.4.2) la fonction H_2 fournit des résultats de convergence qui semblent moins évidents à démontrer avec la fonction H_1. D'autres arguments plus techniques en faveur de H_2 sont donnés dans [107].

Nous avons mis en évidence deux fonctions macroscopiques (entropies) H_1 et H_2 qui contrôlent l'écart d'une solution à la solution stationnaire, décroissante au cours du temps. Dans la littérature, ce phénomène est connu sous le nom *dissipation d'entropie* [107, 3] ou *théorème H*. Il entraîne en particulier que la dynamique de l'équation de Fokker-Planck est *irréversible* : partant d'une distribution donnée, il est impossible d'évoluer et d'y revenir ultérieurement.

Remarque 21. *La notion d'entropie définie en thermodynamique correspond en réalité à l'opposé de l'entropie utilisée en physique statistique [71]. Pour cette raison certains travaux parlent plutôt de création d'entropie et non de dissipation d'entropie*[7].

Pour conclure sur l'équation de Fokker-Planck linéaire : nous avons montré que cette équation est bien posée, dans le sens qu'elle admet des solutions physiquement acceptables. De plus, la dynamique de ces solutions est simple :

- ou bien le système est initialisé selon la distribution stationnaire, et reste dans cette distribution.
- ou bien le système est initialisé selon une autre distribution, et dans ce cas il converge vers la distribution stationnaire à vitesse exponentielle.

Dans la section suivante, nous montrons comment cette information peut être utilisée pour résoudre le problème de l'orientation des agents vers une distribution ciblée.

5.4 Stratégies optimales

Dans cette partie, nous déterminons deux stratégies optimales au sens de la satisfaction globale définie à la partie 5.1.2. L'optimisation que nous effectuons consiste à choisir judi-

7. On rencontre même les deux terminologies dans les travaux de C. Villani, qui explique ce changement de vocabulaire sur son site `http://cedricvillani.org/for-mathematicians/presentation-of-my-research/`

cieusement les coefficients c et d de sorte à maximiser la récompense globale $R(T)$ pour un instant T lointain.

5.4.1 Une stratégie dépendant localement de la répartition ciblée

Nous nous plaçons dans le cas linéaire, où les coefficients c et d ne dépendent que de x. Une manière simple pour maximiser la satisfaction globale à long terme est d'identifier le régime stationnaire u_∞ à la répartition optimale f. Cette démarche revient à fixer l'état d'équilibre du système. Si l'on parvient à trouver des coefficient c et d qui ne dépendent que de x, et qui sont compatibles avec cet équilibre, alors les résultats de la section 5.3.3 garantissent la convergence à vitesse exponentielle vers la répartition optimale.

En remplaçant u_∞ par f dans l'équation stationnaire (5.12) dans le cas où le flux à l'équilibre est nul, on obtient

$$-c(x)f(x) + \frac{d}{dx}(d(x)f(x)) = 0$$

qui est une condition nécessaire pour l'optimalité des coefficients c et d.

Une solution évidente à ce problème est

$$c(x) = \frac{f'(x)}{f(x)} = \frac{d}{dx}\ln f(x) \quad d(x) = 1. \tag{5.21}$$

Nous faisons plusieurs observations sur ces expressions. La solution stationnaire pour ces coefficients (Cf. (5.13)) correspond à

$$u_\infty(x) = \frac{f(x)}{\int_{\mathcal{C}} f(z)dz}$$

c'est-à-dire la distribution objectif f avec une intégrale normalisée à 1.

Ensuite, si la fonction f possède un ou plusieurs points d'annulation, alors l'expression proposée pour c n'est plus valable. Le sens à donner à cette limitation est que la vitesse moyenne des particules en ces points est infinie. Il est bien sûr possible de définir la fonction c presque partout, et de considérer l'équation de Fokker-Planck au sens faible [8].

Nous constatons que le coefficient de convection optimal [9] c est proportionnel à la dérivée logarithmique de la densité des ressources $\frac{d}{dx}\ln f$. Ainsi, pour récolter de manière optimale, les agents doivent observer l'ordre de grandeur des quantités de ressource environnantes, et la sensibilité varie comme le logarithme de l'excitation. Ce phénomène est connu sous le nom de *loi de Weber-Fechner* [56, 108] en psychophysique. Une conséquence de cette loi est que le fait de multiplier la répartition f par un nombre strictement positif n'a aucune influence sur la valeur de c (et donc sur le comportement des agents).

Au-delà de ces remarques portant sur la forme des coefficients (5.21), cette stratégie montre un fait important : il est possible de piloter les agents vers la ressource avec un

8. Par exemple, pour le cas $f(x) = |sin(x/2)|$ le coefficient c admet deux singularités pouvant être interprétées comme des barrières de hauteur infinie. La résolution de ce cas à l'aide d'une méthode spectrale peut être trouvée dans [127].

9. À diffusion constante, égale à 1.

contrôle local. En effet, les coefficients c et d ne dépendent que de la fonction f et sa dérivée au point x.

Un autre fait intéressant est que sous cette stratégie, les agents n'ont aucune interaction. L'objectif global peut donc être réalisé par des agents indépendants.

Par contre, le coefficient de convection c est calculé à partir de la répartition visée f. Nous avons donc implicitement admis que les agents perçoivent cette fonction et ses dérivées (localement). Dans un contexte plus général où l'on souhaite répartir des agents selon une certaine distribution, ils ne sont pas nécessairement capables de la percevoir.

En vertu de ces remarques, la stratégie définie par les coefficients (5.21) correspond à une influence locale exercée sur les agents par la fonction f. Cette influence correspond plutôt à l'action d'un contrôleur global qu'au résultat d'un raisonnement et d'une décision individuels. On ne peut donc pas réellement parler d'*autonomie* pour cette stratégie.

Pour ces raisons, nous proposons une autre stratégie dans la partie suivante. Cette stratégie fait intervenir des interactions entre les agents, et se base sur la connaissance locale des récompenses.

5.4.2 Une stratégie basée sur la satisfaction locale

Une stratégie intéressante, inspirée par celle de la section 5.4.1, est donnée par les paramètres c et d suivants

$$c(u,x) = \frac{\partial}{\partial x} \ln \left(u(x) e^{1-\frac{u(x)}{f(x)}} \right) \qquad d(u,x) = 1. \tag{5.22}$$

où $r(x,u) = u(x)\, e^{1-\frac{u(x)}{f(x)}}$ est la fonction de récompense locale.

Le coefficient de diffusion d est choisi constant égal à 1. Le coefficient de convection c est choisi comme la dérivée logarithmique de la récompense locale. De manière intuitive, ce choix des paramètres correspond à une tendance de déplacement vers les lieux où la récompense est la plus élevée. Comme le suggère la loi de Weber-Fechner (Cf. paragraphe précédent) l'intensité des déplacements varie comme la dérivée logarithmique de l'excitation, soit ici la récompense locale.

Les paramètres définis par (5.22) mènent à l'équation non linéaire

$$\frac{\partial u}{\partial t} = -\frac{\partial}{\partial x} \left[u \frac{\partial}{\partial x} \ln \left[u(x,t) e^{1-\frac{u(x,t)}{f(x)}} \right] \right] + \frac{\partial^2 u}{\partial x^2} \tag{5.23}$$

munie de la condition initiale

$$u(x,0) = u_0(x)$$

(où u_0 est une densité sur le cercle $\mathcal{C}$), et la condition au bord périodique

$$u(x,t) = u(x+2\pi,t).$$

La nonlinéarité de cette équation est due au coefficient c qui dépend de la densité u, ainsi que de sa dérivée spatiale $\frac{\partial u}{\partial x}$, à travers la fonction de récompense locale $r(u,x)$. Les résultats établis à la section 5.3 sont limités au cas linéaire, et donc inutilisables ici.

5.4.2.1 Premières propriétés de l'équation d'évolution

Nous commençons par une simplification de l'équation :

Proposition 13. *L'équation* (5.23) *est équivalente à*

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left[u\frac{\partial}{\partial x}\left(\frac{u}{f}\right)\right] \tag{5.24}$$

Démonstration

Il suffit de constater que

$$c = \frac{\partial}{\partial x}\ln\left[ue^{1-\frac{u}{f}}\right] = \frac{\frac{\partial u}{\partial x}}{u} - \frac{\partial}{\partial x}\left(\frac{u}{f}\right).$$

En substituant cette expression dans (5.23) on obtient

$$\frac{\partial u}{\partial t} = -\frac{\partial}{\partial x}\left[\left[\frac{\frac{\partial u}{\partial x}}{u} - \frac{\partial}{\partial x}\left(\frac{u}{f}\right)\right]u\right] + \frac{\partial^2 u}{\partial x^2} = \frac{\partial}{\partial x}\left[u\frac{\partial}{\partial x}\left(\frac{u}{f}\right)\right]$$

□

À présent nous vérifions que l'équation (5.24) constitue un «bon» modèle, en montrant qu'elle admet des solutions et que ces solutions correspondent à des densités.

Théorème 12. *Si la condition initiale u_0 et la fonction f sont de classe C^∞ et strictement positives, alors l'équation* (5.24) *admet une solution de classe C^∞ strictement positive.*

Démonstration

Nous effectuons la transformation suivante $v(x,t) = \frac{u(x,t)}{f(x)}$. L'équation (5.24) devient alors :

$$\frac{\partial v}{\partial t} = \frac{1}{f}\frac{\partial}{\partial x}\left[fv\frac{\partial v}{\partial x}\right] = \frac{\partial}{\partial x}\left[v\frac{\partial v}{\partial x}\right] + \frac{f'}{f}v\frac{\partial v}{\partial x}, \tag{$*$}$$

avec la condition initiale $v_0(x) = \frac{u_0(x)}{f(x)}$. L'équation $(*)$ est une équation parabolique nonlinéaire, de la forme générale

$$\frac{\partial v}{\partial t} = \frac{\partial}{\partial x}\left[a\left(x,t,v,\frac{\partial v}{\partial x}\right)\right] + b\left(x,t,v,\frac{\partial v}{\partial x}\right). \tag{$\dagger$}$$

Dans notre cas

$$a(x,t,v,\frac{\partial v}{\partial x}) = v\frac{\partial v}{\partial x}$$

et

$$b\left(x,t,v,\frac{\partial v}{\partial x}\right) = \frac{f'(x)}{f(x)}\,v\,\frac{\partial v}{\partial x}$$

La théorie générale des équation de cette forme [61, 101, 90] montre que si a et b sont infiniment dérivables en leurs arguments, et si

$$A_1 \leq \frac{\partial a(x,t,v,p)}{\partial p} \leq A_2$$

pour deux constantes $0 < A_1 < A_2$, alors (†) possède une unique solution de classe C^∞ (Cf. [145] sect. 3.1). Malheureusement, dans notre cas, cette condition n'est pas vérifiée, puisque $\frac{\partial a(x,t,v,p)}{\partial p} = v$ peut s'annuler (nous avons affaire à une équation parabolique *dégénérée*.).

Ce problème est réglé en montrons que, puisque la solution initiale est strictement positive, la solution v est strictement positive pour tout $t \geq 0$. Il s'avère donc que notre équation est *nondégénérée*.

Puisque u_0 et f sont supposées régulières et strictement positives, la condition initiale

$$v_0(x) = \frac{u_0(x)}{f(x)}$$

vérifie

$$\varepsilon \leq v_0(x) \leq \frac{1}{\varepsilon}$$

pour une certaine constante $\varepsilon > 0$. À présent nous introduisons une fonction φ vérifiant

$$\varphi(z) = z \text{ si } \varepsilon \leq z \leq \frac{1}{\varepsilon}$$

qui est prolongée en une fonction régulière sur $\mathbb{R}$ de sorte que $A_1 < \varphi(z) < A_2$ pour deux constantes $0 < A_1 < A_2$. Ce prolongement peut être réalisé par des fonctions constantes au voisinage de $z = 0$ et $z = \infty$, et raccordant ces constantes de manière régulière en $z = \varepsilon$ et en $z = \frac{1}{\varepsilon}$ (voir figure 5.5).

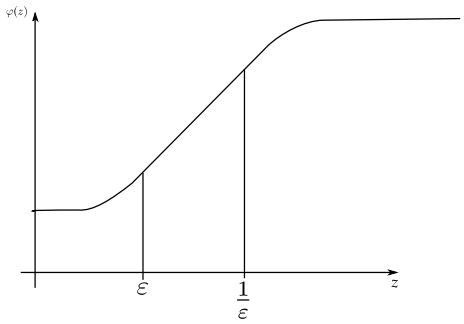


Figure 5.5 — Fonction φ

L'équation $(*)$ est remplacée par :

$$\frac{\partial v}{\partial t} = \frac{\partial}{\partial x}\left[\varphi(v)\frac{\partial v}{\partial x}\right] + \frac{f'}{f} v \frac{\partial v}{\partial x}. \tag{††}$$

Cette équation à la forme (†), i.e. parabolique, avec les coefficients

$$a(x,t,v,\frac{\partial v}{\partial x}) = \varphi(v)\frac{\partial v}{\partial x},\ b(x,t,v,\frac{\partial v}{\partial x}) = \frac{f'(x)}{f(x)}\ v\ \frac{\partial v}{\partial x}.$$

Par contre, comme

$$\frac{\partial a(x,t,v,p)}{\partial p} = \varphi(v)$$

et

$$A_1 \leq \frac{\partial a(x,t,v,p)}{\partial p}) \leq A_2,$$

il s'avère que (††) est nondégénérée. Par conséquent, cette équation a une solution v de classe $\mathcal{C}^\infty$. De plus, nous avons le théorème de comparaison suivant [145] :

Si v et $\tilde{v}$ sont deux solutions régulières de (††), et si leurs conditions initiales vérifient

$$v_0(x) \leq \tilde{v}_0(x)$$

alors $v(x,t) = \tilde{v}(x,t),\ \forall x,t$ ou bien $v(x,t) < \tilde{v}(x,t),\ \forall x,t$.

Puisque les deux constantes ε et $\frac{1}{\varepsilon}$ sont solutions de (††), et parce que $\varepsilon \leq v_0 \leq \frac{1}{\varepsilon}$ on peut dire que

$$\varepsilon \leq v(x,t) \leq \frac{1}{\varepsilon}$$

pour tout (x,t). Par conséquent $\varphi(v(x,t)) = v(x,t)$ pour tout (x,t), et v est donc une solution régulière de $(*)$. Le théorème 12 est obtenu en constatant que

$$u(x,t) = v(x,t)f(x)$$

est une solution régulière strictement positive de (5.23). □

Proposition 14. *Si, en plus des hypothèses du théorème 12, on suppose que la condition initiale u_0 vérifie $\int_{\mathcal{C}} u_0(x)dx = 1$. Alors toute solution de* (5.24) *vérifie*

$$\int_{\mathcal{C}} u(x,t)dx = 1,\ \forall t \geq 0.$$

Démonstration

Une dérivation sous le signe intégral donne

$$\frac{\partial}{\partial t}\int_{\mathcal{C}} u = \int_{\mathcal{C}} \frac{\partial u}{\partial t} = \int_{\mathcal{C}} \frac{\partial}{\partial x}\left[u\frac{\partial}{\partial x}\left(\frac{u}{f}\right)\right] = 0$$

grâce à la périodicité des fonctions en jeu. Par conséquent, l'intégrale de u est constante au cours du temps, égale à sa valeur initiale $\int_{\mathcal{C}} u_0 = 1$. □

5.4.2.2 Régime stationnaire et convergence des solutions

Nous avons démontré que l'équation (5.24) admet des solutions qui correspondent à des densités. À présent, nous démontrons que toutes ces solutions convergent vers l'équilibre optimal f.

Proposition 15. *(i) L'équation stationnaire*

$$\frac{\partial}{\partial x}\left[u\frac{\partial}{\partial x}\left(\frac{u}{f}\right)\right] = 0$$

admet une unique solution périodique vérifiant $\int_{\mathcal{C}} u = 1$, et cette solution est la fonction f.

(ii) Toute solution de (5.24) *converge vers f en norme L^1 à vitesse exponentielle*

$$\|u - f\|_{L^1} \leq Ke^{-\frac{1}{2C_1}t}$$

où K est une constante positive et $C_1 = 2\pi\|\frac{1}{f}\|_{L^\infty}$.

Démonstration

Pour l'assertion (i), l'équation stationnaire

$$\frac{\partial}{\partial x}\left[u\frac{\partial}{\partial x}\left(\frac{u}{f}\right)\right] = 0$$

est directement intégrée en

$$u\frac{\partial}{\partial x}\left(\frac{u}{f}\right) = C,\ C \in \mathbb{R}$$

puis

$$\frac{u}{f}\frac{\partial}{\partial x}\left(\frac{u}{f}\right) = \frac{1}{2}\frac{\partial}{\partial x}\left[\left(\frac{u}{f}\right)^2\right] = \frac{C}{f}.$$

Ainsi, par périodicité de u et f,

$$0 = \int_{\mathcal{C}} \frac{1}{2}\frac{\partial}{\partial x}\left[\left(\frac{u}{f}\right)^2\right] = C\int_{\mathcal{C}} \frac{1}{f}$$

donc $C = 0$ et par conséquent $\frac{u}{f}$ est constant. Comme $\int_{\mathcal{C}} u = \int_{\mathcal{C}} f$ cela signifie que les fonction u et f sont identiques.

Pour l'assertion (ii), on utilise la fonction d'entropie

$$H_2(t) = \int_{\mathcal{C}} u \ln\left(\frac{u}{f}\right)$$

Nous savons déjà que $H_2(t) \geq 0$ et $\|u - f\|_{L^1} \leq \sqrt{2H(t)}$ (Cf. l'inégalité de Csiszar-Kullback : assertion (ii) de la proposition 12). En dérivant sous le signe intégral, il vient

$$\dot{H}_2(t) = \int_{\mathcal{C}} \frac{\partial u}{\partial t} \ln\left(\frac{u}{f}\right) + \frac{\partial u}{\partial t} = \int_{\mathcal{C}} \frac{\partial u}{\partial t} \ln\left(\frac{u}{f}\right).$$

Comme u est solution de (5.24),

$$\dot{H}_2(t) = \int_{\mathcal{C}} \left[\frac{\partial}{\partial x}\left(\frac{u}{f}\right) u\right] \ln\left(\frac{u}{f}\right).$$

Une intégration par parties (avec un terme au bord nul) donne

$$\dot{H}_2(t) = -\int_{\mathcal{C}} \frac{\partial}{\partial x}\left(\frac{u}{f}\right) u \frac{\partial}{\partial x}\left[\ln\left(\frac{u}{f}\right)\right] = -\int_{\mathcal{C}} \left[\frac{\partial}{\partial x}\left(\frac{u}{f}\right)\right]^2 f.$$

On utilise à présent l'inégalité de Wirtinger (lemme 2) avec $f = \frac{u}{f}$ et $\rho = f$. Elle donne

$$\dot{H}_2(t) \leq -\frac{1}{C_1} \int_{\mathcal{C}} \left(\frac{u}{f} - 1\right)^2 f$$

avec $C_1 = 2\pi \|\frac{1}{f}\|_{L^\infty}$. L'inégalité réelle

$$X \ln(X) - X + 1 \leq (X-1)^2,\ X > 0$$

fournit, en substituant $X = \frac{u}{f}$,

$$\frac{u}{f} \ln\left(\frac{u}{f}\right) - \frac{u}{f} + 1 \leq \left(\frac{u}{f} - 1\right)^2$$

puis

$$\int_{\mathcal{C}} \left(\frac{u}{f} \ln\left(\frac{u}{f}\right) - \frac{u}{f} + 1\right) f \leq \int_{\mathcal{C}} \left(\frac{u}{f} - 1\right)^2 f.$$

Par conséquent

$$\dot{H}_2(t) \leq -\frac{1}{C_1} \int_{\mathcal{C}} \left(\frac{u}{f} \ln\left(\frac{u}{f}\right) - \frac{u}{f} + 1\right) f = -\frac{1}{C_1} \int_{\mathcal{C}} u \ln\left(\frac{u}{f}\right) - u + f$$

et comme $\int_{\mathcal{C}} u = \int_{\mathcal{C}} f$, il reste

$$\dot{H}_2(t) \leq -\frac{1}{C_1} \int_{\mathcal{C}} u \ln\left(\frac{u}{f}\right) = -\frac{1}{C_1} H_2(t).$$

Le lemme de Gronwall permet alors de dire que

$$H_2(t) \leq H_2(0) e^{-\frac{1}{C_1} t}.$$

En conclusion

$$\|u - f\|_{L^1} \leq \sqrt{2H_2(t)} \leq \sqrt{2H_2(0)} e^{-\frac{1}{2C_1} t}.$$

ce qui correspond à l'assertion (ii) en posant $K = \sqrt{2H_2(0)}$. □

En résumé, nous avons prouvé que le choix des paramètres

$$c(u,x) = \frac{\partial}{\partial x} \ln\left(u(x,t) e^{1 - \frac{u(x,t)}{f(x)}}\right), \quad d(u,x) = 1 \tag{5.25}$$

constitue une stratégie intéressante pour orienter la distribution des agents vers la distribution optimale.

Nous avons démontré que l'équation non linéaire qui en résulte est bien posée : elle admet des solutions qui correspondent à des densités. La dynamique de ces solutions est celle escomptée : toutes convergent à vitesse exponentielle vers la répartition optimale f.

La vitesse de convergence est quantifiée par le coefficient $\frac{1}{2C_1}$, où C_1 est la constante de l'inégalité de Poincaré-Wirtinger (lemme 2). Cette valeur est à comparer à la vitesse de convergence dans le cas linéaire (section 5.3.3.3) qui vaut $\frac{2}{C_2}$, où C_2 est la constante de l'inégalité de Sobolev logarithmique.

Les constantes C_1 et C_2 dépendent a priori de la topologie du domaine considéré (ici, le cercle) et de la solution stationnaire u_∞. On sait que $C_2 \geq 2C_1$ [47] mais cette inégalité ne permet pas de comparer $\frac{1}{2C_1}$ et $\frac{2}{C_2}$. Les coefficients C_1 et C_2 sont généralement difficiles à évaluer, et nous n'avons pas trouvé de résultat simple permettant de quantifier C_2.

Par ailleurs, il n'est pas certain que la valeur $C_1 = 2\pi\|\frac{1}{u_\infty}\|_{L^\infty}$ donnée dans ce manuscrit soit optimale (vis-à-vis de l'inégalité de Poincaré-Wirtinger). Par conséquent, la comparaison des taux de convergence est difficile et il n'est pas sûr qu'elle offre un résultat valable pour tous les domaines et toutes les positions d'équilibre u_∞ possibles.

Les simulation numériques montrent que dans le cas unidimensionnel du cercle (section 5.5), la seconde stratégie est plus performante, mais que dans le cas bidimensionnel du tore (section 5.6) aucune des deux stratégies ne domine l'autre.

Il est intéressant de noter que les paramètres définis par (5.25) dépendent seulement de la récompense locale (et sa dérivée spatiale), et n'exigent pas que les agents connaissent la distribution optimale f. La convergence vers cette répartition est donc émergente, puisque les agents n'ont pas connaissance de cet objectif. Le fait d'ignorer cet objectif, et de baser ses actions sur des informations locales issues d'autres agents (la satisfaction locale) est une forme d'autonomie.

5.5 Validation numérique

À présent, nous allons valider et illustrer les résultats théoriques de ce chapitre par des simulations numériques. Les deux aspects mis en avant sont la validité du modèle proposé et l'optimalité des paramètres macroscopiques déterminés à la section 5.4. Pour cela, nous nous limitons à la stratégie centralisée établie à la section 5.4.1.

5.5.1 Validation du modèle

La dérivation de l'équation de Fokker-Planck est réalisée par l'intermédiaire d'un modèle à champ moyen. Nous allons donc comparer trois systèmes différents :

- Un **système stochastique** formé d'un grand nombre d'agents se qui se déplacent de manière aléatoire sur l'ensemble discret $\mathcal{C}_\delta$.
- Le **champ moyen** correspondant à la limite $N \to \infty$ du système précédent.
- Le **système continu** en temps et espace, régi par l'équation de Fokker-Planck. Ce système correspond à la limite $\delta, \tau \to 0$ du champ moyen.

Pour toutes les simulations, la ressource est répartie selon la fonction

$$f(x) \propto \sin\left(\frac{x}{2}\right)^{10} + 0,01$$

d'intégrale unitaire sur le cercle. Sa courbe est représentée sur la figure 5.6, et sera indiquée en pointillés bleus sur les figures suivantes.

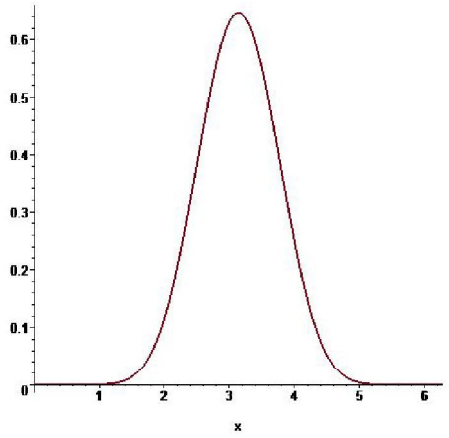


Figure 5.6 — Répartition de la ressource.

Dans les trois sections suivantes, nous présentons les résultats des simulations sans commentaires. Une analyse générale est faite dans la section 5.5.1.4.

5.5.1.1 Le système discret

Le système discret, représenté sur la figure 5.7, est composé de $N = 500$ agents en déplacement aléatoire. Le cercle $\mathcal{C}$ est divisé de façon régulière en 20 positions. Ainsi, le pas spatial est $\delta = \frac{2\pi}{20} \simeq 0,3$. L'intervalle de temps est fixé à $\tau = 0,1$. Les probabilités des déplacements sont

$$p_{\delta,\tau}(x,u) \simeq \frac{\tau}{2\delta^2}\left(2d_{\delta,\tau}(x) + \delta c_{\delta,\tau}(x)\right) \quad \text{et} \quad q_{\delta,\tau}(x,u) \simeq \frac{\tau}{2\delta^2}\left(2d_{\delta,\tau}(x) - \delta c_{\delta,\tau}(x)\right)$$

où les fonctions $c_{\delta,\tau}$ et $d_{\delta,\tau}$ sont définies par

$$c_{\delta,\tau}(x,u) = \frac{f(x+\delta) - f(x-\delta)}{2\delta f(x)} \quad \text{et} \quad d_{\delta,\tau}(x,u) = 1$$

Ces agents sont initialement répartis aléatoirement, de manière uniforme, sur l'espace des positions $\mathcal{C}_\delta$.

