HAL
open science

Auto-optimisation des réseaux sans fil. Une approche par la théorie des jeux

Pierre Coucheney

To cite this version:

Pierre Coucheney. Auto-optimisation des réseaux sans fil. Une approche par la théorie des jeux. Autre [cs.OH]. Université de Grenoble, 2011. Français. NNT : 2011GRENM031 . tel-00647296

HAL Id: tel-00647296

https://theses.hal.science/tel-00647296v1

Submitted on 1 Dec 2011

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

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

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Informatique**

Arrêté ministériel : 7 août 2006

Présentée par

Pierre Coucheney

Thèse dirigée par **M. Bruno Gaujal**
et codirigée par **Mlle. Corinne Touati**

préparée au sein **du Laboratoire d'Informatique de Grenoble**
et de **l'École Doctorale Mathématiques, Sciences et Technologies de l'Information**

Auto-optimisation des réseaux sans fil : une approche par la théorie des jeux

Thèse soutenue publiquement le **31 Août 2011**,
devant le jury composé de :

M. Olivier Bournez
Professeur, Polytechnique, Rapporteur
M. Bruno Tuffin
Chargé de recherche, INRIA, Rapporteur
M. Mérouane Debbah
Professeur, Supélec, Examinatrice
M. Yannick Viossat
Maitre de conférence, Université Paris Dauphine, Examinateur
M. Denis Trystram
Professeur, Université de Grenoble, Examinateur
M. Laurent Roullet
Chercheur, Bell Labs, Examinateur
M. Bruno Gaujal
Directeur de recherche, INRIA, Directeur de thèse
Melle. Corinne Touati
Chargé de Recherche, INRIA, Co-Directeur de thèse

Remerciements

L'aboutissement de la thèse doit beaucoup à plusieurs personnes que je tiens à remercier, en commençant par mes deux directeurs de thèse : Bruno Gaujal et Corinne Touati. Je mesure encore plus aujourd'hui la chance que j'ai eu de les rencontrer, et je les remercie sincèrement pour la confiance qu'ils m'ont accordée en me proposant ce sujet de thèse. Leur encadrement a été d'une très grande qualité que ce soit par leur disponibilité que par leur goût pour la science qu'ils m'ont fait partager. Bien que leurs caractères soient différents, ils partagent une curiosité naturelle dans des domaines très variés. C'est cette curiosité qui a été le moteur de ces trois années exceptionnelles.

Je remercie tous les membres du jury, et particulièrement Bruno Tuffin et Olivier Bournez pour avoir accepté de rapporter mon manuscript. Leurs commentaires, ainsi que ceux des autres membres, et notamment Yannick Viossat, ont largement contribué à améliorer la qualité du document.

J'ai effectué plusieurs séjours au sein des Bell Labs à Villarceaux. Le travail y a été stimulant, et également très agréable au sein de l'équipe dirigée par Yacine El Mghazli. Je remercie les personnes avec qui j'ai collaboré pour leur investissement, en particulier Barbara Orlandi qui a permis d'aboutir à une mise œuvre des algorithmes. Laurent Roullet m'a également beaucoup appris sur les réseaux sans fil.

Merci à tous les membres des équipes Mescal et Moais pour l'ambiance de travail et les discussions passionantes que ce soit lors des groupes de travail qu'à la cafet, lieu central de la vie du bâtiment. Merci à Nicolas et Rémi pour m'avoir fait découvrir les joies du vélo dans la magnifique région Grenobloise.

Je remercie les chercheurs avec qui j'ai eu le plaisir de travailler durant ces trois années, et qui m'ont beaucoup appris. En particulier Emmanuel Hyon, Jean-Marc Kélif, et Alexandre Laugier.

En arrivant à Grenoble pour la thèse, j'ai eu la plaisir d'y rencontrer des personnes formidables qui m'ont beaucoup apporté humainement, et soutenu dans les périodes difficiles. Cette thèse est, de fait, liée à ces amitiés.

Quelques mots sont vains afin d'exprimer ma reconnaissance pour ma compagne, Julie, qui a su éclairer par sa présence ces années à Grenoble, et qui m'a constamment soutenu dans mon travail.

TABLE DES MATIÈRES

Remerciements **3**

1 Introduction **9**
1.1 Contexte de la thèse : le laboratoire commun INRIA / Alcatel-Lucent Bell Labs . . . 9
1.2 Optimisation des réseaux sans fil . . . 10
1.2.1 Les réseaux de communication sans fil . . . 10
1.2.2 Contrôle des réseaux sans fil . . . 12
1.2.3 Objectifs de la thèse . . . 13
1.3 Organisation de la thèse et contributions . . . 14

2 Mécanismes d'incitation entre plusieurs entités indépendantes **17**
2.1 Présentation générale du problème d'incitation . . . 18
2.1.1 Modèle et notations . . . 18
2.1.2 Formulation du problème d'incitation . . . 21
2.1.3 Exemples . . . 22
2.2 Une approche en stratégies dominantes . . . 24
2.2.1 Description de l'approche . . . 24
2.2.2 Les enchères généralisées . . . 25
2.2.3 Limite de l'incitation en stratégies dominantes . . . 26
2.3 Une approche par fonction de potentiel . . . 27
2.3.1 Description de l'approche . . . 28
2.3.2 Construction d'un mécanisme d'incitation par fonction de potentiel dans les jeux finis . . . 30
2.3.3 Application à des problème d'allocation de ressources . . . 30
2.4 Un exemple où les deux approches échouent : inciter des joueurs à participer à une coalition . . . 32

2.4.1 Présentation du problème 32
2.4.2 Mécanisme d'incitation en stratégies dominantes 34
2.4.3 Mécanisme d'incitation par fonction de potentiel 34

3 Jeux de potentiel et modèles d'apprentissage 37
3.1 Définitions et résultats généraux sur les jeux 38
3.1.1 Jeux finis 38
3.1.2 Jeux continus et extension mixte des jeux finis 39
3.2 Jeux de potentiel 47
3.2.1 Fonction de potentiel dans les jeux finis 47
3.2.2 Fonction de potentiel dans les jeux continus 49
3.3 L'apprentissage dans les jeux 55
3.3.1 Quantifier le coût de l'apprentissage 57
3.3.2 Robustesse des résultats 58
3.4 L'algorithme de meilleure réponse 62
3.5 L'apprentissage par la règle du jeu fictif 69

4 Le modèle stochastique de meilleure réponse 73
4.1 Algorithme stochastique de meilleure réponse asynchrone 74
4.1.1 Interprétations de l'algorithme stochastique de meilleure réponse à partir de modèles d'apprentissage 75
4.1.2 Analyse de l'algorithme dans les jeux de potentiel 76
4.2 Utilisation de l'algorithme stochastique de meilleure réponse pour l'optimisation du routage dans les réseaux ad hoc de mobiles 79
4.2.1 Modélisation du réseau ad hoc 80
4.2.2 Implémentation de l'algorithme stochastique de meilleure réponse . 83
4.2.3 Étude numérique 85
4.3 Robustesse de la dynamique stochastique de meilleure réponse aux processus de révision des stratégies 87
4.3.1 Caractérisation des états stochastiquement stables pour les processus de révision généraux 90
4.3.2 Convergence vers les équilibres de Nash 95
4.3.3 Contre-exemples sur la sélection des équilibres optimaux dans les jeux de potentiel 97

5 Extension mixte du modèle de meilleure réponse 101
5.1 Dynamique de meilleure réponse dans l'extension mixte des jeux finis . . . 102
5.1.1 Construction d'une métrique qui garantit l'existence de solutions . . 102
5.1.2 Dynamique de meilleure réponse 105
5.1.3 Propriétés de la dynamique de meilleure réponse 109
5.1.4 Convergence dans les jeux de potentiel 117
5.2 Implémentation de la dynamique de meilleure réponse 120
5.2.1 Résultats généraux sur les approximation stochastiques 121

5.2.2 Approximation stochastique de la dynamique de meilleure réponse . 124
5.2.3 Prise en compte du processus de révision et de l'incertitude sur les gains dans l'approximation stochastique 129
5.2.4 Convergence de l'approximation stochastique dans les jeux de potentiel 130
5.3 Application au problème d'association de mobiles à des réseaux hétérogènes 138
5.3.1 Présentation du problème général 138
5.3.2 Implémentation de l'algorithme d'association des mobiles aux cellules 140
5.3.3 Simulation de l'algorithme 144

Conclusion et extensions **155**

Bibliographie **161**

CHAPITRE 1

INTRODUCTION

1.1 Contexte de la thèse : le laboratoire commun INRIA / Alcatel-Lucent Bell Labs

La thèse présentée dans ce document s'est déroulée dans le cadre d'un laboratoire commun entre l'INRIA et Alcatel Lucent, dont les Bell Labs sont la division de recherche et de développement. Alcatel Lucent est une entreprise dont l'un des cœurs de métier est le développement et la construction d'infrastructures de télécommunication. Dans le domaine des télécommunications, Alcatel Lucent est l'un des principaux constructeurs de réseaux sans fil, et participe activement, à travers les organismes de normalisation, à leur développement.

Le laboratoire commun entre l'INRIA et Alcatel Lucent est articulé autour de trois axes de recherche recouvrant chacun plusieurs sujets de recherche et plusieurs thèses. Le travail présenté dans ce document s'intègre dans l'action de recherche "Selfnets" (Self-optimized mobile cellular networks), dont l'objectif premier est de développer des méthodes et des algorithmes distribués, c'est-à-dire qui reposent sur des information et des décisions locales, dans les réseaux sans fil afin d'optimiser l'utilisation globale de ces réseaux.

Actuellement, les technologies sans fil se complexifient sans cesse, et la gestion des paramètres devient une tache de plus en plus lourde. L'auto-optimisation vise à automatiser le choix des paramètres en fonction de l'évolution de l'environnement. De plus, l'augmentation soutenue des volumes de données qui passent par les réseaux sans fil tend à les saturer. Il est donc nécessaire d'établir des stratégies afin de, par exemple, répartir la charge sur les différents points d'accès sans fil. Actuellement, cela est réalisé par des politiques statiques dans lesquelles les mobiles se connectent prioritairement via des points d'accès Wifi. Cela ne tient compte ni de la qualité de service requise par les mobiles, ni de l'état global, qui change au cours du temps, des autres réseaux auxquels le mobile peut se connecter.

Cette thèse s'inscrit dans cette problématique d'équilibrage de charge dynamique, et

automatique des mobiles sur les points d'accès sans fil, et de manière plus générale, dans les problématiques liées au routage dans les réseaux.

1.2 Optimisation des réseaux sans fil

La spécificité des réseaux sans fil réside non seulement dans le support des communications, les ondes électromagnétiques, dont la gestion des ressources est critique pour les performances, mais également dans la mobilité des usagers que les communications sans fil rendent possible. Il y a donc plusieurs niveaux d'optimisation.

1.2.1 Les réseaux de communication sans fil

Le terme "réseau sans fil" regroupe l'ensemble des réseaux dont *une partie* au moins des communications est assurée par des liaisons radio [1]. Les communications sans fil présentent plusieurs avantages par rapport aux communications filaires. D'une part, elles autorisent la mobilité des utilisateurs, et d'autre part, leur infrastructure est beaucoup plus légère et rapide à déployer. Mais leur capacité est généralement inférieure à celle des réseaux filaires. De ce fait, les liaisons sans fil constituent les points critiques du réseau, c'est-à-dire que ce sont les liaisons qui limitent le débit des communications. Elles constituent ce qu'on appelle les *goulots d'étranglement* du réseau, et, de leur bonne gestion dépend les performances du réseau dans sa globalité.

Le réseau sans fil le plus connu est certainement le réseau de téléphonie mobile, également appelé *réseau cellulaire*. Il regroupe une grande variété de technologies comme GSM, UMTS, LTE. Mais, de plus en plus, les réseaux cellulaires sont utilisés pour des applications autrefois réservées aux communications filaires, parmi lesquelles on peut citer les applications du web, le téléchargement de fichiers, le streaming vidéo. Il en résulte une augmentation croissante des débits observés sur ces réseaux depuis ces dernières années. Dans les réseaux cellulaires, la partie sans fil sert principalement à relier les usagers au coeur du réseau, qui est filaire, par l'intermédiaire d'une station de base, que nous appelons de façon générique un point d'accès [2].

Les communications sans fil servent également à l'établissement de réseaux, reposant exclusivement sur des liaisons sans fil, qui n'ont aucune infrastructure préalable, et qui autorisent la topologie à changer au cours du temps. C'est ce que l'on appelle les *réseaux ad hoc de mobiles*, qui sont utilisés dans de nombreuses applications : ils permettent notamment de reconstituer rapidement des communications après une catastrophe naturelle.

1. En fait, peu de réseaux de communications sans fil sont constitués exclusivement de liaisons radio.
2. Nous appelons "point d'accès" toute antenne, que ce soit pour le Wifi ou les réseaux cellulaires, qui permet de connecter un mobile au reste du réseau.

Les communications sans fil

Les ondes électromagnétiques sont le support des communications sans fil. Contrairement aux liaisons filaires, les ondes électromagnétiques se propagent dans toutes les directions s'il n'y a pas d'obstacles. En fonction de l'environnement, l'onde subit plusieurs altérations dues principalement aux phénomènes de diffraction, de réflexion et d'atténuation. Il en résulte une grande variabilité du signal au niveau d'un récepteur, même si ce dernier a une position géographique fixe. Le phénomène associé aux variations rapides du signal porte le nom de *fading*, alors que celui associé aux variations lentes est appelé *shadowing*, chacun résultant d'un phénomène physique différent.

Plusieurs ressources interviennent de façon critique dans les communications sans fil :

- Les fréquences employées : en plus de la variabilité du signal reçu d'une antenne, plusieurs antennes peuvent interférer si elles émettent sur des fréquences proches. La capacité des réseaux sans fil est donc limité par l'ensemble des fréquences utilisables.
- L'énergie des mobiles : les communications sans fil nécessitent une énergie plus importante que les communications filaires. Or les mobiles sont généralement de petits appareils dont la batterie a une capacité limitée. Cette énergie doit donc être utilisée à bon escient.
- Les ressources du point d'accès : en fonction de la technologie de multiplexage employée, les ressources (fréquences, temps, ou codes) sont partagées entre les différents mobiles connectés au même point d'accès. Il faut noter que, contrairement aux liaisons filaires, le partage des ressources dégrade les performances globales. Par exemple, le débit d'un point d'accès est une fonction qui est largement sous-additive en fonction du nombre d'utilisateurs.

La mobilité

La possibilité de maintenir une communication tout en se déplaçant, est l'un des principaux avantages des communications sans fil. Mais cela représente également une source de difficultés pour la gestion des réseaux.

On distingue en fait deux types de mobilité. D'une part, il y a les mobiles qui se déplacent, et d'autre part, ceux qui initient ou terminent une communication. La mobilité est intrinsèquement un phénomène aléatoire qui vient s'ajouter, mais sur une échelle de temps plus large, aux fluctuations aléatoires du signal des communications sans fil. La modélisation de la mobilité est en soi-même un sujet difficile qui fait partie des axes de recherche du laboratoire commun.

De manière générale, la modélisation fine des communications sans fil est très complexe. En plus de la mobilité, il est nécessaire de modéliser le système physique, en particulier la propagation du signal dans un environnement donné, mais également les interactions entre les différentes couches de protocole utilisées pour les communications. À cela s'ajoute la modélisation des différents types d'applications supportées par les communications (téléphonie, téléchargement de données...).

Dans cette thèse, nous évaluons la qualité des solutions que nous proposons par des sim-

ulations sur des modèles simples qui nous permettent d'analyser la sensibilité des résultats à certains paramètres. Néanmoins, un travail est en cours actuellement pour étendre ces tests sur des prototypes et des systèmes réels en collaboration avec des chercheurs des Bell Labs. Nous ne présentons pas le détail de ces implémentations dans ce document, car elles reposent sur des informations confidentielles.

1.2.2 Contrôle des réseaux sans fil

La connexion d'un mobile à un point d'accès au réseau implique d'une part l'usager qui décide d'établir une connexion, et d'autre part l'opérateur du point d'accès qui gère ensuite la communication. L'opérateur dispose de différents moyens d'actions pour satisfaire au mieux les requêtes des usagers.

Les critères des usagers : la qualité de service

La qualité de service d'un mobile est une notion subjective qui dépend en fait de l'usager et de l'application qu'il utilise. Bien souvent, la qualité de service intègre plusieurs critères. Pour les communications ayant des contraintes temporelles fortes, comme la téléphonie, le délai est un critère prépondérant, alors que pour le téléchargement de fichiers, il s'agit plutôt du débit.

Les critères des opérateurs

L'objectif des opérateurs est de satisfaire au mieux la qualité de service de ses clients. Cela implique de gérer les ressources du réseau le plus efficacement possible. Les mesures de performance couramment utilisées sont le temps de séjour moyen des mobiles, la probabilité qu'un mobile ne puisse pas se connecter au réseau (pour cause de saturation), ou encore, la probabilité qu'un mobile en cours de communication voit sa connexion s'interrompre. Il s'agit donc de critères dynamiques qui reposent sur des moyennes temporelles.

Moyens d'action

L'opérateur qui gère un ensemble de points d'accès sans fil agit sur le réseau à différents niveaux[3] :

1. Par la construction et le dimensionnement des infrastructures nécessaires aux communications. Cela se fait sur une échelle de temps large, et résulte à la fois d'une étude statistique pour anticiper la demande, et de la résolution de problèmes de combinatoire complexes afin de répartir au mieux les antennes (et également les fréquences utilisées par les antennes).
2. Par la mise en place d'une tarification et de services différentiés.

3. Les deux derniers points ne sont pas réellement contrôlés par l'opérateur. Les décisions sont codées dans le matériel qui est vendu par le constructeur. L'opérateur peut néanmoins régler certains paramètres.

3. Par le choix d'un point d'accès pour la communication d'un mobile. Bien souvent, un mobile peut se connecter via plusieurs antennes et ce choix est géré de façon automatique par le réseau. Cela permet de répartir la charge sur l'ensemble des antennes, ce qui constitue une forme particulière de routage. La gestion du choix du point d'accès se fait à chaque communication.
4. Par le partage des ressources d'un point d'accès (fréquences, puissance...). La gestion des ressources se fait à l'échelle de temps de l'émission d'un paquet.

Notons que les usagers peuvent également intervenir dans le choix du point d'accès au réseau, notamment par le choix de la technologie. L'opérateur doit alors inciter les usagers à agir d'une manière qui est globalement efficace.

1.2.3 Objectifs de la thèse

Dans cette thèse on aborde le problème de l'optimisation des préférences de l'opérateur par le contrôle du routage (ou le choix d'un point d'accès) des communications des mobiles. On suppose que les autres moyens d'action (dimensionnement, tarification, gestion des ressources) sont fixés.

Optimisation dynamique ou approche gloutonne

L'optimisation des préférences de l'opérateur peut être vue comme un problème d'optimisation dynamique avec plusieurs critères. Durant la thèse, nous avons travaillé sur des méthodes d'optimisation dynamique avec contraintes qui reposent sur les semi-processus de décision markoviens. Les résultats que nous avons obtenus ont été soumis dans les actes d'une conférence [CHT11].

Le problème de ces méthodes est d'une part que leur complexité est exponentielle en la taille du système, et, de plus, qu'elles nécessitent une connaissance et un contrôle global du système (état du système en chaque instant, statistiques sur la mobilité...). C'est pourquoi nous les avons comparées avec des méthodes gloutonnes [CHTG09], c'est-à-dire basées sur l'optimisation d'un critère instantané, qui ne dépendent pas d'un contrôleur centralisé. En l'occurrence, nous avons montré que, si le système n'est pas trop chargé, alors l'optimisation du débit global du système à chaque instant (ou au moins à chaque événement) donne des performances quasiment optimales en terme de temps moyen de séjour des mobiles.

Par soucis de cohérence du document, ces résultats ne sont pas présentés ici. Néanmoins, ils justifient l'approche que nous développons dans la suite, qui repose sur l'optimisation *instantanée* des performances du système.

Approche par la théorie des jeux

La théorie des jeux analyse le résultat de situations dans lesquelles plusieurs entités prennent des décisions en vue de maximiser leur propre intérêt. La théorie des jeux est à l'interface de nombreuses disciplines, notamment les mathématiques, l'économie, la biologie,

et l'informatique. Le résultat théorique d'un jeu dépend de la modélisation du comportement des joueurs, et notamment de l'information dont ils disposent. Le comportement des joueurs est modélisé par des hypothèses sur leur rationalité, et également par leur manière de s'adapter à la répétition du jeu (qui est appelée "modèle d'apprentissage").