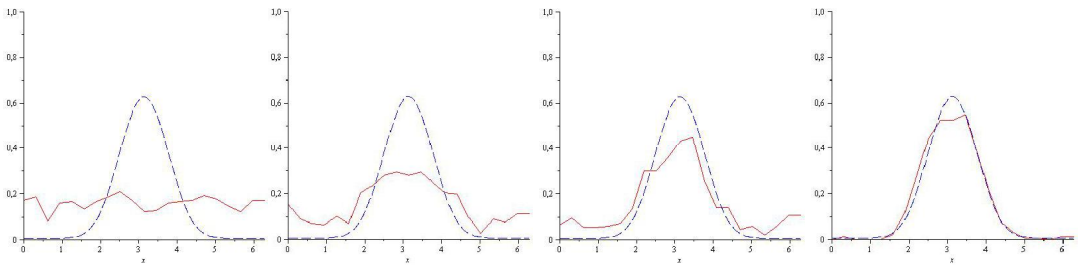


Figure 5.7 — Système discret aux instants $t = 0$, $t = 0,05$, $t = 0,5$ et $t = 2,5$

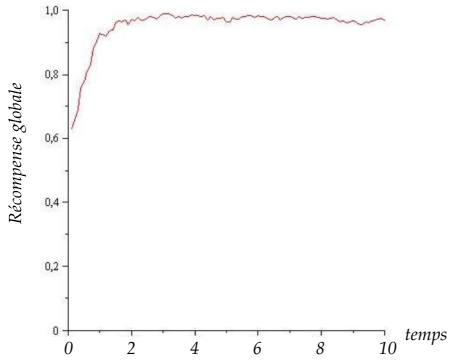


Figure 5.8 — Évolution de la récompense du système discret au cours du temps

5.5.1.2 Le champ moyen

Le champ moyen, représenté sur la figure 5.9, est initialisé selon la distribution uniforme discrète $u_0 = (\frac{1}{20}, \ldots, \frac{1}{20})$ et évolue selon la récurrence donnée dans la proposition 6.

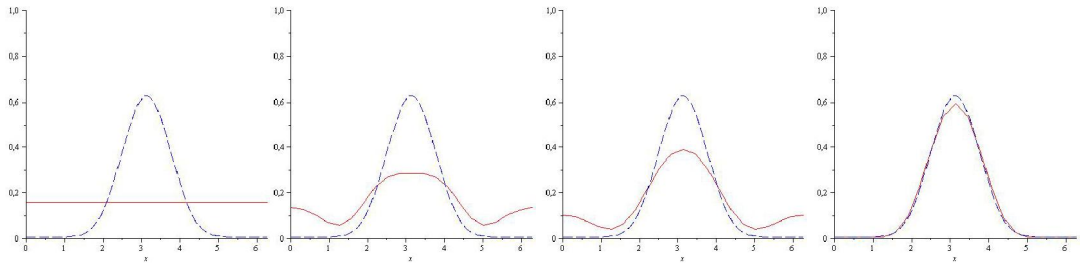


Figure 5.9 — Champ moyen aux instants $t = 0$, $t = 0,05$, $t = 0,5$ et $t = 2,5$

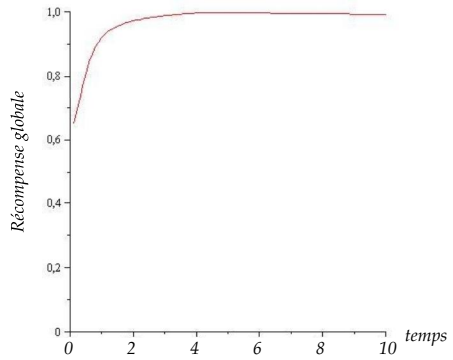


Figure 5.10 — Évolution de la récompense du champ moyen au cours du temps

5.5.1.3 Le système continu

Le système continu, en figure 5.11, évolue selon l'équation de Fokker-Planck (5.11) avec

$$c(x,u) = \frac{d}{dx} \ln\left(f(x)\right) \quad \text{et} \quad d(x,u) = 1$$

et est initialisé selon la distribution uniforme continue sur $\mathcal{C}_\delta$. Comme l'équation de Fokker-Planck n'admet pas de solution analytique sous forme fermée, nous présentons une solution numérique approchée obtenue à l'aide de la méthode de Crank-Nicholson [54].

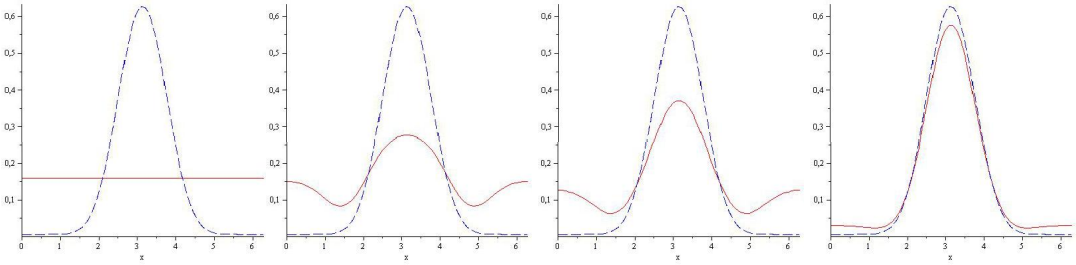


Figure 5.11 — Système continu aux instants $t = 0$, $t = 0,05$, $t = 0,5$ et $t = 2,5$

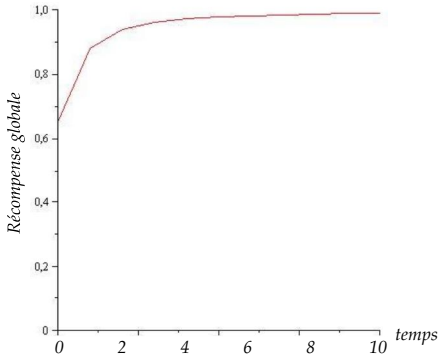


Figure 5.12 — Évolution de la récompense du champ moyen au cours du temps

5.5.1.4 Analyse des résultats

Les figures 5.7, 5.9 et 5.11 montrent que les trois modèles ont des dynamiques similaires. La densité du système stochastique, sujette à des fluctuations aléatoires est similaire à celle du champ moyen ainsi qu'à la dynamique du système continu. De plus, les fonctions de satisfaction des trois systèmes évoluent de manière similaire, la consistance des différentes approximations avec la fonction de récompense.

5.5.2 Validation des stratégies optimales

Afin illustrer l'optimalité des paramètres c et d trouvés dans la section 5.4, nous étudions seulement la dynamique du système continu. Les résultats présentés sont des approximations numériques, toujours obtenues grâce à la méthode de Crank-Nicholson [54].

Dans ces simulations, les agents et la ressource sont répartis sur le cercle $\mathcal{C}$. La ressource est répartie selon la densité

$$f(x) \propto \cos\left(\frac{x-\pi}{2}\right)^{10} + 0.01$$

et les agents sont initialement distribués selon la densité

$$u_0 \propto \cos\left(\frac{x-3\pi}{2}\right)^{10},$$

où les coefficients de proportionnalité sont choisis de sorte que $\int_{\mathcal{C}} f(x)dx = \int_{\mathcal{C}} u_0(x)dx = 1$. Les fonction f et g sont représentées sur les figures 5.13, 5.15 en pointillés verts et bleus respectivement. Notons que initialement, les agents sont situés de façon diamétralement opposée à la ressource.

Stratégie dépendant localement de la répartition ciblée La première stratégie, établie au paragraphe 5.4.1, consiste à paramétrer le déplacement de sorte que

$$c(x) = \frac{d}{dx}\ln\left(f(x)\right), \quad d(x) = 1$$

La dynamique du système et de la récompense globale est représentée sur les figures 5.13 et 5.14.

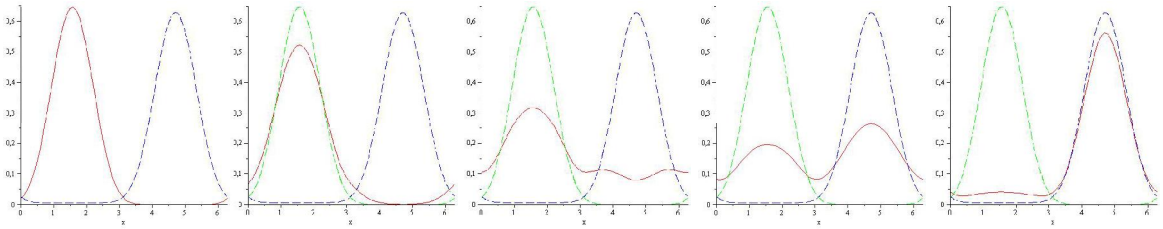


Figure 5.13 — État du système aux instants $t = 0$, $t = 0,1$, $t = 0,5$, $t = 1$ et $t = 3$

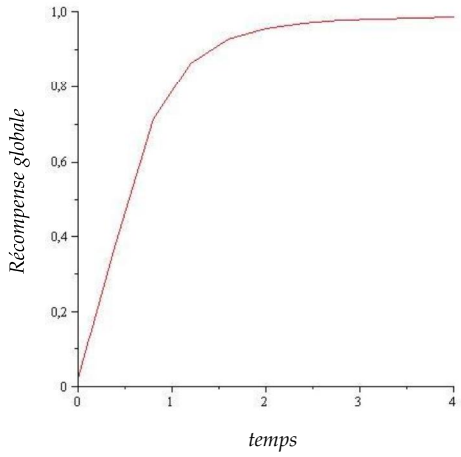


Figure 5.14 — Récompense globale du système au cours du temps

Stratégie dépendant de la satisfaction locale La seconde stratégie, établie au paragraphe 5.4.2, correspond au choix des paramètres

$$c(u,x) = \frac{\partial}{\partial x}\ln\left(u(x,t)e^{1-\frac{u(x,t)}{f(x)}}\right), \quad d(u,x) = 1.$$

La dynamique du système et de la récompense globale sont représentées sur les figure 5.15 et 5.16.

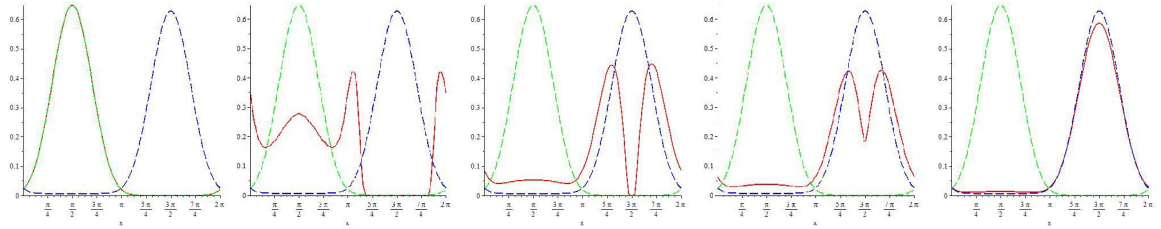


Figure 5.15 — État du système aux instants $t = 0$, $t = 0,02$, $t = 0,2$, $t = 0,3$ et $t = 1$

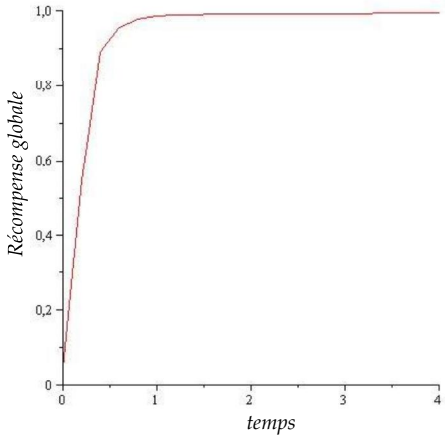


Figure 5.16 — Récompense globale du système au cours du temps

Analyse des résultats En ce qui concerne le mouvement collectif des agents, on observe le comportement souhaité pour les deux stratégies : la densité des agents tend vers la répartition f. On observe également que la fonction de satisfaction R croît en tendant vers sa valeur maximale 1.

De plus, au cours de ces simulations la seconde stratégie s'est avérée plus performante en termes de vitesse de convergence. Malheureusement, les bornes de convergence établies théoriquement ne permettent pas de démontrer ce fait. De plus, les simulations dans le cas bidimensionnel qui suivent (section 5.6) montrent que cette seconde stratégie est moins performante pour un domaine différent.

La figure 5.15 montre toutefois un phénomène intéressant, comparant l'allure de son évolution avec celle de la figure 5.13. La seconde stratégie favorise largement la convection, notamment en début de simulation, ce qui est illustré par la progression «en vague» des agents.

On observe également que sous cette stratégie, les agents ne rejoignent pas directement le centre de la distribution ciblée. Ce phénomène pourrait s'expliquer par le fait que les agents ont une satisfaction élevée aux alentours du centre de la distribution ciblée, et qu'ils ne sont pas encouragés à se déplacer.

5.6 Étude numérique dans le cas bidimensionnel

Avant de conclure ce chapitre, nous proposons une extension rapide du problème considéré au cas bidimensionnel. Dans cette extension, les N agents se déplacent sur le tore $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z} \times \mathbb{R}/2\pi\mathbb{Z}$, c'est-à-dire une espace bidimensionnel périodique en x et en y. Cet espace est divisé de manière régulière en une grille $n_p \times n_p$ positions (voir fig. 5.17), et noté

$$\mathbb{T}_\delta = \left\{ \left(\frac{2k\pi}{n_p}, \frac{2l\pi}{n_p} \right) \ / \ k, l \in \{0, \ldots, n_p - 1\} \right\} = \left\{ (k\delta, l\delta) \ / \ k, l \in \{0, \ldots, n_p - 1\} \right\}$$

en notant $\delta = \frac{2\pi}{n_p}$ le pas spatial.

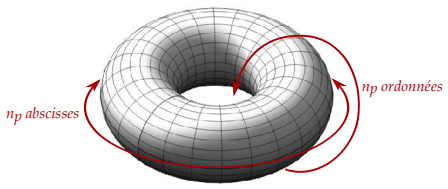


Figure 5.17 — Espace des positions des agents

La proportion d'agents au point (x, y) est notée $u^N_{\delta,\tau}(x, y, t)$, et le vecteur de *densité d'occupation* est donné par

$$u^N_{\delta,\tau}(t) = \left(u^N_{\delta,\tau}(x, y, t) \right)_{(x,y) \in \mathbb{T}_\delta}.$$

Durant une unité de temps τ, chaque agent a la possibilité de se déplacer de

$$(-\delta, 0), \ (\delta, 0), \ (0, -\delta), \ (0, \delta)$$

ou de rester sur place avec des probabilités qui dépendent de sa position courante et de la position des autres agents. Les probabilités des déplacements sont respectivement notées

$$p^N_{\delta,\tau}(x, y, u), \ q^N_{\delta,\tau}(x, y, u), \ r^N_{\delta,\tau}(x, y, u), \ s^N_{\delta,\tau}(x, y, u)$$

pour un agent situé au point (x, y) lorsque les agents sont distribués selon une densité u (voir figure 5.18).

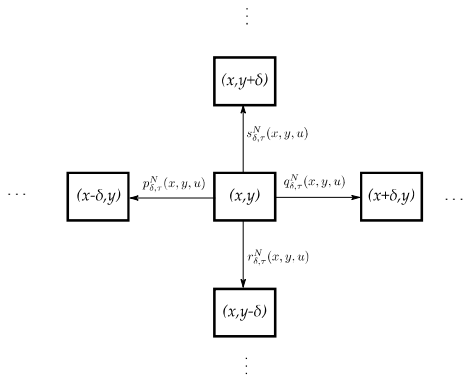


Figure 5.18 — Probabilités des déplacements des agents

5.6.1 Approximation continue

Avec un raisonnement similaire à celui de la section 5.2, nous montrons que ce modèle peut être approché par une équation de type Fokker-Planck. Nous présentons ici donc seulement les grandes étapes de la dérivation du modèle continu.

La limite champ moyen Dans une premier temps, nous supposons que :
- la répartition initiale des agents $u^N_{\delta,\tau}(0)$ converge presque sûrement vers une grandeur $u_{\delta,\tau}(0)$,
- les probabilités $p^N_{\delta,\tau}(x,y,u)$, $q^N_{\delta,\tau}(x,y,u)$, $r^N_{\delta,\tau}(x,y,u)$, $s^N_{\delta,\tau}(x,y,u)$ convergent uniformément en u vers des fonctions continues

$$u \mapsto p_{\delta,\tau}(x,y,u),\ u \mapsto q_{\delta,\tau}(x,y,u),\ u \mapsto r_{\delta,\tau}(x,y,u),\ u \mapsto s_{\delta,\tau}(x,y,u)$$

lorsque $N \to \infty$. En d'autres termes, les distributions initiales des agents, ainsi que leurs probabilités de déplacement convergent lorsque le nombre d'agents tend vers l'infini.

Alors on peut dire que le vecteur de densité d'occupation $u^N_{\delta,\tau}(t)$ converge presque sûrement vers $u_{\delta,\tau}(t) = (u_{\delta,\tau}(x,y,t))_{(x,y)\in\mathbb{T}_\delta}$ défini récursivement par :

$$\begin{aligned}
&u_{\delta,\tau}(x,y,t+\tau) = \\
&u_{\delta,\tau}(x,y,t)(1 - p_{\delta,\tau}(x,y,u_{\delta,\tau}(t)) - q_{\delta,\tau}(x,y,u_{\delta,\tau}(t)) - r_{\delta,\tau}(x,y,u_{\delta,\tau}(t)) - s_{\delta,\tau}(x,y,u_{\delta,\tau}(t))) \\
&\qquad + u_{\delta,\tau}(x-\delta,y,t)q_{\delta,\tau}(x-\delta,y,u_{\delta,\tau}(t)) \\
&\qquad + u_{\delta,\tau}(x+\delta,y,t)p_{\delta,\tau}(x+\delta,y,u_{\delta,\tau}(t)) \\
&\qquad + u_{\delta,\tau}(x,y-\delta,t)r_{\delta,\tau}(x,y-\delta,u_{\delta,\tau}(t)) \\
&\qquad\qquad + u_{\delta,\tau}(x,y+\delta,t)s_{\delta,\tau}(x,y+\delta,u_{\delta,\tau}(t)). \qquad (5.26)
\end{aligned}$$

Ce résultat est l'équivalent de la proposition 6 dans le cas bidimensionnel. La quantité $u_{\delta,\tau}(x,y,t)$ est le *champ moyen*, et correspond à la densité d'agents à la position (x,y) au temps t lorsque le nombre d'agents tend vers l'infini.

Tout comme dans le cas unidimensionnel (équation (5.2)) l'équation (5.26) peut être interprétée comme un bilan au point (x,y) au temps $t+\tau$. Si le vecteur des densités d'occupation vaut u à l'instant t, alors un agent situé en (x,y) à l'instant $t+\tau$ peut
- être resté sur place, avec probabilité

$$1 - p_{\delta,\tau}(x,y,u) - q_{\delta,\tau}(x,y,u) - r_{\delta,\tau}(x,y,u) - s_{\delta,\tau}(x,y,u)$$

- être arrivé par $(x+\delta,y)$ avec probabilité $p_{\delta,\tau}(x,y,u)$
- être arrivé par $(x-\delta,y)$ avec probabilité $q_{\delta,\tau}(x,y,u)$
- être arrivé par $(x,y+\delta)$ avec probabilité $r_{\delta,\tau}(x,y,u)$
- être arrivé par $(x,y-\delta)$ avec probabilité $s_{\delta,\tau}(x,y,u)$

La limite spatiotemporelle Comme dans la section 5.2.2, nous considérons une solution régulière u de (5.26), et nous effectuons des développements limités à l'ordre 1 en τ et à l'ordre 2 en δ. En faisant tendre les pas de temps δ et le pas d'espace τ vers 0 de sorte que $\frac{\delta^2}{\tau}$

reste borné, et en supposant que les limites suivantes existent, uniformément en u,

$$c_1(x,y,u) = \lim_{\delta,\tau} \left(\frac{q_{\delta,\tau}(x,y,u) - p_{\delta,\tau}(x,y,u)}{\tau} \delta \right),$$

$$d_1(x,y,u) = \lim_{\delta,\tau} \left(\frac{q_{\delta,\tau}(x,y,u) + r_{\delta,\tau}(x,y,u)}{2\tau} \delta^2 \right),$$

$$c_2(x,y,u) = \lim_{\delta,\tau} \left(\frac{s_{\delta,\tau}(x,y,u) - r_{\delta,\tau}(x,y,u)}{\tau} \delta \right)$$

et

$$d_2(x,y,u) = \lim_{\delta,\tau} \left(\frac{s_{\delta,\tau}(x,y,u) + r_{\delta,\tau}(x,y,u)}{2\tau} \delta^2 \right)$$

on aboutit à l'équation de Fokker-Planck bidimensionnelle

$$\begin{aligned} \frac{\partial u}{\partial t}(x,t) = &-\frac{\partial}{\partial x}\left(c_1(x,y,u)\, u(x,y,t)\right) + \frac{\partial^2}{\partial x^2}\left(d_1(x,y,u)\, u(x,y,t)\right) \\ &- \frac{\partial}{\partial y}\left(c_2(x,y,u)\, u(x,y,t)\right) + \frac{\partial^2}{\partial y^2}\left(d_2(x,y,u)\, u(x,y,t)\right). \end{aligned} \tag{5.27}$$

Il s'agit simplement de la superposition d'une équation de Fokker-Planck horizontale (en x) et une équation de Fokker-Planck verticale (en y).

Les *coefficients de convection* c_1 et c_2 correspondent respectivement aux tendances de déplacement selon les axes x et y, tandis que les *coefficients de diffusion* d_1 et d_2 quantifient la dispersion dans les directions des axes.

5.6.2 Étude numérique

Dans cette section, nous proposons quelques simulations numériques afin de confirmer les résultats obtenus dans le cas unidimensionnel. Plus précisément, nous prolongeons les stratégies optimales de la section 5.4 au cas bidimensionnel, et montrons que ces stratégies sont toujours optimales.

5.6.2.1 Paramètres des simulations

Dans les simulations qui suivent, la distribution initiale des agents u_0 est proportionnelle à

$$\left(\cos\left(\frac{x}{2} - \frac{\pi}{4}\right) \cdot \cos\left(\frac{y}{2} - \frac{\pi}{4}\right)\right)^{20}$$

et normalisée de sorte que $\int_{\mathbb{T}} u_0(x,y)dxdy = 1$. La distribution ciblée est donnée par la fonction f proportionnelle à

$$\left(\cos\left(\frac{x}{2} - \frac{5\pi}{4}\right) \cdot \cos\left(\frac{x}{2} - \frac{5\pi}{4}\right)\right)^{20} + 0,01$$

et normalisée de sorte que $\int_{\mathbb{T}} f(x,y)dxdy = 1$. Les fonctions f et g sont représentées sur la figure 5.19.

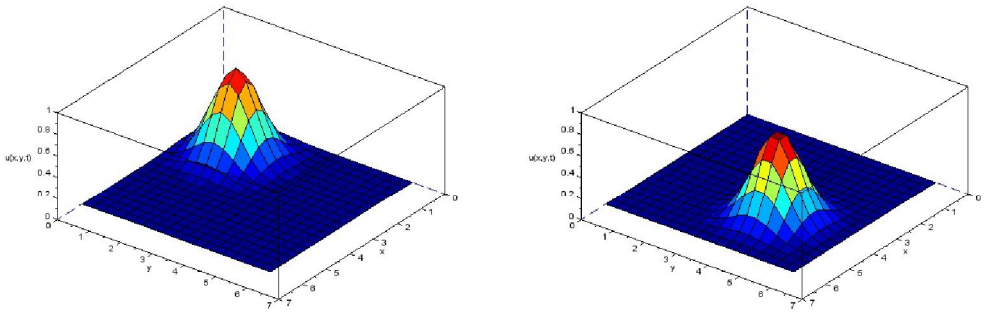


Figure 5.19 — Fonctions u_0 (Base, à gauche) et f (Ressource, à droite)

Toutes les simulations numériques sont effectuées à l'aide de la méthode de Crank-Nicholson [54].

5.6.2.2 Stratégie dépendant localement de la répartition ciblée

L'équivalent bidimensionnel de la stratégie établie à la section 5.4.1 est

$$c_1(x,y) = \frac{\partial}{\partial x} \ln f(x,y),\ d_1(x,y) = 1$$

et

$$c_2(x,y) = \frac{\partial}{\partial y} \ln f(x,y),\ d_2(x,y) = 1.$$

En d'autres termes, la diffusion est constante, égale à 1, tandis que la convection est choisie comme la dérivée logarithmique de la quantité de ressources, selon chacune des directions.

Les résultats des simulations sont représentés sur la figure 5.20.

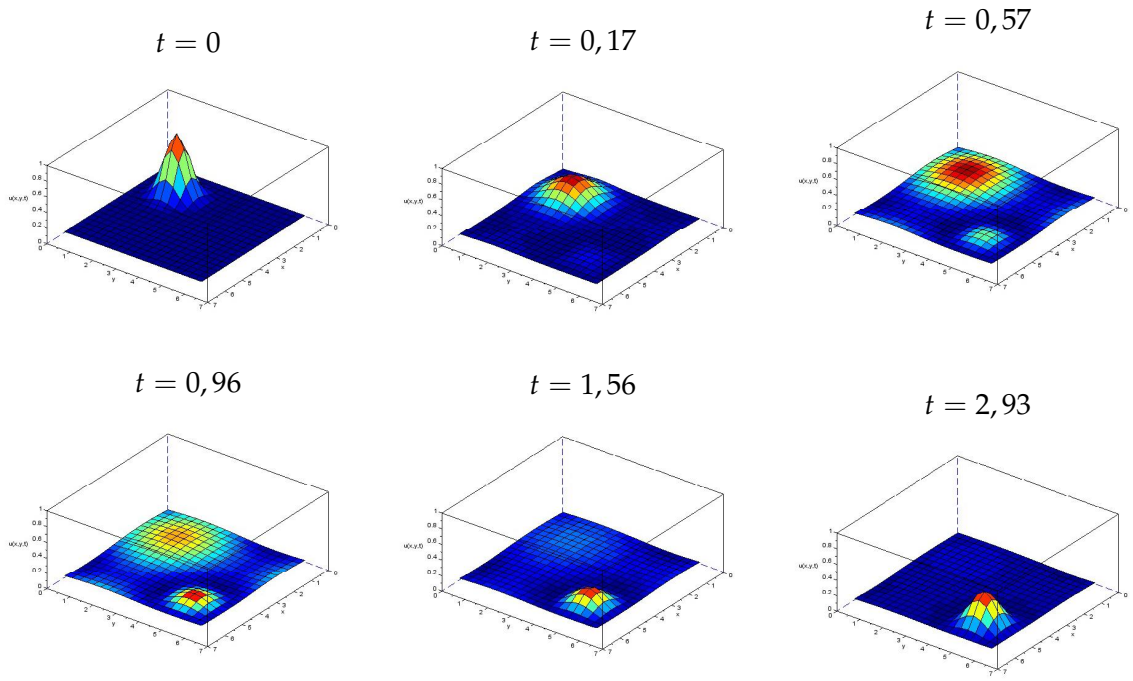


Figure 5.20 — Évolution du système

Ces simulations montrent que le choix des paramètres c_1, c_2, d_1 et d_2 de cette section donne lieu à une dynamique satisfaisante : la distribution des agents tends vers la distribution souhaitée f.

5.6.2.3 Stratégie dépendant de satisfaction locale

L'équivalent bidimensionnel de la stratégie autonome établie à la section 5.4.2 est

$$c_1(x,y,u) = \frac{\partial}{\partial x} \ln r(x,y,u),\ d_1(x,y,u) = 1$$

et

$$c_2(x,y,u) = \frac{\partial}{\partial y} \ln r(x,y,u),\ d_2(x,y,u) = 1.$$

où r désigne la fonction de récompense définie ici par

$$r(x,y,u) = u(x,y)e^{1-\frac{u(x,y)}{f(x,y)}}.$$

En d'autres termes, la diffusion est constante, égale à 1, tandis que la convection est choisie comme la dérivée logarithmique de la récompense locale, selon chacune des directions.

Les résultats des simulations sont représentés sur la figure 5.21.

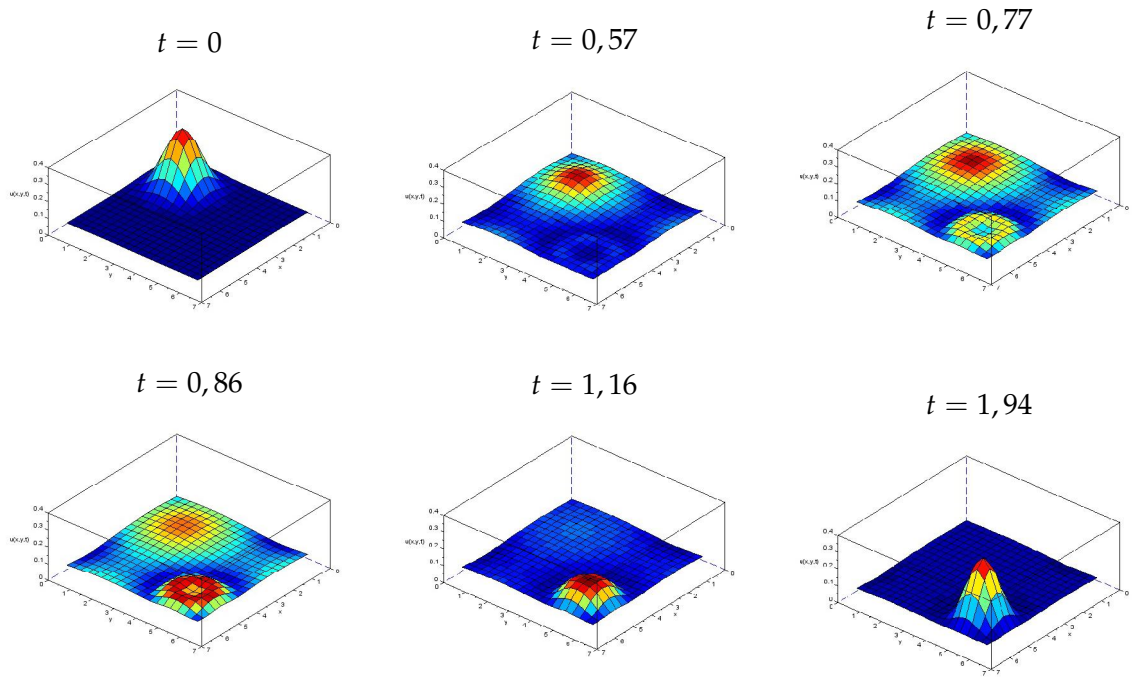


Figure 5.21 — Évolution du système

Ces simulations montrent que le choix des paramètres c_1, c_2, d_1, d_2 de cette section est satisfaisant : la distribution des agents tends vers la distribution souhaitée f.

De plus la figure 5.21 montre que, comme dans le cas unidimensionnel (section 5.5.1.4) sous cette stratégie, les agents montrent une certaine hésitation pour joindre le centre des ressources. Ce fait se manifeste par une concentration «en couronne» autour de ce point.

5.6.2.4 Analyse de la performance

Nous comparons les performance des stratégies des sections 5.6.2.2 et 5.6.2.3 en termes de récompense globale. Cette récompense est donnée par la fonction

$$R(t) = \int_{\mathbb{T}} r(x, y, u(x, y, t))\, dxdy$$

c'est-à-dire la somme des satisfactions locales. Pour simplifier les annotations et explications qui suivent, la stratégie de la section 5.6.2.2 sera appelée *stratégie centralisée* et la stratégie de la section 5.6.2.3 sera appelée *stratégie autonome*.

Au cours des simulations des sections 5.6.2.2 et 5.6.2.3, nous avons mesuré l'évolution de la récompense globale. Ces évolutions sont représentées sur la figure 5.22.

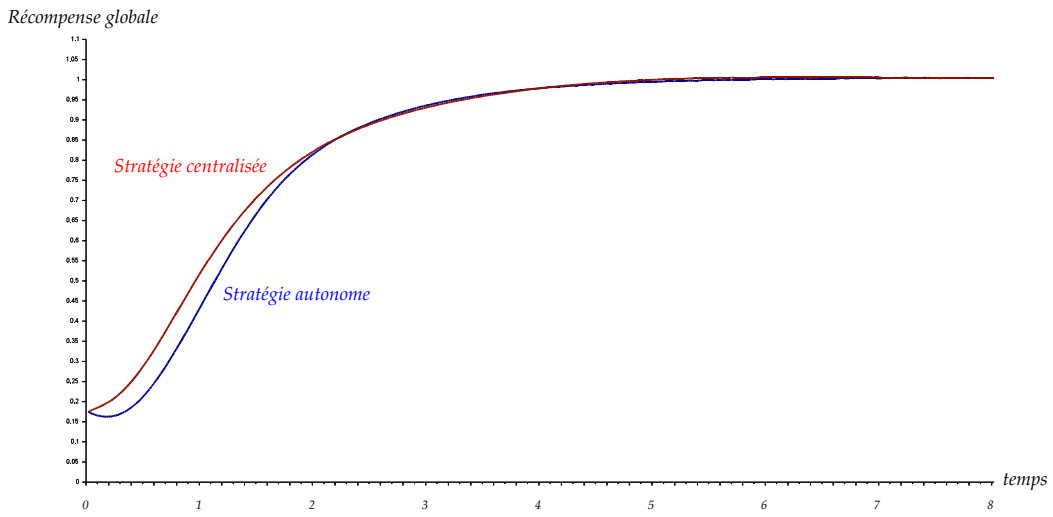


Figure 5.22 — Évolution de la récompense globale du système sous chacune des stratégies

La figure 5.22 confirme que les deux stratégies sont optimales à long terme. Pour chacune des stratégies, la récompense globale converge vers sa valeur maximale 1.

De plus, on note une légère différence vis-à-vis des résultats du cas unidimensionnel dans l'évolution de la récompense globale associée à la stratégie autonome. Contrairement au cas unidimensionnel, cette récompense n'est pas strictement croissante. Elle décroît dans un premier temps, puis croît en tendant vers 1. En raison de cette décroissance initiale, elle s'avère moins performante que la stratégie centralisée en début de simulation.

Il est difficile d'interpréter ce phénomène en termes de comportement d'agents. Une explication possible est que les agents mécontents se contentent de se disperser dans un premier temps, sans s'orienter précisément vers la distribution ciblée. De ce fait, leur performance peut être sous-optimale en début de simulation.

5.7 Conclusion

Dans cette partie nous avons étudié un processus d'advection-diffusion dérivé d'une marche aléatoire discrète. Dans un premier temps, nous faisons tendre le nombre d'individus vers l'infini et obtenons un champ moyen discret déterministe. Ensuite, les pas d'espace et de temps tendent également vers 0 et nous obtenons un modèle continu en espace et en temps. L'objectif collectif est représenté par une fonction de satisfaction locale. Cette fonction est compatible avec les limites effectuées, et a une expression analogue pour chacun des trois modèles. Des résultats théoriques et des simulations numériques montrent la consistance de l'approche. Les systèmes discrets, champ moyen et continu ont des dynamiques similaires et il en est de même pour leurs satisfactions globales.

De plus, nous avons déterminé deux stratégies qui aboutissent à une répartition optimale (au sens de la fonction de satisfaction) à long terme. L'une ne dépend que de la distribution ciblée, et exige un contrôleur global, et l'autre est autonome, et dépend de la satisfaction locale des individus. Conformément à la loi psychophysique de Weber-Fechner, les

individus perçoivent ces stimuli de manière logarithmique. Des résultats théoriques et numériques montrent que ces deux stratégies tendent vers des répartitions optimales à vitesse exponentielle.

Ces résultats permettent de modéliser un comportement multi-agent spatial par un processus d'advection-diffusion. L'approximation n'est exacte que sous des hypothèses théoriques ($N \to \infty$ et $\delta, \tau \to 0$) mais permet de réaliser une étude formelle précise. Dans l'exemple étudié nous avons pu l'exploiter pour

- déterminer la configuration à l'équilibre du système.
- démontrer la convergence exponentielle du système vers cette configuration.
- mesurer la performance du système en termes de satisfaction vis-à-vis d'un objectif fixé.

Ces trois faits constituent des réponses satisfaisantes vis-à-vis des objectifs que nous avons fixés dans cette thèse.

Le problème général de paramétrer les déplacements des agents de façon à atteindre une répartition optimale vis-à-vis d'une ressource immobile a pu être résolu grâce au modèle continu. Les méthodes utilisées s'inscrivent donc également dans une démarche prescriptive, où l'on souhaite programmer les agents afin de réaliser un objectif collectif.

Diverses variations de ce problème sont envisageables. Premièrement, le domaine considéré dans ce chapitre est périodique. Remplacer cette condition par une condition de type Neumann [10] n'exige que quelques légères modifications dans notre approche.

Une deuxième extension intéressante est de considérer une distribution cible f dynamique. Si la dynamique de la ressource est indépendante du système [11] notre approche permet de répondre en partie à la question suivante

- Peut-on paramétrer les déplacements des agents de sorte qu'ils se répartissent à chaque instant selon la distribution $f(t)$ qui dépend du temps ?

En effet, les stratégies déterminées aux sections 5.4 convergent à vitesse exponentielle vers la distribution ciblée. Elles sont donc susceptibles d'avoir un comportement satisfaisant si la ressource évolue suffisamment lentement [12].

Une troisième extension intéressante est de considérer des ressources dynamiques, où la dynamique dépend de celle du système. Dans ce cas, il faut joindre une équation d'évolution de la ressource à celle du système d'agents. Étudier une telle extension exige naturellement un travail mathématique plus important que celui proposé dans ce chapitre.

Pour finir sur les extensions possibles, il serait intéressant d'étudier un problème similaire avec deux populations d'agents (situés sur un même domaine) qu'on souhaite orienter vers deux distributions distinctes. Une approche similaire à celle utilisée dans ce chapitre mène à un système d'équations de la forme

$$\begin{cases} \frac{\partial u}{\partial t} = -\frac{\partial}{\partial x}\Big(c_1(x,u,v)\,u\Big) + \frac{\partial^2}{\partial x^2}\Big(d_1(x,u,v)\,u\Big) \\ \frac{\partial v}{\partial t} = -\frac{\partial}{\partial x}\Big(c_2(x,u,v)\,v\Big) + \frac{\partial^2}{\partial x^2}\Big(d_2(x,u,v)\,v\Big) \end{cases}.$$

10. I.e. le flux sortant est nul aux frontières du domaine.
11. C'est-à-dire si elle est représentée par une fonction de distribution $f(t)$ dépendant du temps.
12. I.e. à vitesse sous-exponentielle.

Ce système d'équation est muni d'une paire de fonctions de récompense locales (r_1, r_2), une pour chaque population, qui mesure la réussite de chaque population d'agents dans son activité. Ces fonctions locales dépendent de la position x, et de la densité de chaque population d'agents en cette position. Il est possible (et intéressant) d'introduire un effet d'aversion dans ces récompenses qui pénalise la présence des agents concurrents. Le problème est de trouver de trouver des coefficients c_1, c_2, d_1, d_2, qui permettent aux deux populations de se répartir selon les distributions qu'elles visent, en évitant au maximum les contacts avec le concurrent.

6 Utilisation d'équations à réaction - diffusion - advection pour l'étude d'une situation de collecte de ressources

6.1 Introduction et position du problème

Le problème étudié prolonge naturellement celui du chapitre 5 : un groupe d'individus se déplace en présence d'une certaine quantité de ressources répartie dans l'espace. Une zone de l'espace où ces agents se déplacent est désignée comme la base. L'objectif que nous fixons est que les individus se déplacent vers la ressource et la rapportent à la base.

Cette situation correspond typiquement à l'activité d'une colonie de fourmis fourrageuses. Le comportement collectif de telles fourmis observé dans la nature a fait l'objet de diverses études [149, 46, 33, 68, 75], car il présente des aspects intéressants parmi lesquels :

- **l'auto-organisation :** À l'issue d'une période transitoire, les fourmis établissent un itinéraire commun pour aller chercher la nourritre. Cet itinéraire est généralement proche du chemin le plus court.
- **l'auto-adaptation :** Le fait d'introduire un obstacle sur le chemin des fourmis va provisoirement arrêter l'acheminement de la nourriture, mais les fourmis finissent par établir collectivement un nouvel itinéraire pour rapporter la nourriture, en contournant l'obstacle.

L'observation de ces phénomènes a motivé la création de systèmes multi-agents capables de reproduire les mêmes comportements émergents [29, 24, 55, 142].

Nous avons utilisé les équations à réaction-diffusion-advection pour l'étude de ce problème, en décomposant le processus de récolte en trois phases :

- **la recherche de la ressource :** Cette phase d'exploration est représentée par un processus d'advection-diffusion.
- **la récupération de la ressource rencontrée :** Les agents qui rencontrent la ressource la ramassent, changent d'état [1] et se déplacent vers la base. Cette transition est représentée par une réaction.
- **le rapatriement de la ressource :** Ce déplacement vers la base est représenté par un

1. Ici, l'état désigne l'activité courante de l'agent.

processus d'advection-diffusion.

Initialement, les agents ignorent la position exacte de la ressource et explorent le milieu dans lequel ils se trouvent. Ce processus correspond à une *diffusion isotrope*, dans le sens où aucune direction n'est privilégiée. Selon les capacités perceptuelles et le degré d'interaction des agents, il est possible que cette diffusion devienne anisotrope, et qu'un champ de déplacements macroscopique (une advection) s'établisse.

Il en est de même pour le rapatriement de la ressource. L'advection permet de représenter la tendance de déplacement vers la base, tandis que la diffusion représente la dispersion des individus. Cette dispersion est liée au caractère aléatoire de leurs déplacements.

À la différence du chapitre 5, nous introduisons une variable de mémoire dans ce modèle. Les agents ont la possibilité de stocker des informations communes, et de communiquer, de manière indirecte, en déposant des marqueurs pour guider leurs congénères vers la ressource. Ce mode de communication à travers l'environnement est appelé *stigmergie* [69], et s'apparente fortement à l'utilisation des phéromones chez certaines espèces animales et végétales. Aussi, tout au long de ce chapitre, nous utilisons le terme *phéromones* pour parler de ces marqueurs.

En introduisant des effets d'évaporation et de diffusion, à l'image des phéromones rencontrées en écologie, deux phénomènes importants sont pris en considération :

- **la diffusion** permet de représenter la propagation de l'information.
  Si une zone est marquée par un agents, les zones proches présentent également un certain intérêt. La diffusion permet de disperser les phéromones localement afin de marquer les zones voisines.
- **l'évaporation** permet de représenter la perte d'informations.
  Si une zone n'est plus considérée intéressante par des agents, les phéromones qui s'y trouvent disparaissent. À l'opposé, les pistes de phéromones qui mènent vers des zones riches en ressources doivent continuellement être signalées par les agents pour persister.

Nous introduisons le cadre formel pour étudier cette situation dans la section 6.2. Ensuite, dans la section 6.3 nous dérivons une approximation continue de ce modèle qui fait partie des *équations à réaction-diffusion-advection*. Dans la section 6.4 qui suit, nous présentons brièvement la littérature existante sur ce type d'équations. Les outils théoriques disponibles sont relativement limités en raison de la difficulté de ces équations. Par conséquent, nous procédons à des simulations numériques dans la section 6.5. Dans la section 6.6, nous proposons, pour diverses raisons, d'étudier une extension au cas bidimensionnel. La section 6.7 termine ce chapitre avec une conclusion qui ouvre sur quelques perspectives envisageables.

6.2 Modèle formel

Le modèle formel utilisé pour représenter le système multi-agents est très similaire à celui du chapitre 5. Nous en présentons les constituants dans cette section.

États Les agents effectuent des marches aléatoires discrètes sur le cercle $\mathcal{C} = \mathbb{R}/2\pi\mathbb{Z}$, divisé de manière régulière en n_p positions (voir fig. 6.1). L'ensemble des positions obtenu

est noté

$$\mathcal{C}_\delta = \left\{ \frac{2k\pi}{n_p} \, , \, k \in \{0, \ldots, n_p - 1\} \right\} = \left\{ k\delta \, , \, k \in \{0, \ldots, n_p - 1\} \right\}$$

en notant $\delta = \frac{2\pi}{n_p}$ le pas spatial.

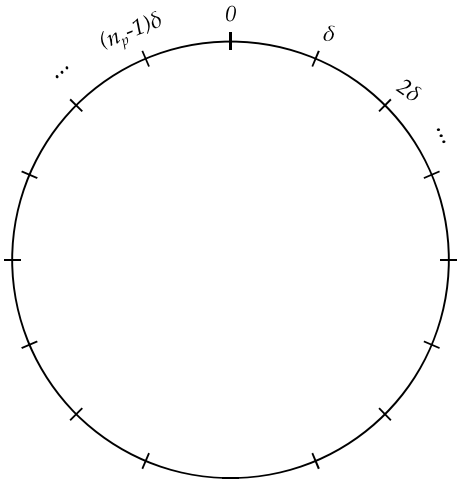


Figure 6.1 — Espace des positions des agents

La nature compacte et périodique de l'espace d'états offre quelques avantages théoriques. Notre approche peut cependant être étendue aux domaines non-périodiques et au cas bidimensionnel sans difficultés.

Les agents du système sont décrits par leur position et leur activité courante qui est l'exploration de la zone ou le rapatriement de la ressource.

- Le vecteur $(X_n(t))_{n=1}^N$ décrit les positions des N agents à l'instant t,
- Le vecteur $(s_n(t))_{n=1}^N$ décrit les états internes des agents, où

$$s_n(t) \in \{Exploration, Rapatriement\}.$$

- La quantité

$$u^N(x,t) = \sum_{n=1}^N \mathbb{1}_{X_n(t)=x, s_n(t)=Exploration}$$

  où

$$\mathbb{1}_{X_n(t)=x, s_n(t)=Exploration} = \begin{cases} 1 & \text{si } X_n(t) = x \text{ et } s_n(t) = Exploration \\ 0 & \text{sinon} \end{cases}$$

  désigne la proportion d'agents explorateurs [2] au point x à l'instant t,
- La quantité

$$v^N(x,t) = \sum_{n=1}^N \mathbb{1}_{X_n(t)=x, s_n(t)=Rapatriement}$$

  désigne la proportion d'agents en cours de rapatriement [2] au point x à l'instant t,
- Les vecteurs $u^N(t) = \left(u^N(x,t)\right)_{x \in \mathcal{C}_\delta}$ et $v^N(t) = \left(v^N(x,t)\right)_{x \in \mathcal{C}}$ sont les vecteurs de densité d'occupation des agents dans chaque état.

2. parmi la totalité des agents.

La ressource est représentée à l'aide d'une fonction régulière positive g, normalisée de sorte que

$$\int_{x\in\mathcal{C}} g(x)dx = 1.$$

En un point x, la valeur $g(x)$ représente la quantité de ressources. De manière analogue, **la base** est représentée à l'aide d'une fonction régulière positive f normalisée de sorte que

$$\int_{x\in\mathcal{C}} f(x)dx = 1.$$

La position de la base ne sera pas réduite à un seul point, mais plutôt à une zone dans laquelle la ressource peut être déposée avec des intensités variables. La fonction f permet de quantifier l'intensité avec laquelle les agents rapatrieurs peuvent déposer la ressource.

La base et la ressource sont représentées comme des fonctions de densité. Avec ce choix, la base et la ressource peuvent être interprétées comme des populations d'agents immobiles, de même effectif que les agents mobiles.

Le fait de traiter la base et les ressources comme des populations d'agents immobiles est intéressant si l'on souhaite étendre le modèle au cas où ces deux éléments sont dynamiques. En effet, pour réaliser une telle extension, il suffit d'adopter un raisonnement identique à celui qui est utilisé pour les agents mobiles.

D'autre part, cette représentation de la ressource et de la base permet de les définir comme distribution cible, selon l'activité de l'agent. De ce fait, nous pouvons réutiliser les résultats du chapitre 5 pour paramétrer les déplacements des agents.

Phéromones Une certaine quantité de phéromones est dispersée sur l'espace des positions des agents. La quantité de phéromones au point x à l'instant t est notée $w^N_{\delta,\tau}(x,t)$, et le vecteur

$$w^N_{\delta,\tau}(t) = \left(w^N_{\delta,\tau}(x,t)\right)_{x\in\mathcal{C}_\delta}$$

représente la répartition des phéromones sur l'espace des positions.

Transitions Durant chaque unité de temps τ, les agents peuvent effectuer les transitions suivantes :

- un déplacement de $-\delta$ ou $+\delta$. Les **agents explorateurs** effectuent ces déplacements avec des probabilités respectives

$$p^N_{\delta,\tau}(x,u,v,w) \quad \text{et} \quad q^N_{\delta,\tau}(x,u,v,w) \tag{6.1}$$

  qui dépendent de la position courante, des densités d'agents et des quantités de phéromones. Pour les **agents en cours de rapatriement**, ces probabilités sont notées

$$r^N_{\delta,\tau}(x,u,v,w) \quad \text{et} \quad s^N_{\delta,\tau}(x,u,v,w). \tag{6.2}$$

- un agent explorateur qui rencontre la ressource peut en ramasser une certaine quantité et changer d'état pour la rapatrier. Cette transition s'effectue avec une probabilité proportionnelle à la densité de ressources

$$b_\tau g(x).$$

Il est supposé que le coefficient de proportionnalité b_τ ne dépend que du pas de temps τ.

– un agent en cours de rapatriement arrive à la base, dépose la ressource et se remet à explorer la zone avec une probabilité

$$a_\tau f(x)$$

proportionnelle à la fonction f qui représente la base. Il est supposé que le coefficient de proportionnalité a_τ ne dépend que du pas de temps τ.

Ces transitions sont résumées par le diagramme suivant :

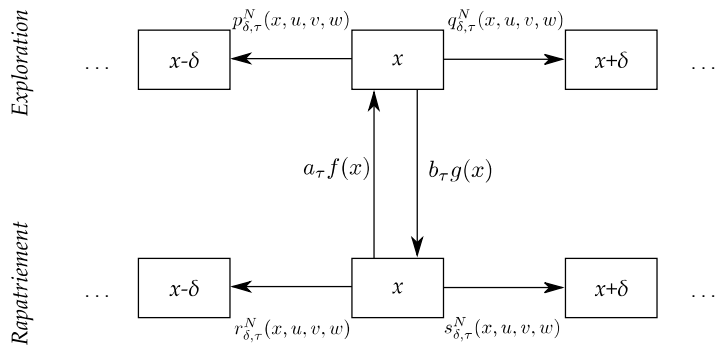


Figure 6.2 — Diagramme des transitions des agents

Évolution des phéromones Durant chaque intervalle de temps τ, les phéromones sont sujettes à trois effets :

– **le dépôt :** la quantité de phéromones au point x à l'instant t est augmentée de

$$d_\tau(x,u,v,w),$$

c'est-à-dire une certaine quantité dépendant des différentes densités d'agents et de la quantité de phéromones courante.

– **la dispersion :** une proportion $\frac{\psi_{\delta,\tau}}{2}$ des phéromones situées en x se déplace vers $-\delta$, et une proportion identique se déplace vers $+\delta$. Le coefficient $\psi_{\delta,\tau}$ est le *coefficient de diffusion des phéromones* et ne dépend, par hypothèse, pas du nombre d'individus.

– **l'évaporation :** une proportion γ_τ des phéromones situées en x disparaît. Le coefficient γ_τ est le *coefficient d'évaporation des phéromones*, et ne dépend par hypothèse que du pas de temps τ.

Dans la partie suivante, nous dérivons une loi d'évolution macroscopique déterministe et continue de ce modèle discret et aléatoire.

6.3 Dérivation de l'équation à réaction-diffusion-advection

Le modèle discret aléatoire que nous avons décrit tend, sous certaines hypothèses, vers un modèle continu et déterministe. La démarche adoptée est très similaire à celle du chapitre 5 : dans un premier temps, nous effectuons la limite statistique (champ moyen) $N \to \infty$ et

obtenons une approximation discrète déterministe. Ensuite, nous effectuons la limite spatiotemporelle $\delta, \tau \to 0$ d'une manière bien choisie, et obtenons une équation aux dérivées partielles de type *réaction-diffusion-advection*.

6.3.1 Limite champ moyen

Le modèle introduit à la section 6.2 correspond exactement au formalisme du chapitre 4, section 4.2.1. En effet, nous avons affaire à un système d'agents en interactions à travers les densités, ainsi qu'à une mémoire commune, sous formes de phéromones, qui évolue en interaction avec la densité.

Notons que, la mémoire commune (qui représente les phéromones) est appelée *variable ressource* dans le chapitre 4, et ne doit pas être confondue avec la ressource que les agents essayent de récolter.

En appliquant directement le théorème 5, on obtient le résultat suivant :

Proposition 16. *Pour chaque position x, on suppose que*
- *La fonction de dépôt $(u, v, w) \mapsto d_\tau(x, u, v, w)$ est continue.*
- *Les fonctions de probabilités définies en* (6.1) *et* (6.2) *convergent uniformément en (u, v, w) vers des fonctions*
$$p_{\delta,\tau},\ q_{\delta,\tau},\ r_{\delta,\tau},\ s_{\delta,\tau}$$
continues en (u, v, w), lorsque N tend vers l'infini.
- *Les conditions initiales $u^N_{\delta,\tau}(x, 0), v^N_{\delta,\tau}(x, 0)$ et $w^N_{\delta,\tau}(x, 0)$ convergent presque sûrement vers des quantités respectives $u^0_{\delta,\tau}(x)$, $v^0_{\delta,\tau}(x)$ et $w^0_{\delta,\tau}(x)$ lorsque N tend vers l'infini.*

Alors les quantités $u^N_{\delta,\tau}$, $v^N_{\delta,\tau}$ et $w^N_{\delta,\tau}$ convergent presque sûrement vers $(u_{\delta,\tau}, v_{\delta,\tau}, w_{\delta,\tau})$ définies par $u_{\delta,\tau}(x, 0) = u^0_{\delta,\tau}(x)$, $v(x, 0) = v^0_{\delta,\tau}(x)$ et $w(x, 0) = w^0_{\delta,\tau}(x)$, et par les relations de récurrence

$$\begin{aligned} u_{\delta,\tau}(x, t+\tau) &= u_{\delta,\tau}(x, t)(1 - p_{\delta,\tau}(x, u_{\delta,\tau}, v_{\delta,\tau}, w_{\delta,\tau}) - q_{\delta,\tau}(x, u_{\delta,\tau}, v_{\delta,\tau}, w_{\delta,\tau})) \\ &\quad + u_{\delta,\tau}(x-\delta, t) q_{\delta,\tau}(x-\delta, u_{\delta,\tau}, v_{\delta,\tau}, w_{\delta,\tau}) + u_{\delta,\tau}(x+\delta, t) p_{\delta,\tau}(x+\delta, u_{\delta,\tau}, v_{\delta,\tau}, w_{\delta,\tau}) \\ &\quad - u_{\delta,\tau}(x, t) b_\tau g(x) + v_{\delta,\tau}(x, t) a_\tau f(x), \end{aligned} \tag{6.3}$$

$$\begin{aligned} v_{\delta,\tau}(x, t+\tau) &= v_{\delta,\tau}(x, t)(1 - r_{\delta,\tau}(x, u_{\delta,\tau}, v_{\delta,\tau}, w_{\delta,\tau}) - s_{\delta,\tau}(x, u_{\delta,\tau}, v_{\delta,\tau}, w_{\delta,\tau})) \\ &\quad + v_{\delta,\tau}(x-\delta, t) s_{\delta,\tau}(x-\delta, u_{\delta,\tau}, v_{\delta,\tau}, w_{\delta,\tau}) + v_{\delta,\tau}(x+\delta, t) r_{\delta,\tau}(x+\delta, u_{\delta,\tau}, v_{\delta,\tau}, w_{\delta,\tau}) \\ &\quad + v_{\delta,\tau}(x, t) b_\tau g(x) - v_{\delta,\tau}(x, t) a_\tau f(x) \end{aligned} \tag{6.4}$$

et

$$\begin{aligned} w_{\delta,\tau}(x, t+\tau) &= w_{\delta,\tau}(x, t)(1 - \gamma_\tau - \psi_{\delta,\tau}) \\ &\quad + \frac{\psi_{\delta,\tau}}{2}(w_{\delta,\tau}(x-\delta, t) + w_{\delta,\tau}(x+\delta, t)) \\ &\quad + d_\tau(x, u_{\delta,\tau}, v_{\delta,\tau}, w_{\delta,\tau}) \end{aligned} \tag{6.5}$$

Les relations (6.3), (6.4) et (6.5) s'interprètent simplement comme des bilans. Commentons la première :

- La première ligne de (6.3) représente la diminution des agents explorateurs au point x due aux déplacements.
- La deuxième ligne de (6.3) représente l'augmentation des agents explorateurs au point x due aux déplacements des positions voisines.
- La troisième ligne de (6.3) représente deux effets :
  - la disparition des agents explorateurs qui rencontrent la ressource et qui changent d'état,
  - l'apparition de nouveaux agents explorateurs, lorsque des agents en cours de rapatriement arrivent à la base.