Dans les problèmes de routage, il est naturel de modéliser les usagers par des joueurs qui cherchent à maximiser leur qualité de service. Les performances (instantanées) du système dépendent uniquement du résultat du jeu. Le problème de l'optimisation distribuée des performances se traduit, ici, par la construction d'un jeu, et par l'implémentation d'un modèle d'apprentissage de manière à ce que le résultat du jeu corresponde à des performances optimales du système.

L'originalité de notre approche réside dans le fait de considérer *conjointement* la construction (que l'on appelle mécanisme d'incitation) *et* le modèle d'apprentissage, alors qu'à notre connaissance, tous les articles, dans le domaine de l'informatique, qui reposent sur la théorie des jeux, supposent donné, a priori, l'un des deux. L'exemple typique est celui du prix de l'anarchie, dans lequel il est considéré comme acquis que le résultat du jeu est un équilibre de Nash.

Nous prenons également en considération, dans les modèles d'apprentissage que nous proposons, la possibilité de leur implémentation dans des systèmes réels. Nous intégrons dans nos modèles les fluctuations aléatoires qui sont inhérentes aux réseaux sans fil. De plus, en raison de l'aspect fortement décentralisé des réseaux, il est impossible d'assurer la synchronisation parfaite de la prise des décisions par les usagers. Nous prenons également en compte ce paramètre dans l'analyse de nos modèles.

1.3 Organisation de la thèse et contributions

La thèse est articulée en quatre chapitres. Les liens entre les travaux de la thèse et les travaux existants sont introduits au fur et à mesure des chapitres. Dans tout le document, les résultats (théorèmes et propositions) existants se distinguent par la référence bibliographique qui les accompagne. Leur démonstration n'est pas donnée, sauf si la technique utilisée est employée dans d'autres démonstrations.

Nous terminons par une conclusion et des extensions des travaux développés dans cette thèse pouvant donner lieu à de futures recherches.

Chapitre 2 : Mécanismes d'incitation entre plusieurs entités indépendantes

Ce chapitre porte sur la construction de mécanismes d'incitation. Habituellement, les mécanismes d'incitation reposent sur la construction d'un jeu en stratégies dominantes. Nous montrons, dans ce chapitre, la limite de cette approche, et nous proposons une autre approche, basée sur la construction d'un jeu de potentiel, qui permet de résoudre le problème de l'anonymat des joueurs dans les jeux de routage. De plus, dans certaines situations, ce mécanisme est complètement distribué. Pour finir, nous montrons les limites de notre approche pour résoudre le problème d'incitation par le contrôle du partage des

gains dans un jeu de coalition.

Chapitre 3 : Jeux de potentiel et modèles d'apprentissage

Le chapitre précédent repose sur l'hypothèse que le résultat d'un jeu de potentiel est connu. À partir de ce chapitre nous allons justifier ce point, en proposant des modèles d'apprentissage qui aboutissent à ce résultat.

Ce chapitre est une introduction aux deux chapitres suivant. Il ne comporte pas de contribution importante. Nous y rappelons les principaux résultats existants sur les jeux en général et les jeux de potentiel en particulier, dans le cas d'espaces de stratégies finis et continus. Ensuite, nous justifions l'usage de modèles d'apprentissage, et nous mettons en lumière les contraintes d'implémentation de ces modèles. Enfin, nous analysons deux modèles d'apprentissage classiques.

Chapitre 4 : Le modèle stochastique de meilleure réponse

Le modèle stochastique de meilleure réponse est un modèle d'apprentissage simple dans lequel les gains des joueurs sont soumis à des perturbations aléatoires. Un résultat classique affirme que, par l'ajout de bruit, le résultat de l'apprentissage dans un jeu de potentiel est un état qui maximise le potentiel. Partant de ce résultat, nous proposons un algorithme pour optimiser le routage dans les réseaux de mobile ad hoc, et nous détaillons une implémentation possible de cet algorithme. Malheureusement, ce résultat n'est plus valable dès lors que les joueurs ne modifient pas, dans le modèle d'apprentissage, leur stratégie de manière asynchrone. Dans la dernière section du chapitre, nous caractérisons les résultats du jeu en fonction du processus de révision, c'est-à-dire de la synchronisation des joueurs.

Chapitre 5 : Extension mixte du modèle de meilleure réponse

Ce chapitre concentre la plus grande partie des contributions de la thèse.

Le modèle d'apprentissage du chapitre précédent n'est pas robuste au processus de révision des stratégies. Une des raisons est que c'est un algorithme qui évolue dans un espace discret (l'ensemble des stratégies d'un jeu fini). En considérant l'extension mixte des jeux, c'est-à-dire des stratégies aléatoires, nous proposons un modèle d'apprentissage qui évolue dans un espace continu, que nous appelons *dynamique de meilleure réponse*. La dynamique de meilleure réponse peut prendre différentes formes en fonction de la métrique employée. Nous donnons des conditions sur la métrique pour que le modèle soit bien défini, et nous montrons que, par un choix particulier de métrique, la dynamique de meilleure réponse correspond à la dynamique de réplication. Nous analysons les propriétés de ces dynamiques, notamment leur convergence. Ensuite, nous en proposons une implémentation reposant sur la théorie des approximations stochastiques dont nous analysons les propriétés. Nous montrons que ces résultats sont robustes à la fois aux fluctuations aléatoires des gains du jeu, et au processus de révision des stratégies. Cette implémentation est enfin illustrée dans un problème d'association de mobiles à des points d'accès sans fil.

CHAPITRE 2

MÉCANISMES D'INCITATION ENTRE PLUSIEURS ENTITÉS INDÉPENDANTES

Résumé du chapitre

Dans ce chapitre, on s'intéresse à la construction de mécanismes qui incitent des entités indépendantes les unes des autres (agents économiques, utilisateurs de ressources dans les réseaux de communication...) à agir de manière à maximiser un critère global. Un tel mécanisme peut être modélisé par un jeu impliquant ces différentes entités dans lequel un opérateur [1] peut imposer des pénalités qui dépendent des actions prises par les joueurs.

Cependant, déterminer le résultat d'un jeu est, en général, un problème difficile qui repose à la fois sur la modélisation du comportement des joueurs et sur l'analyse de ce modèle. La plupart des mécanismes d'incitation reposent sur la construction d'un jeu ayant des *stratégies dominantes*. Leur utilisation est conditionnée par le fait que les joueurs vont effectivement choisir une telle stratégie. Néanmoins, ces mécanismes atteignent leur limite lorsque l'on contraint les pénalités à être *anonymes*. Nous montrons alors que l'on peut, dans certaines situations comme, par exemple, les jeux de routage, imposer des pénalités qui tiennent compte de l'anonymat des joueurs, et telles que le jeu ainsi obtenu est un *jeu de potentiel*.

L'organisation du chapitre est la suivante : nous commençons en détaillant notre modèle et ce que l'on entend par "mécanisme d'incitation". Dans la deuxième section, nous rappelons le principal résultat concernant la construction de stratégies dominantes, qui est le mécanisme VCG. Nous montrons également la limite de l'approche en stratégies dominantes. Dans la troisième section, nous proposons une deuxième approche qui repose sur les jeux de potentiel. Enfin, dans la dernière section, nous montrons que ces deux approches échouent à résoudre le problème dans lequel on cherche à inciter des joueurs à participer à une coalition à partir du contrôle du partage des gains des coalitions.

1. Aussi appelé régulateur ou planificateur. Néanmoins, nous optons pour "opérateur" afin de garder la terminologie des réseaux.

2.1 Présentation générale du problème d'incitation

Un des objectifs de l'analyse économique est la construction de mécanismes incitant des agents à agir selon l'intérêt d'une organisation. Ces mécanismes se traduisent par des lois au niveau des états, comme par exemple celles visant à réguler les monopoles sur les marchés, ou, sur une échelle plus petite, par des règles au niveau des entreprises.

Les mécanismes d'incitation, c'est-à-dire les lois ou les règles qui sont édictées, visent un but qui a été fixé par l'organisation qui les met en place. L'instauration d'un mécanisme d'incitation a un coût pour l'organisation car il implique des contrôles pour vérifier que les règles sont bien respectées. Il faut donc pouvoir évaluer le gain d'un tel mécanisme, ce qui suppose de disposer de modèles de comportement des agents suffisamment précis.

On retrouve en informatique le même type de problématiques en ce qui concerne le partage des ressources, et plus particulièrement les ressources de calcul et de communication. Dans les réseaux informatiques, il y a d'une part des utilisateurs dont les décisions reposent sur des critères individuels de qualité de service et une vision très locale du réseau, et d'autre part, un opérateur qui doit faire en sorte que le réseau fonctionne globalement de façon efficace. Il est clair que, si chaque utilisateur désire le débit le plus large possible et qu'aucune restriction n'existe, le réseau deviendra rapidement saturé et inutilisable.

Parmi les moyens d'action de l'opérateur, il y a d'abord la possibilité de dimensionner le réseau en fonction des besoins des utilisateurs, ce qui implique d'évaluer les performances et de renforcer le réseau au niveau des points de congestion. Ensuite, l'opérateur peut utiliser des mécanismes d'incitation qui peuvent prendre la forme de tarification ou de contraintes d'utilisation (limitation du téléchargement par exemple). Mais la plupart des mécanismes de contrôle du réseau sont en fait intégrés directement dans les terminaux des utilisateurs sous forme de protocole de communication. C'est le cas, par exemple, avec le protocole TCP pour le contrôle de congestion ou le protocole CSMA/CA pour l'accès au média pour les communications en Wifi. En quelque sorte, ces protocoles ont été construits comme des mécanismes visant à inciter [2] les terminaux à agir localement de manière à avoir un usage du réseau globalement efficace.

Dans ce chapitre, nous nous plaçons dans un modèle plus général que le cadre informatique, mais nous illustrons les résultats avec des applications à des problèmes d'allocation de ressources dans les réseaux.

2.1.1 Modèle et notations

Considérons un système général qui peut être dans plusieurs états (une façon de partager des ressources de communication entre plusieurs utilisateurs, un routage de différents flux de communication...). L'état du système est en partie contrôlé par un opérateur, et par des entités indépendantes les unes des autres [3], que nous appelons des *joueurs*. L'opérateur

2. Le terme "inciter" est un peu abusif ici car les terminaux n'ont, contrairement aux utilisateurs, aucun intérêt dans la communication.

3. Par indépendant, on entend que ces entités n'ont pas de possibilités de communication entre elles, et donc pas de moyens directs d'entente et de coordination.

et les joueurs ont des préférences individuelles sur les états du système. En général, la préférence de l'opérateur est liée à celle des joueurs, par exemple si les préférences des joueurs sont données par des valeurs pour chaque état, et si la préférence de l'opérateur est la somme de ces valeurs. Le problème général consiste, pour l'opérateur, a inciter les joueurs à agir de manière à maximiser *sa* préférence. Cela passe par la construction d'un *mécanisme d'incitation*.

Les deux principales difficultés auxquelles l'opérateur peut être confronté sont, d'une part, qu'il ne connaît pas les préférences des joueurs, on parle alors de préférences privées, et d'autre part, qu'il ne contrôle pas le choix de l'état final (mais peut néanmoins imposer des pénalités).

Plus formellement, on note $\mathcal{U}$ l'ensemble des joueurs et $\mathcal{E}$ l'ensemble des états du système. Pour éviter les problèmes d'implémentation des mécanismes que nous allons présenter, nous supposons que ces ensembles sont finis.

Préférences sur l'ensemble des états

L'opérateur et les joueurs ont des *préférences* sur l'ensemble $\mathcal{E}$ des états. Ces préférences peuvent prendre la forme soit d'un ordre total sur les états, soit (ce qui implique également un ordre total) une fonction qui à chaque état associe une valeur. Dans ce deuxième cas, les préférences sont appelées *valuations*. On note v_u la valuation du joueur u, et V la valuation de l'opérateur. Les valuations sont donc des fonctions de $\mathcal{E}$ dans $\mathbb{R}$. Il se peut que la valuation de l'opérateur dépende de la valuation des joueurs : par exemple si $V = \sum_{u \in \mathcal{U}} v_u$. Dans ce cas, le problème du choix de l'état qui maximise V est communément appelé problème du *choix social* (voir par exemple le chapitre 9 dans [Nis07]).

Les valuations peuvent représenter une valeur monétaire (ex : le prix d'un bien), mais également une grandeur physique (ex : le débit d'une communication), ou de façon plus abstraite une utilité. On suppose qu'il est cohérent d'additionner et de comparer les valuations des joueurs. Dans le cas de grandeurs physiques, cela signifie que les valeurs sont exprimées dans la même unité.

Il existe essentiellement deux situations : l'une pour laquelle les préférences des joueurs sont publiques, c'est-à-dire connues par l'opérateur, et l'autre où les préférences sont complètement privées [4].

Construction d'un mécanisme

Nous nous plaçons ici dans le cas de préférences données par des valuations.

4. Il est également possible que les préférences des joueurs soient en partie publiques et en partie privées comme dans [San07]. L'exemple classique est le choix d'une technologie de communication : plus le nombre de personnes utilisant une technologie de communication est grand, plus le nombre de contacts que l'on a, et donc la préférence à choisir cette technologie est grande. Ceci est une préférence publique. Néanmoins, certains individus, pour des considérations diverses (morales, esthétiques...), peuvent personnellement préférer une technologie à une autre même si elle ne donne pas accès au plus grand nombre de contacts. Il s'agit là d'une préférence privée.

Le choix de l'état dans l'ensemble $\mathcal{E}$ résulte d'une procédure en deux temps. D'abord, l'opérateur établit les règles d'un jeu, puis, connaissant les règles du jeu, les joueurs agissent en conséquence. La donnée des règles du jeu définit un *mécanisme* qui est incitatif si le *résultat* du jeu correspond à un état qui maximise la valuation de l'opérateur. Nous reviendrons dans le paragraphe suivant sur ce que l'on entend par "résultat du jeu".

Nous détaillons comment le jeu est défini par l'opérateur dans un cadre général. Il faut toutefois noter que les possibilités de l'opérateur sont, selon les situations considérées, sujettes à des restrictions. Tout d'abord, il définit un ensemble d'actions $\mathcal{S}_u$ pour chaque joueur. Cet ensemble ne supporte a priori aucune restriction, il peut par exemple être continu. Si l'on note s_u l'action dans $\mathcal{S}_u$ choisie par le joueur u, un *profil d'actions* est la donnée d'une action pour chaque joueur et se note $s = (s_u)_{u\in\mathcal{U}}$. Celui-ci appartient à l'ensemble des profils d'actions noté $\mathcal{S} \stackrel{\text{def}}{=} \underset{u\in\mathcal{U}}{\times} \mathcal{S}_u$. Ensuite, l'opérateur donne une fonction de choix qui, à chaque profil d'action, associe un état. Nous notons $f : \mathcal{S} \to \mathcal{E}$ cette fonction de choix. Enfin, pour chaque joueur, l'opérateur définit une fonction de pénalité $p_u : \mathcal{S} \to \mathbb{R}$, de telle manière que le gain du joueur u sous le profil d'action s vaut :

$$c_u(s) \stackrel{\text{def}}{=} v_u(f(s)) - p_u(s).$$

Si la pénalité est négative, cela revient à augmenter le gain par rapport à la valuation initiale de l'état qui a été choisi par la fonction de choix. La fonction de pénalité peut s'interpréter de différentes manières selon les applications : il peut s'agir d'un transfert d'argent comme d'une grandeur virtuelle transmise aux joueurs afin qu'il puissent agir en fonction de l'intérêt général. Cela sera illustré dans les applications des chapitres suivants.

Nous constatons que la fonction de gain comporte une partie publique qui est la pénalité, et une partie qui peut être privée et donc inconnue de l'opérateur, qui est $v_u \circ f$. De ce fait, l'opérateur ne contrôle pas complètement les gains des joueurs.

Finalement, le jeu construit par l'opérateur est défini par :
- l'ensemble des joueurs $\mathcal{U}$,
- les ensembles d'actions pour chaque joueur $\mathcal{S}_u$,
- les fonctions de gain pour chaque joueur $c_u : \mathcal{S} \to \mathbb{R}$.

Il faut bien faire la différence entre le profil d'action s et l'état correspondant $f(s)$. Du point de vue des joueurs, ce qui importe est leur gain dans le jeu et donc le profil d'action, mais du point de vue de l'opérateur il s'agit de l'état. On suppose ici que les pénalités infligées aux joueurs ne sont pas répercutées sur l'opérateur, seule sa valuation V compte[5].

Résultat du jeu

Le résultat du jeu est un profil d'actions du jeu. Ce profil d'action peut être sélectionné de différentes manières. Soit le jeu est joué une fois et une seule : la prévision du résultat s'avère difficile en général, sauf dans le cas où il existe des stratégies dominantes. Cela fait

5. Cela n'est pas réaliste si les pénalités sont des sommes d'argent qui sont versées ou prélevées par l'opérateur.

l'objet de la section 2.2. Soit le jeu est répété, et dans ce cas, les joueurs adaptent leur action à ce qu'ils apprennent des répétitions du jeu. Il y a donc un processus d'apprentissage qu'il faut modéliser et analyser (plusieurs modèles seront étudiés dans les chapitres suivants). Dans ce cas, le résultat du jeu est le profil d'actions qui est asymptotiquement sélectionné si le processus d'apprentissage converge. Dans tous les cas, le résultat du jeu dépend d'hypothèses faites a priori sur le comportement des joueurs.

Que le jeu soit joué une fois ou bien répété, de nombreux modèles de comportement des joueurs prévoient, sous certaines hypothèses, que le résultat sera un *équilibre de Nash*, c'est-à-dire un profil d'action dans lequel aucun joueur ne peut gagner à modifier son action *unilatéralement* :

Définition 2.1 (*Équilibre de Nash*)

Le profil d'action $s \in \mathcal{S}$ est un équilibre de Nash du jeu $(\mathcal{U}, \mathcal{S}, (c_u)_{u \in \mathcal{U}})$ si, pour tout joueur u et toute action $s'_u \in \mathcal{S}_u$, on a :

$$c_u(s_u, s_{-u}) \geq c_u(s'_u, s_{-u}).$$

La notation classique $-u$ désigne l'ensemble des joueurs sauf u, et par abus de notation, on écrira $s = (s_u, s_{-u})$ lorsque l'on veut distinguer l'action du joueur u (il ne s'agit pas du déplacement de s_u à la première position du vecteur).

Notons qu'un équilibre de Nash n'est pas nécessairement stable par déviation de deux joueurs ou plus. Cependant, on suppose que les joueurs ne communiquent pas, et ne peuvent donc pas s'accorder, ou en d'autres termes former une coalition en vue d'augmenter conjointement leur gain.

Notons enfin que, en général, il n'y a aucune raison pour que le résultat d'un jeu soit un équilibre de Nash, premièrement parce qu'il n'en existe pas nécessairement, et deuxièmement, même s'il en existe un et qu'il est unique, les joueurs peuvent gagner plus en choisissant une autre action (cf. le dilemme du prisonnier dans les exemples qui suivent).

2.1.2 Formulation du problème d'incitation

Finalement, le problème d'incitation consiste à construire un jeu de la manière décrite précédemment de manière à ce que le résultat s du jeu (ou tous les résultats possibles), soit tel que l'état correspondant $f(s)$ maximise la valuation V de l'opérateur.

Notons que, s'il n'y a aucune contrainte sur la construction du jeu, l'opérateur peut choisir n'importe quel état $e \in \mathcal{E}$ en posant $f(s) = e$ pour tout $s \in \mathcal{S}$. Mais, d'une part, l'opérateur ne connaît pas forcément l'état optimal, en particulier si celui-ci dépend des valuations privées des joueurs, ou bien si la complexité rend son calcul impossible. D'autre part, dans les cas pratiques que nous étudions par la suite, soit l'opérateur choisit f mais les valuations sont privées, soit l'ensemble des états et des profils d'actions coïncident, c'est-à-dire $\mathcal{E} = \mathcal{S}$ et la fonction de choix est $f(s) = s$, ce qui signifie que l'état choisi est le résultat du jeu.

2.1.3 Exemples

Citons quelques exemples classiques pour lesquels il est important d'établir des mécanismes d'incitation.

Intérêt individuel contre intérêt collectif

Il est assez courant que l'intérêt individuel n'aboutisse pas au choix d'un état qui maximise l'intérêt collectif. Illustrons cela sur deux exemples classiques dans lesquels $\mathcal{E} = \mathcal{S}$.

Commençons par l'exemple du dilemme du prisonnier. Le scénario est le suivant : deux suspects sont arrêtés et interrogés séparément. Ils peuvent soit dénoncer l'autre, soit ne rien avouer. Si aucun n'avoue, alors chaque suspect écope d'une peine de prison minimale (disons 6 mois). Si les deux avouent, alors ils écopent d'une peine moyenne (disons 5 ans), et si un seul dénonce, alors celui qui a dénoncé est libéré tandis que l'autre écope d'une peine lourde (disons 10 ans). Dans ce jeu, l'intérêt collectif des suspects est qu'aucun d'eux n'avoue. Cependant il est tentant, individuellement, de dénoncer pour être libéré immédiatement, d'autant plus qu'il est risqué de se taire si l'autre nous dénonce. Ici, le seul équilibre de Nash qui est un résultat possible du jeu (et même en stratégies dominantes comme nous le verrons) est le profil d'actions où les deux suspects dénoncent, ce qui ne correspond pas à l'intérêt collectif[6].

Le deuxième exemple porte sur le problème de routage suivant. N joueurs veulent transmettre un paquet du sommet source s au sommet destination d (voir figure 2.1). Pour cela ils choisissent l'une des deux routes possibles en cherchant à minimiser leur coût, le délai par exemple. L'état du système est ici la manière dont se répartissent les joueurs sur les deux routes. La route du bas a un délai constant qui vaut N, ce qui est long, mais indépendant de la charge, la charge étant le nombre de joueurs ayant choisi la route. La route du haut a un délai qui est égal à la charge ℓ. Donc, plus de paquets passent par cette route, plus le délai de chaque paquet est grand.

On constate ici que les seules situations d'équilibre sont les profil d'actions où tous les paquets ou tous les paquets sauf un passent par la route du haut, si bien que tous les joueurs ont un coût qui vaut au moins $N-1$, et le coût social vaut $\sum_{i=1}^{N}(N-1) = N(N-1)$. Supposons maintenant que les joueurs se répartissent à moitié en bas et en haut (si le nombre de joueur est pair). Alors le coût social, qui est l'optimum, vaut $N/2 \times N + N/2 \times N/2 = 3/4N^2$ et est inférieur strictement au coût social à l'équilibre si N est suffisamment grand.

Le problème dans ces deux exemple est de trouver un mécanisme d'incitation afin d'atteindre l'état optimal en terme de coût social, en pénalisant les joueurs en fonction de l'action ou de la route qu'ils ont choisie.

6. Notons que le fait qu'il n'y ait qu'un unique équilibre de Nash ne signifie pas pour autant que les joueurs humains vont choisir de dénoncer, voir à ce sujet les résultats expérimentaux dans [AM93]. Cependant, en l'absence de modèle fin du comportement des humains, on se contente du modèle de rationalité classique qui implique que l'équilibre de Nash est le résultat du jeu.

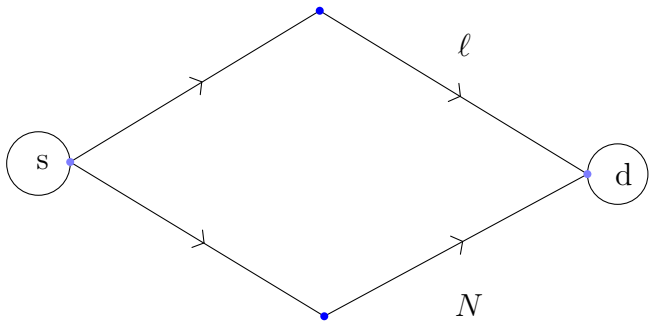


Figure 2.1 – Routage compétitif entre N joueurs. La route du bas a un délai constant N, et le délai sur la route du haut est égal à la charge ℓ.

Mentir sur ses préférences

Nous proposons maintenant un exemple dans lequel les valuations des joueurs sont privées, l'opérateur cherche à maximiser la somme des valuations, et l'action de chaque joueur est l'annonce d'une valuation. L'objectif de l'opérateur est d'inciter les joueurs à annoncer leur vraie valuation afin de choisir l'état qui maximise la somme des valuations.

L'exemple typique est celui de la mise aux enchères d'un bien où plusieurs joueurs proposent simultanément une offre afin de l'acquérir. L'ensemble $\mathcal{E}$ des états du système est l'ensemble des façons de distribuer le bien parmi l'ensemble des joueurs. On suppose que la valeur attribuée à un état par chaque joueur est positive si le joueur a le bien dans l'état, et nulle sinon. Au final, annoncer une valuation (c'est-à-dire proposer une enchère) revient à annoncer la valeur que le joueur attribue à l'état dans lequel il possède le bien.

L'opérateur définit une fonction de choix qui attribue le bien à l'un des joueurs, et instaure des pénalités qui dépendent des annonces. Supposons ici que la fonction de choix attribue le bien au joueur qui a fait l'offre la plus élevée et que celui-ci paye le montant de son offre (la pénalité étant nulle pour les joueurs qui n'ont pas obtenu le bien).

Pour simplifier les notations, supposons qu'il n'y ait que deux joueurs. Chaque joueur évalue le bien respectivement à v^- et v^+, avec $v^- < v^+$. Le joueur le plus intéressé par le bien est donc le joueur qui a la valuation v^+. Le joueur le moins intéressé fait l'offre s^- et le plus intéressé l'offre s^+. Alors, leur gain dans le jeu vaut respectivement $c^- = \big(v^- - s^-\big)\mathbf{1}_{s^->s^+}$ et $c^+ = \big(v^+ - s^+\big)\mathbf{1}_{s^+>s^-}$. Dans le cas où les offres sont égales, le gain est nul pour les deux joueurs.

Si les joueurs annoncent leur vraie valuation, ils sont assurés d'avoir un gain nul : en effet s'ils obtiennent le bien, la valuation du bien est compensée par le prix qu'ils payent, autrement ils n'obtiennent rien et ne payent rien. Les joueurs gagnent toujours plus à annoncer leur valuation plutôt qu'à faire une offre supérieure. Par conséquent, aucun joueur n'a intérêt à proposer une offre strictement supérieure à sa valuation.

Si le joueur le plus intéressé connaît le choix s^- de l'autre joueur, alors il a intérêt à proposer une offre s^+ légèrement supérieur, l'offre optimale n'existant pas ici (car annoncer $s^+ = s^-$ procure un gain nul). Dans ce cas, le bien revient au joueur le plus intéressé, mais

celui-ci n'a pas déclaré sa vraie valuation. Il résulte du fait de ne pas annoncer sa vraie valuation que l'opérateur n'a pas, au final, la garantie que le bien a effectivement été donné au joueur le plus intéressé.

Considérons maintenant l'exemple du vote dans lequel les préférences des joueurs ne sont plus données par des valuations, mais uniquement par un ordre total sur l'ensemble des candidats. Il est connu par le théorème de Gibbard–Satterthwaite [Gib73], que, s'il y a au moins trois candidats, alors la seule fonction de choix qui fasse en sorte que les joueurs n'aient jamais intérêt à ne pas annoncer leur vraie préférence est le *choix dictatorial*. Cela signifie qu'il existe un joueur tel que le candidat choisi est le candidat préféré de ce joueur, indépendemment de l'action des autres joueurs.

Nous montrerons que, dans le cas où les préférences sont données par des valuations, dont le problème des enchères est un cas particulier, il existe un mécanisme d'incitation [7].

2.2 Une approche en stratégies dominantes

Dans cette section, nous présentons un mécanisme d'incitation qui repose sur la construction d'un jeu ayant des stratégies dominantes.

2.2.1 Description de l'approche

Le but de l'opérateur est de construire un jeu, à partir d'une fonction de choix et de l'instauration de pénalités, tel que le résultat du jeu corresponde à un état qui maximise sa valuation.

Déterminer le résultat d'un jeu est cependant un problème : en effet, comme le gain d'un joueur ne dépend pas uniquement de son action mais du profil d'action global, il n'y a, a priori, pas de stratégie optimale. Un cas de figure pour lequel l'action d'un joueur peut être anticipée, c'est lorsqu'il existe une *stratégie dominante*, et que celle-ci est unique :

Définition 2.2 (*Stratégie dominante*)

L'action $s_u \in \mathcal{S}_u$ est une stratégie dominante pour le joueur u dans le jeu $(\mathcal{U}, \mathcal{S}, (c_u)_{u\in\mathcal{U}})$ si pour tout $s'_u \in \mathcal{S}_u$ et pour tout $s_{-u} \in \mathcal{S}_{-u}$:

$$c_u(s_u, s_{-u}) \geq c_u(s'_u, s_{-u}).$$

Une stratégie dominante est toujours meilleure que n'importe quelle autre stratégie quelles que soient les actions des autres joueurs. Quand une stratégie dominante existe, on suppose que le joueur ne choisira pas une stratégie qui ne l'est pas : cela constitue notre hypothèse de comportement. Notons qu'une telle stratégie peut ne pas exister, et si elle existe, ne pas être unique.

7. Notons également les résultats positifs et négatifs donnés dans le problème du mariage [GS62] où les préférences sont des ordres totaux (voir le théorème 6 dans [BR97]). En utilisant une certaine fonction de choix, on peut faire en sorte que au moins l'une des parties (homme ou femme dans le cas du mariage) soit incitée à annoncer ses vraies préférences.

Lorsqu'il existe une stratégie dominante pour chaque joueur, le résultat du jeu est nécessairement un profil d'action constitué de stratégies dominantes, et c'est alors un équilibre de Nash. On parle alors d'*équilibre de Nash en stratégies dominantes* (tous les équilibres de Nash n'étant pas constitués de stratégies dominantes, même si l'équilibre est unique).

Finalement, le problème d'incitation en stratégies dominantes peut s'énoncer de la manière suivante : construire un jeu, c'est-à-dire donner les ensembles d'action des joueurs, la fonction de choix $f : \mathcal{S} \to \mathcal{E}$, et les fonctions de pénalité de telle manière que, si un profil d'action $s \in \mathcal{S}$ est un équilibre de Nash en stratégies dominantes du jeu ainsi défini, alors $f(s)$ est un état optimal pour la valuation V de l'opérateur.

Ce problème n'a en général pas de solution. Néanmoins, il est résolu dans la classe des problèmes liés à la révélation des valuations dans les situations d'enchère au sens large.

2.2.2 Les enchères généralisées

Ici, on considère une situation dans laquelle les valuations des joueurs sur l'ensemble des états sont privées, et la valuation de l'opérateur est la somme des valuations des joueurs : $V = \sum_{u \in \mathcal{U}} v_u$.

L'ensemble d'action des joueurs est donné (il n'est pas choisi par l'opérateur) : il s'agit de l'annonce d'une valuation. Formellement, l'action du joueur u est donc une fonction $s_u : \mathcal{E} \to \mathbb{R}$. En pratique, cela suppose que l'ensemble d'état $\mathcal{E}$ n'est pas trop grand. Le but de l'opérateur est d'inciter les joueurs à annoncer leur vraie valuation afin de choisir l'état optimal, donc de faire en sorte que $s_u = v_u$ soit une stratégie dominante dans le jeu.

Notons que ce problème apparaît typiquement dans la situation où plusieurs biens sont simultanément mis aux enchères et sont ensuite répartis entre des joueurs. L'ensemble des états du système est l'ensemble des répartitions des biens parmi les joueurs. La valuation des joueurs ne dépend généralement que de l'ensemble des biens qu'ils acquièrent. La valuation des joueurs étant privée, établir des règles pour les enchères de manière à ce que les joueurs soient incités à déclarer leur vraie valuation rentre dans le cadre précédent.

Ce problème est bien connu et résolu par le mécanisme VCG (Vickrey Clarke Groves). Ce résultat à la fois étonnant et puissant a été exploité dans de nombreux problèmes économiques liés au réseau, comme la taxation du routage entre plusieurs domaines [8]. Voici les principes de ce mécanisme.

Le jeu construit par le mécanisme VCG est déterminé par le choix particulier de fonctions de pénalité et d'une fonction de choix $f : \mathcal{S} \to \mathcal{E}$ données par :

- l'état choisi est l'un des états qui maximisent la somme des valuations annoncées, *i.e.* $f(s) \in \operatorname{argmax}_{e \in \mathcal{E}} \sum_{u \in \mathcal{U}} s_u(e)$,

8. Voir à ce sujet [MT07] qui présente plusieurs applications du mécanisme VCG ainsi que ses limitations.

– la pénalité du joueur u vaut $p_u(s) = -\sum_{i \neq u} s_i(f(s)) + h_u(s_{-u})$, où $h_u(s_{-u})$ est une fonction quelconque qui ne dépend pas de s_u.

On a alors le théorème bien connu :

Théorème 2.3 (*Vickrey Clarke Groves*)

Le jeu défini par le mécanisme VCG est tel que l'annonce de sa vraie valuation est une stratégie dominante pour chaque joueur.

Le théorème implique alors que l'état qui est choisi par le mécanisme VCG est celui qui maximise la somme des valuations des joueurs.

La démonstration du théorème se fait par vérification que si le joueur u annonce s_u alors son gain est forcément inférieur à celui qu'il obtient en annonçant v_u quelle que soit l'annonce s_{-u} des autres joueurs. Notons $e = f(v_u, s_{-u})$ et $e' = f(s_u, s_{-u})$. Il faut donc montrer que $v_u(e) + \sum_{i \neq u} s_i(e) - h_u(s_{-u})$ est plus grand que $v_u(e') + \sum_{i \neq u} s_i(e') - h_u(s_{-u})$. Mais le choix de f implique que e maximise $v_u(e'') + \sum_{i \neq u} s_i(e'')$, d'où le résultat.

Les choix particuliers de la fonction h permettent ensuite d'obtenir des propriétés supplémentaires sur la fonction de pénalité. Par exemple, le choix de $h_u(s) = \max_{e \in \mathcal{E}} \sum_{i \neq u} s_u(e)$, appelé pivot de Clarke, assure que les gains finaux des joueurs sont positifs (donc que les joueurs ne sont jamais perdant quand ils participent au jeu), et en plus que les pénalités sont strictement positives, ce qui se traduit par le fait qu'il n'y pas de transfert d'argent vers les joueurs [9].

Par exemple, s'il n'y a qu'un seul bien mis aux enchères, le mécanisme VCG avec le pivot de Clarke revient à attribuer le bien au joueur qui a fait l'offre la plus haute, mais au lieu de payer le montant de son offre, ce joueur paye le montant de la deuxième offre la plus élevée.

2.2.3 Limite de l'incitation en stratégies dominantes

Reprenons le problème de routage correspondant à la figure 2.1, en prenant $N = 4$. Le routage optimal pour l'opérateur, *i.e.* celui qui minimise le coût social, est celui pour lequel deux joueurs choisissent la route du bas, et les deux autres la route du haut. Afin d'inciter les joueurs à se répartir de cette manière, l'opérateur peut inciter *individuellement* chaque joueur à choisir une certaine action. Si les joueurs sont numérotés, l'opérateur peut, par exemple, mettre une pénalité très élevée sur la route du haut et une pénalité nulle sur la route du bas pour les deux premiers joueurs, et le contraire pour les deux derniers. Ce faisant, la stratégie dominante pour les deux premiers joueurs est le choix de la route du bas, et la route du haut pour les autres. Il en résulte que l'équilibre de Nash en stratégies dominantes correspond effectivement à l'état optimal.

9. D'autres propriétés des mécanismes VCG obtenues par les différents choix de fonction h ont été étudiées dans [TT10].

Supposons maintenant que les fonctions de pénalité imposées par l'opérateur soient contraintes par *l'anonymat* des joueurs[10]. Par anonymat, on entend que deux joueurs *identiques* ont les mêmes fonctions de pénalité. Les joueurs u_1 et u_2 sont identiques si :

– ils ont le même ensemble d'actions :

$$\mathcal{S}_{u_1} = \mathcal{S}_{u_2},$$

– leurs gains sont les mêmes :

$$\forall s \in \mathcal{S}, c_{u_1}(s) = c_{u_2}(s'),$$

où s' est le profil d'action obtenu à partir de s après permutation des actions de u_1 et de u_2.

– ils ont le même impact sur les autres joueurs :

$$\forall u \in \mathcal{U} \backslash \{u_1, u_2\}, \forall s \in \mathcal{S}, c_u(s) = c_u(s'),$$

où s' est le profil d'action obtenu à partir de s après permutation des actions de u_1 et de u_2.

En pratique, deux joueurs sont identiques s'ils partagent les mêmes caractéristiques, ou encore s'ils sont du même type. L'anonymat contraint donc les pénalités à être les mêmes pour tous les joueurs du même type.

Dans le problème de routage de la figure 2.1, tous les joueurs sont identiques. S'il existe une fonction de pénalité qui est la même pour chaque joueur et qui implémente une stratégie dominante, alors cette stratégie est la même pour chaque joueur. En particulier, elle ne peut pas répartir les joueurs à moitié en haut et à moitié en bas, à moins que les deux actions soient dominantes. Dans ce dernier cas, n'importe quelle répartition des joueurs est un équilibre en stratégie dominante, même les états qui ne sont pas optimaux.

On voit donc que, sous la contrainte d'anonymat, il n'existe aucun mécanisme d'incitation en stratégies dominantes. Ce problème sera résolu par notre deuxième approche, qui repose sur les fonctions de potentiel. Cependant cette approche suppose que le résultat du jeu est le produit d'un processus d'apprentissage quand le jeu est répété (et même infiniment répété), ce qui n'était pas nécessaire pour les mécanismes en stratégies dominantes.

2.3 Une approche par fonction de potentiel

L'approche consiste ici à construire un jeu qui ne possède pas nécessairement de stratégies dominantes, mais tel que, s'il est répété et que les joueurs suivent une certaine règle d'apprentissage, alors le résultat asymptotique correspond à un état optimal. Cela est le cas si la valuation de l'opérateur est une *fonction de potentiel* du jeu.

10. La contrainte d'anonymat peut aussi être vue comme une simplification du calcul des fonctions de pénalités, l'opérateur n'ayant pas nécessairement la capacité de déterminer l'état optimal et les pénalités qui vont avec, ce qui est courant dans les problèmes de routage.

Dans toute cette section, on suppose que les préférences des joueurs sont publiques, que l'ensemble des états est l'ensemble des profils d'actions du jeu $(\mathcal{U}, \mathcal{S}, (v_u)_{u\in\mathcal{U}})$ (les valuations des joueurs sont donc les gains du jeu), et que la fonction de choix est l'identité, *i.e.* $f(s) = s$ pour tout $s \in \mathcal{S}$.

2.3.1 Description de l'approche

L'objectif est de construire un jeu de manière à ce que la fonction [11] V, la valuation de l'opérateur, soit une fonction de potentiel du jeu :

Définition 2.4 (*Fonction de potentiel*)

Soit un jeu $(\mathcal{U}, \mathcal{S}, (c_u)_{u\in\mathcal{U}})$. La fonction $P : \mathcal{S} \to \mathbb{R}$ est une fonction de potentiel si pour tout $s \in \mathcal{S}$, pour tout $u \in \mathcal{U}$ et pour tout $s'_u \in \mathcal{S}_u$:

$$c_u(s_u, s_{-u}) - c_u(s'_u, s_{-u}) = P(s_u, s_{-u}) - P(s'_u, s_{-u}).$$

On voit immédiatement que si P est une fonction de potentiel du jeu, alors les gains des joueurs s'écrivent nécessairement sous la forme $c_u(s) = P(s) + h_u(s_{-u})$, si bien que, du point de vue des joueurs, tout se passe comme si les fonctions de gain du jeu étaient remplacées par la fonction de potentiel (le terme $h_u(s_{-u})$ étant indépendant de leur action). Cela implique en particulier que le profil d'actions qui maximise la fonction de potentiel est un équilibre de Nash du jeu.

Finalement, un mécanisme d'incitation par fonction de potentiel est la construction d'un jeu qui repose sur l'établissement de pénalités tel que V est une fonction de potentiel.

Justification de l'approche par fonction de potentiel

Si V est une fonction de potentiel du jeu, et que s est un profil d'actions tel que $f(s)$ maximise V, alors s est un équilibre de Nash. Cependant, s n'est pas nécessairement en stratégies dominantes.

La justification de l'utilisation d'une fonction de potentiel repose en fait sur l'existence d'un modèle d'apprentissage appelé modèle stochastique de meilleure réponse (que nous étudions au chapitre 4) dont le résultat asymptotique est un profil d'actions qui maximise la fonction de potentiel. De plus, les dynamiques d'apprentissage continues vérifiant les conditions de corrélations positives [San10], convergent des ensembles d'équilibres de Nash.

Notons que cela suppose que les joueurs vont suivre exactement ce processus d'apprentissage ce qui peut apparaître comme une hypothèse trop restrictive. Cependant, comme pour les mécanismes en stratégies dominantes, cela est possible à partir du moment où les joueurs sont des machines dont le comportement est codé de manière à suivre cette règle d'apprentissage.