Les suites déterministes $u_{\delta,\tau}(t) = (u_{\delta,\tau}(x,t))_{x\in\mathcal{C}_\delta}$, $v_{\delta,\tau}(t) = (v_{\delta,\tau}(x,t))_{x\in\mathcal{C}_\delta}$ représentent les densités d'agents dans le contexte théorique où ils sont en nombre infini, tandis que $w_{\delta,\tau}(t) = (w_{\delta,\tau}(x,t))_{x\in\mathcal{C}_\delta}$ représente la distribution de phéromones dans ce même contexte.

L'étude numérique de ces suites est sans difficulté, mais l'étude formelle peut être délicate si le nombre de positions n_p est grand. Pour cette raison, nous effectuons une seconde approximation dans la section 6.3.2.

6.3.2 Loi d'évolution macroscopique continue

Un modèle continu est obtenu en faisant tendre les intervalle de temps τ et espace δ vers 0. Considérons des solutions u, v et w suffisamment régulières des (6.3), (6.4) et (6.5), au moins deux fois dérivables en x et une fois en t. Un développement limité à l'ordre 1 en τ et l'ordre 2 en δ donne, après simplifications

$$\begin{aligned}\tau\frac{\partial}{\partial t}u(x,t)+o(\tau) = \delta\frac{\partial}{\partial x}&\left((p_{\delta,\tau}(x,u,v,x)-q_{\delta,\tau}(x,u,v,x))u(x,t)\right)\\ &+\frac{\delta^2}{2}\frac{\partial^2}{\partial x^2}\left((p_{\delta,\tau}(x,u,v,x)+q_{\delta,\tau}(x,u,v,x))u(x,t)\right)\\ &\qquad -u(x,t)b_\tau g(x)+v(x,t)a_\tau f(x)+o(\delta^2)\end{aligned}$$

$$\begin{aligned}\tau\frac{\partial}{\partial t}v(x,t)+o(\tau) = \delta\frac{\partial}{\partial x}&\left((r_{\delta,\tau}(x,u,v,x)-s_{\delta,\tau}(x,u,v,x))v(x,t)\right)\\ &+\frac{\delta^2}{2}\frac{\partial^2}{\partial x^2}\left((r_{\delta,\tau}(x,u,v,x)+s_{\delta,\tau}(x,u,v,x))v(x,t)\right)\\ &\qquad +u(x,t)b_\tau g(x)-v(x,t)a_\tau f(x)+o(\delta^2)\end{aligned}$$

$$\tau\frac{\partial}{\partial t}w(x,t)+o(\tau) = \psi_{\delta,\tau}\frac{\delta^2}{2}\frac{\partial^2 w}{\partial x^2}(x,t)-\gamma_\tau w(x,t)+d_\tau(x,u,v,w)+o(\delta^2)$$

A présent, nous faisons tendre $\delta,\tau\to 0$ de sorte que $\frac{\delta^2}{\tau}$ reste borné, et on suppose que les limites suivantes existent, uniformément en (u,v,w),

$$\begin{cases} c_1(x,u,v,w) = \lim_{\delta,\tau}\left(\frac{q_{\delta,\tau}(x,u,v,w)-p_{\delta,\tau}(x,u,v,w)}{\tau}\,\delta\right)\\ d_1(x,u,v,w) = \lim_{\delta,\tau}\left(\frac{q_{\delta,\tau}(x,u,v,w)+p_{\delta,\tau}(x,u,v,w)}{2\tau}\,\delta^2\right),\end{cases} \tag{6.6}$$

$$\begin{cases} c_2(x,u,v,w) = \lim_{\delta,\tau}\left(\frac{s_{\delta,\tau}(x,u,v,w)-r_{\delta,\tau}(x,u,v,w)}{\tau}\,\delta\right)\\ d_2(x,u,v,w) = \lim_{\delta,\tau}\left(\frac{s_{\delta,\tau}(x,u,v,w)+r_{\delta,\tau}(x,u,v,w)}{2\tau}\,\delta^2\right)\end{cases}, \tag{6.7}$$

puis

$$\begin{cases} \alpha = \lim_\tau \frac{a_\tau}{\tau} \\ \beta = \lim_\tau \frac{b_\tau}{\tau} \end{cases}, \tag{6.8}$$

et

$$\begin{cases} \psi = \lim_{\delta,\tau} \left(\frac{\delta^2}{2\tau} \psi_{\delta,\tau} \right) \\ \gamma = \lim_\tau \frac{\gamma_\tau}{\tau} \\ d(x,u,v,x) = \lim_\tau \frac{d_\tau(x,u,v,x)}{\tau} \end{cases}. \tag{6.9}$$

Les relations (6.6), (6.7), (6.8), (6.9) constituent les *conditions de scaling* dans ce contexte. Dans le cas où elles sont vérifiées, nous obtenons les équations limites (sous forme abrégée)

$$\begin{cases} \frac{\partial u}{\partial t} = -\frac{\partial}{\partial x}(c_1 u) + \frac{\partial^2}{\partial x^2}(d_1 u) - \beta g u + \alpha f v \\ \frac{\partial v}{\partial t} = -\frac{\partial}{\partial x}(c_2 v) + \frac{\partial^2}{\partial x^2}(d_2 v) + \beta g u - \alpha f v \\ \frac{\partial w}{\partial t} = \psi \frac{\partial^2 w}{\partial x^2} - \gamma w + d \end{cases}. \tag{6.10}$$

Les coefficients $c_1, c_2, d_1, d_2, \alpha, \beta, \psi, \gamma, d$ ont des interprétations physiques simples :

- c_1 est le *coefficient d'advection* (ou *coefficients de convection*) des agents explorateurs, et représente la vitesse moyenne de ces agents au point x (idem pour c_2, et les agents rapatrieurs).
- d_1 est *coefficients de diffusion* des agents explorateurs, et indique la tendance des agents au point x à se disperser (idem pour d_2, et les agents rapatrieurs).
  Par construction ces coefficients sont positifs. De même le *taux de diffusion des phéromones* ψ est supposé positif.
- α et β désignent les *taux de collecte et de dépôt de phéromones*.
- γ est le *taux d'évaporation des phéromones* par unité de temps.
- d désigne l'*intensité de dépôt de phéromones* par unité de temps.

Le système d'équations (6.10) peut s'écrire sous la forme

$$\begin{aligned} \frac{\partial}{\partial t}\begin{pmatrix} u \\ v \\ w \end{pmatrix} = & -\frac{\partial}{\partial x}\left(\begin{pmatrix} c_1 & 0 & 0 \\ 0 & c_2 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} u \\ v \\ w \end{pmatrix}\right) \\ & + \frac{\partial^2}{\partial x^2}\left(\begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & \psi \end{pmatrix}\begin{pmatrix} u \\ v \\ w \end{pmatrix}\right) \\ & + \begin{pmatrix} -\beta g & \alpha f & 0 \\ \beta g & -\alpha f & 0 \\ 0 & 0 & -\gamma \end{pmatrix}\begin{pmatrix} u \\ v \\ w \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ d(x,u,v,w) \end{pmatrix} \end{aligned}$$

Le premier terme correspond à un effet d'*advection*, c'est-à-dire une tendance ou un déplacement macroscopique. Le deuxième est un terme de *diffusion*, et représente la dispersion des agents. La troisième ligne (somme du troisième terme et du quatrième terme) est le terme de *réaction*, et modélise les changements d'état des agents et le dépôt de phéromones.

6.3.3 Conditions au bord et condition initiale

Le système d'équations (6.10) fournit seulement une loi d'évolution approchée du système d'agents à l'intérieur du domaine considéré. Pour assurer une définition correcte de

problème, il faut y joindre des conditions initiales

$$\begin{cases} u(x,0) = u_0(x) \\ v(x,0) = v_0(x) \\ w(x,0) = w_0(x) \end{cases}$$

et préciser les conditions au bord du domaine. Comme les agents agents se déplacent sur le cercle, nous avons des conditions au bord périodiques : les agents se déplacent sur le cercle $\mathcal{C} = \mathbb{R}/2\pi\mathbb{Z}$. Les coefficients c_1, c_2, d_1, d_2, d sont tous supposés 2π-périodiques en x, et nous cherchons des solutions u, v, w périodiques en x.

6.4 À propos des équations à réaction-diffusion-advection

En partant d'un modèle discret, aléatoire, nous avons obtenu le système d'équations à réaction-diffusion-advection suivant

$$\begin{cases} \frac{\partial u}{\partial t} = -\frac{\partial}{\partial x}(c_1 u) + \frac{\partial^2}{\partial x^2}(d_1 u) - \beta g u + \alpha f v \\ \frac{\partial v}{\partial t} = -\frac{\partial}{\partial x}(c_2 u) + \frac{\partial^2}{\partial x^2}(d_2 v) + \beta g u - \alpha f v \\ \frac{\partial w}{\partial t} = \psi \frac{\partial^2 w}{\partial x^2} - \gamma w + d \end{cases} .$$

La forme générale de ce système, sous forme vectorielle, est

$$\frac{\partial U}{\partial t} = -\frac{\partial}{\partial x} \cdot (C(x,U) \cdot U) + \frac{\partial^2}{\partial x^2}(D(x,U) \cdot U) + R(x,U) \tag{6.11}$$

où C et D sont des matrices dépendant de la position courante, et des différentes coordonnées de U, et où R est une fonction à valeurs vectorielles. Notre objectif est d'étudier ces équations à l'aide d'outils d'analyse mathématique.

Bien que la littérature sur les équations à réaction-diffusion-advection soit importante, les résultats théoriques rencontrés sont généralement limités à des cadres spécifiques.

Par exemple, une étude quantitative du comportement à long terme est proposé dans [38] dans le cas particulier où les termes de réaction et diffusion sont linéaires et identiques, et où le terme de réaction est de type Lotka-Volterra. Dans ce contexte, les déplacements de tous les agents sont supposés identiques.

Dans les travaux [128, 21, 97] des résultats sur l'existence de solutions intégrables et positives sont exposés, dans le cas particulier où le terme d'advection est nul, le terme de diffusion est non linéaire, et où le terme de réaction est quadratique.

De façon générale, les solutions de l'équation (6.11) peuvent avoir des comportements irréguliers. Par exemple, la simple existence de solutions régulières (ou même bornées) de cette équation n'est généralement pas garantie.

L'équation générale (6.11) possède par ailleurs une grande richesse. Elle possède trois termes distincts (réaction-diffusion-advection) dont chaque paire possède déjà de l'intérêt. Le cas $R = 0$ correspond à *l'équation de Fokker-Planck* que nous avons déjà rencontrée.

Le cas $D = 0$ correspond aux *équations d'advection-réaction*. Ces équations peuvent être utilisées pour étudier la propagation d'une rumeur [35], l'écoulement d'un liquide dans un réservoir [1] ou encore les diverses situations rencontrées en biochimie citées dans [147].

Dans le cas particulier où $C = 0$ on obtient les *équations à réaction-diffusion*, qui sont sans doute le cas particulier le plus étudié dans la littérature. Cet intérêt peut être expliqué par l'article historique [144] dans lequel Turing utilise des équations à réaction-diffusion pour décrire le processus de développement chez l'embryon (morphogénèse). Depuis ces travaux, il y a eu de nombreuses études sur les équations à réaction-diffusion. Un panorama très large d'applications et de méthodes théoriques dans le cadre de la dynamique des populations se trouve dans [83]. Pour d'autres travaux récents on peut consulter [27, 13].

Si l'on souhaite utiliser les arguments du chapitre 5, et démontrer la convergence des solutions de l'équation (6.11) vers une éventuelle solution stationnaire avec des méthodes d'entropie, le travail le plus approfondi à ce jour est [48]. Les résultats énoncés dans ces travaux sont cependant limités aux équations linéaires, et exigent des conditions de symétrie entre la matrice de diffusion et la matrice de réaction. L'impossibilité de définir une entropie dissipative peut s'expliquer par la potentielle réversibilité des équations.

Cependant, un résultat qui peut être obtenu sans difficulté est le suivant :

Proposition 17. *Toute solution* (u, v, w) *dérivable du système* (6.10) *vérifie la condition de conservation d'effectif*

$$\int_{\mathcal{C}} (u(x,t) + v(x,t))dx = \int_{\mathcal{C}} (u_0(x) + v_0(x))dx, \ \forall t \geq 0$$

Démonstration

Le théorème de dérivation des intégrales à paramètres donne

$$\frac{d}{dt}\int_{\mathcal{C}} u + v = \int_{\mathcal{C}} \frac{\partial}{\partial t}(u+v) = \int_{\mathcal{C}} \frac{\partial}{\partial x}\left(c_1 u + c_2 v + \frac{\partial}{\partial x}(d_1 u + d_2 v)\right)$$

or cette dernière intégrale est nulle grâce à la périodicité des fonctions en jeu. □

Pour conclure, la littérature montre que l'étude formelle des équations à réaction-diffusion-advection est délicate, et que les résultats existants sont limités à des cadres spécifiques. Pour cette raison, nous bornons notre approche à une étude numérique (approchée) et qualitative du système (6.10). Cette étude est effectuée dans la section suivante.

6.5 Étude numérique

Le système multi-agent étudié est représenté à l'aide du système d'équations à réaction-diffusion-advection (6.10). Dans cette section, nous fixons les paramètres de cette équation et réalisons une étude numérique ensuite.

6.5.1 Paramètres des simulations

Tout d'abord, nous fixons le cadre des simulations en affectant des valeurs aux différents paramètres du système d'équations (6.10).

Les conditions initiales La base est représentée par la fonction f proportionnelle à

$$\cos\left(\frac{x}{2} - \frac{\pi}{4}\right)^{10}$$

où le coefficient de proportionnalité est choisi de sorte que $\int_{\mathcal{C}} f(x)dx = 1$, et la ressource est représentée par la fonction g proportionnelle à

$$\cos\left(\frac{x}{2} - \frac{5\pi}{8}\right)^{10}$$

avec un coefficient de proportionnalité choisi de sorte que $\int_{\mathcal{C}} g(x)dx = 1$. Les fonctions f et g sont représentées sur la figure 6.3 et figurent sur toutes les simulations en pointillés verts et rouges respectivement.

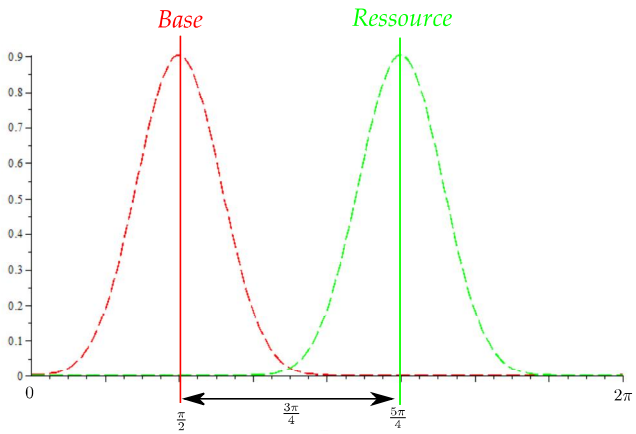


Figure 6.3 — Fonctions f (base, en rouge) et g (ressource, en vert)

La base et la ressource sont centrées en $\frac{\pi}{2}$ et $\frac{5\pi}{4}$ respectivement. Leurs centres sont éloignés de $\frac{3\pi}{4}$ et le chemin le plus court reliant ces centres est le segment $[\frac{\pi}{2}, \frac{5\pi}{4}]$.

Les agents sont initialement distribués selon la fonction f qui définit la base, et tous sont dans l'état $Exploration$, c'est-à-dire

$$u_0(f) = f(x) \quad \text{et} \quad v_0(x) = 0,$$

tandis que la quantité initiale de phéromones est nulle : $v_0(x) = 0$.

Paramètres du système Les déplacements des **agents explorateurs** sont décrits par un terme de convection égal à la dérivée logarithmique de la quantité de phéromones, et un terme de diffusion constant égal à 1.

Ainsi, les agents explorateurs sont attirés par les fortes concentrations de phéromones. Conformément à la loi de Weber-Fechner, l'intensité de cette attraction varie comme le logarithme du stimulus, c'est-à-dire la quantité de phéromones.

Les coefficients d'intensité de collecte et d'intensité de dépôt de ressource α et β sont fixés à 2.

L'équation continue décrivant l'évolution des agents explorateurs est :

$$\frac{\partial u}{\partial t} = -\frac{\partial}{\partial x}\left(u\frac{\partial}{\partial x}\ln(w + 0,01)\right) + \frac{\partial^2 u}{\partial x^2} - 2ug + 2vf.$$

Nous avons légèrement augmenté la quantité de phéromones (telle qu'elle est perçue par les agents) dans le terme de convection des agents explorateurs, afin de garantir la bonne définition du logarithme.

En ce qui concerne les déplacements **des agents en cours de rapatriement**, nous supposons que ces agents ont connaissance de la position de la base, et s'y dirigent par un déplacement qui est proche du chemin le plus court. Le coefficient de convection de ce déplacement est choisi comme $c_2(x) = -\sin(x - \frac{\pi}{2})$ et le coefficient de diffusion est $d_2(x) = 0,1$. Concrètement, les agents se déplacent vers la base avec une convection qui varie en fonction de leur distance à la base, et avec de faibles fluctuations aléatoires.

La convection est représentée sur la figure 6.4.

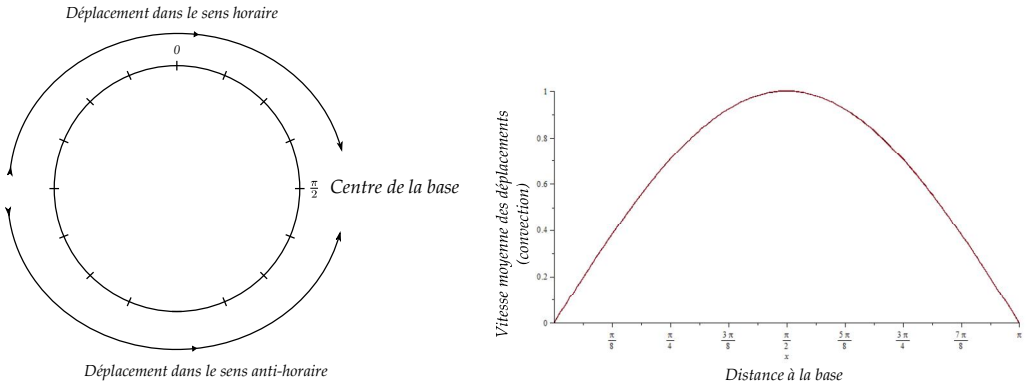


Figure 6.4 — Convection des agents rapatrieurs

L'allure de cette courbe montre que la vitesse de déplacement des agents croît à mesure qu'on s'éloigne de la base, puis décroît pour être nulle lorsque les agents sont à une distance maximale (π). Cette allure est satisfaisante : lorsque les agents sont proches de la base ils doivent diminuer leur vitesse afin d'éviter de se retrouver tous concentrés au centre de la base. Si une telle concentration devait se produire, la densité correspondante ne serait clairement pas régulière. Or, un minimum de régularité est nécessaire pour réaliser un étude numérique.

Par ailleurs, la faible vitesse de déplacement des agents à distance maximale de la base correspond au fait qu'en ces positions, les deux sens de circulation donnent des trajets aussi courts vers la base. Il est donc légitime que les agents soient «hésitants» et se déplacent lentement en ces positions.

Le faible terme de diffusion (avec un coefficient de $0,1$) dans ce déplacement a plusieurs motivations. D'une part, nous savons que la diffusion a un effet «régularisant» sur les solutions, ce qui favorise l'étude numérique que nous souhaitons effectuer. D'autre part, ce terme de diffusion permet d'introduire des petites fluctuations aléatoires dans les déplacements des agents. Ces fluctuations permettent d'éviter que les agents situés aux point x où la convection s'annule restent immobiles, et d'éviter des phénomènes de concentration.

L'équation continue décrivant l'évolution des agents rapatrieurs est donc :

$$\frac{\partial v}{\partial t} = \frac{\partial}{\partial x}\left(v \sin(x - \frac{\pi}{2})\right) + 0,1\frac{\partial^2 v}{\partial x^2} + 2ug - 2vf.$$

Paramètres des phéromones La diffusion des phéromones est supposée «lente» devant celle des agents explorateurs. Nous fixons $\psi = 0,1$.

Le coefficient d'évaporation des phéromones est fixé à $\gamma = 0,2$, ce qui signifie que durant chaque unité de temps la quantité de phéromones en chaque lieu est réduite de 20% par évaporation.

Nous avons choisi de fixer le coefficient de dépôt des phéromones d à :

$$d(x, u, v, w) = 2ug.$$

Concrètement, les phéromones augmentent avec un taux égal au double du nombre de rencontres entre agents explorateurs et ressource. Pour le système discret, cela signifie que chaque agent qui rencontre une ressource dépose deux unités de phéromones.

L'équation continue qui décrit l'évolution des phéromones est :

$$\frac{\partial w}{\partial t} = 0,1\,\frac{\partial^2 w}{\partial x^2} - 0,1w + 2\,u\,g.$$

En déposant des phéromones au niveau de la ressource, les agents comptent sur la diffusion des phéromones pour informer leurs congénères lointains de la présence de la ressource. Ce comportement n'est pas exactement celui des algorithmes de type fourmi, où les agents créent des traces de phéromones en déposant ces marqueurs tout au long du rapatriement. Il est possible de considérer d'autres stratégies de dépôt de phéromones, en modifiant convenablement le coefficient d. D'éventuelles extensions de ce type seront discutées en conclusion.

6.5.2 Simulations numériques dans le cas unidimensionnel

Nous avons fixé les paramètres du système étudié dans la section 6.5.1. Le système correspondant s'écrit :

$$\left\{\begin{array}{lllll} \frac{\partial u}{\partial t} = & -\frac{\partial}{\partial x}\left(u\frac{\partial}{\partial x}\ln(w+0,01)\right) & +\frac{\partial^2 u}{\partial x^2} & -2ug+2vf \\ \frac{\partial v}{\partial t} = & \frac{\partial}{\partial x}\left(v\sin(x-\frac{\pi}{2})\right) & +0,1\frac{\partial^2 v}{\partial x^2} & +2ug-2vf \\ \frac{\partial w}{\partial t} = & & 0,1\,\frac{\partial^2 w}{\partial x^2} & -0,1w+2\,u\,g \end{array}\right. . \tag{6.12}$$

Nous avons réalisé des simulation numériques à l'aide de la méthode de Crank-Nicholson. Les résultats de ces simulations sont représentées sur la figure 6.5.

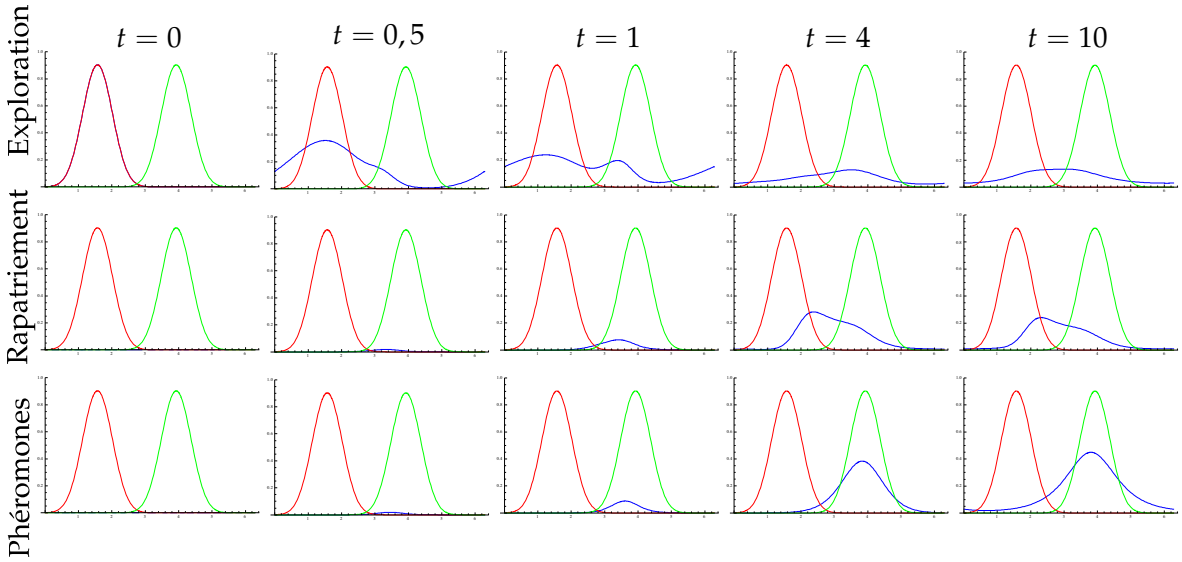


Figure 6.5 — Évolution du système

À l'instant initial $t = 0$ les agents sont tous en phase d'exploration et distribués selon le fonction f qui définit la base, et il n'y a aucune ressource. Ensuite, les agents se diffusent et rencontrent la ressource vers $t = 0,5$. À partir de ce moment-là, des agents chargés et des phéromones commencent à apparaître ($t = 1$). Le système tend ensuite progressivement ($t = 4$, $t = 10$) vers un état stable où

- les agents explorateurs sont omniprésents, mais de manière plus marquée sur le chemin le plus court $[\frac{\pi}{2}, \frac{5\pi}{8}]$,
- les agents rapatrieurs sont essentiellement positionnés sur le chemin le plus court,
- la quantité de phéromones est maximale en un point près du centre des ressources, et diminue en s'éloignant de ce point.

L'utilisation des phéromones a donc permis de rendre l'exploration plus précise. En effet, la figure 6.5 (ligne «Exploration» à $t = 10$) montre que la densité des agents explorateurs est plus importante sur le chemin le plus court entre la base et la ressource.

Pour mettre cette observation en évidence, nous utilisons la fonction ρ définie par

$$\rho(t) = \frac{\int_{\frac{\pi}{2}}^{\frac{5\pi}{4}} u(x,t)dx}{\int_{\mathcal{C}} u(x,t)dx},$$

qui mesure le ratio d'agents explorateurs sur le cadran de cercle $[\frac{\pi}{2}, \frac{5\pi}{4}]$ parmi la totalité des agents explorateurs. Son évolution est représentée sur la figure 6.6.

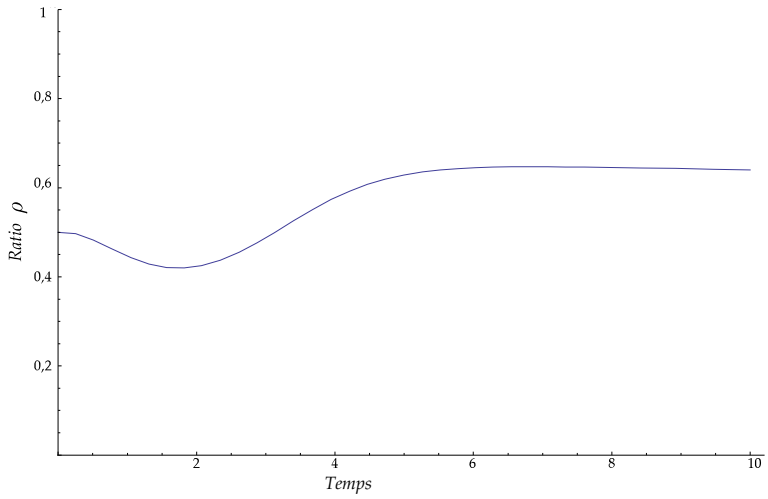


Figure 6.6 — Évolution de la fonction ρ

Dans un premier temps le ratio ρ décroît jusqu'à une valeur minimale d'environ $0,4$, ensuite il croît jusqu'à une valeur maximale d'environ $0,6$. Les agents explorateurs favorisent donc légèrement le cadran de cercle $[\frac{\pi}{2}, \frac{5\pi}{4}]$, c'est-à-dire le chemin le plus court vers la ressource, en l'occupant avec 60% de leur effectif.

Nous finissons l'analyse numérique en évaluant la performance en termes de récolte. La vitesse de récolte instantanée est mesurée par l'intégrale

$$\int_C 2\, v(x,t) f(x) dx$$

qui représente le nombre de ressources déposées à la base par unité de temps. Son évolution est représentée sur la figure 6.7.

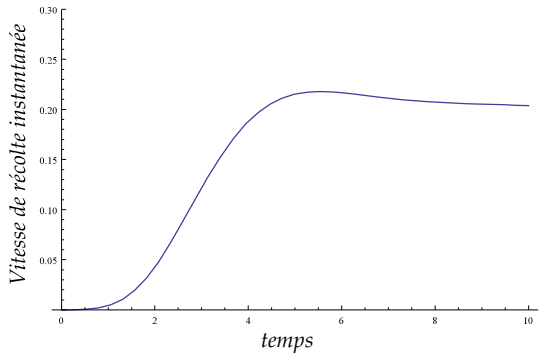


Figure 6.7 — Évolution de la vitesse de récolte instantanée au cours du temps

La vitesse de récolte instantanée croît dans un premier temps, jusqu'à atteindre une valeur maximale d'environ $0,23$. Ensuite, elle décroît légèrement en convergeant vers une valeur d'environ $0,2$.

Cette valeur peut être interprétée comme le taux d'agents chargés qui déposent une ressource à la base par unité de temps. Au bout d'une longue période ce taux converge vers $0,2$. Étant donné la distance qui sépare le centre de la base du centre des ressources, qui est

pratiquement la plus grande distance possible sur le cercle, nous estimons que ce résultat est satisfaisant.

6.6 Simulations numériques dans le cas bidimensionnel

Les simulateurs multi-agents [142, 153, 29, 24, 55] cités à la section 6.1 fonctionnent tous dans le cas bidimensionnel. Les résultats numériques proposés dans la section 6.5 ne proposent donc pas d'éléments de comparaison.

De plus, le domaine périodique étudié dans le cas unidimensionnel n'offre que deux alternatives aux agents pour joindre leur destination. Prenons les agents explorateurs : leur exploration ne peut s'effectuer que dans deux sens de circulation (horaire ou antihoraire), et le mouvement collectif observé est réduit à un sens de circulation. Le cas unidimensionnel est donc un peu trop simple pour mettre des mouvements collectifs complexes intéressants en évidence.

Pour finir, dans le cas unidimensionnel périodique, il était parfaitement envisageable pour les agents de réaliser l'exploration par un processus de convection unilatéral, en choisissant un sens de circulation commun[3]. De telles stratégies peuvent s'avérer très performantes, sans être le résultat d'un réel processus de raisonnement. De ce fait, le domaine du cercle risque de dévaloriser des stratégies intéressantes.

Pour toutes ces raisons, nous proposons dans cette section une étude numérique du système multi-agent considéré en prolongeant la situation au cas bidimensionnel.

6.6.1 Le système d'équations à réaction-diffusion-advection

Dans cette section, nous établissons le système d'équations à réaction-diffusion-advection dans le cas bidimensionnel. Nous reprenons la configuration de la section 5.6, où les agents sont positionnés sur le tore $\mathbb{T}$. La démarche adoptée est très similaire à celle de la section 5.6. Nous expliquons donc seulement les équations résultantes, avec un paramétrage similaire au cas unidimensionnel (section 6.5.1).

La base est représentée par la fonction f proportionnelle à

$$\left(\cos\left(\frac{x}{2}-\frac{\pi}{4}\right)\cdot\cos\left(\frac{y}{2}-\frac{\pi}{4}\right)\right)^{10}$$

et normalisée de sorte que $\int_{\mathbb{T}} f(x,y)dxdy = 1$. La ressource est répartie selon la fonction g proportionnelle à

$$\left(\cos\left(\frac{x}{2}-\frac{5\pi}{8}\right)\cdot\cos\left(\frac{x}{2}-\frac{5\pi}{8}\right)\right)^{10}$$

normalisée de sorte que $\int_{\mathbb{T}} g(x,y)dxdy = 1$. Les fonctions f et g sont représentées sur la figure 6.8.

3. Sous réserve que les agents soient capables de parvenir à un tel consensus avant la simulation.

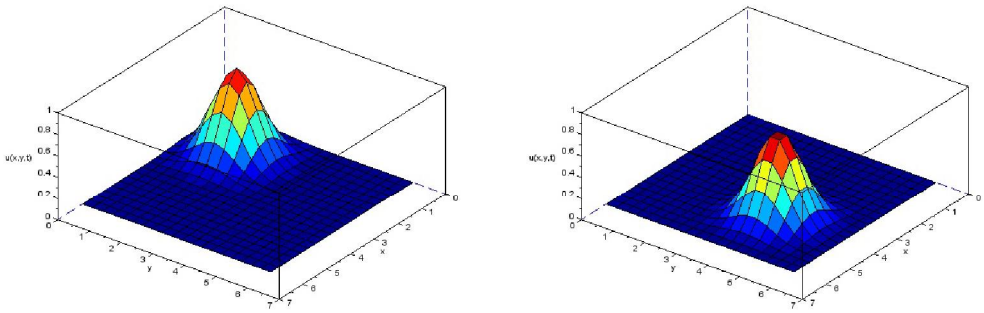


Figure 6.8 — Fonctions f (base, à gauche) et g (ressource, à droite)

Les agents explorateurs se déplacent selon un processus d'advection-diffusion, dont le coefficient de convection est donné par la dérivée logarithmique des quantités de phéromones environnantes. Dans le cas bidimensionnel, cette convection s'exprime par des dérivées partielles relativement aux directions des deux axes. Le coefficient de diffusion est constant égal à 1, tandis que les coefficients de collecte et le dépôt des ressources sont fixés à 2.

L'équation continue correspondant à cette description est

$$\frac{\partial u}{\partial t} = -\frac{\partial}{\partial x}\left(u\frac{\partial}{\partial x}\ln(w+0,01)\right) - \frac{\partial}{\partial y}\left(u\frac{\partial}{\partial y}\ln(w+0,01)\right) + \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} - 2ug + 2vf.$$

Nous avons légèrement augmenté la quantité de phéromones telle qu'elle est perçue par les agents, afin de garantir la bonne définition des dérivées logarithmiques dans les coefficients de convection.

Les agents rapatrieurs se dirigent vers la base selon un processus d'advection-diffusion, avec un champ de convection donné par

$$C_2(x,y) = \left(\frac{-\sin(x-\frac{\pi}{2})}{\sqrt{\sin(x-\frac{\pi}{2})^2+\sin(y-\frac{\pi}{2})^2+1}}, \frac{-\sin(y-\frac{\pi}{2})}{\sqrt{\sin(x-\frac{\pi}{2})^2+\sin(y-\frac{\pi}{2})^2+1}}\right).$$

Le champ de convection C_2 est représenté sur la figure 6.9.

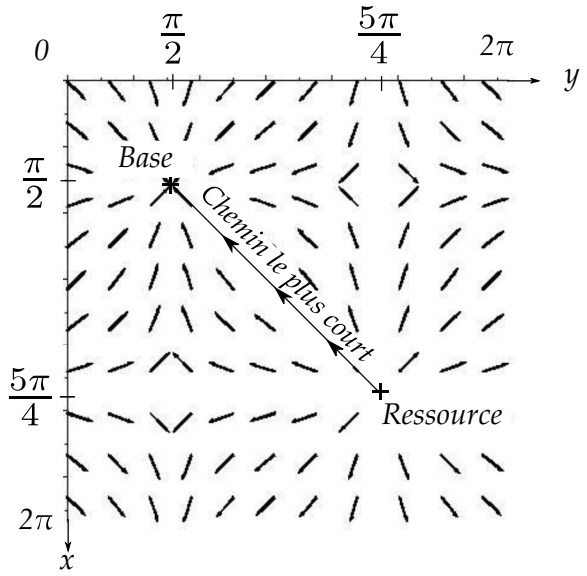


Figure 6.9 — Champ de convection des agents rapatrieurs

Les vecteurs sur la figure 6.9 représentent la vitesse moyenne macroscopique des agents rapatrieurs. Ce champ de vecteurs oriente les mouvements vers le point de coordonnées $(\frac{\pi}{2}, \frac{\pi}{2})$, qui est précisément le centre de la base. De plus, la direction indiquée correspond approximativement au chemin le plus court vers ce centre.

Pour des raisons de régularité des solutions, et pour éviter la formation de phénomènes de concentration (Cf. section 6.5.1) le mouvement de rapatriement est doté d'un petit terme de diffusion, avec un coefficient de $0,1$.

L'équation continue correspondant à cette description est

$$\frac{\partial v}{\partial t} = \frac{\partial}{\partial x}\left(\frac{\sin(x-\frac{\pi}{2})}{\sqrt{\sin(x-\frac{\pi}{2})^2+\sin(y-\frac{\pi}{2})^2+1}}v\right) + \frac{\partial}{\partial y}\left(\frac{\sin(y-\frac{\pi}{2})}{\sqrt{\sin(x-\frac{\pi}{2})^2+\sin(y-\frac{\pi}{2})^2+1}}v\right) + 0,1\frac{\partial^2 v}{\partial x^2} + 0,1\frac{\partial^2 v}{\partial y^2} + 2ug - 2vf$$

La ressource se diffuse de manière isotrope, avec un petit coefficient de $0,1$. Son coefficient d'évaporation est fixé à $0,1$, et l'intensité de dépôt de phéromones est fixé à 2.

L'équation continue correspondant à cette description est

$$\frac{\partial w}{\partial t} = 0,1\frac{\partial^2 w}{\partial x^2} + 0,1\frac{\partial^2 w}{\partial y^2} - 0,1w + 2ug$$

En résumé, l'évolution du système multi-agent complet est décrit par le système d'équations

$$\begin{cases} \frac{\partial u}{\partial t} = & -\frac{\partial}{\partial x}\left(u\frac{\partial}{\partial x}\ln(w+0,01)\right) - \frac{\partial}{\partial y}\left(u\frac{\partial}{\partial y}\ln(w+0,01)\right) \\ & +\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} - 2ug + 2vf \\ \frac{\partial v}{\partial t} = & \frac{\partial}{\partial x}\left(\frac{\sin(x-\frac{\pi}{2})}{\sqrt{\sin(x-\frac{\pi}{2})^2+\sin(y-\frac{\pi}{2})^2+1}}v\right) + \frac{\partial}{\partial y}\left(\frac{\sin(y-\frac{\pi}{2})}{\sqrt{\sin(x-\frac{\pi}{2})^2+\sin(y-\frac{\pi}{2})^2+1}}v\right) \\ & +0,1\frac{\partial^2 v}{\partial x^2} + 0,1\frac{\partial^2 v}{\partial y^2} + 2ug - 2vf \\ \frac{\partial w}{\partial t} = & 0,1\frac{\partial^2 w}{\partial x^2} + 0,1\frac{\partial^2 w}{\partial y^2} - 0,1w + 2ug \end{cases} . \tag{6.13}$$

Nous ajoutons la condition au bord périodique à ce système d'équations, ainsi que les conditions initiales suivantes :

$$\begin{cases} u(x,y,0) = f(x,y) \\ v(x,y,0) = 0 \\ w(x,y,0) = 0 \end{cases} .$$

Ces conditions signifient qu'à l'instant initial, tous les agents sont en phase d'exploration, et distribués selon la fonction f qui définit la base. De plus, il n'y a pas de phéromones à l'instant initial.

6.6.2 Les simulations numériques

Dans cette section, nous proposons d'étudier le système (6.13) par des simulations numériques, obtenues à l'aide de la méthode de Crank-Nicholson. Les figures 6.10, 6.11 et 6.12 contiennent les résultats des simulations.

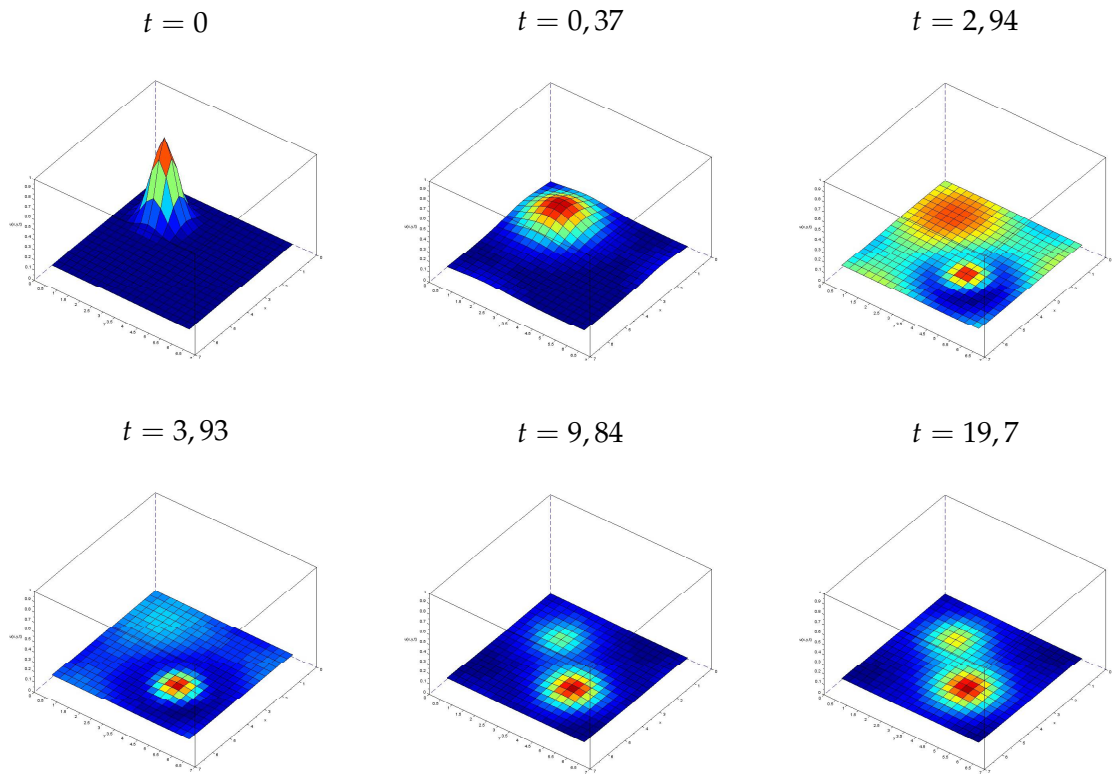


Figure 6.10 — Évolution des agents explorateurs

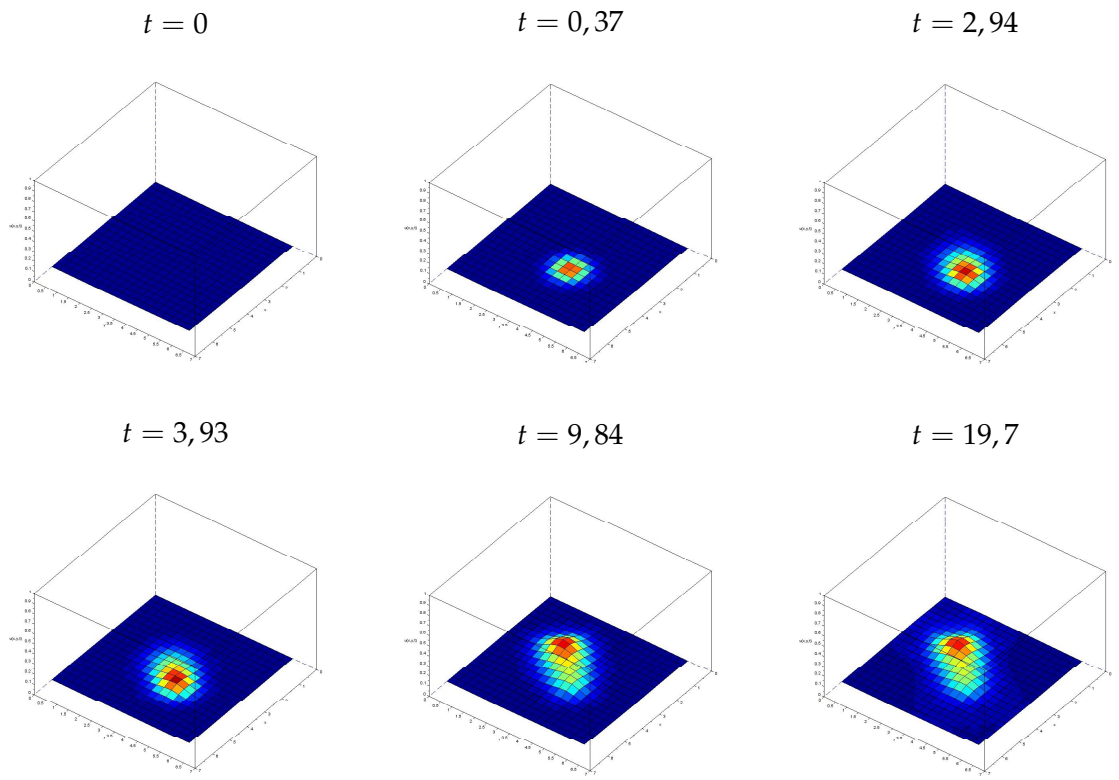


Figure 6.11 — Évolution des agents rapatrieurs

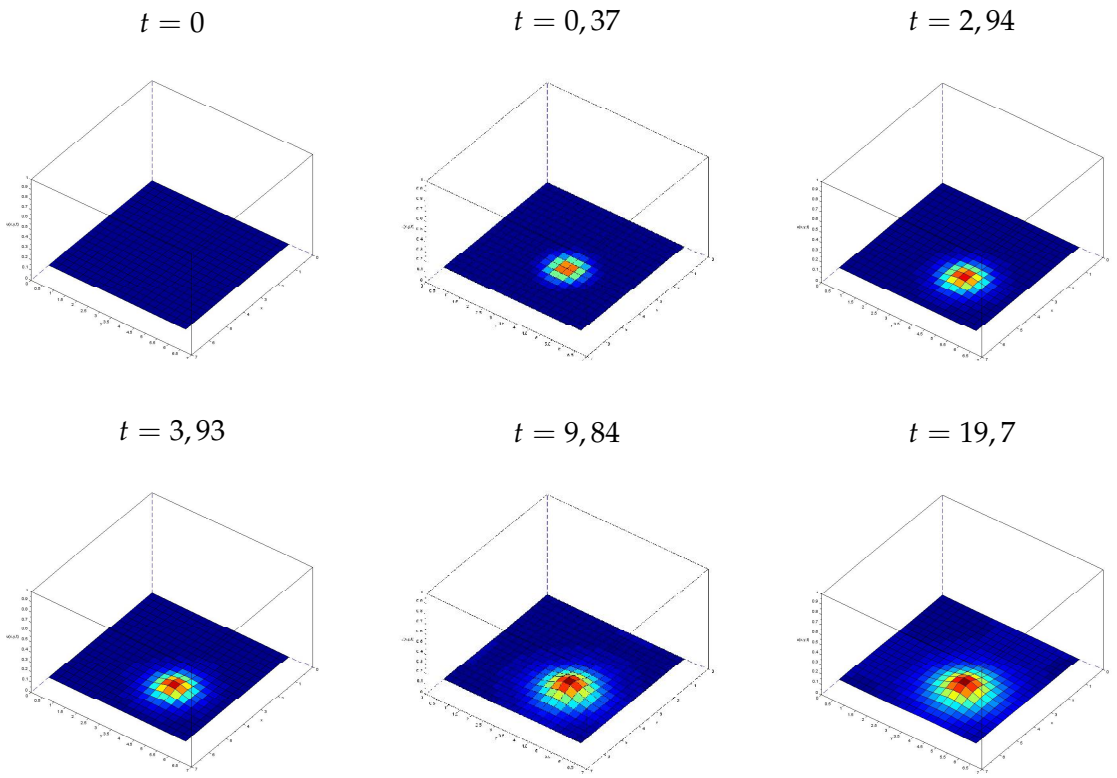


Figure 6.12 — Évolution des phéromones

Nous constatons des faits similaires au cas unidimensionnel. Dans un premier temps, les agents explorateurs se dispersent dans toutes les directions jusqu'à rencontrer la ressource à $t = 0,37$. À ce moment-là, les agents chargés et les phéromones font leur apparition. Ensuite, le système tend progressivement ($t = 2,94$, $t = 3,93$, $t = 9,84$) vers un état d'équilibre ($t = 19,7$) où :

- les agents explorateurs et les agents chargés sont essentiellement situées sur une zone qui relie le centre de la base au centre des ressources,
- les phéromones ont un effectif maximal proche du centre des ressources, et diminuent en s'éloignant de ce point.

La zone qui contient l'essentiel des agents semble proche du chemin le plus court entre la base et les ressources. Pour illustrer cette observation, nous introduisons une fonction ρ qui mesure la proportion de fourmis exploratrices proches du chemin le plus court.

Nous définissons la région $\mathcal{R}$ comme l'ensemble des positions qui sont à une distance d'au plus $\frac{\pi}{2}$ du chemin le plus court (voir figure 6.13). L'aire occupée par cette région représente environ 46% du domaine complet.

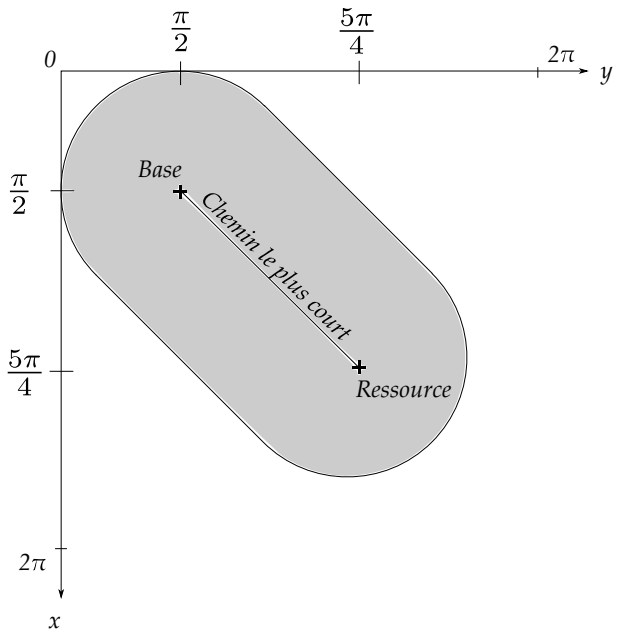


Figure 6.13 — Région $\mathcal{R}$ qui englobe le chemin leplus court

La fonction ρ, qui mesure le ratio d'agents explorateurs positionnés à proximité du chemin le plus court, est définie par :

$$\rho(t) = \frac{\int_{\mathcal{R}} u(x,y,t)dxdy}{\int_{\mathbb{T}} u(x,y,t)dxdy}.$$

Son évolution est représentée sur la figure 6.14.

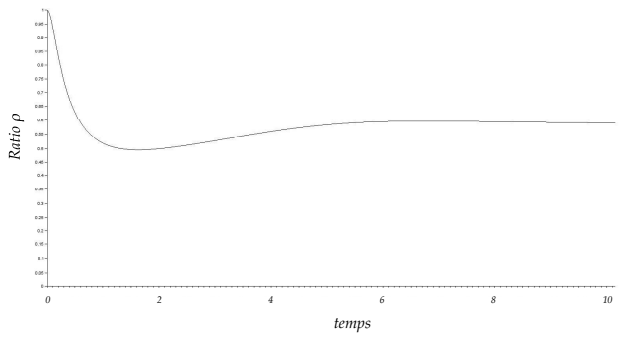


Figure 6.14 — Évolution du ratio ρ

Le ratio ρ diminue dans un premier temps, pour atteindre une valeur minimale d'environ $0,5$. Ensuite, il croît légèrement en tendant vers une valeur maximale d'environ $0,6$. Ce résultat est similaire au cas unidimensionnel (section 6.5.2), et nous tirons la même conclusion.

Les agents explorateurs expriment donc (légèrement) leur préférence pour le chemin le plus court : 60% de ces agents s'y trouvent à proximité lorsque le système est stabilisé.

La performance du système est mesurée par la vitesse de récolte instantanée

$$\int_{\mathbb{T}} 2\, v(x,y,t)\, f(x,y)\, dxdy,$$

c'est-à-dire la quantité de ressources déposées à la base par unité de temps. L'évolution de cette valeur est représentée sur la figure 6.15.

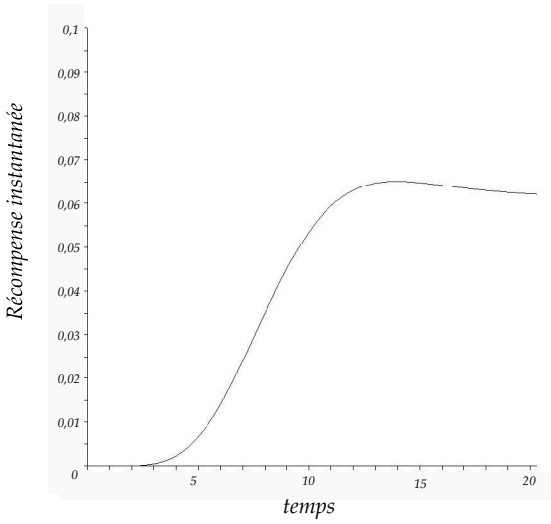


Figure 6.15 — Évolution de la récompense instantanée

Dans un premier temps, la récompense instantanée croît jusqu'à une valeur maximale d'environ $0,065$. Ensuite, elle décroît légèrementen tendant vers une valeur d'environ $0,06$.