Remarquons également qu'il existe de nombreux modèles d'apprentissage pour lesquels, dans certaines classes de jeu (les jeux stables et les jeux super-modulaires par exemple), on

11. En toute rigueur, on devrait écrire $V \circ f$, mais ici, f est l'identité.

connaît le résultat, et pour lesquels il serait possible de construire un mécanisme de façon à ce que ce résultat corresponde à un état optimal. L'approche par jeu de potentiel n'est donc pas la seule envisageable.

Retour sur l'exemple de routage

Reprenons l'exemple du jeu de routage de la figure 2.1, et construisons un mécanisme d'incitation par fonction de potentiel qui tienne compte de la contrainte d'anonymat.

On suppose que, dans le jeu de routage, l'ensemble des joueurs est de cardinal N, chaque joueur ayant les actions H (route du haut) et B (route du bas). On note $\ell(s)$ le nombre de joueurs qui choisissent H, ce nombre dépendant du profil d'action s. Le coût pour un joueur (qui est sa valuation) vaut N (resp. $\ell(s)$) s'il choisit la route du bas (resp. du haut).

Ce jeu est un *jeu de congestion*, classe de jeu sur laquelle nous reviendrons, et dont on sait depuis les travaux dans [Ros73] qu'elle possède une fonction de potentiel qui est ici donnée par :

$$P(s) = \sum_{i=1}^{\ell(s)} i + \sum_{i=1}^{N-\ell(s)} N = \frac{\ell(s)^2 + \ell(s)}{2} + N(N - \ell(s)).$$

Cette fonction de potentiel ne coïncide pas avec la valuation de l'opérateur dans le cas où celui-ci cherche à minimiser la somme des coûts, qui vaut donc :

$$V(s) = \ell(s)^2 + N(N - \ell(s)).$$

Il faut alors faire en sorte d'ajouter une pénalité sur chaque route de façon à ce que le jeu ainsi construit ait pour fonction de potentiel V. Notons $p_H(\ell(s))$ et $p_B(N - \ell(s))$ la pénalité infligée à un joueur sur chacune des routes (que l'on suppose dépendre uniquement du nombre de joueur choisissant la route). Notons que ces pénalités sont identiques pour tous les joueurs et donc qu'elles satisfont la contrainte d'anonymat. Le coût sur la route du haut (resp. bas) devient donc $\ell(s) + p_H(\ell(s))$ (resp. $N + p_B(N - \ell(s))$).

On doit avoir :

$$\forall s, \sum_{i=1}^{\ell(s)} (i + p_H(i)) + \sum_{i=1}^{N-\ell(s)} (N + p_B(N - i)) = V(s).$$

En omettant la dépendance en s afin de clarifier le calcul, ceci équivaut à :

$$\begin{aligned} & \frac{\ell^2 + \ell}{2} + N(N - \ell) + \sum_{i=1}^{\ell} p_H(i) + \sum_{i=1}^{N-\ell} p_B(N - i) = \ell^2 + N(N - \ell) \\ \Leftrightarrow \quad & \sum_{i=1}^{\ell} p_H(i) + \sum_{i=1}^{N-\ell} p_B(N - i) = \frac{\ell^2}{2} - \frac{\ell}{2}. \end{aligned}$$

On voit alors que le choix de $p_H(i) = i - 1$ et $p_B(N - i) = 0$ donne le résultat escompté. En introduisant ces pénalités, le coût sur la route du haut devient $2\ell(s) - 1$ et celui sur la route du bas est inchangé et vaut N. Le jeu ainsi construit admet comme potentiel la fonction V.

Nous généralisons maintenant cette construction.

2.3.2 Construction d'un mécanisme d'incitation par fonction de potentiel dans les jeux finis

Soit un jeu fini $(\mathcal{U}, \mathcal{S}, (v_u)_{u\in\mathcal{U}})$. La fonction de valuation de l'opérateur $V : \mathcal{S} \to \mathbb{R}$ est quelconque, c'est-à-dire qu'elle ne dépend pas nécessairement des valuations des joueurs. L'opérateur peut alors établir des pénalités $(p_u)_{u\in\mathcal{U}}$ de manière à ce que V soit une fonction de potentiel du jeu fini $(\mathcal{U}, \mathcal{S}, (c_u)_u)$, où $c_u(s) = v_u(s) - p_u(s)$ en posant :

$$p_u(s) = v_u(s) - V(s) + h_u(s_{-u}), \tag{2.1}$$

où h_u est une fonction quelconque.

En effet, dans ce cas $c_u(s) = V(s) - h_u(s_{-u})$, et V est bien une fonction de potentiel.

Il est intéressant de constater la ressemblance avec la fonction de pénalité du mécanisme VCG qui vaut :

$$s_u(f(s)) - V(f(s)) + h_u(s_{-u}).$$

Néanmoins, dans le mécanisme VCG, la fonction $V \circ f$ n'est pas nécessairement une fonction de potentiel.

Le choix des fonctions h_u est libre et permet, dans certaines situations, d'obtenir des fonctions de pénalité ayant des propriétés intéressantes, comme la possibilité d'être calculée localement (nous illustrons cela dans la suite).

2.3.3 Application à des problème d'allocation de ressources

Nous donnons deux cas d'application dans des problèmes d'allocation de ressources. Nous supposerons ici que la valuation de l'opérateur est la somme des valuations des joueurs, c'est-à-dire $V(s) = \sum_{u\in\mathcal{U}} v_u(s)$.

Association de joueurs à un ensemble de ressource

Ici, chaque joueur doit choisir une et une seule ressource parmi un ensemble $\mathcal{R}$. Cet ensemble constitue donc l'ensemble des actions de chaque joueur. Un état du système, ou de façon équivalente un profil d'actions, est l'association de chaque joueur à une ressource unique.

Étant donné un profil d'actions s, chaque ressource a une charge que nous notons $\ell_r(s)$ qui dépend uniquement des joueurs qui ont choisi cette ressource. Contrairement aux jeux de congestion, la charge n'est pas uniquement le nombre de joueurs qui ont choisi cette ressource, les caractéristiques des joueurs étant ici prises en compte : la charge est un vecteur binaire $\ell_r = (\ell_r^u)_{u\in\mathcal{U}}$ tel que $\ell_r^u = 1$ si u a choisi r et 0 sinon.

On suppose enfin que la valuation de chaque joueur est une fonction qui dépend uniquement de la charge de la ressource qu'il a choisie. Si le joueur u a choisi la ressource r, donc si $s_u = r$, alors sa valuation est $v_u(s) = d_r^u(\ell_r(s))$, où d_r^u est une fonction quelconque. Notons également que, contrairement aux jeux de congestion, la valuation est une fonction qui peut être spécifique à chaque joueur.

Ce jeu modélise en particulier la situation dans laquelle des mobiles ont le choix de se connecter à une cellule dans un ensemble donné de cellules. La valuation de chaque mobile est une fonction de la qualité de service qu'il obtient, celle-ci étant par exemple le débit ou le délai. Dire que la valuation dépend uniquement de la charge de la cellule signifie que l'on ne prend pas en compte les interférences entre les différentes cellules. Cela se justifie si les cellules sont de technologie différente, ou opèrent sur des bandes de fréquence séparées. En général, la qualité de service ne dépend pas uniquement du nombre de mobiles connectés, car elle dépend des caractéristiques (de priorité, d'intensité de trafic pour en citer quelques unes) qui sont propres à chaque mobile.

Notons $(\emptyset, s_{-u})$ le profil d'actions en l'absence du joueur u. La charge sur chaque ressource en l'absence du joueur u s'écrit alors $(0, \ell_r^{-u})$, et $v_i(\emptyset, s_{-u}) = d^i_{s_i}(0, \ell_r^{-u}(s))$. Finalement, le choix de la fonction $h_u(s_{-u}) = \sum_{i \neq u} v_i(\emptyset, s_{-u})$ dans l'équation (2.1) donne :

$$p_u(s) = \sum_{i \neq u} \delta_{i,u}(s_{-u}),$$

où $\delta_{i,u}(s_{-u}) \stackrel{\text{def}}{=} \big(v_i(\emptyset, s_{-u}) - v_i(s)\big)$ est la différence de valuation pour le joueur i sans et avec le joueur u. En général c'est une valeur négative car chaque ressource est partagée en autant de parts que le nombre de joueurs l'ayant choisie. Notons également que $\delta_{i,u}(s_{-u})$ est l'impact du joueur u sur le joueur i.

Pour les joueurs qui n'ont pas choisi la même ressource que u, cet impact est nul, si bien que $\delta_{i,u}(s_{-u}) = 0$. Au final, la fonction de pénalité s'écrit :

$$p_u(s) = \sum_{i \neq u | s_i = s_u} \delta_{i,u}(s_{-u}).$$

On voit que, par le choix de cette fonction de pénalité, c'est-à-dire de cette fonction h, il suffit de connaître uniquement la valuation des joueurs qui ont choisi la même ressource. Le calcul des pénalités peut s'effectuer localement ce qui constitue un gain en complexité. Reprenons l'exemple des mobiles qui doivent s'attacher à des cellules. Ce dernier résultat signifie que le calcul peut être effectué au niveau de chaque cellule. Si ce n'était pas le cas, par exemple si on choisit $h_u = 0$, il faudrait que les cellules puissent communiquer entre elles, ce qui est particulièrement complexe si elles ne sont pas de la même technologie.

Jeu de routage

Généralisons le modèle précédent au choix d'un sous-ensemble de ressources dans $\mathcal{R}$. Cela modélise un jeu de routage dans lequel les ressources sont les arcs d'un graphe, et les joueurs choisissent un ensemble d'arcs de manière à joindre deux sommets du graphe. Chaque joueur a une valuation sur chacune des ressources qu'il utilise. Nous notons $d^u_r(\ell_r)$ la valuation du joueur u sur la ressource r avec la charge ℓ_r.

Considérons le cas où la valuation de chaque joueur est la somme des valuations sur chacune des ressources qu'il a choisies, à l'image de la fonction du délai dans les

réseaux qui est la somme du délai sur chacune des arêtes empruntées. Cela se traduit par $v_u(s) = \sum_{r \in s_u} d_r^u(\ell_r(s))$. Comme dans le problème d'association précédent, on choisit $h_u(s_{-u}) = \sum_{i \neq u} v_i(\emptyset, s_{-u})$ ce qui donne finalement :

$$p_u(s) = \sum_{r \in s_u} \sum_{i \neq u | r \in s_u} \delta_{i,u}^r(s_{-u}),$$

avec $\delta_{i,u}^r(s_{-u}) \stackrel{\text{def}}{=} \big(d_r^i(0, \ell_r^{-u}(s_{-u})) - d_r^i(\ell_r(s))\big)$.

On remarque que la fonction de pénalité est la somme de fonctions de pénalité calculées sur chacune des ressources choisies, ce qui réduit une nouvelle fois la complexité de calcul. Cela repose de façon critique sur le fait que les valuations des joueurs sont la somme des valuations sur chaque ressource utilisée.

Supposons maintenant que la valuation des joueurs soit le minimum des valuations sur les ressources utilisées, *i.e.* $v_u(s) = \min_{r \in s_u} d_r^u(\ell_r(s))$. Cela modélise la fonction de débit dans les réseaux, le débit d'un chemin étant limité par le plus petit débit sur les arcs du chemin. Jusqu'à présent, nous ne sommes pas parvenus à trouver des fonctions (h_u) qui simplifient le calcul des fonctions de pénalité par rapport au choix $h_u = 0$.

Il est intéressant de remarquer que, dans les applications que nous venons de présenter, la pénalité que nous obtenons est une généralisation de la taxe pigouvienne [Pig52] : il s'agit d'intégrer dans les fonctions de gain des joueurs le coût qu'ils induisent pour les autres (mais pas uniquement à l'état optimal).

2.4 Un exemple où les deux approches échouent : inciter des joueurs à participer à une coalition

Nous terminons ce chapitre en étudiant le problème d'incitation dans le cadre des jeux de participation. Cette section vient illustrer les deux approches que nous avons proposées, mais n'est toutefois pas utilisée dans la suite du document.

2.4.1 Présentation du problème

Le problème est le suivant. Considérons un ensemble $\mathcal{U}$ de joueurs ayant la possibilité de participer à une *coalition*. Une coalition est un sous-ensemble des joueurs que nous notons $C \in \mathcal{P}(\mathcal{U})$. Si un joueur u choisit de ne pas participer à la coalition, alors il est assuré d'obtenir un gain v_u. Sinon, à chaque coalition est associée une valeur $\mathring{\mathrm{v}}(C)$, et l'on suppose que $\mathring{\mathrm{v}}(\emptyset) = 0$. La valeur de la coalition est ensuite partagée entre les joueurs. Le gain du joueur u dans la coalition C est ce qu'il reçoit du partage. On note $v_u(C)$ ce gain qui doit satisfaire la contrainte que la somme des gains ne dépasse pas la valeur de la

coalition :

$$\sum_{u \in C} v_u(C) \leq \mathring{\mathrm{v}}(C). \tag{2.2}$$

On dit que le partage est *efficace* si $\sum_{j \in C} v_j(C) = \mathring{\mathrm{v}}(C)$, c'est-à-dire si toute la valeur est redistribuée. Remarquons qu'un partage qui n'est pas efficace n'est pas non plus Pareto optimal (il existe un partage qui augmente strictement le gain d'un joueur sans détériorer celui des autres). On s'attachera donc à construire des mécanismes reposant sur des partages efficaces. Notons également que, aussi bien la valeur des coalitions que les gains des joueurs peuvent être négatifs.

La donnée des valeurs des coalitions, c'est-à-dire de la fonction $\mathring{\mathrm{v}} : \mathcal{P}(\mathcal{U}) \to \mathbb{R}$ définit un jeu de coalition avec des *utilités transférables*, comme introduit dans [MVN47], en raison du fait que la valeur de la coalition peut être répartie de n'importe quelle manière entre ses membres. On peut voir la valeur d'une coalition comme une somme d'argent que les joueurs peuvent se partager. La fonction de valeur $\mathring{\mathrm{v}}$ est générale : on ne suppose pas qu'elle est super-additive, ce qui impliquerait que la *grande coalition*, c'est-à-dire la coalition qui contient tous les joueurs, soit socialement optimale. Notons également que une et une seule coalition peut être formée : des joueurs qui ne sont pas dans la coalition courante ne peuvent pas se grouper pour former une deuxième coalition.

Une question centrale posée par les jeux de coalition est de répartir les gains de la valeur des coalitions de façon "équitable". Le terme équitable doit bien entendu être précisé : la première idée est d'attribuer à un joueur sa contribution marginale à la coalition C qui vaut $\mathring{\mathrm{v}}(C) - \mathring{\mathrm{v}}(C \backslash \{u\})$, c'est-à-dire la valeur ajoutée par le joueur u à la coalition. Néanmoins, il est possible que la somme $\sum_{u \in C} \mathring{\mathrm{v}}(C) - \mathring{\mathrm{v}}(C \backslash \{u\})$ soit plus grande que la valeur de la coalition, ou si elle satisfait cette contrainte, que la répartition ne soit pas efficace. Dans le cas d'utilités transférables, le partage des gains selon la valeur de Shapley fait consensus. Nous le détaillons dans la suite.

L'utilisation des jeux de coalition dans des problèmes liés aux systèmes multi-agents en général, et les systèmes de télécommunications en particulier a été croissante ces dernières années. La récente thèse [Saa10] présente de nombreuses applications. Alors que les questions soulevées dans cette thèse portent en grande partie sur la caractérisation des coalitions stables, et la formation de ces coalitions, nous soulevons la question de l'existence de mécanismes d'incitation comme nous les avons définis (en stratégies dominantes ou par fonction de potentiel) qui permettent d'atteindre une coalition donnée.

Afin d'inciter les joueurs à former une coalition donnée, l'opérateur peut intervenir sur les partages des gains des coalitions. Le jeu de coalition peut alors être vu comme un jeu fini dans lequel chaque joueur a deux actions qui consistent à participer ou ne pas participer à la coalition. Nous appelons ce jeu un *jeu de participation*.

Dans un jeu de participation, le gain du joueur u est soit v_u si il ne participe pas, soit $v_u(C(s))$, où $C(s)$ est la coalition formée à partir du profil d'actions s. Du point de vue de l'opérateur, les gains doivent être partagés de telle manière que soit la coalition visée correspond à une stratégie dominante du jeu de participation, soit le jeu de participation

admet une fonction de potentiel qui est maximale en cette coalition. Notons que ce problème d'incitation est très proche de celui des jeux finis de la section 2.3.2, à deux différences près :

- l'opérateur doit établir les gains en satisfaisant la contrainte (2.2),
- l'opérateur ne peut pas pénaliser les joueurs s'ils ne participent pas à la coalition.

2.4.2 Mécanisme d'incitation en stratégies dominantes

Tout d'abord, notons que l'on peut toujours implémenter la coalition vide en stratégies dominantes en fixant les valeurs des joueurs qui participent à la coalition à un niveau suffisamment bas pour qu'il soit toujours plus intéressant de ne pas participer. Dans ce cas, le partage des gains des coalitions a de bonnes chances de ne pas être efficace.

Mais en général, n'importe quel profil d'actions ne peut pas correspondre à un équilibre en stratégies dominantes, en particulier celui qui maximise la somme des gains des joueurs, comme le montre l'exemple qui suit.

Considérons trois joueurs pouvant former une coalition, les valeurs des coalitions étant données par :

- s'ils sont seuls ou en dehors de la coalition, les joueurs ont un gain de 1,
- toute coalition constituée de deux joueurs a la valeur 0,
- et la grande coalition, c'est-à-dire avec les trois joueurs, a une valeur supérieure ou égale à 3.

Alors, il n'existe pas de répartition de la valeur des coalitions de manière à ce que la grande coalition soit un équilibre en stratégies dominantes du jeu. Cela résulte tout simplement du fait qu'au moins un joueur a un gain négatif s'il participe à une coalition de deux joueurs, ce qui est inférieur à son gain lorsqu'il ne participe pas.

2.4.3 Mécanisme d'incitation par fonction de potentiel

Nous montrons ici que l'on peut établir un mécanisme d'incitation par fonction de potentiel, mais que cela n'est possible que pour une fonction de potentiel particulière. Le résultat repose sur le partage des gains d'une coalition selon la valeur de Shapley.

Valeur de Shapley d'une coalition

La valeur de Shapley correspond à une façon de partager la valeur d'une coalition entre ses membres en fonction de la fonction de valeur $\mathring{\mathrm{v}} : \mathcal{P}(\mathcal{U}) \to \mathbb{R}$ dans sa totalité.

Dans l'article initial [Sha97], Shapley donne trois axiomes qu'un partage de gain équitable doit vérifier. De façon surprenante, ces trois axiomes simples définissent de façon unique le partage des gains. De plus, ce partage est calculable. Les trois axiomes sont les suivants :

1. le partage est efficace,
2. le partage est anonyme, *i.e.* toute permutation sur les joueurs de la coalition donne lieu au même partage (en particulier, deux joueurs du même type auront le même gain dans une coalition),

3. le partage de la somme de deux fonctions de valeur est la somme des partages de chacune des fonctions.

L'unique fonction de partage, appelée valeur de Shapley, qui vérifie ces trois axiomes est donnée pour tout $C \in \mathcal{P}(\mathcal{U})$ par :

$$v_u(C) = \sum_{D \subseteq C | u \in S} \frac{(d-1)!(c-d)!}{c!} \big(\mathring{\mathrm{v}}(D) - \mathring{\mathrm{v}}(D \backslash \{u\})\big), \tag{2.3}$$

avec $c = |C|$ et $d = |D|$. La valeur de Shapley fait donc intervenir la valeur marginale des joueurs de la coalition $\big(\mathring{\mathrm{v}}(D) - \mathring{\mathrm{v}}(D \backslash \{u\})\big)$, pour chaque sous-coalition D avec le coefficient $\dfrac{(d-1)!(c-d)!}{c!}$ qui définit une loi de probabilité sur l'ensemble des sous-coalitions.

Existence et unicité d'une fonction de potentiel

Aucun mécanisme d'incitation en stratégies dominantes ne peut être obtenu en général. Pour l'implémentation via les jeux de potentiel on a le résultat suivant :

Théorème 2.5 (*Existence d'une fonction de potentiel*)

Le jeu de participation est un jeu de potentiel *si et seulement si* le partage des coalitions est la valeur de Shapley (2.3) de la fonction de valeur $\mathring{\mathrm{v}}$. Dans ce cas, le potentiel vaut (à une constante additive près) pour tout $s \in \mathcal{S}$:

$$P(s) = \sum_{D \subseteq C(s)} \frac{(d-1)!(c-d)!}{c!} \mathring{\mathrm{v}}(D) - \sum_{u \in C(s)} v_u,$$

avec $c = |C(s)|$ et $d = |D|$.

La preuve de ce résultat repose sur le théorème 6.1 dans [MS96b], dont la preuve découle elle-même du théorème A dans [HMC89]. En effet, ce dernier théorème caractérise la valeur de Shapley d'une coalition comme la contribution marginale des joueurs dans la coalition pour une certaine fonction qui s'avère être notre fonction de potentiel.

Ce résultat est à la fois positif car il donne un mécanisme d'incitation grâce à la fonction de potentiel, et à la fois négatif car la seule fonction de potentiel pour laquelle il existe un mécanisme d'incitation est celle donnée dans le théorème.

Notons que, étant donnée une coalition C, si l'on choisit au hasard une sous-coalition $D \subseteq C$ de la façon suivante :

– on choisit le cardinal d de D uniformément entre 1 et $c = |C|$,

– on choisit un ensemble de d joueurs dans C uniformément, parmi les $\binom{C}{d}$ possibilités,

alors la fonction de potentiel se réécrit :

$$P(s) = \sum_{u \in C(s)} \mathbb{E}[\frac{v(D)}{d} - v_u].$$

Le terme $\mathbb{E}[\frac{v(D)}{d} - v_u]$ s'interprète comme la moyenne de la différence du gain moyen par joueur dans une sous coalition D tirée au hasard avec la valeur du joueur quand il choisit de ne pas participer. Le potentiel d'une coalition fait donc intervenir la valeur de toutes les sous-coalitions, et on ne peut pas s'attendre à ce que la valeur maximale du potentiel corresponde à la coalition qui maximise la somme des gains, comme le montre l'exemple suivant.

Supposons qu'il y ait trois joueurs. Leur gain lorsqu'ils ne participent pas à la coalition ou lorsqu'ils sont seuls vaut 1, vaut 0 s'ils sont deux dans la coalition, et $3k$ pour la grande coalition, avec $k \geq 0$. Comme les joueurs sont symétriques, la répartition des gains dans chaque coalition, donnée par la valeur de Shapley, est équitable entre les joueurs de la coalition. Elle vaut donc 1, 0 ou k selon le cardinal de la coalition, et le potentiel correspondant vaut 0, -1 et $k-2$. Par conséquent, la valeur maximale du potentiel est 0 si $k \leq 2$ qui est obtenue pour la coalition vide. Si $k = 2 - \epsilon$ alors le gain social de la coalition vide vaut 3 et vaut $6 - 3\epsilon$ pour la grande coalition, ce qui est supérieure à 3 si $\epsilon > 0$ est suffisamment petit. Donc l'état qui maximise le potentiel et celui qui maximise le gain social ne sont pas les mêmes.

CHAPITRE 3

JEUX DE POTENTIEL ET MODÈLES D'APPRENTISSAGE

Résumé du chapitre

Lorsque l'on modélise une situation de compétition (routage compétitif, partage de ressources) par un jeu, c'est dans le but d'en connaître la ou les issues possibles. La détermination des résultats d'un jeu est un sujet central de la théorie des jeux, et nécessite de modéliser le comportement des joueurs.

Dans ce chapitre, on aborde cette question par l'apprentissage c'est-à-dire l'adaptation des joueurs à la répétition du jeu. Les résultats du jeu sont alors les profils d'actions qui sont joués après un grand nombre d'itérations. L'existence d'une fonction de potentiel pour le jeu, que l'on peut obtenir par des mécanismes d'incitation (voir le chapitre précédent), permet souvent de caractériser précisément les résultats de l'apprentissage.

L'objectif de ce chapitre est de donner les résultats de base sur les jeux de potentiel et les modèles d'apprentissage qui seront développés dans les chapitres suivants. Une part importante du chapitre est consacrée aux jeux avec des espaces de stratégies continus, en particulier à l'extension mixte des jeux finis. Cette extension permet de définir des modèles d'apprentissage sur des stratégies aléatoires, et également de prendre en compte les incertitudes sur les gains du jeu lorsque ceux-ci correspondent à des mesures physiques.

L'organisation du chapitre est la suivante : les deux premières sections sont consacrées aux résultats généraux sur les jeux à espace de stratégies finis et continus, et en particulier sur les jeux de potentiel. La troisième section est une introduction aux modèles d'apprentissage. Nous terminons par l'étude de deux modèles simples d'apprentissage et de leurs résultats dans le cas des jeux de potentiel.

3.1 Définitions et résultats généraux sur les jeux

Dans cette section, nous introduisons les notations et les premières définitions utilisées dans la suite du document.

3.1.1 Jeux finis

Nous considérons des jeux finis donnés sous leur forme normale, c'est-à-dire des jeux qui s'expriment comme un triplet $(\mathcal{U}, \mathcal{S}, c)$ où :

- $\mathcal{U}$ désigne un ensemble fini de joueurs,
- $\mathcal{S} \stackrel{\text{def}}{=} (\mathcal{S}_u)_{u \in \mathcal{U}}$, où $\mathcal{S}_u$ est l'ensemble des actions du joueur u que l'on suppose non vide et fini. De plus, on note n_u le cardinal de $\mathcal{S}_u$,
- $c \stackrel{\text{def}}{=} (c_u)_{u \in \mathcal{U}}$, où $c_u : \mathcal{S} \to \mathbb{R}$ est la fonction de gain du joueur u.

$s \in \mathcal{S}$ est appelé un *profil d'actions*. Lorsque le profil d'actions résulte d'une stratégie des joueurs, on parle directement de *stratégie*. La fonction de gain de chaque joueur associe une valeur réelle à chaque profil d'actions. Nous supposons dans toute la suite du chapitre (et du document) que les joueurs cherchent à *maximiser* leur gain.

Notons qu'il existe de nombreuses situations de compétition en informatique, et particulièrement dans le domaine des réseaux, dans lesquelles les joueurs ont plusieurs critères de préférences (par exemple la qualité de service dans les réseaux repose sur le débit, le délai, etc...). Cependant, lorsque le jeu présente plusieurs critères par joueur, il n'existe pas de notion générale d'équilibre. Nous nous bornons donc à un critère unique (éventuellement, plusieurs critères peuvent être agrégés en un seul à l'aide de pondérations).

L'hypothèse que $\mathcal{U}$ est fini se justifie dans la plupart des applications que nous développerons par la suite, même si le nombre de joueurs peut être grand. On parle alors de jeu *atomique* : cela signifie que la décision de chaque joueur peut avoir un impact sur le gain des autres joueurs. Néanmoins, lorsque le nombre de joueurs est grand, cet impact tend à devenir nul. Sous certaines hypothèses, les jeux *non-atomiques* avec un continuum de joueurs, dans lesquels la décision de chaque joueur ne modifie pas les gains des autres joueurs, fournissent une bonne approximation. Les jeux non-atomiques sont généralement plus faciles à analyser que les jeux atomiques avec un grand nombre de joueurs, par l'utilisation d'outils d'analyse fonctionnelle (voir la thèse [Bou04]).

Comme dans le chapitre précédent, nous utilisons la notation classique $-u$ pour désigner l'ensemble des joueurs excepté le joueur u. Pour distinguer l'action d'un joueur particulier, on notera parfois (a, s_{-u}) au lieu de s avec $s_u = a$.

On définit la correspondance (fonction multivaluée) de *meilleure réponse* du joueur u à un profil de stratégies des autres joueurs $\mathrm{BR}_u : \mathcal{S}_{-u} \to \mathcal{S}_u$ par $\mathrm{BR}_u(s_{-u}) \stackrel{\text{def}}{=} \operatorname*{argmax}_{a \in \mathcal{S}_u} c_u(a, s_{-u})$.

Dans les jeux, chaque joueur cherche à maximiser son gain. Cependant, il ne contrôle qu'une partie du résultat final. La notion de stratégie optimale n'est pas clairement définie, excepté dans le cas d'un joueur unique. Aussi, le résultat du jeu dépend des hypothèses sur le comportement des joueurs. La notion d'*équilibre de Nash* est fondamentale. Rappelons la définition :

Définition 3.1 (*Equilibre de Nash en stratégies pures*)
$s \in \mathcal{S}$ est un équilibre de Nash en stratégies pures du jeu $(\mathcal{U}, \mathcal{S}, c)$ si $\forall u \in \mathcal{U}, s_u \in \mathrm{BR}_u(s_{-u})$.

En d'autres termes, un profil d'actions est un équilibre de Nash si aucun des joueurs ne peut améliorer son gain en changeant d'action *unilatéralement*. Un profil d'actions qui n'est pas un équilibre de Nash n'est pas stable car au moins un joueur a intérêt à changer d'action. Notons que l'on pourrait imposer plus que la stabilité par rapport aux déviations unilatérales dans la définition de l'équilibre. Cependant, lorsque les joueurs ne peuvent pas communiquer, ils n'ont pas la possibilité de s'entendre pour améliorer conjointement leur gain.

Un équilibre de Nash est défini comme un point fixe de la correspondance de meilleure réponse. Ce point fixe n'existe pas en général. Prenons l'exemple d'un jeu à deux joueurs. Il peut se représenter sous la forme d'une matrice $m \times n$ où l'action du premier joueur est le choix de l'une des m lignes et l'action de l'autre joueur est le choix de l'une des n colonnes d'une matrice. Les éléments de la matrice sont des couples de valeurs, la première valeur étant le gain du premier joueur (le joueur qui choisit les lignes par exemple) et la deuxième étant le gain du deuxième joueur. Par exemple, si chaque joueur doit choisir parmi deux actions, le jeu se présente sous la forme suivante :

Matrice de gains

(a, A)	(b, B)
(c, C)	(d, D)

Ce jeu n'a pas d'équilibre de Nash en stratégies pures si, par exemple, $c > a$, $D > C$, $b > d$, et $A > B$. En effet, dans ce cas, aucune stratégie n'est un point fixe de la correspondance de meilleure réponse.

Comme nous l'avons vu au chapitre précédent, les jeux dans lesquels chaque joueur a une stratégie dominante ont au moins un équilibre de Nash. Néanmoins, déterminer s'il existe un équilibre de Nash (en stratégies pures) dans un jeu général donné sous sa forme normale est un problème NP-complet (voir [GGS03]). Par contre, lorsque l'on étend le jeu à des stratégies aléatoires, ce que nous appelons l'extension mixte du jeu, il existe toujours un équilibre de Nash. On est alors dans le cadre des jeux continus.

3.1.2 Jeux continus et extension mixte des jeux finis

Il existe des situations d'accès à des ressources qui impliquent des prises de décision dans un ensemble continu d'actions. En informatique, l'exemple typique est le routage compétitif dans les réseaux. Chaque joueur doit choisir une ou plusieurs routes afin de transmettre un flux d'informations d'un point du réseau à un autre point (voir, par exemple, le modèle dans [ORS93]). En général, plusieurs, voire même un grand nombre de routes joignent ces deux points. Deux cas de figure apparaissent :

1. les flux sont indivisibles, et doivent transiter sur une route unique. C'est le cas lorsque les paquets d'information doivent arriver dans le même ordre que celui dans lequel ils ont été envoyés, comme par exemple, dans les communications téléphoniques. Dans cette situation, l'ensemble des décisions est fini, et la situation de compétition se modélise par un jeu fini.
2. les flux peuvent être divisés, et chaque portion envoyée sur une route différente. Dans ce cas, les joueurs choisissent la proportion du flux qu'ils envoient sur chacune des routes. Cette prise de décision se fait donc dans un ensemble continu.

Nous commençons par décrire le cadre général des jeux à espace de stratégies continu, et les conditions suffisantes pour qu'il existe un équilibre de Nash. Ensuite, nous détaillons une classe particulière de jeux continus utilisée dans la suite du chapitre : il s'agit des jeux obtenus comme extension mixte des jeux finis. L'extension mixte peut être vue comme une extension des stratégies déterministes à des stratégies aléatoires. Cela nous permettra ultérieurement de définir des modèles d'apprentissage qui reposent sur des stratégies aléatoires, ce qui autorise plus de souplesse et de robustesse (au bruit par exemple) que les seuls modèles en stratégies déterministes.

Jeux continus

Comme précédemment, l'ensemble des joueurs $\mathcal{U}$ est fini.

On suppose qu'il existe des entiers $(n_u)_{u\in\mathcal{U}}$ tels que l'ensemble des actions du joueur u est compris dans un ensemble convexe et compact de $\mathbb{R}^{n_u}$ noté $\mathcal{X}_u$. Dans le cas du routage, n_u correspond au nombre de routes possibles, et $\mathcal{X}_u$ est l'ensemble des façons de partager un flux sur ces routes :

$$\mathcal{X}_u = \{x \in \mathbb{R}^{n_u} | x \geq 0 \text{ et } \sum_{r=1}^{n_u} x_r = 1\}.$$

On note $\mathcal{X}$ l'espace global des stratégies. En général, cet espace n'est pas le produit des espaces de stratégies de chaque joueur c'est-à-dire $\underset{u\in\mathcal{U}}{\times} \mathcal{X}_u$. En effet, des contraintes peuvent interdire certains profils d'actions. Dans le cas du routage de flux de données, les contraintes proviennent, par exemple, de la capacité limitée sur les liens ou les nœuds du réseau. De ce fait, l'ensemble des actions d'un joueur dépend des actions prises par les autres joueurs. Finalement, un profil d'actions est un vecteur $x = (x_u)_{u\in\mathcal{U}} \in \mathcal{X}$, et $\forall u \in \mathcal{U}, x_u \in \mathcal{X}_u$.

Pour chaque joueur, une fonction de gain $f_u : \mathcal{X} \to \mathbb{R}$ est donnée, et on note $f(x) = (f_u(x))_{u\in\mathcal{U}}$ [1]. La fonction f est l'équivalent de la fonction de gain c dans les jeux discrets. Le triplet $(\mathcal{U}, \mathcal{X}, f)$ définit un jeu continu.

La fonction de meilleure réponse d'un joueur par rapport à la stratégie des autres joueurs est définie par $\mathrm{BR}_u(x_{-u}) \overset{\text{def}}{=} \underset{y\in\mathcal{X}_u|(y,x_{-u})\in\mathcal{X}}{\operatorname{argmax}} f_u(y, x_{-u})$, où l'on suppose que le maximum est

1. Dans le cas de l'extension mixte des jeux finis, la fonction de gain est l'espérance des gains du jeu fini que nous avons noté c. Pour distinguer les valeurs du jeu fini de leur espérance, nous utilisons une notation différente, c'est-à-dire f, pour désigner les gains sur des espaces continus.

atteint, ce qui est le cas, par exemple, si les fonctions de gain $f_u(x)$ sont semi-continues supérieurement.

La définition des équilibres de Nash dans les jeux finis (aussi appelés équilibres de Nash en stratégies pures), se généralise naturellement de la façon suivante :

Définition 3.2 (*Equilibre de Nash*)

$x \in \mathcal{X}$ est un équilibre de Nash du jeu $(\mathcal{U}, \mathcal{X}, f)$ si $\forall u \in \mathcal{U}, x_u \in \mathrm{BR}_u(x_{-u})$.

Les équilibres de Nash du jeu sont les points fixes de la correspondance de meilleure réponse. Pour des espaces de stratégies continus, une condition suffisante d'existence de point fixe est donnée par le théorème de Kakutani [Kak41], qui est lui-même une généralisation du théorème de Brouwer pour les correspondances.

Certaines conditions d'existence d'un équilibre reposent sur la propriété de quasi-concavité des fonctions de gain. On dit qu'une fonction $g : \mathcal{X} \to \mathbb{R}$ est *quasi-concave* si ses sur-niveaux sont convexes, où le sur-niveau $t \in \mathbb{R}$ est l'ensemble $\{x \in \mathcal{X}$ tels que $g(x) \geq t\}$. En particulier, une fonction concave est quasi-concave. Le théorème de Kakutani se traduit par le théorème suivant [2] :

Théorème 3.3 (*Existence des équilibres de Nash [Nas50, Nas51]*)

Soit $(\mathcal{U}, \mathcal{X}, f)$ un jeu tel que $\mathcal{X}$ est convexe compact, pour tout joueur u, f_u est continue, et pour tout $x_{-u} \in \mathcal{X}_{-u}$ la fonction $y \mapsto f_u(y, x_{-u})$ est quasi-concave. Alors ce jeu admet au moins un équilibre de Nash.

Les hypothèses du théorème sont par exemple vérifiées dans les jeux de routage qui font intervenir des fonctions de gain concaves (ou de manière équivalente des fonctions de coût convexes), comme par exemple, les fonctions de délais données par des files d'attente M/M/1. Les hypothèses sont aussi vérifiées dans le cas de l'extension mixte des jeux finis que nous détaillons par la suite.

Notons également que, dans certaines classes de jeu, les équilibres de Nash peuvent être caractérisés autrement que comme solution d'une équation de point fixe. Par exemple, dans les jeux de potentiel que nous avons introduits au chapitre précédent, les équilibres de Nash sont les points où la dérivée du potentiel s'annule (ou, pour prendre en compte les contraintes de l'espace des stratégies, les points qui vérifient les conditions de Karush Kuhn Tucker). Dans ce cas, l'existence d'un équilibre découle de l'existence d'une valeur maximale de la fonction de potentiel qui est atteinte. Nous reviendrons sur ce point dans la section 3.2.

Un cas particulier de jeu continu : l'extension mixte des jeux finis

Dans la suite du document, nous nous intéressons particulièrement aux jeux continus obtenus par convexification de l'espace des stratégies d'un jeu fini, que l'on appelle communément *extension mixte* des jeux finis. La convexification permet de modéliser des

2. Le théorème suivant n'apparaît sous cette forme aussi générale dans les articles de Nash, bien que celui-ci utilise le théorème de Kakutani. En particulier l'extension aux ensembles de stratégies qui ne sont pas des ensembles produits provient de [Ros65].

stratégies aléatoires. En effet, toute loi de probabilité sur un l'ensemble fini des actions peut être vue comme des coordonnées barycentriques sur cet ensemble. Plus formellement, on appelle convexification d'un ensemble fini $\mathcal{E}$ à n éléments, l'ensemble noté $\Delta(\mathcal{E})$ défini par $\Delta(\mathcal{E}) \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \text{ tels que } \sum_{a\in\mathcal{E}} x_a = 1 \text{ et } \forall a \in \mathcal{E}, x_a \geq 0\ \}$.

Si l'on considère un jeu fini $(\mathcal{U}, \mathcal{S}, c)$, la convexification de l'ensemble fini des stratégies $\mathcal{S}_u$ du joueur u est appelé ensemble des *stratégies mixtes*. Cet ensemble est donc $\Delta(\mathcal{S}_u) = \{x_u \in \mathbb{R}^{n_u} \text{ tels que } \sum_{a\in\mathcal{S}_u} x_{u,a} = 1 \text{ et } \forall a \in \mathcal{S}_u, x_{u,a} \geq 0\}$. L'ensemble des stratégies mixtes d'un joueur est indépendant des actions des autres joueurs. Par conséquent l'espace global des stratégies noté $\Delta(\mathcal{S})$ [3] est défini par $\Delta(\mathcal{S}) \stackrel{\text{def}}{=} \times_{u\in\mathcal{U}} \Delta(\mathcal{S}_u)$.

Pour chaque joueur, l'ensemble $\Delta(\mathcal{S}_u)$ est l'intersection du sous-espace affine de $\mathbb{R}^{n_u}$, $\mathcal{H}_u \stackrel{\text{def}}{=} \{x_u \in \mathbb{R}^{n_u} \text{ tels que } \sum_{a\in\mathcal{S}_u} x_{u,a} = 1\}$ avec l'orthant positif $\{x_u \in \mathbb{R}^{n_u} \text{ tels que } x_u \geq 0\}$. Par conséquent, c'est un polytope convexe et compact, et l'ensemble produit $\Delta(\mathcal{S})$ est également un polytope convexe et compact.

En tant que polytope, $\Delta(\mathcal{S})$ possède des sommets [4]. Ces sommets sont appelés *stratégies pures*. Les stratégies pures sont en bijection naturelle avec l'ensemble des profils d'actions du jeu fini. En effet, il est facile de voir qu'une stratégie pure est un vecteur $x = (x_u)_{u\in\mathcal{U}} \in \Delta(\mathcal{S})$ tel que chaque vecteur x_u vaut zéro partout excepté en une composante qui vaut 1. On construit alors la bijection Φ entre l'ensemble des stratégies pures et l'ensemble des profils d'actions $\mathcal{S}$ du jeu fini en posant : pour toute stratégie pure x, $\Phi(x) = s \in \mathcal{S}$ si et seulement si pour tout $u \in \mathcal{U}$, $x_{u,a} = 1 \Leftrightarrow s_u = a$. Par la suite, on identifie, sans le préciser, chaque stratégie pure au profil d'actions obtenu par la bijection Φ.