On peut comprendre ce résultat simplement : lorsque le système a convergé, le taux d'agents chargés qui déposent une ressource à la base à chaque pas de temps est d'environ $0,06$. Ce résultat est moins satisfaisant que dans le cas unidimensionnel où ce taux de dépôt était de $0,2$.

Cette chute de performance peut s'expliquer par plusieurs raisons. D'une part, la distance entre le centre de la base et celui des ressources est plus importante dans les simulations bidimensionnelles. D'autre part, il existe une multitude de trajectoires possibles entre la base et la ressource, là où il n'y en avait que deux dans le cas unidimensionnel (horaire ou anti-horaire). De ce fait, il est possible que de nombreux agents explorateurs parcourent un long trajet, et fassent un grand détour par rapport au chemin le plus court, avant de rencontrer la ressource.

6.7 Conclusion

Dans ce chapitre, nous avons introduit un modèle continu afin d'étudier un problème de collecte de ressources.

Dans un premier temps, nous avons représenté le problème à l'aide d'une famille de marches aléatoires corrélées, en interaction à travers les densités en chaque position ainsi

que des de marqueurs spatiaux et environnementaux qui permettent aux agents de communiquer de manière indirecte.

Ensuite nous avons dérivé une approximation continue de ce modèle en faisant tendre le nombre d'agents vers l'infini, puis en faisant tendre les pas de temps et espace vers 0. L'équation limite obtenue est de type *réaction-diffusion-advection*.

En parcourant la littérature sur les équations à réaction-diffusion-advection, il s'est avéré que la résolution de ces équations est délicate, et que peu de résultats mathématiques sont à disposition pour une étude en toute généralité. Même certaines questions très simples, comme la simple existence de solutions physiques (positives, régulières) sont délicates à traiter. Pour cette raison, nous avons eu recours à des simulations numériques.

Ces simulations numériques, paramétrées de manière convenable, montrent un phénomène intéressant. Le dépôt de phéromones permet aux agents d'établir un itinéraire vers la ressource, qui est proche du chemin le plus court.

Les équations utilisées dans ce chapitre exhibent donc un phénomène émergent intéressant. Démontrer ces phénomènes de manière formelle semble être une tâche délicate. Il est sans doute possible, dans des situations particulières, d'y parvenir à l'aide de méthodes spécifiques comme des changements de variable. L'étude approfondie d'un de ces cas particuliers peut constituer une perspective intéressante à court terme.

Un problème qui n'a pas été abordé dans ce chapitre est la recherche de la meilleure stratégie de collecte, en formulant le problème étudié comme un MDP. Pour ce faire il faut

- définir les paramètres de contrôle, qui sont les probabilités des déplacements des agents, et/ou le coefficient de dépôt de phéromones d,
- munir le modèle discret aléatoire de la fonction de récompense instantanée

$$\int_{\mathbb{T}} 2u(x,y,t)g(x,y)dxdy.$$

Nous avons déjà vu des conditions sous lesquelles le MDP obtenu est compatible avec la limite champ moyen (section 4.5.3). Nous conjecturons que, sous certaines hypothèses, la limite spatiotemporelle aboutisse à un problème de type *jeu à champ moyen* [92]. Ce problème limite est la simple réunion du système à réaction diffusion (6.10), et d'une équation d'optimalité [4] qui exprime l'optimalité des paramètres de contrôle.

Nous n'avons pas approfondi cette dernière idée. Des difficultés dans sa mise en oeuvre sont à anticiper. Les jeux à champ moyen sont des problèmes techniques, généralement difficiles à résoudre, analytiquement comme numériquement [5].

4. De type *Hamilton-Jacobi-Bellman*.
5. Des méthodes de résolution numérique spécifiques ont été développées dans [2].

Conclusion

Bilan

Dans cette thèse, nous nous sommes intéressés à l'étude mathématique des systèmes multi-agents à fonctionnalité émergente. Dans cette section, nous résumons les idées et résultats qui ont été présentés dans ce manuscrit.

La modélisation des systèmes complexes Le choix d'une représentation mathématique convenable des systèmes complexes à fonctionnalité émergente est une tâche délicate. Nous avons étudié la pertinence d'un ensemble de modèles envisageables vis-à-vis d'une classe de systèmes spécifiques. Deux critères importants ont été identifiés :

- **l'expressivité du modèle**, c'est-à-dire l'ensemble des traits que le modèle prend en compte,
- **la maîtrise mathématique du modèle**, c'est-à-dire la facilité avec laquelle le modèle peut être étudié, et la portée des résultats que l'on peut démontrer.

La modélisation consiste à établir un compromis entre ces deux critères. Malheureusement ils sont difficilement quantifiables, et le compromis est donc difficile à mesurer pour chaque modèle.

L'étude des systèmes multi-agents à l'aide des processus de Markov décisionnels Les processus décisionnels markoviens permettent d'intégrer la finalité du système dans sa dynamique, et d'exprimer sans ambiguïté le problème de la recherche d'une stratégie optimale. En revanche, ce formalisme aboutit à une impasse pour de grands systèmes, en raison d'une complexité algorithmique explosive.

De plus, les techniques de calcul d'une stratégie optimale d'un MDP font appel à des informations importantes sur le système. Ce fait les rend inutilisables en pratique, et ne correspond clairement pas au caractère local et incomplet des perceptions des agents dans un contexte multi-agent. Les adaptations des MDP aux systèmes multi-agents proposées ne sont efficaces que dans des contextes spécifiques, très éloignés d'une situation réaliste.

Modélisation des systèmes multi-agents par des méthodes de champ moyen Les méthodes de champ moyen permettent, moyennant une approximation qui peut être discutée, de simplifier la représentation du système et de négliger les fluctuations statistiques dans

son évolution.

À travers quelques exemples nous avons illustré l'utilisation de ces méthodes pour représenter, et pour contrôler de manière optimale, un grand système d'agents en interaction. Les résultats montrent que le calcul de la stratégie optimale peut être mise en oeuvre simplement. Même si cette stratégie optimale exige des connaissances globales qui ne sont pas à la portée des agents, il est possible de l'utiliser dans les stratégies heuristiques afin d'obtenir des performances intéressantes.

Par ailleurs, nous avons pu étudier certaines stratégies heuristiques grâce à l'approximation champ moyen, et comparé leur performance à celle de la stratégie optimale. Ces stratégies heuristiques avaient toutefois la particularité d'être compatibles avec la limite du champ moyen, ce qui n'est pas toujours le cas.

Utilisation d'équations à réaction-diffusion-advection pour la représentation de systèmes multi-agents situés réactifs Les chapitres 5 et 6 montrent que la modélisation continue d'un système multi-agent situé réactif à l'aide d'équations aux dérivées partielles peut fournir des résultats intéressants. Les outils mathématiques utilisés pour étudier ces équations ne sont pas influencés par la taille du système, et peuvent donc être plus avantageux que des méthodes discrètes lorsqu'il s'agit de démontrer des propriétés formelles.

Dans la situation du chapitre 5, nous avons déterminé deux stratégies locales qui permettent de réaliser un objectif global dans le contexte spécifique d'un système multi-agent situé réactif. Des résultats théoriques montrent que sous ces stratégies le système converge vers une distribution optimale, et permettent de quantifier la vitesse de convergence. Ces résultats donnent également une mesure de la robustesse du système vis-à-vis des perturbations, et laissent entrevoir une capacité d'adaptation lorsque le système évolue en environnement dynamique.

En revanche, pour la situation du chapitre 6, les méthodes mathématiques existantes ne permettent pas encore de réaliser une étude formelle en toute généralité. Des simulations numériques permettent cependant de constater un phénomène émergent intéressant.

Une disparité dans les méthodes formelles Les problèmes étudiés dans les chapitres 5 et 6 montrent que, moyennant une approximation continue, on peut étudier les propriétés dynamiques d'un grand système multi-agent à l'aide de méthodes d'analyse mathématique. Ces méthodes permettent de contourner les difficultés liées à la complexité algorithmique des systèmes étudiés.

Il existe donc une disparité dans les méthodes formelles : l'étude mathématique d'un très petit système, formé de quelques agents et d'un très petit espace d'états est réalisable par des méthodes classiques. À l'autre extrême, un très grand système avec un grand espace d'états peut être approché par une modèle continu dont on peut démontrer des propriétés formelles. Entre ces deux extrêmes l'étude formelle peut être très délicate, et malheureusement une grande partie des systèmes utilisés de nos jours se trouvent dans cet intervalle.

Perspectives

Les résultats de cette thèse ouvrent sur quelques perspectives envisageables.

Exploration des limites des méthodes discrètes Dans le deuxième chapitre, certains modèles discrets comme les systèmes de particules ou les jeux sont écartés pour des raisons de complexité algorithmique. Le troisième chapitre consacré aux MDP reprend cet argument et quantifie les limites d'applicabilité des MDP en termes de complexité algorithmique.

Ces type d'argument montre essentiellement que dans le pire cas de figure, les modèles exigent des calculs trop importants pour être exploitables. En revanche, rien n'exclut la possibilité de mener ces calculs à terme dans des cas particuliers.

Par exemple, les automates cellulaires peuvent exhiber des comportements surprenants, qui sont difficiles voire impossibles à démontrer de manière formelle. Cela ne signifie pas qu'il n'existe aucun résultat formel sur les automates cellulaires. Dans certains cas particuliers, comme par exemple les fourmis de Langton, on peut démontrer des propriétés formelles avec des méthodes ad-hoc.

De la même manière, la théorie des jeux a été écartée en raison de la NP-complétude d'un ensemble important de jeux. Il est clair que tous les jeux ne sont pas concernés par cette complexité algorithmique, et il est possible que de nombreux exemples intéressants puissent être résolus en pratique.

De manière générale, les modèles que nous avons écartés peuvent s'avérer utiles pour l'étude de systèmes particuliers. L'étude de tels cas particuliers pourrait, à terme, faire apparaître les limites concrètes de ces modèles de manière plus précise.

Utilisation de la théorie des jeux pour guider le processus de conception Comme nous l'avons mentionné dans le chapitre 2, la théorie des jeux peut être utilisée au cours du processus de conception d'un système multi-agent. L'objectif global du système est représenté par une fonction de satisfaction (ou gain) g. L'idée est de créer un jeu pour lequel les joueurs vont naturellement avoir tendance à choisir les actions qui maximisent le gain g. Dans la littérature, une telle démarche est appelée *mécanisme d'incitation* ou *reverse game theory*.

La création d'un jeu en adéquation avec l'objectif global passe par une étape de décomposition du gain global en gains individuels :

$$g \mapsto (g_1, \dots, g_N), \tag{6.14}$$

où g_n est le gain individuel du joueur n. En effectuant cette décomposition de manière convenable, il est possible que :

- chaque gain individuel g_n ne dépende que d'un petit nombre de joueurs, et par conséquent chaque joueur n'a besoin de coordonner sa stratégie qu'avec un petit nombre d'autres joueurs,
- des stratégies individuelles égoïstes aboutissent à une stratégie collective optimale [6].

6. Au sens du gain global g.

Pour des systèmes multi-agents simples, comme celui utilisé dans [121] pour la résolution du problème des N reines, il est possible d'écrire la décomposition (6.14) de manière explicite.

Il serait intéressant d'étudier la forme générale que prend cette décomposition pour des méthodologies multi-agents comme la théorie AMAS [31, 64]. Cela pourrait être fait en généralisant le raisonnement utilisé dans [121], qui est relativement générique.

Dans un second temps, il pourrait être intéressant d'étudier les qualités des jeux obtenus par ces décompositions. L'existence d'équilibres de Nash est une question qui vient naturellement à l'esprit. On peut également s'interroger sur l'existence de stratégies simples qui permettent d'aboutir à ces équilibres de Nash.

Étalissement d'un critère de sélection des modèles Dans le deuxième chapitre nous avons vu que plusieurs modèles peuvent être des candidats intéressants pour représenter un système donné, et tenté de donner des arguments pour discriminer ces modèles. Parallèlement, dans les chapitres 4, 5 et 6 nous avons proposé différents modèles pour un même système en nous basant sur les grandeurs caractéristiques du système.

Dans l'exemple du chapitre 5 nous avons caractérisé le système par sa «taille». Pour N agents en marche aléatoire sur une grille discrète qui effectuent des déplacements d'un pas élémentaire δ pendant chaque unité de temps τ, la taille est mesurée par le triplet (N, δ, τ). Lorsque ce triplet est proche du point limite $(\infty, 0, 0)$, le modèle le plus intéressant est une équation de Fokker-Planck. En revanche, si N est grand devant les deux autres grandeurs δ et τ, le modèle du champ moyen est le plus pertinent.

Il serait intéressant d'établir une méthode de sélection systématique qui indique le modèle le plus pertinent selon la valeur du triplet (N, δ, τ). Cette démarche revient à déterminer des zones dans l'espace des valeurs de (N, δ, τ), dont chacune correspond à un modèle. Il serait par ailleurs intéressant d'étudier les frontières de ces zones, pour lesquelles plusieurs modèles ont la même pertinence. Ces questions de sélection des modèles peuvent être généralisées à tout système complexe.

Poursuite de l'étude des modèles continus Dans le chapitre 5, nous avons déterminé deux stratégies optimales en étudiant une représentation continue d'un système multi-agent spécifique. Il est fort possible que d'autres stratégies, également basées sur des informations locales, soient optimales à long terme. Nous n'avons pas l'intention d'explorer la totalité de ces stratégies, mais il serait intéressant d'en chercher d'autres et de les interpréter en termes de comportements d'agents.

Une perspective très intéressante est de prolonger le problème considéré dans le chapitre 5 (orienter un système d'agents vers une distribution ciblée) en introduisant plusieurs populations d'agents. Ces populations peuvent avoir des distributions ciblées identiques ou différentes, et interagir de manière positive ou négative au cours de leurs rencontres.

Pour une extension à deux populations, nous nous attendons à un système d'équations

de la forme :

$$\begin{cases} \frac{\partial u}{\partial t} = -\frac{\partial}{\partial x}\left(c_1(x,u,v)u\right) + \frac{\partial^2}{\partial x^2}\left(d_1(x,u,v)u\right) \\ \frac{\partial v}{\partial t} = -\frac{\partial}{\partial x}\left(c_2(x,u,v)v\right) + \frac{\partial^2}{\partial x^2}\left(d_2(x,u,v)v\right) \end{cases},$$

où

- u et v sont les densités des deux populations d'agents
- c_1 et c_2 sont leurs coefficients de convection respectifs, et dépendent de la position courante, ainsi que des densités de chaque population d'agents,
- d_1 et d_2 sont les coefficients de diffusion des population, et dépendent également de la position courante, ainsi que des densités de chaque population d'agents.

Une première approche possible est d'ajouter un couple de récompenses locales (r_1, r_2) à ces équations qui

- récompensent les agents de chaque population de manière croissante à mesure que leur densité s'approche de la densité ciblée,
- pénalisent le surencombrement du à la présence des deux populations.

Nous pouvons intégrer un effet d'aversion entre les deux populations d'agents dans ces fonctions de récompense, en les faisant baisser de manière importante avec la densité de la population concurrente.

Un point de départ simple est de fixer un jeu de paramètres (i.e. une stratégie) c_1, c_2, d_1, d_2, et de simuler l'évolution des systèmes ainsi que l'évolution de leurs récompenses. Comme nous sommes en présence de deux populations d'agents, avec des objectifs potentiellement divergents, les stratégies doivent comporter un compromis entre :

- **égoïsme :** simple optimisation de sa propre fonction de satisfaction,
- **altruisme :** faire augmenter la satisfaction de l'autre population d'agents à son détriment.

Dans ce contexte, la coopération est un compromis entre ces deux attitudes extrêmes.

Revenons sur système d'équations à réaction-diffusion-advection utilisé dans le chapitre 6 pour modéliser un système d'agents en cours de collecte d'une ressource. Les travaux présentés au cours de l'atelier «*New trends in modeling, control and inverse problems*» à l'Université Paul Sabatier en juin 2014 [7] montrent qu'il existe une forte activité mathématique récente sur les équations à réaction-diffusion-advection. Il est possible que des résultats généraux sur l'existence de solution régulières, et sur la convergence de ces solutions, soient publiés dans un avenir proche.

Utilisation de la théorie des jeux à champ moyen Une piste très intéressante, est d'utiliser les jeux à champ moyen [92] pour représenter les systèmes multi-agents. Cette théorie mathématique récente étudie le système comme la limite d'un jeu noncoopératif dont le nombre de joueurs tend vers l'infini. De plus, dans un jeu à champ moyen les joueurs déterminent leur stratégie optimale en interagissant avec la totalité de joueurs [8] plutôt qu'avec chacun d'eux.

Pour un jeu à champ moyen, l'optimalité [9] d'une stratégie est exprimée par un sys-

7. `http://www.math.univ-toulouse.fr/~ervedoza/WebpageCIMI-Enrique/`
8. Donc le *champ moyen*, à la limite.
9. Au sens de Nash.

tème d'équations formé d'une équation de Fokker-Planck et d'une équation d'optimalité. Cette dernière est appelée équation de *Hamilton-Jacobi-Bellmann*, et correspond à l'équivalent continu des équations d'optimalité discrètes que nous avons rencontrées dans le cadre des MDP (chapitre 3).

La recherche d'une stratégie optimale, c'est-à-dire la résolution du système d'équations associé au jeu à champ moyen, est généralement difficile. Il est cependant possible d'obtenir de résultats précis dans divers cas particuliers [73].

En exploitant la piste des jeux à champ moyen il pourrait être intéressant, dans un premier temps, d'étudier un certain nombre de systèmes multi-agents spécifiques. Un objectif à plus long terme est d'inscrire des théories multi-agents générales, comme la théorie AMAS [31, 64], dans le cadre des jeux à champ moyen. Ce travail permettrait de justifier ces démarches d'un point de vue théorique et de proposer, dans un second temps, des garanties formelles de leur adéquation.

Quatrième partie

Annexe

A Recherche de structure dans les graphes pour la certification d'un circuit électrique avec le plan spécifié.

Cette annexe contient les résultats obtenus à la semaine SEME 2012 [1], en collaboration avec D. Slutskiy et D. Stépanova. Le sujet intitulé *"Comparaison d'un circuit électrique avec le plan spécifié"* a été proposé par S. Valette et J.-C. Courrège, et s'est déroulé sous la responsabilité académique de J.-M. Couveignes.

A.1 Présentation du problème

Les appareils électriques sont omniprésents dans le monde actuel, et leur complexité croît de manière importante. La loi de Moore illustre ce propos, en conjecturant une croissance exponentielle du nombre de transistors par appareil au cours du temps.

Le même constat vaut pour les circuits intégrés. Ces circuits forment les briques élémentaires de tous les appareils électriques, et contiennent un nombre croissant de transistors et de connecteurs. Le circuit imprimé classique (mono-couche) n'est plus une solution envisageable pour des contraintes de volume, et a été remplacé depuis plusieurs décennies par les circuits multi-couches.

Les circuits multi-couches sont une superposition de plusieurs circuits simples, reliés entre eux par des connecteurs appelés *vias*. Le fabrication de ces circuits est réalisée en plusieurs étapes :

- une base ultrafine de transistors est déposée,
- des couches successives d'isolant et conducteurs sont déposés sur cette base,
- la gravure des pistes électriques est réalisée par vernissage à haute précision et rayonnement ultraviolet.

Ce processus est extrêmement délicat, et est soumis à des normes de sécurité et d'hygiène

1. http://www.math.univ-toulouse.fr/SEME/

très élevées [2]. D'autre part, l'extrême précision nécessaire aux différentes étapes du processus fait qu'une grande partie des productions initiales est détruite (les pertes peuvent aller jusqu'à 80%).

Le résultat de ce processus est néanmoins très impressionnant. Actuellement
- un circuit multi-couches peut contenir plusieurs dizaines de couches,
- la taille des couches peut atteindre un ordre de grandeur de $10nm$,
- un seul appareil électrique peut contenir des millions de transistors [3].

En raison des difficultés liées à la production et du faible rendement initial, les usines de fabrication des circuits multi-couches (appelés *fab*) se raréfient. *La loi empirique de Rock* énonce que le prix d'une *fab* est multipliée par deux tous les 4 ans. À l'heure actuelle, la construction d'une fab atteint un prix de plusieurs milliards de dollars [4]. En conséquence, la conception du plan d'un circuit et sa réalisation physique sont généralement réalisées dans des lieux différents.

En définitive le concepteur transmet le plan de spécification d'un circuit à la fonderie et reçoit, à l'issue de la conception, un ensemble de circuits physiques. Ces circuits sont soumis à un ensemble de vérifications pour valider leur fonctionnement, mais ces vérifications ne permettent pas certifier la conformité complète du circuit physique au plan de spécification. Si le concepteur effectue des vérifications dans les limites de l'utilisation standard du circuit, il ne pourra détecter des fonctionnalités supplémentaires. Un fabricant malveillant a donc la possibilité d'introduire des fonctionnalités supplémentaires qui peuvent perturber le fonctionnement du circuit, et qui ne sont pas discernées par les tests du concepteur.

En plus des tests de fonctionnement, il est possible d'effectuer des certifications physiques. Une certification physique consiste à analyser, par imagerie haute précision, différentes régions du circuit et à les comparer au plan de spécification. Afin d'exposer les parties intérieures du circuit il faut procéder à une découpe laser, toujours avec une précision extrême. La démarche de certification physique est donc une tâche longue et onéreuse.

La problématique étudiée dans cette annexe est la suivante : peut-on accélérer ce processus de vérification physique en identifiant les zones les plus sensibles ? En partant d'un fichier de spécification, nous allons étudier certaines régions d'un circuit en termes physiques (densité de cuivre) mais aussi en termes de connexité. Cette approche permet d'identifier les zones les plus en proie à une modification malveillante.

Contrairement aux apparences, le thème et la problématique de ce travail sont liés aux systèmes multi-agents. Il y a un certain nombre de points communs aux circuits électriques et aux systèmes multi-agents :
- un grand nombre d'entités en interaction, qui peuvent être les transistors, les connecteurs extérieurs, ou encore les pistes de cuivre,
- des interactions locales entre ces entités,
- une fonctionnalité globale qui dépend de l'organisation de ces différents éléments, et qui n'est pas simplement déductible d'informations locales.

2. Bien plus élevées que pour l'industrie alimentaire.
3. Certaines cartes graphiques en contiennent plus de 6 millions `http://en.wikipedia.org/wiki/Transistor_count`
4. `http://en.wikipedia.org/wiki/Semiconductor_fabrication_plant`

Le problème étudié est également lié à des problématiques multi-agents. En effet, les systèmes multi-agents peuvent être modifiés de manière locale, au niveau des agents, afin de produire une comportement global différent. Ces modifications peuvent être bénéfiques, mais aussi néfastes pour le système. Il est par exemple envisageable qu'un système contienne des agents "saboteurs" qui visent à faire échouer la fonctionnalité globale. Pour ce faire, ils ont besoin d'accéder aux parties les plus importantes et vulnérables du système, ce qui nous ramène à la problématique des circuits intégrés.

En outre, les algorithmes proposés dans cette annexe permettent d'obtenir une connaissance locale de l'importance d'un agent dans le système. Cette information peut avoir de nombreuses applications. Elle donne par exemple la possibilité de hiérarchiser les agents par ordre d'influence dans le système, ce qui permet de quantifier les risques de défaillance d'un agent, et l'effet d'une telle défaillance sur le système. Dans le paradigme multi-agent AMAS [64] ces situations de défaillance jouent un rôle important dans le comportement des agents.

La suite de cette annexe est organisée en quatre parties. Dans un premier temps (section A.2), nous présentons l'architecture typique d'un fichier de spécification. Ensuite nous donnons quelques algorithmes pour lire ces fichiers (section A.3). La partie suivante (section A.4) présente des algorithmes qui permettent d'analyser la structure du circuit spécifié. Nous concluons à la section A.5 en proposant quelques perspectives envisageables.

A.2 Architecture d'un fichier de spécification

Un fichier de spécification contient l'ensemble des informations nécessaires à la conception d'un circuit. Ces informations sont essentiellement d'ordre géométrique, indiquant la position souhaitée de chaque piste de métal.

Il est important de noter que le concepteur n'a pas un contrôle complet sur ce fichier. En effet, le fichier de spécification est obtenu à l'aide d'un compilateur qui dresse le plan physique d'un circuit à partir d'un plan électrique (logique). Aussi, il existe différents formats de fichier de spécification. Nous nous concentrons sur le format *cif (Common Intermediate Format)*. Un fichier de spécification de ce format se présente sous la forme :

```
L CWP ;
   B 448 424 400 628 ;
   B 642 288 400 188 ;
L CMS ;
   ⋮

C 43 R 1 0 T 20 340 ;
```

On identifie

- un ensemble de paragraphes commençant par *L (Layer)* indiquant les différentes couches,
- des lignes commençant par *B (Box)* dans chaque paragraphe. Ces lignes spécifient la position et la forme des pistes de cuivre,

– (éventuellement) un ensemble de lignes en fin de fichier contenant des opérations géométriques. Le but de ces lignes est de recopier certaines parties du circuit qui se répètent, en effectuant des translations *(Lettre T)*, rotations *(Lettre R)* ou éventuellement des symétries.

Les pistes de cuivre du fichier étudié sont rectangulaires, et leur position est indiquée par 4 coordonnées. Les deux premières indiquent la position du centre, et les deux dernières indiquent la largeur et longueur du rectangle (voir fig. A.2).

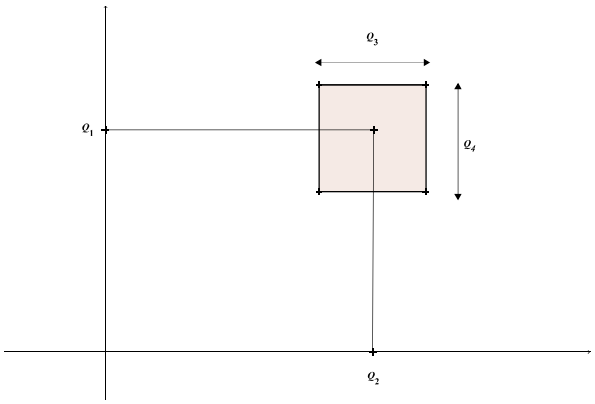


Figure A.1 — Spécification de la position d'un rectangle (Q_1, Q_2, Q_3, Q_4).

Le fichier de spécification que nous avions à disposition pour effectuer des essais contient 500 000 lignes. Les différentes couches sont organisées selon

CMS : 319			
CVA : 162988			
CMF : 3668			
CCA : 188852			
CPG : 5782	*ou*	CMS	Couche 2
CSP : 494		CVA	Vias
CSN : 568		CMF	Couche 1
CWP : 220		CCA + CPG	Vias
CWN : 256		CSN + CSP	Transistors
		CWN + CWP	Transistors

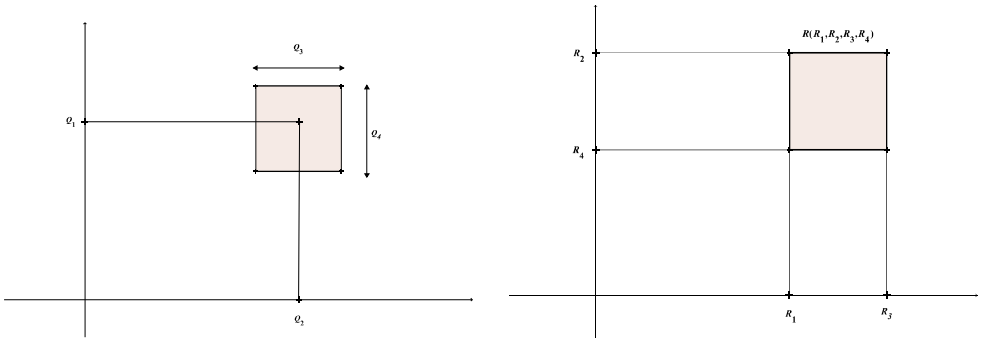
.