Enfin, les fonctions de gain des joueurs dans l'extension mixte sont définies par :

$$f_u(x) = \sum_{a\in\mathcal{S}_u} x_{u,a} f_{u,a}(x), \text{ et } f_{u,a}(x) = \sum_{s_{-u}\in S_{-u}} c_u(a, s_{-u}) \prod_{v\in\mathcal{U}\setminus\{u\}} x_{v,s_v}. \tag{3.1}$$

Remarquons que si x est une stratégie pure correspondant au profil d'actions s du jeu fini, alors $f_u(x) = c_u(s)$. Il s'agit bien d'une extension du jeu fini.

Les stratégies mixtes peuvent également être vues comme une loi de probabilité sur l'ensemble des actions du jeu fini. Dans ce cas, le gain d'un joueur dans l'extension mixte s'interprète comme l'espérance de gain dans le jeu fini. Notons S_u l'action choisie aléatoirement par le joueur u sous la stratégie x_u, c'est-à-dire $\mathbb{P}[S_u = a] = x_{u,a}$, et $S = (S_u)_{u\in\mathcal{U}}$. On suppose que les variables aléatoires S_u sont indépendantes [5]. Ainsi, la

3. La notation $\Delta(\mathcal{S})$ est abusive car elle ne désigne pas le convexifié de l'ensemble $\mathcal{S}$, mais le produit des ensembles de stratégies mixtes. Néanmoins, elle permet de souligner qu'il s'agit de l'extension mixte d'un jeu fini.

4. Formellement, x est un sommet d'un polytope $\mathcal{X}$ si pour tout y et z dans $\mathcal{X}$, et pour tout $\lambda \in (0, 1)$, $x = \lambda y + (1 - \lambda)z \Rightarrow y = z = x$.

5. Cela suppose que chaque joueur dispose d'un générateur aléatoire indépendant de celui des autres. Dans un cadre plus général, on peut définir des stratégies corrélées et la notion d'équilibre adaptée (se

loi de probabilité de S est le produit des lois de S_u, *i.e.* $\mathbb{P}[S = s] = \prod_{u \in \mathcal{U}} x_{u,s_u}$.

Le gain du joueur u dépend du profil d'actions S qui a été choisi aléatoirement selon x. Ce gain est donc lui-même aléatoire, et son espérance est une fonction de x. Cette espérance vaut $\mathbb{E}[c_u(S)] = \sum_{a \in \mathcal{S}_u} \mathbb{P}[S_u = a]\mathbb{E}[c_u(S)|S_u = a]$. En remarquant que $\mathbb{E}[c_u(S)|S_u = a] = \sum_{s_{-u} \in S_{-u}} c_u(a, s_{-u}) \prod_{v \in \mathcal{U} \setminus \{u\}} x_{v,s_v}$, il vient, par l'équation (3.1), que $\mathbb{E}[c_u(S)] = f_u(x)$. Notons également que $\mathbb{E}[c_u(S)|S_u = a] = f_{u,a}(x)$, et que $f_{u,a}(x)$ ne dépend pas de x_u. La fonction $f_{u,a}(x)$ s'interprète comme l'espérance de gain du joueur u s'il choisit l'action a.

D'après la définition 3.2, les équilibres de Nash d'un jeu continu sont les points fixes de la correspondance de meilleure réponse. Dans l'extension mixte d'un jeu fini, cela se traduit par :

Proposition 3.4 (*Equilibre de Nash en stratégie mixte*)

$x^* \in \Delta(\mathcal{S})$ est un équilibre de Nash de l'extension mixte $(\mathcal{U}, \Delta(\mathcal{S}), f)$ du jeu fini $(\mathcal{U}, \mathcal{S}, c)$ si et seulement si $\forall u \in \mathcal{U}, x^*_{u,a} > 0 \Rightarrow f_{u,a}(x^*) = \max_{b \in \mathcal{S}_u} f_{u,b}(x^*)$.

Par conséquent, en situation d'équilibre, les joueurs ne jouent avec une probabilité non nulle que la ou les actions qui lui assurent une espérance de gain maximale. Il s'agit exactement de la définition d'un équilibre de Wardrop [War52] dans les jeux de routage non-atomique : à l'équilibre, les chemins empruntés sont ceux qui maximisent l'utilité des joueurs. Il s'agit d'une interprétation de l'extension mixte non plus en terme de probabilité dans les jeux avec un nombre fini de joueurs, mais en terme de proportion d'une population infinie de joueurs se répartissant sur les routes d'un réseau [6].

D'après l'équation (3.1), la fonction $y \mapsto f_u(y, x_{-u})$ est quasi-concave car linéaire. De plus, $f_u(x)$ est une fonction multi-affine (affine en chacune des variables), donc continue en x. Les hypothèses du théorème 3.3 sont vérifiées si bien que :

Corollaire 3.5 (*Existence d'un équilibre de Nash dans l'extension mixte*)

Dans tout jeu fini, il existe un équilibre de Nash en stratégie mixte.

Exemple : Soit le jeu à deux joueurs, tel que chaque joueur à deux actions possibles, dont la matrice des gains est donnée par :

	G	D
H	$(1,4)$	$(0,0)$
B	$(0,0)$	$(4,1)$

(3.2)

référer par exemple à la thèse [Aum87]). Les équilibres corrélés qui ne sont pas des équilibres de Nash nécessitent d'établir des signaux entre les joueurs (l'exemple classique de corrélation étant celui du feu tricolore aux carrefours : il s'agit d'un signal qu'il est préférable de suivre).

6. Voir également [San01] pour une interprétation des stratégies mixtes dans les jeux de population.

Dénotons par x (resp. y) la probabilité pour le joueur qui choisit les lignes (resp. colonnes) de jouer la ligne du haut H (resp. colonne de droite D). À la figure 3.1, on a tracé les correspondances de meilleure réponse du joueur "ligne" $\mathrm{BR}_1(y)$ et du joueur "colonne" $\mathrm{BR}_2(x)$. Les équilibres de Nash, définis comme points fixes des correspondances de meilleure réponse sont les points d'intersection des deux courbes, à savoir $(0, 1)$, $(1, 0)$ et $(0.2, 0.2)$.

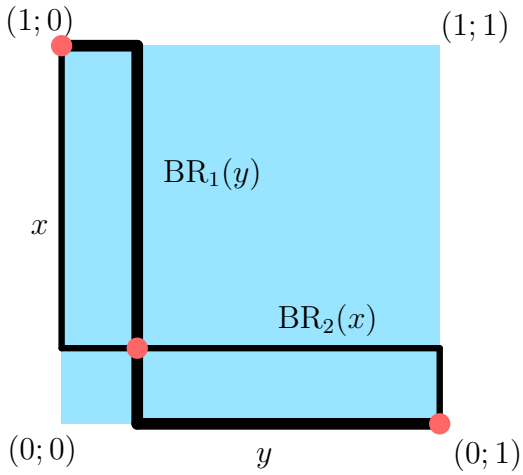


Figure 3.1 – Correspondances de meilleure réponse du jeu (3.2). Les équilibres de Nash sont les points d'intersection (en rouge) des deux courbes.

Caractérisation géométrique des équilibres de Nash dans l'extension mixte

Parallèlement à la définition en terme de point fixe, les équilibres de Nash peuvent être caractérisés par une relation géométrique. Cette caractérisation s'avérera utile pour étendre la notion de potentiel aux jeux continus. Elle repose sur la notion de *cône normal*, et du lien qui existe entre ces cônes et les conditions nécessaires d'optimalité.

Dans le cas où la frontière de l'ensemble est lisse, le cône normal en un point de la frontière est l'ensemble des directions orthogonales au plan tangent qui sortent de l'ensemble. En général :

Définition 3.6 (*Cône normal à un ensemble*)

On appelle cône normal en $x \in \mathbb{R}^n$ à un ensemble convexe $\mathcal{X}$ de $\mathbb{R}^n$, dont le produit scalaire est noté $\langle \cdot, \cdot \rangle$, l'ensemble défini par :

$$N_{\mathcal{X}}(x) \stackrel{\text{def}}{=} \{d \in \mathbb{R}^n \text{ tels que } \forall y \in \mathcal{X}, \langle d, y - x \rangle \leq 0\}.$$

On vérifie immédiatement que l'on a toujours $0 \in N_{\mathcal{X}}(x)$, que $N_{\mathcal{X}}(x)$ est un cône, c'est-à-dire que si $d \in N_{\mathcal{X}}(x)$, alors $\lambda d \in N_{\mathcal{X}}(x)$ pour tout $\lambda > 0$, et même, que c'est un cône convexe. Des exemples dans $\mathbb{R}^2$ sont donnés à la figure 3.2. Le cône normal en x est aussi

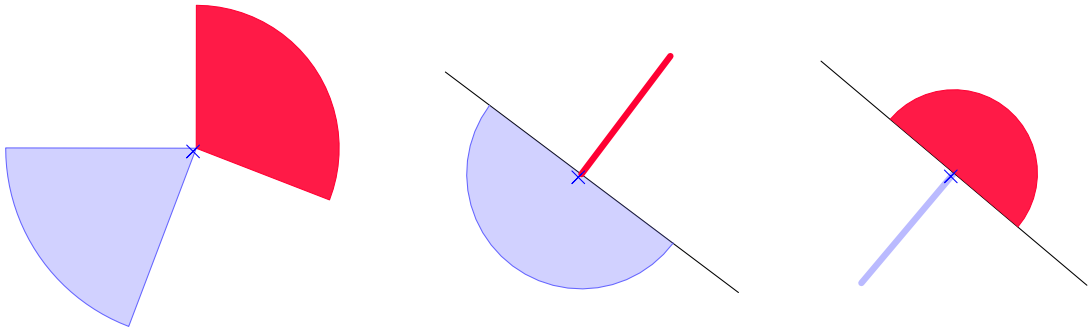


Figure 3.2 – Exemples de cônes normaux dans $\mathbb{R}^2$ pour le produit scalaire canonique. En bleu clair l'ensemble convexe, et en rouge foncé son cône. De gauche à droite : le cône normal au sommet d'un secteur angulaire est un secteur angulaire, le cône normal au centre d'un demi-disque est une demi-droite, le cône normal à l'extrémité d'un segment est un demi-plan.

égal (à une translation par x près) à l'ensemble des points dont la projection orthogonale est x.

Dans le cas d'espaces continus, le calcul différentiel permet de donner des conditions d'optimalité d'une fonction différentiable. Si une fonction f est différentiable en x, on note par $\nabla f(x)$ son gradient pour un produit scalaire donné. Rappelons que le gradient $\nabla f(x)$ est l'unique vecteur représentant [7] l'application linéaire différentielle pour ce produit scalaire. Dans le cas où les dérivées partielles de f existent en x, le gradient associé au produit scalaire canonique est le vecteur des dérivées partielles.

La proposition suivante est une version géométrique des conditions nécessaires d'optimalité de Kuhn Tucker, qui repose sur la notion de cône normal.

Proposition 3.7 (*Conditions d'optimalité du premier ordre*)

Si x^* est un maximum local sur un ensemble $\mathcal{X}$ convexe d'une fonction réelle f définie sur espace euclidien, et si f est différentiable en x^*, alors $\nabla f(x^*) \in N_{\mathcal{X}}(x^*)$, où $N_{\mathcal{X}}(x^*)$ est le cône normal à $\mathcal{X}$ en x^*.

Cela nous permet de donner une caractérisation géométrique des équilibres de Nash dans l'extension mixte d'un jeu (la preuve est directe pour vérifier que les conditions sont suffisantes).

Proposition 3.8 (*Caractérisation géométrique des équilibres de Nash*)

x^* est un équilibre de Nash de l'extension mixte $(\mathcal{U}, \Delta(\mathcal{S}), f)$ si et seulement si

$$\left(f_{u,a}(x^*)\right)_{u\in\mathcal{U}, a\in\mathcal{S}_u} \in N_{\Delta(\mathcal{S})}(x^*).$$

Lorsque $f : \underset{u\in\mathcal{U}}{\times} \mathbb{R}^{n_u} \to \mathbb{R}$, on note $\nabla_u f(x)$ le gradient de la fonction dans les directions

7. L'existence et l'unicité du gradient sont une conséquence du théorème de représentation de Riesz.

incluses dans $\mathbb{R}^{n_u}$, c'est-à-dire le gradient de la fonction $x_u \mapsto f(x_u, x_{-u})$.

Exemple : Reprenons l'exemple du jeu (3.2). À la figure 3.3, on représente les cônes normaux à l'espace des stratégies mixtes (que l'on identifie à $[0;1] \times [0;1]$). Les vecteurs $\nabla_1 f_1(x)$ et $\nabla_2 f_2(x)$ sont chacun de dimension deux et leur produit est de dimension quatre. Pour résoudre ce problème de représentation, on a tracé en plusieurs points de l'espace des stratégies le produit des projections de $\nabla_1 f_1(x)$ et $\nabla_2 f_2(x)$ sur la droite d'équation $x_1 + x_2 = 1$. Nous justifierons ultérieurement que la proposition 3.8 reste valide dans ce cas (c'est-à-dire que x^* est un équilibre de Nash si et seulement si le gradient projeté des fonctions de gain en x^* est dans le cône normal en x^*).

Finalement, les équilibres de Nash correspondent aux points où les vecteurs représentés appartiennent au cône normal. On retrouve bien les mêmes équilibres que ceux donnés par les points fixes de la correspondance de meilleure réponse à la figure 3.1.

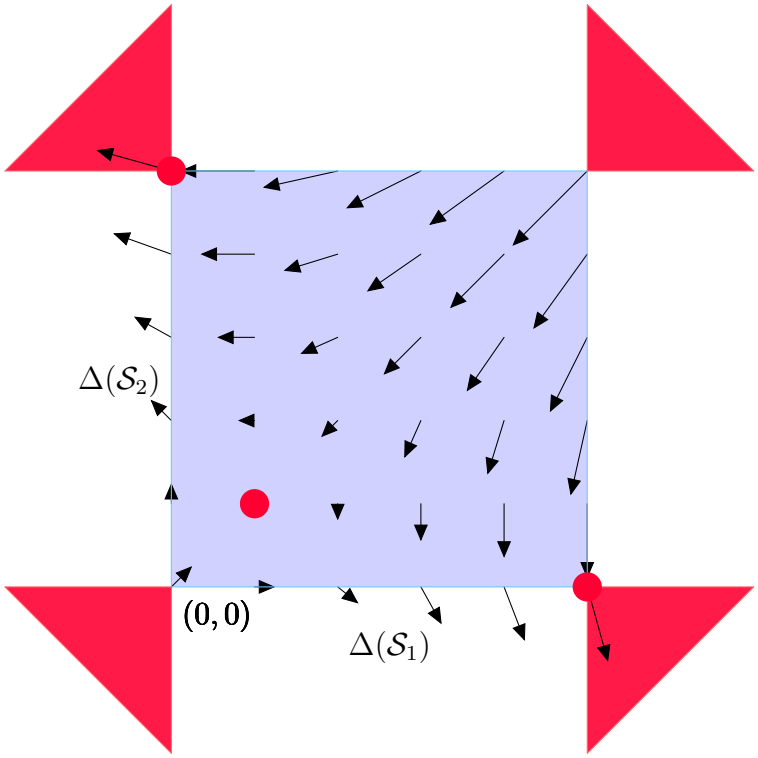


Figure 3.3 – Jeu (3.2) à deux joueurs, chacun ayant 2 actions. L'espace des stratégies de chaque joueur peut s'identifier au segment $[0,1]$. On représente en bleu clair l'ensemble des stratégies, et en rouge foncé les cônes normaux en plusieurs points : aux sommets du carré, il s'agit de quarts de plan, aux arêtes (entre les sommets) ce sont des demi-droites, et pour les points à l'intérieur, le cône est réduit à $\{0\}$. Les vecteurs représentent le gradient (en fait la projection des gradients) des fonctions de gain. Les équilibres de Nash $(1,0)$, $(0,1)$ et $(0.2, 0.2)$ sont les points où le vecteur est inclus dans le cône normal.

3.2 Jeux de potentiel

Les jeux de potentiel sont caractérisés par l'existence d'une fonction réelle définie sur l'espace des stratégies qui agrège les fonctions de gain de chaque joueur : les gains d'un joueur peuvent être interprétés comme la contribution marginale de ce joueur au potentiel. Intuitivement, le fait d'agréger tous les gains en une seule fonction ramène le jeu à un problème d'optimisation classique. Cela laisse présager le fait que les jeux de potentiel forment une classe de jeux ayant des propriétés fortes de convergence dans la plupart des modèles d'apprentissage.

De plus, comme nous l'avons vu à la section 2.3.2, les jeux de potentiels peuvent être obtenus à partir de n'importe quel jeu fini à l'aide de mécanismes d'incitation, éventuellement distribués, qui reposent sur l'instauration de pénalités.

Dans cette section, nous détaillons les propriétés principales des jeux de potentiel finis. Nous proposons ensuite une extension aux jeux continus.

3.2.1 Fonction de potentiel dans les jeux finis

Rappelons la définition d'un jeu de potentiel :

Définition 3.9 (*Jeu de potentiel [MS96b]*)

$(\mathcal{U}, \mathcal{S}, c)$ est un jeu de potentiel si $\exists P : \mathcal{S} \to \mathbb{R}$ telle que :

$$\forall u \in \mathcal{U}, \forall s_{-u} \in \mathcal{S}_{-u}, \forall a, b \in \mathcal{S}_u, c_u(a, s_{-u}) - c_u(a, s_{-u}) = P(a, s_{-u}) - P(b, s_{-u}). \quad (3.3)$$

Une classe de jeux de potentiel : les jeux de congestion

C'est dans les jeux de congestion [Ros73] que la notion de fonction de potentiel apparaît pour la première fois. Dans les jeux de congestion, un ensemble fini de ressources $\mathcal{R}$ est donné, et l'action de chaque joueur est le choix d'un sous-ensemble de ressources, *i.e.* $s_u \in \mathcal{P}(\mathcal{R})$. Le gain d'un joueur est la somme des gains sur chaque ressource choisie, et le gain provenant d'une ressource r ne dépend que du nombre de joueurs $\ell_r(s)$ ayant choisi cette ressource. En particulier, le gain ne dépend pas du joueur. Si le gain s'écrit $g_r(\ell_r(s))$ sur la ressource r, alors $c_u(s) = \sum_{r \in s_u} g_r(\ell_r(s))$.

Il est connu, et cela se vérifie directement, que ce jeu possède une fonction de potentiel donnée par :

$$P(s) = \sum_{r \in \mathcal{R}} \sum_{i=1}^{\ell_r(s)} g_r(i).$$

La classe des jeux de congestion est particulièrement importante pour la modélisation du routage dans les réseaux : les ressources correspondent aux nœuds et aux liens du réseau (routeurs et canaux de communication), et chaque joueur choisit un sous-ensemble de ces ressources de manière à former une route entre une source et une destination. Le gain d'un

joueur est par exemple le délai pour transmettre un message de la source à la destination qui est la somme des délais sur chacun des liens du chemin emprunté.

Notons que la complexité du calcul des équilibres dans les jeux de congestion est connue pour être PLS-complet (voir [Voc06] pour un aperçu des résultats à ce sujet). Tout jeu de congestion est un jeu de potentiel. A l'inverse, il est connu que tout jeu de potentiel est isomorphique à un jeu de congestion (voir le théorème 8 dans [Pot06]). Cependant, les bijections connues ne sont pas des transformations polynomiales, si bien que les résultats de complexité des jeux de congestion ne s'appliquent pas directement aux jeux de potentiel généraux.

Chemins d'amélioration

Une propriété importante des jeux de potentiel porte sur les *chemins d'amélioration*. Un chemin d'amélioration est une suite de profils d'actions $(s^1, s^2, \dots)$ tels que s^i et s^{i+1} diffèrent en au plus une composante, disons u, et tels que le gain du joueur u est *strictement* amélioré : $c_u(s^{i+1}) > c_u(s^i)$. Si un chemin d'amélioration est fini, le dernier profil d'actions est nécessairement un équilibre de Nash (en stratégies pures). Néanmoins il se peut que certains chemins d'amélioration soient finis et que d'autres ne le soient pas (ce qui implique l'existence de cycles).

Définition 3.10 (*Propriété des chemins d'amélioration finis [MS96b]*)

Un jeu fini a la propriété des chemins d'amélioration finis si *tous* les chemins d'amélioration sont finis.

Tout jeu qui a la propriété des chemins d'amélioration finis a donc un équilibre de Nash en stratégies pures. La réciproque n'est en revanche pas vrai.

Il est intéressant de noter que, ici, l'existence des équilibres de Nash ne repose pas sur un argument local de point fixe, mais sur l'étude des points limites d'une trajectoire donnée par des améliorations strictes du gain des joueurs. Une telle trajectoire peut être vue comme le résultat d'un processus d'apprentissage.

Il est clair que les jeux de potentiel vérifient la propriété des chemins d'amélioration finis, et de ce fait admettent toujours un équilibre de Nash. Il existe des jeux qui vérifient cette propriété mais qui ne sont pas des jeux de potentiel. La classe des jeux caractérisée par cette propriété est la classe des jeux de potentiel généralisés (nous reviendrons sur ce point dans la section 3.4).

Une fonction de potentiel pour les coalitions

Jusqu'à présent, nous avons défini les équilibres de Nash, les fonctions de potentiel et les chemins d'amélioration à partir de contraintes sur les déviations unilatérales. Imposer la même contrainte sur des déviations de groupes d'au plus k joueurs aboutit à des propriétés de stabilité des équilibres plus fortes (si de tels équilibres existent). En particulier, quand $k = 2$, on parle de stabilité par paires (voir par exemple [AJM07] dans le cas d'un jeu de formation de réseau).

Le cas extrême est lorsqu'un jeu de potentiel vérifie l'égalité (3.3) (ou une forme un peu plus faible connue sous le nom de potentiel ordinal) par déviation de n'importe quel sous-ensemble de joueurs. Dans ce cas, il existe un équilibre de Nash fort (défini dans [Aum99]) qui est stable par déviation de n'importe quelle coalition de joueurs.

Cette classe de jeux de potentiel est très restreinte. Il existe cependant une classe intéressante de jeux qui possèdent une telle fonction de potentiel : il s'agit des jeux de congestion à goulot d'étranglement. De la même manière que dans les jeux de congestion, chaque joueur choisit un ensemble de ressources. La différence réside dans le fait que le gain d'un joueur n'est pas la somme des gains sur les ressources choisies mais le minimum. En particulier, ces jeux modélisent des situations de routage dans lesquelles le gain d'un joueur est son débit. Contrairement au délai qui est une fonction additive sur les liens du réseau, le débit d'un chemin est égal au débit minimal sur les liens qui le composent.

Comme cela est démontré dans [HHKS10], les jeux de congestion à goulot d'étranglement admettent une fonction de potentiel qui vérifie la contrainte d'égalité du potentiel par déviation de n'importe quel ensemble de joueurs. Ces jeux possèdent donc une structure plus forte que les jeux de congestion classiques. Il est par contre plus complexe d'établir des mécanismes d'incitation distribués dans ce cas (voir la remarque à la fin de la section 2.3.3).

3.2.2 Fonction de potentiel dans les jeux continus

La notion de jeu de potentiel se généralise de manière naturelle, moyennant quelques détails techniques, aux jeux dont l'espace des stratégies est continu, en particulier les jeux obtenus par extension mixte des jeux finis. Intuitivement, un jeu admet un potentiel s'il existe une fonction réelle différentiable sur l'espace des stratégies dont la dérivée dans une direction de l'espace des stratégies d'un joueur est égale à la dérivée de la fonction de gain de ce joueur dans la même direction. Par conséquent, tout déplacement unilatéral d'un joueur dans la direction de plus grande pente se fait également dans la direction de plus grande pente de la fonction de potentiel. Il s'agit d'une généralisation du cas fini dans lequel la dérivée (discrète) est la différence de valeur entre deux profils d'actions qui diffèrent en une composante.

La différentielle d'une fonction réelle f de classe $\mathcal{C}^1$ sur un espace euclidien tel $\mathbb{R}^n$ en un point x est l'unique application linéaire notée $df(x)$ de $\mathbb{R}^n$ dans $\mathbb{R}$ telle que :

$$\forall h \in \mathbb{R}^n, f(x+h) = f(x) + df(x)(h) + o(h),$$

quand h tend vers 0 (indépendamment de la norme utilisée). Par le théorème de représentation de Riesz, on sait qu'il existe un unique vecteur de $\mathbb{R}^n$ appelé gradient de f en x, et noté $\nabla f(x)$, tel que $\forall h \in \mathbb{R}^n, df(x)(h) = \langle \nabla f(x), h \rangle$. Le gradient dépend du produit scalaire utilisé, et lorsqu'il s'agit du produit scalaire canonique, le gradient est le vecteur des dérivées partielles. Quand f est définie sur l'espace produit $\underset{u \in \mathcal{U}}{\times} \mathbb{R}^{n_u}$, on note $\nabla_u f(x)$ le gradient au point x de la fonction $y \mapsto f(y, x_{-u})$, c'est-à-dire la fonction f restreinte à la variable x_u.

Soit un jeu continu $(\mathcal{U}, \mathcal{X}, f)$ tel que $\mathcal{X} = \underset{u \in \mathcal{U}}{\times} \mathcal{X}_u$, avec $\mathcal{X}_u \subset \mathbb{R}^{n_u}$, et tel que $f = (f_u)_{u \in \mathcal{U}}$

est de classe $\mathcal{C}^1$ sur $\underset{u\in\mathcal{U}}{\times} \mathbb{R}^{n_u}$. La manière classique (voir [San01, BC09]) de définir un jeu de potentiel dans ce cas est :

Définition 3.11 (*Jeu de potentiel continu*)

$(\mathcal{U}, \mathcal{X}, f)$ est un jeu de potentiel s'il existe une fonction $F : \underset{u\in\mathcal{U}}{\times} \mathbb{R}^{n_u} \to \mathbb{R}$ de classe $\mathcal{C}^1$ telle que

$$\forall x \in \mathcal{X}, \forall u \in \mathcal{U}, \nabla_u F(x) = \nabla_u f_u(x).$$

Cette définition implique que les dérivées directionnelles de F et de f_u sont égales quelle que soit la direction dans $\mathbb{R}^{n_u}$. Cependant, cela peut être trop restrictif dans le sens où l'ensemble des directions dans $\mathcal{X}_u$ est plus petit que dans $\mathbb{R}^{n_u}$, par exemple dans le cas où c'est un sous-espace affine de dimension inférieure à n_u. On peut en fait réduire l'ensemble des directions au plus petit sous-espace affine contenant $\mathcal{X}_u$. Cela pose néanmoins un problème car notre définition de la différentielle suppose que l'espace de définition est ouvert, ce qui n'est plus le cas avec un sous-espace affine.

Définition du potentiel à partir des directions de plus grande pente

Indépendamment de ce problème de différentiabilité, une condition raisonnable pour qu'une fonction soit le potentiel d'un jeu continu est que ses directions de plus grande pente coïncident avec les directions de plus grande pente des fonctions de gain des joueurs. En un sens, la direction de plus grande pente est l'analogie de la correspondance de meilleure réponse dans les jeux finis.

Définition 3.12 (*Direction de plus grande pente*)

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction de classe $\mathcal{C}^1$, $\mathcal{H}_0$ un sous-espace vectoriel de $\mathbb{R}^n$, et soit $x \in \mathbb{R}^n$. On dit que $d(x) \in \mathcal{H}_0$ est une direction de plus grande pente de f en x dans $\mathcal{H}_0$ si $\exists \eta > 0$ tel que $\forall \varepsilon \in (0, \eta)$, et $\forall v \in \mathcal{H}_0$ de même norme que $d(x)$, on a $f(x + \varepsilon d(x)) \geq f(x + \varepsilon v)$.

Considérons un sous-espace affine $\mathcal{H}$ général. On peut toujours le définir par $\mathcal{H} = \{x \in \mathbb{R}^n$ tels que $Ax = b\}$, où A est une matrice réelle de taille $m \times n$ avec $m < n$, de rang plein, et $b \in \mathbb{R}^m$. Notons également $\mathcal{H}_0 \overset{\text{def}}{=} \mathrm{Ker}(A) = \{x \in \mathbb{R}^n$ tels que $Ax = 0\}$ le sous-espace vectoriel correspondant. On note $\mathrm{Proj}_{\mathcal{H}_0} : \mathbb{R}^n \to \mathcal{H}_0$ la projection orthogonale sur $\mathcal{H}_0$ pour le produit scalaire courant.

Le choix naturel de direction de plus grande pente en x dans le sous espace affine $\mathcal{H}$, que nous notons par $\nabla f_{|\mathcal{H}}(x)$ est [8] :

$$\forall x \in \mathbb{R}^n, \nabla f_{|\mathcal{H}}(x) = \mathrm{Proj}_{\mathcal{H}_0}(\nabla f(x)). \tag{3.4}$$

8. Cette définition correspond à la définition du gradient $\nabla f_{|\mathcal{H}}(x)$ de la fonction f restreinte au sous-espace affine $\mathcal{H}$, ce qui justifie la notation. Comme le sous-espace tangent de $\mathcal{H}$ est $\mathcal{H}_0$, le gradient est caractérisé par les deux conditions suivantes :
   - pour tout $x \in \mathcal{H}$, $\nabla f_{|\mathcal{H}}(x) \in \mathcal{H}_0$,
   - pour tout $x \in \mathcal{H}$ et $v \in \mathcal{H}_0$, $df(x)(v) = \langle \nabla f_{|\mathcal{H}}(x), v\rangle$.

Ces deux conditions impliquent que $\nabla f_{|\mathcal{H}}(x) = \mathrm{Proj}_{\mathcal{H}_0}(\nabla f(x))$.

Bien évidemment, $\mathrm{Proj}_{\mathcal{H}_0}(\nabla f(x)) \in \mathcal{H}_0$. Vérifions que $\mathrm{Proj}_{\mathcal{H}_0}(\nabla f(x))$ est bien une direction de plus grande pente. Posons $d = \mathrm{Proj}_{\mathcal{H}_0}(\nabla f(x))$ et $g_d(\varepsilon) = f(x+\varepsilon d) - f(x) - \langle \nabla f(x), d\rangle = o(\varepsilon)$. Si $d = 0$ alors c'est une direction de plus grande pente d'après la définition. Sinon, comme f est différentiable en x, si $v \in \mathcal{H}_0$ est de même norme que d et $v \neq d$, on a :

$$\begin{aligned}
& \langle \mathrm{Proj}_{\mathcal{H}_0}(\nabla f(x)), d\rangle > \langle \mathrm{Proj}_{\mathcal{H}_0}(\nabla f(x)), v\rangle \\
\Rightarrow\ & \langle \nabla f(x), d\rangle > \langle \nabla f(x), v\rangle \\
\Rightarrow\ & \exists \eta > 0, \forall \varepsilon \in (0,\eta), \langle \nabla f(x), d\rangle + \frac{g_d(\varepsilon)}{\varepsilon} \geq \langle \nabla f(x), v\rangle + \frac{g_v(\varepsilon)}{\varepsilon} \\
\Rightarrow\ & \exists \eta > 0, \forall \varepsilon \in (0,\eta), f(x) + \varepsilon\langle \nabla f(x), d\rangle + g_d(\varepsilon) \geq f(x) + \varepsilon\langle \nabla f(x), v\rangle + g_v(\varepsilon) \\
\Rightarrow\ & \exists \eta > 0, \forall \varepsilon \in (0,\eta), f(x+\varepsilon d) \geq f(x+\varepsilon v),
\end{aligned}$$

la dernière ligne étant ce qu'il fallait démontrer.

La matrice A étant de rang plein, la projection sur $\mathcal{H}_0$ est donnée par $\mathrm{Proj}_{\mathcal{H}_0}(x) = \big(I - A^T(AA^T)^{-1}A\big)x$, où I est la matrice identité de taille n, et A^T désigne la matrice transposée de A. Finalement :

$$\forall x \in \mathcal{H}, \nabla f_{|\mathcal{H}}(x) = \big(I - A^T(AA^T)^{-1}A\big)\nabla f(x). \tag{3.5}$$

Retournons aux jeux continus. Par abus de notation, lorsque F est définie sur l'espace produit $\underset{u\in\mathcal{U}}{\times} \mathbb{R}^{n_u}$, on écrit $\nabla F_{|\mathcal{H}_u}$ au lieu de $\nabla_u F_{|\mathcal{H}}$. En toute rigueur, $\nabla F_{|\mathcal{H}_u}$ est le vecteur $\nabla_u F_{|\mathcal{H}}$ complété par des zéros sur ses autres composantes (les deux vecteurs n'ont pas la même taille). De même, pour la lisibilité, on écrira $\nabla f_{|\mathcal{H}_u}$ au lieu de $\nabla_u f_{u|\mathcal{H}}$.

Cela nous amène à introduire une nouvelle définition des jeux de potentiel pour les jeux continus :

Définition 3.13 (*Jeu de potentiel au sens faible*)

Soit $(\mathcal{U}, \mathcal{X}, f)$ un jeu continu. C'est un jeu de potentiel au sens faible s'il existe une fonction $F : \underset{u\in\mathcal{U}}{\times} \mathbb{R}^{n_u} \to \mathbb{R}$ de classe $\mathcal{C}^1$ telle que $\forall u \in \mathcal{U}$, il existe un sous-espace affine $\mathcal{H}_u$ qui contient $\mathcal{X}_u$ tel que $\forall x \in \mathcal{X}$:

$$\nabla F_{|\mathcal{H}_u}(x) = \nabla f_{|\mathcal{H}_u}(x). \tag{3.6}$$

Si l'égalité précédente est vérifiée avec $\mathcal{H}_u = \mathbb{R}^{n_u}$ pour tout u, cela s'écrit finalement $\nabla_u F(x) = \nabla_u f(x)$ comme dans la définition classique[9]. D'après le théorème de Schwarz, on a :

Proposition 3.14 (*Caractérisation des jeux de potentiel*)

Le jeu $(\mathcal{U}, \mathcal{X}, f)$, avec f de classe $\mathcal{C}^2$, est un jeu de potentiel continu si et seulement si :

$$\forall x \in \mathcal{X}, \forall u, v \in \mathcal{U}, \forall a \in \mathcal{S}_u, \forall b \in \mathcal{S}_v, \frac{\partial^2 f_u}{\partial x_{v,b}\partial x_{u,a}}(x) = \frac{\partial^2 f_v}{\partial x_{u,a}\partial x_{v,b}}(x).$$

9. Dans ce cas, l'égalité (3.6) est vérifiée pour tout sous-espace affine contenant $\mathcal{X}$.

Dans le cas continu, le terme "fonction de potentiel" prend son sens physique si on interprète le vecteur $\nabla_u f_u(x)$ comme une force qui s'exerce sur la stratégie x_u de chaque joueur. Dans ce cas, l'existence d'une fonction de potentiel revient à dire que la force en question est conservative, et donc que le travail produit par cette force entre deux points est indépendant du chemin suivi. Ici, si le point $x \in \mathcal{X}$ se déplace à la vitesse $\dot{x}$, le travail de la force est $\langle \nabla f(x), \dot{x} \rangle$. De façon formelle, si on note $\nabla f_{|\mathcal{H}} = (\nabla f_{|\mathcal{H}_u})_{u \in \mathcal{U}}$, on a :

Théorème 3.15 (*Calcul d'une fonction de potentiel*)

Si $(\mathcal{U}, \mathcal{X}, f)$ est un jeu de potentiel continu, si y est un point quelconque de $\mathcal{X}$, et $\mathcal{H}$ le plus petit sous-espace affine qui contient $\mathcal{X}$, on pose :

$$F(x) = K + \int_0^1 \langle \nabla f_{|\mathcal{H}}(z(t)), \dot{z}(t) \rangle dt,$$

où $z : [0 : 1] \to \mathcal{X}$ est une fonction (chemin) de classe $\mathcal{C}^1$ telle que $z(0) = y$ et $z(1) = x$, et K est un nombre réel quelconque. Alors $F(x)$ ne dépend pas du choix du chemin z, et c'est une fonction de potentiel du jeu.

Démonstration : $(\mathcal{U}, \mathcal{X}, f)$ étant un jeu de potentiel, et $\mathcal{H}$ le plus petit sous-espace affine contenant $\mathcal{X}$, il existe une fonction réelle G telle que $\forall z \in \mathcal{X}, \nabla G_{|\mathcal{H}}(z) = \nabla f_{|\mathcal{H}}(z)$. Or $\langle \nabla G_{|\mathcal{H}}(z(t)), \dot{z}(t) \rangle = \frac{d}{dt}\big(G_{|\mathcal{H}}(z(t))\big)$, si bien que $\int_0^1 \langle \nabla f_{|\mathcal{H}}(z(t)), \dot{z}(t) \rangle dt = G_{|\mathcal{H}}(x) - G_{|\mathcal{H}}(y) = G(x) - G(y)$. Donc, en choisissant F comme indiqué dans le théorème, on a $F(x) = G(x) + K - G(y)$, où $K - G(y)$ est constant, d'où l'on déduit que $\nabla F_{|\mathcal{H}}(z) = \nabla G_{|\mathcal{H}}(z)$ et le fait que F est une fonction de potentiel. ■

Cette démonstration prouve également que toute fonction de potentiel s'écrit comme l'intégrale, le long de n'importe quel chemin, du travail de la force $\nabla f_{|\mathcal{H}}(z)$. Cela implique que la fonction de potentiel est unique à une constante additive près (la valeur de K).

De plus, le théorème donne un critère qui permet de montrer qu'un jeu n'est pas un jeu de potentiel. Il suffit de montrer qu'il existe deux chemins z et z' de x à y dans $\mathcal{X}$, tels que $\int_0^1 \langle \nabla f_{|\mathcal{H}}(z(t)), \dot{z}(t) \rangle dt \neq \int_0^1 \langle \nabla f_{|\mathcal{H}}(z'(t)), \dot{z'}(t) \rangle$. Ce critère s'applique également pour les jeux finis, mais pour des chemins discrets.

Le potentiel dans l'extension mixte des jeux finis

Nous étudions maintenant le cas particulier de l'extension mixte des jeux finis, notamment comment l'équation (3.6) s'écrit dans ces jeux.

Notons que les fonctions de gain de l'extension mixte ne sont définies que sur le produit de simplexes $\Delta(\mathcal{S})$. Cependant, elles s'étendent naturellement au produit des sous-espaces affines $\mathcal{H} = \underset{u \in \mathcal{U}}{\times} \mathcal{H}_u$, où $\mathcal{H}_u = \{y \in \mathbb{R}^{n_u} \text{ tels que } Ay = 1\}$, avec A une matrice ligne de taille n_u dont toutes les composantes valent 1. D'après l'équation (3.5),

$\forall x \in \mathcal{X}, \forall u \in \mathcal{U}, \mathrm{Proj}_{\mathcal{H}}(\nabla_u f_u(x)) = \left(f_{u,a}(x) - \frac{1}{n_u} \sum_{b \in \mathcal{S}_u} f_{u,b}(x) \right)_{a \in \mathcal{S}_u}$. La projection du gradient de la fonction de gain du joueur u sur $\mathcal{H}$ est donc la différence entre le gain espéré $f_{u,a}(x)$ en choisissant l'action a et la moyenne arithmétique $\frac{1}{n_u} \sum_{b \in \mathcal{S}_u} f_{u,b}(x)$ des gains espérés sur l'ensemble des actions. Par conséquent :

Proposition 3.16

L'extension mixte $(\mathcal{U}, \Delta(\mathcal{S}), f)$ d'un jeu fini est un jeu de potentiel (au sens faible) si et seulement s'il existe $F : \underset{u \in \mathcal{U}}{\times} \mathbb{R}^{n_u} \to \mathbb{R}$ de classe $\mathcal{C}^1$ telle que :

$$\forall x \in \Delta(\mathcal{S}), \forall u \in \mathcal{U}, \nabla F_{|\mathcal{H}_u}(x) = \left(f_{u,a}(x) - \frac{1}{n_u} \sum_{b \in \mathcal{S}_u} f_{u,b}(x) \right)_{a \in \mathcal{S}_u}.$$

Dans les jeux finis, on a vu que les profils d'actions qui maximisent la fonction de potentiel sont des équilibres de Nash. On a une propriété équivalente dans le cas de l'extension mixte :

Théorème 3.17 (*Maximums locaux de la fonction de potentiel*)

Soit $(\mathcal{U}, \Delta(\mathcal{S}), f)$ l'extension mixte d'un jeu fini $(\mathcal{U}, \mathcal{S}, c)$. Si l'extension mixte admet une fonction de potentiel F de classe $\mathcal{C}^1$ sur le sous-espace affine $\mathcal{H}$, et si x^* est un maximum local de F sur $\mathcal{X}$, alors x^* est un équilibre de Nash de l'extension mixte.

Pour cela, nous utiliserons le lemme suivant :

Lemme 3.18

Soit $F : \mathbb{R}^n \to \mathbb{R}$. $\nabla F(x) \in N_{\Delta(\mathcal{S})}(x)$ si et seulement si $\nabla F_{|\mathcal{H}}(x) \in N_{\Delta(\mathcal{S})}(x)$, où $\mathcal{H} \supseteq \Delta(\mathcal{S})$ est un sous-espace affine de $\mathbb{R}^n$.