La représentation de droite tient compte de la disposition géométrique de ces couches. En effet, les paires de couches $CCA + CPG$, $CSN + CSP$ et $CWN + CWP$ sont situées sur un même niveau, mais sont formés d'un métal différent.

Contrairement aux apparences (à gauche) ce circuit ne contient que deux couches : CMF et CMS. Les autres font office de d'interconnecteurs entre niveaux *(vias)*, ou de transistors. Notons également que les transistors sont formés de deux types de métaux N ou P. La superposition de ces métaux forme des transistors $N - P - N$ ou $P - N - P$ ce qui explique les couches $CNS + CSP$ et $CWN + CWP$.

A.3 Lecture d'un fichier de spécification

Nous avons choisi d'utiliser le logiciel Maple 17 pour le traitement numérique du fichier de spécification. Dans ce logiciel, les rectangle sont représentés par un quadruplet (R_1, R_2, R_3, R_4) où (R_1, R_2) sont les coordonnées du coin supérieur gauche du rectangle, et (R_3, R_4) les coordonnées du coin inférieur droit. Il a donc fallu transformer les informations issues du fichier de spécification afin de les mettre au format de Maple (voir fig. A.3)



B 448 424 400 628 $\longrightarrow$ $[448, 424, 400, 628, niveau]$

Figure A.2 — Rectangle au format cif (à gauche) et au format Maple (à droite).

Une cinquième composante est jointe au vecteur obtenu afin de préciser le niveau auquel se trouve le rectangle.

Cette opération permet de visualiser les couches les moins denses du circuit.

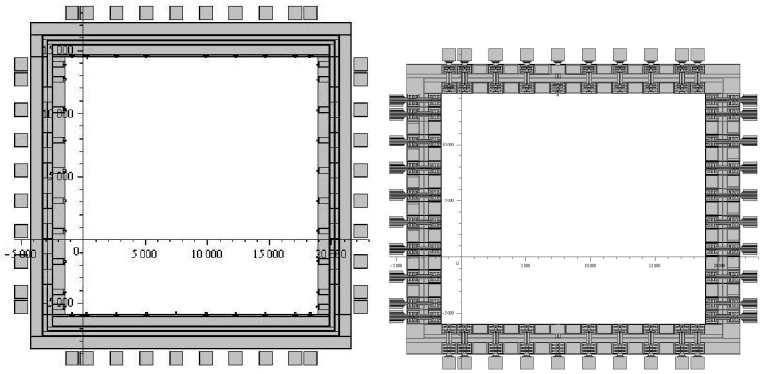


Figure A.3 — La couche CMS (à gauche) et CMF (à droite)

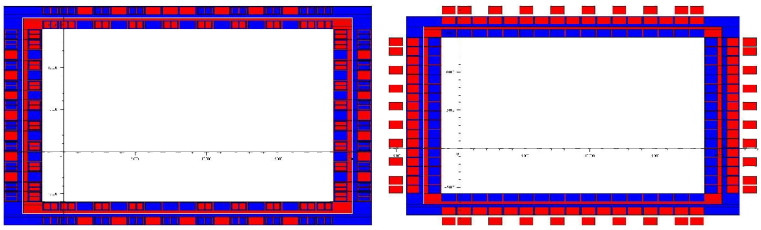


Figure A.4 — La couche CSN+CSP (à gauche) et CWN+CWP (à droite)

Sur la figure A.4 on trouve les deux couches qui constituent les transistors par superposition. Les cristaux **P** sont représentés en rouge et les cristaux **N** en bleu.

La forme spécifique de ce circuit, creuse au centre, est vraisemblablement liée à son fonctionnement que nous ignorons. L'analyse que nous faisons dans les parties suivantes n'utilise d'ailleurs pas cette information.

Les 40 petits rectangles extérieurs servent à connecter le circuit.

A.4 Analyse du fichier de spécification

Nous partageons les caractéristiques du plan de spécification en deux catégories :
- les propriétés géométriques, portant sur la position des pistes de cuivre,
- les propriétés topologiques, caractérisant la connexité des parties du circuit.

Une opération indispensable pour cette analyse est de vérifier l'intersection entre deux rectangles donnés. Nous utilisons de manière intensive l'observation suivante (voir fig. A.5) :

Deux rectangles (représentés sous format Maple) $R = (R_1, R_2, R_3, R_4)$ *et* $S = (S_1, S_2, S_3, S_4)$ *s'intersectent si et seulement si* $[R_1, R_3]$ *intersecte* $[S_1, S_3]$ ***et*** $[R_4, R_2]$ *intersecte* $[S_4, S_2]$.

Si tel est le cas, leur intersection est le rectangle

$$[R_1, R_3] \cap [S_1, S_3] \times [R_4, R_2] \cap [S_4, S_2]$$

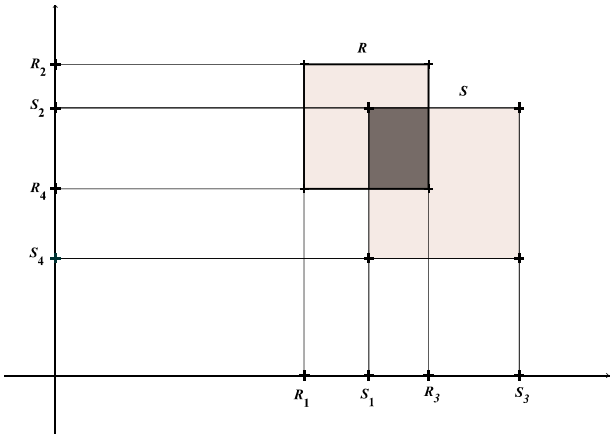


Figure A.5 — Intersection de deux rectangles R et S au format Maple.

L'algorithme en langage naturel correspondant à cette vérification est :

Algorithme A.1 — Intersection de deux rectangles

Entrées : Deux rectangles R et S

Sorties : Rectangle M (éventuellement réduit à un segment, un point, ou l'ensemble vide)

1 L $\leftarrow [R_1, R_3] \cap [S_1, S_3]$
2 H $\leftarrow [R_4, R_2] \cap [S_4, S_2]$
3 **si** $L \neq \emptyset$ *et* $H \neq \emptyset$ **alors**
4 | **retourner** $[L_1, H_2, L_2, H_1]$
5 **sinon**
6 | **retourner** $\emptyset$
7 **fin**

Grâce à cet algorithme, nous pouvons effectuer une première analyse du fichier de spécification. La zone complète du circuit est partagé en $20 \times 20 = 400$ zones carrées. Dans chacun de ces carrés, l'algorithme A.1 permet de calculer la densité de cuivre, ainsi que le nombre de carrés qui sont en contact avec la zone. Les résultats sont représentés sur la figure A.6. Les zones les plus sombres sont plus denses ou ont un nombre de rectangles plus important selon la figure.

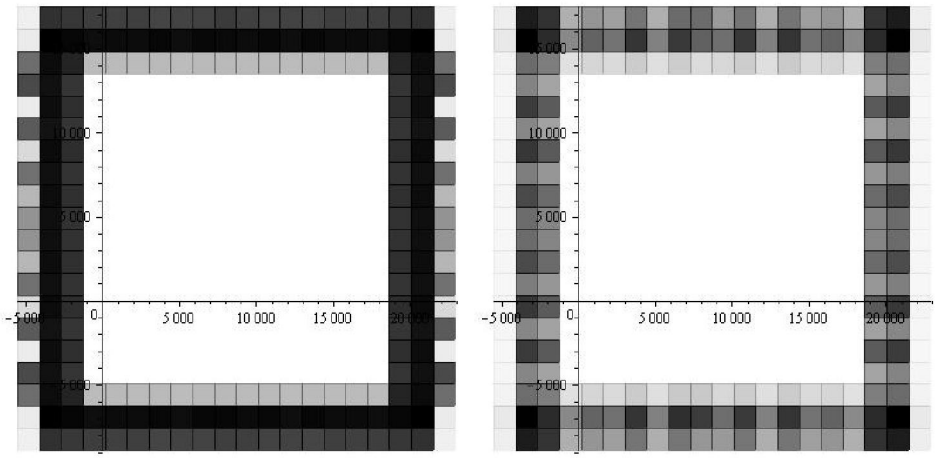


Figure A.6 — Densité de cuivre (à gauche) et nombre de rectangles (à droite)

Ces figures mettent en évidence des régions intéressantes. On identifie clairement des régions peu denses en cuivre et peu denses en rectangles. Ces régions sont plus accessibles pour l'implantation de fonctionnalités supplémentaires.

L'algorithme A.1 permet également de vérifier le contact entre deux rectangles donnés. L'algorithme A.2 réalise cette opération.

Algorithme A.2 — Test de contact entre deux rectangles

Entrées : Deux rectangles R et S, munis d'un numéro de niveau en cinquème coordonnée

Sorties : Valeur Booléenne : *vrai* ou *faux*

1 **si** $IntersectionRectangles(R, S) \neq \emptyset$ ***et*** $|R_5 - S_5| \leq 1$ **alors**
2 | **retourner** *vrai*
3 **sinon**
4 | **retourner** *faux*
5 **fin**

Il est important de noter que cet algorithme ne permet pas d'établir de manière triviale la liste d'adjacence[5] complète de tous les rectangles. En effet, tester les contacts de toutes les paires de rectangles demande de $O(N^2)$ opérations ce qui est irréalisable pour des gros fichiers.

Il est cependant possible de dresser la liste de tous les voisins d'un rectangle donné en temps $O(N)$. En exploitant la structure géométrique du fichier nous pouvons même réduire ce temps. Cette astuce utilise la remarque suivante (voir fig. A.7 pour une illustration) :

Si tous les rectangles sont petits, alors il n'est nécessaire de tester le contact entre deux rectangles que si leurs coordonnées ne sont pas trop écartées.

De plus, si les listes de rectangles sont triées par coordonnées en amont, on peut rapidement éliminer les rectangles trop éloignés.

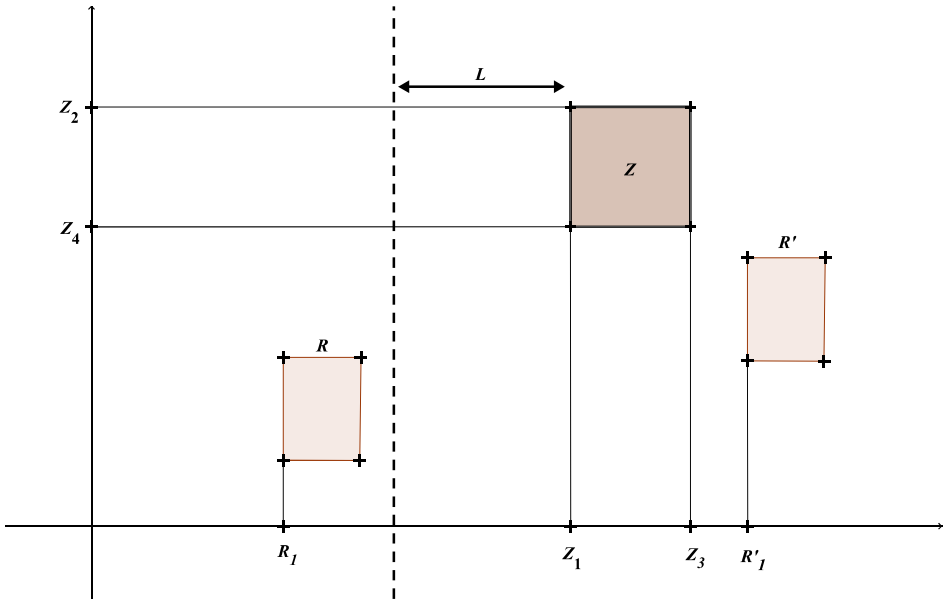


Figure A.7 — Test de contact rapide.

5. i.e. la liste de tous les voisins de chaque rectangle

De manière plus précise, on fixe un rectangle Z dont on veut trouver les voisins (les rectangles en contact) :

- Si tous les rectangles ont une largeur $< L$, alors il n'est pas nécessaire de tester de contact avec les rectangles R pour lesquels $R_1 < Z_1 - L$. Ces rectangles sont situés "trop à gauche".
- De la même manière, les rectangles R pour lesquels $R_1 > Z_3$ sont situés "trop à droite" et ne peuvent donc être en contact avec Z.

Si les rectangles sont triés par leurs premières coordonnées, il est possible de trouver rapidement les rectangles R pour lesquels $Z_1 - L \leq R_1 \leq Z_3$ en procédant par dichotomie. L'algorithme A.3 met ces différentes remarques en oeuvre.

Algorithme A.3 — Voisins rapide

Entrées : Une liste l de rectangles de largeur $\leq L$, triés par première coordonnée et un rectangle R.

Sorties : Ensemble de rectangles M

```
1  M ← {R}
2  S_inf ← R_1 − L
3  S_sup ← R_3
4  i_g ← 1
5  i_d ← |l|
6  tant que i_g < i_d − 1 faire
7  |  i_m ← E((i_g + i_d)/2)
8  |  si L[i_m]_1 < S_inf alors
9  |  |  i_g ← i_m
10 |  sinon
11 |  |  i_d ← i_m
12 |  fin
13 fin
14 i ← i_g
15 tant que L[i]_1 < S_sup faire
16 |  si Contact(R, l[i]) alors
17 |  |  M ← M ∪ {l[i]}
18 |  fin
19 |  i ← i + 1
20 fin
21 retourner M
```

Remarque 22. *Il est évidemment possible d'appliquer ce même raisonnement aux hauteurs des triangles, et d'établir un algorithme similaire pour les hauteurs.*

L'algorithme A.3 travaille sur des listes de rectangles triées, de taille inférieure à L. Le tri des listes, réalisé en amont, peut être réalisé en temps $O(N \log(N))$. Une fois effectué ce tri, on peut identifier les rectangles de taille $> L$ en temps logarithmique en procédant par dichotomie. De plus ce tri n'a besoin d'être effectué qu'une seule fois par fichier de

spécification.

Durant la SEME, nous avons pris la décision d'éliminer les 10% des rectangles les plus larges. L'algorithme accéléré *Voisins* qui donne les voisins d'un rectangle donné procède de la manière suivante :

- tester le contact avec les 10% des rectangles les plus larges,
- utiliser l'algorithme rapide A.3 pour déterminer les petits rectangles voisins,
- renvoyer la liste de voisins obtenus.

L'algorithme correspondant à ces opérations découle immédiatement de cette description.

Notons que l'algorithme *Voisins* s'effectue en temps linéaire vis-à-vis du nombre de rectangles, en raison de la vérification des 10% des rectangles les plus larges. Le nombre de tests de contact vaut environ $0,1 \cdot N + \log(N)$. Le facteur $0,1$ explique pourquoi en pratique cet algorithme est beaucoup plus efficace que la vérification de contact exhaustive avec tous les rectangles.

À ce stade nous pouvons donner, de manière relativement rapide, tous les voisins d'un rectangle donné. Cette information nous renseigne localement sur la connexité du circuit. Il est possible, pour une zone de taille raisonnablement petite, de donner le degré moyen qui nous renseigne sur l'importance de cette zone dans le circuit.

En utilisant l'algorithme *Voisins* de manière répétée, on construit un algorithme simple pour donner les voisins d'ordre n d'un rectangle donné :

Algorithme A.4 — Voisins d'ordre n

Entrées : Une liste de rectangles L, un rectangle R de cette liste, et un entier n.

Sorties : Un ensemble de rectangles O

1 $O \leftarrow \{R\}$
2 $F \leftarrow \{R\}$
3 $G \leftarrow \{R\}$
4 **pour** $i = 2..n$ **faire**
5 $\quad G \leftarrow G \cup \left(\bigcup_{f \in F} Voisins(f)\right)$
6 $\quad F \leftarrow G \setminus O$
7 $\quad O \leftarrow O \cup G$
8 **fin**
9 **retourner** O

L'algorithme A.4 nous renseigne également sur l'importance d'un rectangle du point de vue connectique, mais de manière plus précise, en prolongeant les contact aux voisins d'ordre n.

Avec ces différents algorithmes, nous pouvons construire un indicateur intéressant pour un rectangle donné. Cet indicateur, appelé *Profil*, indique pour des valeurs croissantes de n

- le nombre de voisins à l'ordre n,
- la distance maximale (rayon) aux voisins d'ordre $\leq n$.

Algorithme A.5 — Profil

Entrées : Une liste de rectangles L, un rectangle R de cette liste, et un entier n.

Sorties : Une liste d'entiers v_i et une liste de distances r_i

1 $v_1 \leftarrow 1$
2 $r_1 \leftarrow 0$
3 $O \leftarrow \{R\}$
4 $F \leftarrow \{R\}$
5 $G \leftarrow \{R\}$
6 **pour** $i = 2..n$ **faire**
7 $\quad G \leftarrow G \cup \left[\bigcup_{f \in F} Voisins(f)\right]$
8 $\quad F \leftarrow G \setminus O$
9 $\quad O \leftarrow O \cup G$
10 $\quad v_i \leftarrow |O|$
11 $\quad r_i \leftarrow \max(r_{i-1}, \max(distance(R, f), f \in F))$
12 **fin**
13 **retourner** O

Le profil nous renseigne de manière assez précise sur l'importance d'un rectangle dans le circuit. Il indique à la fois le nombre de connexions d'un rectangle donné à des ordres croissants, mais aussi la portée géographique de ces connexions. Avec ces informations, il est possible de détecter si un rectangle a beaucoup de contacts et si ces contacts sont éloignés, ce qui lui confère une certaine importance dans le circuit.

Afin d'illustrer ces propos, nous dressons le profil d'un connecteur extérieur (fig. A.8) et d'un transistor (fig. A.9).

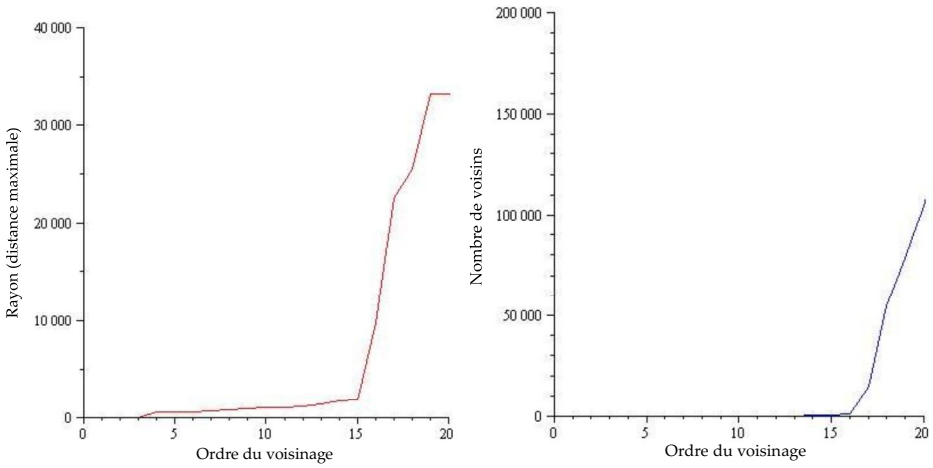


Figure A.8 — Profil d'un connecteur : Rayon (à gauche) et voisins (à droite)

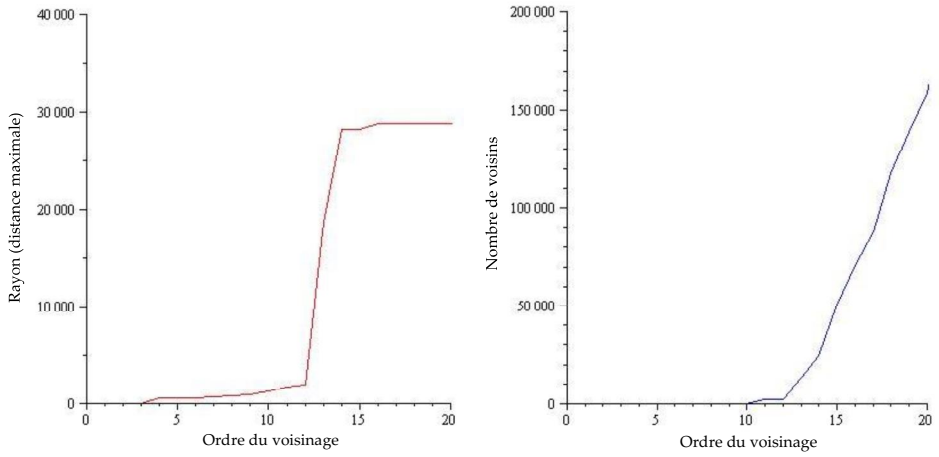


Figure A.9 — Profil d'un transistor : Rayon (à gauche) et voisins (à droite)

Ces figures sont clairement opposées : les voisinage progressif du connecteur se propage lentement , sur une petite distance avec peu de voisins, dans un premier temps, puis croît rapidement sur une petite période.

À l'opposé, les voisinages progressifs du transistor se déplacent rapidement en distance, mais pas en nombre de voisins. Ensuite, le rayon stagne tandis que le nombre de voisins augmente presque linéairement. De manière globale, ce transistor est connecté rapidement sur de longues distances, tandis que le connecteur reste dans un espace relativement confiné.

Pour finir, un objectif que nous n'avions pas envisagé est d'établir le graphe de connexions complet du circuit. Il s'est avéré que ce circuit était suffisamment simple pour parvenir à cette fin [6].

D'une part, nous avons constaté que les deux couches supérieures étaient formées de 40 composantes connexes, ce qui coïncide avec les 40 connecteurs extérieurs. Ces composantes connexes sont représentées en figure A.10.

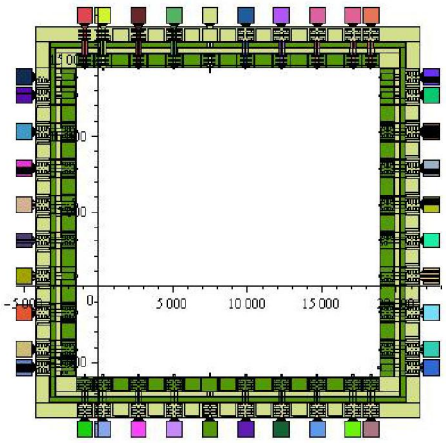


Figure A.10 — Composantes connexes des couches supérieures.

6. Avec un fichier de spécification de 10 Mo contre plus de 100 Go pour des circuits récents.

Ces composantes connexes ont plusieurs aspects. La plupart ont une des formes suivantes

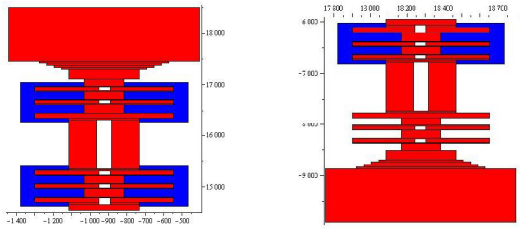
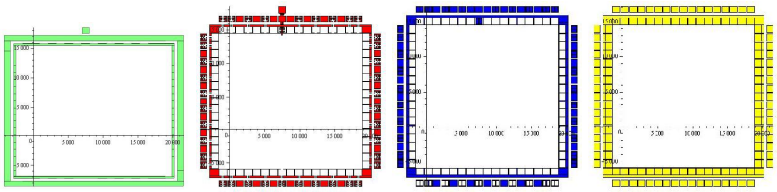


Figure A.11 — Forme courante des composantes connexes.

Deux connecteurs échappent à cette règle, et sont de la forme suivante



Les connecteurs associés à ces composantes connexes jouent un rôle important pour le circuit, car ils sont omniprésents. Il est fort possible qu'il s'agisse de connecteurs d'alimentation ou de masse.

Le schéma logique du circuit est représenté de manière schématique sur la figure A.12.

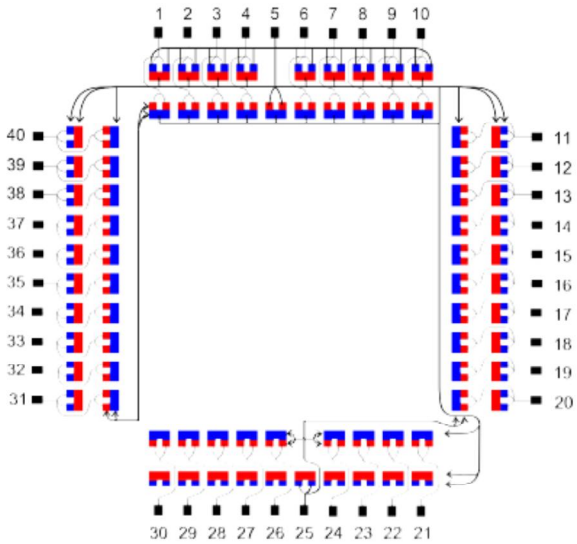


Figure A.12 — Schéma logique du circuit.

Il s'avère donc que pour ce fichier, les méthodes présentées permettent de retrouver le schéma logique. Cette *retro-engineering* ne figure cependant pas parmi les objectifs de ce travail.

A.5 Conclusion

En résumé, nous avons proposé quelques algorithmes simples qui, à partir d'un fichier de spécification, permettent d'étudier de manière locale la structure d'un circuit intégré. Les indicateurs proposés mettent en contraste deux aspects du circuit : sa géométrie et sa connexité. À défaut d'avoir pu essayer ces algorithmes sur des fichiers de taille raisonnable, ils ont donné des résultats satisfaisants sur un circuit simple.

Une piste envisageable, qui n'a pas été explorée faute de temps, est d'utiliser des méthodes de détection de communauté dans les graphes. Ces méthodes permettraient d'identifier les différentes parties en forte interaction. Les interconnexions entre ces "clusters" jouent un rôle important dans le circuit et forment donc une zone vulnérable aux modifications.

Une forte limitation de cette méthode d'exploration locale du circuit est la faible garantie donnée aux vérifications. Comme il est impossible d'explorer le circuit dans son intégralité, on ne peut donner une garantie de certification totale.

Nous n'avons pas eu le temps d'explorer la connexion de ce travail avec les systèmes multi-agent. Les différents indicateurs proposés peuvent s'avérer pertinents pour les systèmes multi-agents, à condition que la topologie du système reste inchangée au cours du temps. Ils donnent un idée locale de l'effet d'un individu sur le système, en termes géométriques et en termes de contacts.

Une perspective intéressante à court terme est d'appliquer les algorithmes à un système multi-agent donné (avec un graphe de communication statique), afin d'établir une hiérarchie entre les agents selon leur importance. Dans un second temps, il serait intéressant de comprendre quel est l'apport de cette hiérarchie sur l'efficacité du système.

De plus, comme nous l'avons mentionné à l'introduction, ces méthodes permettent d'identifier les zones vulnérables aux attaques. Il est possible qu'un système multi-agent soit perturbé localement, par des agent saboteurs ou une interaction bruitée, afin d'en influencer le fonctionnement. Une perspective possible est d'étudier à quel point les méthodes proposées parviennent à détecter une telle attaque sur un système donné.

Bibliographie

[1] Jamal Hussein Abou-Kassem, SM Farouq-Ali, and M Rafiq Islam. *Petroleum Reservoir Simulations*. Elsevier, 2013.

[2] Yves Achdou and Italo Capuzzo-Dolcetta. Mean field games : Numerical methods. *SIAM Journal on Numerical Analysis*, 48(3) :1136–1162, 2010.

[3] Anton Arnold, Peter Markowich, Giuseppe Toscani, and Andreas Unterreiter. *On logarithmic Sobolev inequalities, Csiszar-Kullback inequalities, and the rate of convergence to equilibrium for Fokker-Planck type equations*. Techn. Univ., Fachbereich Mathematik (3), 1998.

[4] W Brian Arthur. Inductive reasoning and bounded rationality. *The American economic review*, pages 406–411, 1994.

[5] François Baccelli, David R. McDonald, and Julien Reynier. A mean-field model for multiple TCP connections through a buffer implementing red. *Performance Evaluation*, 49(1) :77–97, 2002.

[6] June Barrow-Green, Florin Diacu, and Philip Holmes. Poincaré and the three body problem. *additive number theory : the classical bases*, 31 :121, 1999.

[7] George Keith Batchelor. *An introduction to fluid dynamics*. Cambridge university press, 2000.

[8] Fabio Bellifemine, Federico Bergenti, Giovanni Caire, and Agostino Poggi. Jade - a java agent development framework. In *Multi-Agent Programming*, pages 125–147. Springer, 2005.

[9] R. Bellman and R.E. Bellman. *Adaptive Control Processes : A Guided Tour*. Rand Corporation. Research studies. Princeton University Press, 1961.

[10] Michel Benaim and Jean-Yves Le Boudec. On mean field convergence and stationary regime. *arXiv preprint arXiv :1111.5710*, 2011.

[11] Michel Benaim and Jean-Yves Le Boudec. A class of mean field interaction models for computer and communication systems. *Performance Evaluation*, 65(11) :823–838, 2008.

[12] A Benaouda, N Zerhouni, C Varnier, et al. Une approche multi-agents coopératifs pour la gestion des ressources matérielles dans un contexte multi-sites de e-manufacturing. In *6ème Conférence Francophone de MOdélisation et SIMulation, MOSIM'06. Modélisation, Optimisation et Simulation des systèmes : défis et opportunités.*, 2006.

[13] H. Berestycki and F. Hamel. *Reaction-Diffusion Equations and Propagation Phenomena*. Applied Mathematical Sciences : Preliminary Entry Series. Springer verlag GMBH, 2007.

[14] Carole Bernon, Marie-Pierre Gleizes, Sylvain Peyruqueou, and Gauthier Picard. Adelfe : A methodology for adaptive multi-agent systems engineering. In *Engineering Societies in the Agents World III*, pages 156–169. Springer, 2003.

[15] Daniel S Bernstein, Robert Givan, Neil Immerman, and Shlomo Zilberstein. The complexity of decentralized control of Markov decision processes. *Mathematics of operations research*, 27(4) :819–840, 2002.

[16] Dimitri P Bertsekas. *Dynamic programming and optimal control*, volume 1. Athena Scientific Belmont, MA, 1995.

[17] Jérémy Boes, Frédéric Migeon, and François Gatto. Self-Organizing Agents for an Adaptive Control of Heat Engines. In *International Conference on Informatics in Control, Automation and Robotics (ICINCO)*. SciTePress, 2013.

[18] François Bolley and Cédric Villani. Weighted Csiszar-Kullback-Pinsker inequalities and applications to transportation inequalities. *Annales de la faculté des sciences de Toulouse*, 14(3) :331–352, 2005.

[19] Charles Bordenave, Venkat Anantharam, et al. Optimal control of interacting particle systems. 2007.

[20] Charles Bordenave, David McDonald, and Alexandre Proutiere. A particle system in interaction with a rapidly varying environment : Mean field limits and applications. *arXiv preprint math/0701363*, 2007.

[21] Dieter Bothe, Michel Pierre, and Guillaume Rolland. Cross-diffusion limit for a reaction-diffusion system with fast reversible reaction. *Communications in Partial Differential Equations*, 37(11) :1940–1966, 2012.

[22] François Bousquet, Innocent Bakam, Hubert Proton, and Christophe Le Page. Cormas : Common-pool resources and multi-agent systems. In *Tasks and Methods in Applied Artificial Intelligence*, pages 826–837. Springer, 1998.

[23] Craig Boutilier. Sequential optimality and coordination in multiagent systems. In *IJCAI*, volume 99, pages 478–485, 1999.

[24] Pierre Bovet and Simon Benhamou. Spatial analysis of animals' movements using a correlated random walk model. *Journal of theoretical biology*, 131(4) :419–433, 1988.

[25] Paolo Bresciani, Anna Perini, Paolo Giorgini, Fausto Giunchiglia, and John Mylopoulos. Tropos : An agent-oriented software development methodology. *Autonomous Agents and Multi-Agent Systems*, 8(3) :203–236, 2004.

[26] Haïm Brezis. *Analyse fonctionelle*, volume 5. Masson, 1983.

[27] Nicholas F Britton et al. *Reaction-diffusion equations and their applications to biology.* Academic Press, 1986.

[28] Olivier Buffet, Alain Dutech, and François Charpillet. Incremental reinforcement learning for designing multi-agent systems. In *Proceedings of the fifth international conference on Autonomous agents*, pages 31–32. ACM, 2001.

[29] Zoe Bukovac, Alan Dorin, and Adrian Dyer. A-bees see : A simulation to assess social bee visual attention during complex search tasks. In *Advances in Artificial Life, ECAL*, volume 12, pages 276–283, 2013.

[30] B Calderhead, M Girolami, and DJ Higham. Is it safe to go out yet ? statistical inference in a zombie outbreak model. *Preprint, University of Strathclyde, Department of Mathematics and Statistics*, 2010.

[31] Davy Capera, JeanYPierre Georgé, M-P Gleizes, and Pierre Glize. The AMAS theory for complex problem solving based on self-organizing cooperative agents. In *Enabling Technologies : Infrastructure for Collaborative Enterprises, 2003. WET ICE 2003. Proceedings. Twelfth IEEE International Workshops on*, pages 383–388. IEEE, 2003.

[32] José Antonio Carrillo, Stéphane Cordier, and Simona Mancini. A decision-making Fokker–Planck model in computational neuroscience. *Journal of mathematical biology*, 63(5) :801–830, 2011.

[33] CR Carroll and Daniel H Janzen. Ecology of foraging by ants. *Annual Review of Ecology and Systematics*, pages 231–257, 1973.

[34] Yaser Chaaban and Christian Müller-Schloer. A survey of robustness in multi-agent systems. In *COGNITIVE 2013, The Fifth International Conference on Advanced Cognitive Technologies and Applications*, pages 7–13, 2013.

[35] Augustin Chaintreau, Jean-Yves Le Boudec, and Nikodin Ristanovic. The age of gossip : spatial mean field regime. In *Proceedings of the eleventh international joint conference on Measurement and modeling of computer systems*, pages 109–120. ACM, 2009.

[36] Hyeong Soo Chang. Localization and a distributed local optimal solution algorithm for a class of multi-agent Markov decision processes. *International Journal of Control Automation and Systems*, 1 :358–367, 2003.

[37] Hyeong Soo Chang and Michael C Fu. A distributed algorithm for solving a class of multi-agent Markov decision problems. In *42nd IEEE Conference on Decision and Control, 2003.*, volume 5, pages 5341–5346. IEEE, 2003.

[38] Xinfu Chen, Richard Hambrock, and Yuan Lou. Evolution of conditional dispersal : a reaction–diffusion–advection model. *Journal of mathematical biology*, 57(3) :361–386, 2008.

[39] Vincent Chevrier, Nazim Fatès, et al. Multi-agent systems as discrete dynamical systems : influences and reactions as a modelling principle. Technical report, inria-00345954, 2008.

[40] Ruslan K Chornei, Hans Daduna, and Pavel S Knopov. *Control of spatially structured random processes and random fields with applications*, volume 86. Springer, 2006.

[41] I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2 :299–318, 1967.

[42] Felipe Cucker and Steve Smale. Emergent behavior in flocks. *Automatic Control, IEEE Transactions on*, 52(5) :852–862, 2007.

[43] James W Daniel. Splines and efficiency in dynamic programming. *Journal of Mathematical Analysis and Applications*, 54(2) :402–407, 1976.

[44] George Bernard Dantzig. *Linear programming and extensions*. Princeton university press, 1965.

[45] Daniela Pucci de Farias and Benjamin Van Roy. The linear programming approach to approximate dynamic programming. *Operations Research*, 51(6) :850–865, 2003.

[46] J-L Deneubourg, Serge Aron, Simon Goss, and Jacques Marie Pasteels. The self-organizing exploratory pattern of the argentine ant. *Journal of insect behavior*, 3(2) :159–168, 1990.

[47] Jean-Dominique Deuschel and Daniel W Stroock. Hypercontractivity and spectral gap of symmetric diffusions with applications to the stochastic Ising models. *Journal of functional analysis*, 92(1) :30–48, 1990.

[48] Marco Di Francesco, Klemens Fellner, and Peter A Markowich. The entropy dissipation method for spatially inhomogeneous reaction–diffusion-type systems. *Proceedings of the Royal Society A : Mathematical, Physical and Engineering Science*, 464(2100) :3273–3300, 2008.

[49] Giovanna Di Marzo Serugendo, Marie-Pierre Gleizes, and Anthony Karageorgos. Self-organization in multi-agent systems. *The Knowledge Engineering Review*, 20(02) :165–189, 2005.

[50] Edsger Wybe Dijkstra. *A short introduction to the art of programming*, volume 4. Technische Hogeschool Eindhoven, 1971.

[51] Robert J Elliott, Lakhdar Aggoun, and John B Moore. *Hidden Markov Models*. Springer, 1994.

[52] Michel Émery and Joseph E Yukich. A simple proof of the logarithmic Sobolev inequality on the circle. In *Séminaire de Probabilités XXI*, pages 173–175. Springer, 1987.

[53] Lawrence C Evans. Partial differential equations : Graduate studies in mathematics. *American Mathematical Society*, 2, 1998.

[54] S. E. Fadugba, Edogbanya O. H, and S. C Zelibe. Crank-Nicolson method for solving parabolic partial differential equations. *International Journal of Applied Mathematics and Modeling IJA2M*, 2013.

[55] Samia Faruq, Peter W McOwan, and Lars Chittka. The biological significance of color constancy : An agent-based model with bees foraging from flowers under varied illumination. *Journal of vision*, 13(10), 2013.

[56] G.T. Fechner. *Elemente der Psychophysik*. Number Bd. 1 in Elemente der Psychophysik. Breitkopf und Härtel, 1860.

[57] Eugene A Feinberg and Fenghsu Yang. On polynomial cases of the unichain classification problem for markov decision processes. *Operations Research Letters*, 36(5) :527–530, 2008.

[58] Jacques Ferber. *Multi-agent systems : an introduction to distributed artificial intelligence*. Addison-Wesley Reading, 1999.

[59] Adolf Fick. Ueber diffusion. *Annalen der Physik*, 170(1) :59–86, 1855.

[60] Joseph Fourier. *Theorie analytique de la chaleur, par M. Fourier*. Chez Firmin Didot, père et fils, 1822.

[61] Avner Friedman. Partial differential equations of parabolic type, 1964. *Holt, Reinhart, and Winston Inc., New York*.

[62] Nicolas Gast. Optimization and control of large systems : Fighting the curse of dimensionality. 2010.

[63] Nicolas Gast and Bruno Gaujal. A mean field approach for optimization in particle systems and applications. In *Proceedings of the Fourth International ICST Conference on Performance Evaluation Methodologies and Tools*, page 39. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009.

[64] Jean-Pierre Georgé, Marie Pierre Gleizes, and Pierre Glize. Conception de systèmes adaptatifs à fonctionnalité émergente : la théorie AMAS. *Revue d'Intelligence Artificielle*, 17(4) :591–626, 2003.

[65] Olivier Gies and Brahim Chaib-draa. Apprentissage de la coordination multiagent : une méthode basée sur le Q-learning par jeu adaptatif. *Revue d'Intelligence Artificielle*, 20(2-3) :383–410, 2006.

[66] Claudia V Goldman and Shlomo Zilberstein. Decentralized control of cooperative systems : Categorization and complexity analysis. *J. Artif. Intell. Res.(JAIR)*, 22 :143–174, 2004.

[67] François Golse. The mean-field limit for the dynamics of large particle systems. *Journées équations aux dérivées partielles*, 9 :1–47, 2003.

[68] Simon Goss, Serge Aron, Jean-Louis Deneubourg, and Jacques Marie Pasteels. Self-organized shortcuts in the argentine ant. *Naturwissenschaften*, 76(12) :579–581, 1989.

[69] Pierre Grassé. La reconstruction du nid et les coordinations interindividuelles chez-bellicositermes natalensis etcubitermes sp. la théorie de la stigmergie : Essai d'interprétation du comportement des termites constructeurs. *Insectes sociaux*, 6(1) :41–80, 1959.

[70] Frederic Green. NP-complete problems in cellular automata. *Complex Systems*, 1(3) :453–474, 1987.

[71] Andreas Greven, Gerhard Keller, and Gerald Warnecke. *Entropy*. Princeton university press, 2003.

[72] Leonard Gross. Logarithmic Sobolev inequalities. *American Journal of Mathematics*, pages 1061–1083, 1975.

[73] Olivier Guéant, Jean-Michel Lasry, and Pierre-Louis Lions. Mean field games and applications. In *Paris-Princeton Lectures on Mathematical Finance 2010*, pages 205–266. Springer, 2011.

[74] Carlos Guestrin, Daphne Koller, and Ronald Parr. Multiagent planning with factored MDPs. In *NIPS*, volume 1, pages 1523–1530, 2001.

[75] Laszlo Gulyas, Laszlo Laufer, and Richard Szabo. Measuring stigmergy : The case of foraging ants. In SvenA. Brueckner, Salima Hassas, MÃ¡rk Jelasity, and Daniel Yamins, editors, *Engineering Self-Organising Systems*, volume 4335 of *Lecture Notes in Computer Science*, pages 50–65. Springer Berlin Heidelberg, 2007.

[76] Malte Henkel. Sur la solution de Sundman du problème des trois corps. *Philosophia scientiae*, 5(2) :161–184, 2001.

[77] Francis Heylighen, Johan Bollen, and Alexander Calderhead. The evolution of complexity. 1999.

[78] M. B. Horowitz, A. Damle, and J. W. Burdick. Linear Hamilton Jacobi Bellman Equations in High Dimensions. *ArXiv e-prints*, April 2014.

[79] Auke Jan Ijspeert, Alcherio Martinoli, Aude Billard, and Luca Maria Gambardella. Collaboration through the exploitation of local interactions in autonomous collective robotics : The stick pulling experiment. *Autonomous Robots*, 11(2) :149–171, 2001.

[80] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4) :620, 1957.

[81] Chang-hua Jiang, Wei Han, and You-hua Hu. REPAST-a multi-agent simulation platform. *Journal of System Simulation*, 8(8) :2319–2322, 2006.

[82] Tom Jorquera, Jean-Pierre Georgé, Marie-Pierre Gleizes, and Christine Régis. A Natural Formalism and a MultiAgent Algorithm for Integrative Multidisciplinary Design Optimization. In *IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT)*. IEEE Computer Society, 2013.

[83] Ansgar Jüngel. Diffusive and nondiffusive population models. In *Mathematical Modeling of Collective Behavior in Socio-Economic and Life Sciences*, pages 397–425. Springer, 2010.

[84] Narendra Karmarkar. A new polynomial-time algorithm for linear programming. In *Proceedings of the sixteenth annual ACM symposium on Theory of computing*, pages 302–311. ACM, 1984.

[85] Donald E Kirk. *Optimal control theory : an introduction*. Courier Dover Publications, 2012.

[86] Victor Klee and George J Minty. How good is the simplex algorithm. Technical report, DTIC Document, 1970.

[87] Daphne Koller and Ronald Parr. Computing factored value functions for policies in structured mdps. In *In Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, pages 1332–1339. Morgan Kaufmann, 1999.

[88] S. Kullback. A lower bound for discrimination information in terms of variation (corresp.). *IEEE Trans. Inf. Theor.*, 13(1) :126–127, September 2006.

[89] Thomas G Kurtz. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes and Their Applications*, 6(3) :223–240, 1978.

[90] Olga Aleksandrovna Ladyzhenskaëìa, NN Ural'Ceva, and VA Solonnikov. *Linear and quasi-linear equations of parabolic type*. American Mathematical Society, 1968.

[91] Oscar E Lanford III. The hard sphere gas in the Boltzmann-Grad limit. *Physica A : Statistical Mechanics and its Applications*, 106(1) :70–76, 1981.

[92] Jean-Michel Lasry and Pierre-Louis Lions. Mean field games. *Japanese Journal of Mathematics*, 2(1) :229–260, 2007.

[93] J-Y Le Boudec, David McDonald, and Jochen Mundinger. A generic mean field convergence result for systems of interacting objects. In *Quantitative Evaluation of Systems, 2007. QEST 2007. Fourth International Conference on the*, pages 3–18. IEEE, 2007.

[94] Claude Le Bris and Pierre-Louis Lions. Existence and Uniqueness of Solutions to Fokker-Planck Type Equations with Irregular Coefficients. *Communications in Partial Differential Equations*, 33(7) :1272–1317, 2008.

[95] Michel Ledoux. *The concentration of measure phenomenon*, volume 89. American Mathematical Soc., 2005.

[96] Julien Lepagnot, Amir Nakib, Hamouche Oulhadj, and Patrick Siarry. A new multiagent algorithm for dynamic continuous optimization. *International Journal of Applied Metaheuristic Computing (IJAMC)*, 1, 2010.

[97] Thomas Lepoutre, Michel Pierre, and Guillaume Rolland. Global well-posedness of a conservative relaxed cross diffusion system. *SIAM Journal on Mathematical Analysis*, 44(3) :1674–1693, 2012.

[98] Kristina Lerman and Aram Galstyan. A general methodology for mathematical analysis of multi-agent systems. *ISI-TR-529, USC Information Sciences Institute, Marina del Rey, CA*, 2001.

[99] Tullio Levi-Civita. Sur la résolution qualitative du problème restreint des trois corps. *Acta Mathematica*, 30(1) :305–327, 1906.

[100] Na Li and Jason R Marden. Designing games for distributed optimization. *Selected Topics in Signal Processing, IEEE Journal of*, 7(2) :230–242, 2013.

[101] G.M. Lieberman. *Second Order Parabolic Differential Equations*. World Scientific, 1996.

[102] Thomas M Liggett. Particle systems. 1985.

[103] Michael L Littman, Thomas L Dean, and Leslie Pack Kaelbling. On the complexity of solving Markov decision problems. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pages 394–402. Morgan Kaufmann Publishers Inc., 1995.

[104] Jiming Liu, Han Jing, and Y.Y. Tang. Multi-agent oriented constraint satisfaction. *Artificial Intelligence*, 136, 2002.

[105] Irene Marcovici. *Automates cellulaires probabilistes et mesures spécifiques sur des espaces symboliques*. PhD thesis, Université Paris-Diderot-Paris VII, 2013.

[106] Jason R Marden, Gürdal Arslan, and Jeff S Shamma. Cooperative control and potential games. *Systems, Man, and Cybernetics, Part B : Cybernetics, IEEE Transactions on*, 39(6) :1393–1407, 2009.

[107] Peter A Markowich and Cédric Villani. On the trend to equilibrium for the Fokker-Planck equation : an interplay between physics and functional analysis. *Mat. Contemp*, 19 :1–29, 2000.

[108] Sergio Cesare Masin, Verena Zudini, and Mauro Antonelli. Early alternative derivations of Fechner's law. *Journal of the History of the Behavioral Sciences*, 45(1) :56–65, 2009.

[109] Melanie Mitchell. *Complexity : A guided tour*. Oxford University Press, 2009.

[110] James R Munkres. *Topology : a first course*, volume 23. Prentice-Hall Englewood Cliffs, NJ, 1975.

[111] James D Murray. Mathematical biology i : An introduction, vol. 17 of interdisciplinary applied mathematics, 2002.

[112] Roger B Myerson. *Game theory*. Harvard university press, 2013.

[113] Jacques Neveu. *Martingales à temps discret*. Masson, 1972.

[114] Cynthia Nikolai and Gregory Madey. Tools of the trade : A survey of various agent based modeling platforms. *Journal of Artificial Societies and Social Simulation*, 12(2) :2, 2009.

[115] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V Vazirani. *Algorithmic game theory*. Cambridge University Press, 2007.

[116] Frans A Oliehoek. Decentralized POMDPs. In *Reinforcement Learning*, pages 471–503. Springer, 2012.

[117] Frans A Oliehoek, Matthijs TJ Spaan, and Nikos A Vlassis. Optimal and approximate q-value functions for decentralized POMDPs. *J. Artif. Intell. Res.(JAIR)*, 32 :289–353, 2008.

[118] Manfred Opper and David Saad. *Advanced mean field methods : Theory and practice*. MIT press, 2001.

[119] Christos H Papadimitriou. *Computational complexity*. John Wiley and Sons Ltd., 2003.

[120] Gauthier Picard, Carole Bernon, and Marie-Pierre Gleizes. Etto : Emergent timetabling by cooperative self-organization. In *Engineering Self-Organising Systems*, pages 31–45. Springer, 2006.

[121] Gauthier Picard and Pierre Glize. Modélisation et expérimentations d'une décision locale basÃ©e sur l'auto-organisation coopérative . In *Journées Francophones sur les Systèmes Multi-Agents (JFSMA'05) , Calais, France, 23/11/2005-25/11/2005*, 2005.

[122] S. Pinsker. *Information and information stability of random variables and processes*. Holden-Day series in time series analysis. Holden-Day, 1964.

[123] Martin L Puterman. *Markov decision processes : discrete stochastic dynamic programming*, volume 414. John Wiley & Sons, 2009.

[124] David V Pynadath and Milind Tambe. The communicative multiagent team decision problem : Analyzing teamwork theories and models. *Journal of Artificial Intelligence Research*, 16 :389–423, 2002.

[125] Julia Radoszycki, Nathalie Peyrard, and Régis Sabbadin. Finding good stochastic factored policies for factored Markov decision processes. In *ECAI 2014 - 21st European Conference on Artificial Intelligence*, pages 1083–1084, 2014.

[126] Linda E Reichl and Ilya Prigogine. *A modern course in statistical physics*, volume 186. University of Texas press Austin, 1980.

[127] Hannes Risken. *Fokker-Planck Equation*. Springer, 1984.

[128] Guillaume Rolland. *Global existence and fast-reaction limit in reaction-diffusion systems with cross effects*. Thè, École normale supérieure de Cachan - ENS Cachan ; Technische Universität (Darmstadt, Allemagne), 2012.

[129] Archie E Roy et al. *Predictability, stability, and chaos in N-body dynamical systems*. Plenum Pub. Corp., 1991.

[130] Stuart Russell and Peter Norvig. A modern approach. *Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs*, 25, 1995.

[131] Régis Sabbadin, Nathalie Peyrard, and Nicklas Forsell. A framework and a mean-field algorithm for the local control of spatial processes. *International Journal of Approximate Reasoning*, 53(1) :66–86, 2012.

[132] Michael Schillo, Hans-Jürgen Bürckert, Klaus Fischer, and Matthias Klusch. Towards a definition of robustness for market-style open multi-agent systems. In *Proceedings of the fifth international conference on Autonomous agents*, pages 75–76. ACM, 2001.

[133] Paul J Schweitzer and Abraham Seidmann. Generalized polynomial approximations in markovian decision processes. *Journal of mathematical analysis and applications*, 110(2) :568–582, 1985.

[134] Giovanna di Marzo Serugendo, Marie Pierre Gleizes, and Anthony Karageorgos. Self-organisation and emergence in MAS : An overview. *Informatica (Slovenia)*, 30(1) :45–54, 2006.

[135] Yoav Shoham and Kevin Leyton-Brown. *Multiagent systems : Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, 2009.

[136] Olivier Sigaud and Olivier Buffet. *Processus décisionnels de Markov en intelligence artificielle*, volume 1. Lavoisier - Hermes Science Publications, 2008.

[137] Antoine Spicher, Nazim Fatès, Olivier Simonin, et al. From reactive multi-agent models to cellular automata-illustration on a diffusion-limited aggregation model. *Proceedings of ICAART'09*, pages 422–429, 2009.

[138] Ian Stewart. The ultimate in anty-particles. *Scientific American*, 271(1) :104–107, 1994.

[139] Alain-Sol Sznitman. Topics in propagation of chaos. In *Ecole d'Eté de Probabilités de Saint-Flour XIX-1989*, pages 165–251. Springer, 1991.

[140] Patrick Taillandier, Duc-An Vo, Edouard Amouroux, and Alexis Drogoul. Gama : a simulation platform that integrates geographical information data, agent-based modeling and multi-scale control. In *Principles and Practice of Multi-Agent Systems*, pages 242–258. Springer, 2012.

[141] Seth Tisue and Uri Wilensky. Netlogo : A simple environment for modeling complexity. In *International Conference on Complex Systems*, pages 16–21, 2004.

[142] Xavier Topin, Vincent Fourcassié, Marie-Pierre Gleizes, Guy Theraulaz, and Christine Régis. Theories and Experiments on Emergent Behaviour : From Natural to Artificial Systems and Back. In *European Conference on Cognitive Science, Siena, 27/10/1999-30/10/1999*.

[143] John N Tsitsiklis. NP-hardness of checking the unichain condition in average cost MDPs. *Operations research letters*, 35(3) :319–323, 2007.

[144] Alan Mathison Turing. The chemical basis of morphogenesis. *Bulletin of mathematical biology*, 52(1) :153–197, 1990.

[145] Juan Luis Vázquez. *The porous medium equation : mathematical theory*. Oxford University Press, 2007.

[146] Cédric Villani. *Optimal transport : old and new*, volume 338. Springer, 2008.

[147] Hong Wang, Richard E Ewing, Guan Qin, Stephen L Lyons, Mohamed Al-Lawatia, and Shushuang Man. A family of Eulerian–Lagrangian localized adjoint methods

for multi-dimensional advection-reaction equations. *Journal of Computational Physics*, 152(1) :120–163, 1999.

[148] Warren Weaver. Science and complexity. In *Facets of Systems Science*, pages 449–456. Springer, 1991.

[149] Rudiger Wehner. Desert ant navigation : how miniature brains solve complex tasks. *Journal of Comparative Physiology A*, 189(8) :579–588, 2003.

[150] G. Weiss. *Aspects and Applications of the Random Walk (Random Materials & Processes S.)*. North-Holland, 1994.

[151] Gerhard Weiss. *Multiagent systems : a modern approach to distributed artificial intelligence*. MIT press, 1999.

[152] Danny Weyns, H Van Dyke Parunak, Fabien Michel, Tom Holvoet, and Jacques Ferber. Environments for multiagent systems state-of-the-art and research challenges. In *Environments for multi-agent systems*, pages 1–47. Springer, 2005.

[153] Uri Wilensky. Netlogo ants model. *Center for Connected Learning and Computer-Based Modeling. Northwestern University, Evanston*, 1997.

[154] Stephen Wolfram. Universality and complexity in cellular automata. *Physica D : Nonlinear Phenomena*, 10(1) :1–35, 1984.

[155] Michael Woolridge and Nicholas R Jennings. Intelligent agents : Theory and practice. *Knowledge Engineering Review*, 10(2) :115–152, 1995.

[156] Yinyu Ye. The simplex and policy-iteration methods are strongly polynomial for the Markov decision problem with a fixed discount rate. *Mathematics of Operations Research*, 36(4) :593–603, 2011.