Démonstration (Lemme) : $\nabla F(x) \in N_{\Delta(\mathcal{S})}(x)$ équivaut à $\forall y \in \Delta(\mathcal{S})$, $\langle \nabla F(x), y - x \rangle \leq 0$. Comme $y - x \in \mathcal{H}$, on a $\langle \nabla F(x), y - x \rangle = \langle \mathrm{Proj}_{\mathcal{H}}(\nabla F(x)), y - x \rangle$, d'où le résultat. ■

Démonstration (Théorème 3.17) : Les conditions géométriques d'optimalité du premier ordre de la proposition 3.7 imposent que si x^* est un maximum local de F sur $\mathcal{X}$, alors $\nabla F(x^*) \in N_{\Delta(\mathcal{S})}(x^*)$ et donc, par le lemme précédent, $\nabla F_{|\mathcal{H}}(x^*) \in N_{\Delta(\mathcal{S})}(x^*)$. Puisque F est une fonction de potentiel, $\forall u \in \mathcal{U}, \nabla f_{\mathcal{H}}(x^*) \in N_{\Delta(\mathcal{S})}(x^*)$. Le lemme précédent implique que pour tout $u \in \mathcal{U}, \nabla_u f_u(x^*) \in N_{\Delta(\mathcal{S})}(x^*)$. Enfin, le résultat découle de la proposition 3.8. ■

Ce théorème n'est pas vrai en général pour les jeux continus. En effet, n'importe quel jeu à un seul joueur admet la fonction de gain du joueur comme fonction de potentiel. Cette fonction peut avoir plusieurs maximum locaux, alors que les équilibres de Nash sont les maximums globaux.

L'existence d'un maximum de la fonction de potentiel F (qui est continue) sur l'ensemble compact $\Delta(\mathcal{S})$ implique, par le théorème précédent, qu'il existe un équilibre de Nash. Il

s'avère en fait que le jeu fini, dont on considère l'extension mixte, est également un jeu de potentiel. Même, la classe des jeux de potentiel finis et la classe des jeux dont l'extension mixte est un jeu de potentiel sont les mêmes.

Proposition 3.19 (*Équivalence du potentiel du jeu fini et de l'extension mixte*)

Soit un jeu fini $(\mathcal{U}, \mathcal{S}, c)$ et $(\mathcal{U}, \Delta(\mathcal{S}), f)$ son extension mixte. Le jeu fini est un jeu de potentiel si et seulement si l'extension mixte est un jeu de potentiel. De plus, si $P : \mathcal{S} \to \mathbb{R}$ est une fonction de potentiel du jeu fini, alors $F : \underset{u\in\mathcal{U}}{\times} \mathbb{R}^{n_u} \to \mathbb{R}$ défini par $F(x) = \sum_{s\in\mathcal{S}} P(s) \prod_{u\in\mathcal{U}} x_{u,s_u}$ (ou par $F(x) = \mathbb{E}[P(S)]$ sur $\mathcal{X}$) est une fonction de potentiel pour l'extension mixte.

Démonstration : Ce résultat généralise la proposition 4 dans [BC09] pour des jeux de potentiel définis de façon plus générale.

Si l'extension mixte est un jeu de potentiel, alors d'après le théorème 3.15, il existe une fonction de potentiel $F(x) = \int_0^1 \langle \nabla f_{|\mathcal{H}}(z(t)), \dot{z}(t)\rangle dt$, où $z(t)$ est un chemin reliant un point quelconque y à x dans $\mathcal{H}$. Ici, cela s'écrit $F(x) = \int_0^1 \sum_{u\in\mathcal{U}} \sum_{a\in\mathcal{S}_u} \big(f_{u,a}(z(t)) - \bar{f}_u(z(t))\big)\dot{z}_{u,a}(t)dt$, où $\bar{f}_u(x) = \frac{1}{n_u} \sum_{b\in\mathcal{S}_u} f_{b,u}(x)$. En particulier, prenons $y = (a, s_{-u})$, $x = (b, s_{-u})$ deux stratégies pures, avec $a, b \in \mathcal{S}_u$, et $z : [0;1] \to \Delta(\mathcal{S})$ un chemin inclus dans l'arête (du polytope $\Delta(\mathcal{S})$) qui joint y à x. Alors $\dot{z}_{v,.} = 0$ pour tout $v \neq u$, et $\dot{z}_{u,\alpha} = 0$ pour tout $\alpha \neq a, b$. Notons $\bar{c}_u(s_{-u})$ la moyenne arithmétique des $\big(c_u(\alpha, s_{-u})\big)_{\alpha\in\mathcal{S}_u}$, alors :

$$F(x) = F(y) + \big(c_u(a, s_{-u}) - \bar{c}_u(s_{-u})\big) \int_0^1 \dot{z}_{u,a}(t)dt + \big(c_u(b, s_{-u}) - \bar{c}_u(s_{-u})\big) \int_0^1 \dot{z}_{u,b}(t)dt,$$

car, pour tout $\alpha \in \mathcal{S}_u$, $f_{u,\alpha}(z(t))$ est constant le long du chemin et vaut $c_u(\alpha, s_{-u})$. Comme $\int_0^1 \dot{z}_{u,a}(t)dt = -1$ et $\int_0^1 \dot{z}_{u,a}(t)dt = 1$, on obtient $F(b, s_{-u}) - F(a, s_{-u}) = c_u(b, s_{-u}) - c_u(a, s_{-u})$. Donc la fonction F restreinte aux stratégies pures est une fonction de potentiel du jeu fini.

Montrons la réciproque. Supposons que le jeu fini ait une fonction de potentiel P. On définit $F(x) = \sum_{s\in\mathcal{S}} P(s) \prod_{u\in\mathcal{U}} x_{u,s_u}$ sur $\underset{u\in\mathcal{U}}{\times} \mathbb{R}^{n_u}$. Alors $F(x) = \sum_{u\in\mathcal{U}} \sum_{a\in\mathcal{S}_u} x_{u,a} F_{u,a}(x)$, où $F_{u,a}(x) = \sum_{s_{-u}} P(a, s_{-u}) \prod_{v\neq u} x_{v,s_v}$ qui ne dépend pas de x_u. De plus, comme P est une fonction de potentiel, on a $F_{u,a}(x) - F_{u,b}(x) = f_{u,a}(x) - f_{u,b}(x)$. Au final :

$$\nabla F_{|\mathcal{H}_u}(x) = \left(\frac{1}{n_u}\Big(\sum_{b\in\mathcal{S}_u} (F_{u,a}(x) - F_{u,b}(x))\Big)\right)_{a\in\mathcal{S}_u} = \nabla f_{|\mathcal{H}_u}(x).$$

■

Par analogie avec les jeux finis, un chemin d'amélioration dans l'extension mixte peut se définir comme une trajectoire continue où, en chaque point, la direction suivie par

chaque joueur est une direction dans laquelle son gain espéré est strictement croissant. Plus formellement, un chemin d'amélioration est une trajectoire $(x(t))$ telle que pour tout t et pour tout $u \in \mathcal{U}$, $\langle \frac{dx_u}{dt}(t), \nabla f_{|\mathcal{H}_u}(x(t)) \rangle > 0$. Dans ce cas, il est clair que, s'il existe une fonction de potentiel, alors le potentiel est strictement croissant et converge. Néanmoins, cela n'implique pas que la trajectoire converge vers un équilibre de Nash, car contrairement au cas discret, le nombre d'états est infini. Il se pose de plus le problème de la construction de telles trajectoires. Une idée consiste à suivre les directions de plus grande en chaque point, mais on se confronte alors au problème que les trajectoires peuvent sortir de l'ensemble des stratégies mixtes. Néanmoins, ce problème peut être résolu en utilisant des métriques appropriées. Cela fait l'objet du chapitre 5 sur les dynamiques de meilleures réponses dans l'extension mixte des jeux finis.

3.3 L'apprentissage dans les jeux

Cette section est une introduction générale aux modèles d'apprentissage dans les jeux.

Alors que la théorie des jeux classique tente de prédire le résultat d'un jeu à partir de la rationalité des joueurs et de leur analyse stratégique du jeu, l'apprentissage modélise le comportement de joueurs dont la stratégie évolue au cours du temps en fonction de l'expérience qu'ils acquièrent lorsque le jeu est répété. Étudier un modèle d'apprentissage revient alors à déterminer les profils d'actions que l'on peut observer à la suite "d'un grand nombre de répétitions". Il s'agit donc d'un résultat asymptotique.

Un modèle d'apprentissage peut se définir de manière informelle comme une règle qui, pour chaque joueur, définit sa stratégie à chaque itération du jeu en fonction de l'historique de ses observations et de sa connaissance du jeu. Bien souvent, cette règle est très simple, et consiste pour chaque joueur à choisir une action qui est une meilleure réponse pour un certain critère, comme par exemple son gain courant.

La littérature sur l'apprentissage s'est beaucoup développée depuis quelques dizaines d'années. Une référence sur le sujet est par exemple [FL98]. Notons que la plupart des résultats portent sur l'analyse du modèle d'apprentissage, et beaucoup moins sur la modélisation en elle-même. Même dans des jeux simples de coordination entre deux joueurs humains, il n'existe pas de modèle d'apprentissage qui soit en adéquation avec ce que l'on observe expérimentalement (voir [DWV07]).

Lien entre l'apprentissage dans les jeux et les algorithmes décentralisés

Dans les applications que nous développerons dans la suite du document, les règles de décision associées à un modèle d'apprentissage sont "codées" dans des machines. De ce fait, le terme "processus" d'apprentissage est plus adapté que celui de "modèle", et dans ce cas, un tel processus peut être apparenté à un *algorithme décentralisé*. Ce processus, à l'instar des modèles d'apprentissage, est contraint par les informations disponibles au niveau des machines/joueurs.

Notre sujet n'est pas l'apprentissage à proprement parler, mais plutôt l'optimisation. Néanmoins, il y a plusieurs raisons de s'intéresser à l'apprentissage dans ce contexte [10] :

- dans certains cas (en particulier dans les jeux de potentiel), les modèles d'apprentissage fournissent une solution algorithmique d'optimisation,
- de plus, les techniques d'analyse peuvent être adaptées pour l'analyse d'algorithmes décentralisés,
- le résultat de l'apprentissage possède souvent des propriétés de stabilité dans le jeu que n'aurait pas nécessairement un algorithme décentralisé quelconque,
- les règles d'apprentissage sont la plupart du temps simples et intuitives, ce qui est favorable pour leur implémentation pratique.

La connaissance du jeu et la capacité d'observation des joueurs

Lorsqu'un jeu est donné sous sa forme normale, on ne précise pas quelles sont les connaissances des joueurs sur ce jeu. En particulier, il se peut que les joueurs ne connaissent même pas le nombre total de joueurs. D'autre part, lors de la répétition du jeu, la capacité des joueurs à observer les actions jouées ainsi que les gains des autres joueurs varie selon les situations. Aussi bien la connaissance du jeu que la capacité d'observation sont cruciales pour bien modéliser le comportement des joueurs (ou pour écrire un processus d'apprentissage).

La situation typique dans laquelle des joueurs sont impliqués dans un jeu répété est le trafic routier. Les travaux pionniers de la théorie des jeux dans les transports datent de Wardrop [War52]. Prenons l'exemple d'un automobiliste qui, tous les jours, se rend de son domicile à son lieu de travail et doit choisir une route pour minimiser son temps de parcours. Celui-ci partage la route avec d'autres automobilistes qui ont également pour but d'avoir le temps de parcours le plus faible possible. Si notre automobiliste connaît, avant de prendre sa voiture, le temps de parcours sur tous les itinéraires possibles, celui-ci empruntera bien évidemment l'itinéraire le plus court indépendemment de son expérience passée. Sinon, il doit choisir son itinéraire à partir d'autres informations, qui relèvent essentiellement de la connaissance qu'il a acquise par expérience. A l'extrême, s'il connaît à l'avance les itinéraires de tous les autres automobiliste et sait que tous les automobilistes ont cette information (et qu'ils savent que tout le monde sait...), alors il pourra calculer, dans la mesure de la complexité du calcul, l'équilibre (de Wardrop) sans avoir à l'apprendre.

Dans le cas du routage de flux d'informations, comme par exemple le choix d'un point d'accès sans fil, l'expérience des usagers s'acquière le temps d'une communication et donc sur une échelle de temps beaucoup plus courte que celle des réseaux routiers. Les usagers peuvent modifier leur point d'accès durant une communication (en utilisant le handover vertical dans le cas de changement de technologie) et donc estimer leur gain (qualité de service) sur chacun. L'observation se limite exclusivement, pour chaque joueur, à l'observation de son gain.

10. Notons que la frontière entre la théorie des jeux et l'informatique est de plus en plus ténue, en témoigne le livre récent [Nis07]. Notre démarche suit cette tendance.

3.3.1 Quantifier le coût de l'apprentissage

L'objectif de l'analyse d'un modèle d'apprentissage est de prévoir la stratégie qui sera jouée après un grand nombre d'itérations. Mais bien souvent, la stratégie n'est pas unique, et de plus, elle dépend des conditions initiales. Dans ce cas, l'analyse tente de donner des propriétés que doit satisfaire une stratégie pour être jouée asymptotiquement : par exemple être un équilibre de Nash, ou bien un extremum de la fonction de potentiel.

Supposons maintenant que l'on connaisse l'ensemble des résultats possibles d'un jeu, la question de savoir par quel mécanisme on y arrive étant mise de côté. Supposons également qu'une fonction réelle sur l'ensemble des profils d'actions soit donnée, et représente, par exemple, les préférences d'un opérateur (comme dans le modèle décrit au chapitre précédent). Une question naturelle est de calculer le coût *maximal* de l'apprentissage pour cette fonction, c'est-à-dire le rapport entre la valeur maximale sur l'ensemble de tous les profils d'actions, qui est obtenue par un contrôle centralisé, et la valeur minimale sur les résultats de l'apprentissage.

Prenons l'exemple du routage dans un réseau, dans lequel les joueurs cherchent à maximiser leur qualité de service (débit, délai...), alors que l'opérateur souhaite maximiser un certain critère de performance (par exemple la somme des débits). Il est tout à fait possible que l'un des résultats du jeu consiste en des performances catastrophiques du point de vue de l'opérateur.

Quantifier la dégradation de performance permet ensuite de décider, pour l'opérateur, s'il est intéressant d'instaurer des mécanismes visant à garantir de meilleures performances, sachant que ces mécanismes ont également un coût. Dans le cas de réseaux informatiques, les mécanismes de contrôle du réseau induisent la plupart du temps du trafic supplémentaire ou des procédures de synchronisation entre différentes machines.

La mesure de dégradation la plus connue est le *prix de l'anarchie* [KP98] [11]. Il s'agit du rapport maximal entre la performance optimale et la performance d'un équilibre de Nash : c'est donc une mesure du pire cas qui ne donne aucune information sur la dégradation moyenne. Utiliser le prix de l'anarchie comme critère de décision pour instaurer des mécanismes de contrôle suppose donc que le processus d'apprentissage va mener à un équilibre de Nash [12]. Plus formellement, soit $(\mathcal{U}, \mathcal{S}, c)$ un jeu, on note $\mathrm{Perf} : \mathcal{S} \to \mathbb{R}$ la fonction de performance, et $\mathcal{E}$ l'ensemble des équilibres de Nash. Le prix de l'anarchie vaut :

$$\mathrm{PoA} = \frac{\max_{s \in \mathcal{S}} \mathrm{Perf}(s)}{\min_{s \in \mathcal{E}} \mathrm{Perf}(s)}. \tag{3.7}$$

Un prix de l'anarchie élevé signifie qu'il existe une stratégie dans $\mathcal{E}$ dont la performance est mauvaise par rapport à la performance optimale. Au contraire, un prix de l'anarchie de 1 signifie que tous les équilibres sont optimaux.

Les mesures de performance classiques sont l'efficacité globale, c'est-à-dire la somme des gains de chaque joueur, l'équité proportionnelle qui est le produit des gains (si on

11. Il semble que le prix de l'anarchie soit la seule mesure de dégradation utilisée alors que, pour de nombreux jeux et modèles d'apprentissage, les résultats ne se résument pas aux seuls équilibres de Nash.

12. Néanmoins, de nombreux articles donnant des bornes du prix de l'anarchie ne justifient pas ce point.

les suppose strictement positifs), et de manière générale, l'ensemble des performances de α-équité [MW98] :

$$\mathrm{Perf}(s) = \frac{1}{1-\alpha} \sum_{u \in \mathcal{U}} c_u(s)^{1-\alpha}.$$

Le cas $\alpha = 0$ correspond à la somme des gains, $\alpha = 1$ à la somme des logarithmes des gains, et $\alpha = +\infty$ à l'équité min max. Maximiser ce dernier critère revient à trouver la stratégie qui maximise le vecteur des gains classés par ordre croissant pour l'ordre lexicographique, c'est-à-dire maximiser le plus petit gain, puis le deuxième plus petit, et ainsi de suite.

De nombreuses bornes du prix de l'anarchie sont connues dans les jeux de congestion atomiques et non-atomiques si la mesure de performance est la somme des gains (voir le chapitre 18 de [Nis07], et les références citées). Citons également l'article plus récent [CCSM09] qui traite du cas mixte atomique et non-atomique. Toutes ces bornes dépendent de l'accroissement maximal des fonctions de gain. Or, dans la plupart des problèmes de routage, cet accroissement est non borné, si bien que le prix de l'anarchie est infini. Cependant, si la topologie du réseau est simple, par exemple arcs parallèles, des bornes existent (voir [ABP10]).

3.3.2 Robustesse des résultats

L'étude d'un modèle d'apprentissage revient à caractériser les profils d'actions qui seront joués asymptotiquement. Si cela est faisable, une question qui se pose alors est d'analyser la sensibilité de ces résultats par rapport à certains paramètres. Ici, nous nous focalisons sur la robustesse par rapport aux processus de révision des stratégies ("Quand les joueurs apprennent-ils ?"), et aux incertitudes sur les gains du jeu. Cela se justifie parce que d'une part, il est difficile de contrôler la synchronisation ou la non synchronisation dans la prise de décision des joueurs, et d'autre part, les gains que les joueurs observent sont des quantités physiques qui reposent sur des mesures, ce qui implique des variations aléatoires.

Nous détaillons ici ce que l'on entend par processus de révision des stratégies, et incertitudes sur les gains. Dans la suite du document, ces deux critères de robustesse sont analysés pour chaque processus d'apprentissage étudié.

Processus de révisions des stratégies

Dans le modèle d'apprentissage par meilleure réponse (voir section 3.4), les révisions des stratégies des joueurs se font de manière *asynchrone*. On montre alors que, dans les jeux de potentiel, le processus converge vers un équilibre de Nash. Asynchrone signifie qu'il n'y a jamais deux joueurs qui modifient leur stratégie en même temps. Or, dans de nombreux protocoles réseaux (par exemple le protocole Aloha discrétisé, le protocole OFDM utilisé dans le Wifi) les terminaux sont synchronisés, et les révisions des stratégies ont lieu simultanément. Est-ce que la propriété de convergence vers les équilibres de Nash est conservée dans ce cas ?

Si elle ne l'est pas, il faut faire en sorte d'assurer l'asynchronisme. Cela se fait, par exemple, en faisant attendre chaque joueur un temps aléatoire distribué selon une loi expo-

nentielle avant de changer de stratégie. Ce faisant, la probabilité que deux joueurs révisent simultanément leur stratégie est nulle[13]. Néanmoins cette procédure induit un surcoût lié aux temps d'attente. Il est donc préférable que la convergence soit indépendante du processus de révision.

Définition 3.20 (*Processus de révision*)

Un processus de révision des stratégies, ou processus de révision, est une mesure de probabilité μ sur l'ensemble $\mathcal{P}(\mathcal{U})$ des sous-ensembles de joueurs telle que :

$$\forall u \in \mathcal{U}, \exists \mathcal{V} \subseteq \mathcal{U} \text{ tel que } u \in \mathcal{V} \text{ et } \mu(\mathcal{V}) > 0,$$

où $\mu(\mathcal{V})$ est la probabilité que tous les joueurs dans $\mathcal{V}$ modifient leur stratégie simultanément.

La définition d'un processus de révision implique que chaque joueur a une probabilité non nulle de modifier sa stratégie à chaque itération. On suppose ici que le processus de révision ne dépend pas du temps. Cela pourrait se généraliser pour prendre en compte le cas où le processus de révision dépend d'un processus exogène périodique.

On note par $\mathcal{R}_\mu \stackrel{\text{def}}{=} \{\mathcal{V} \subseteq \mathcal{U} | \mu(\mathcal{V}) > 0\}$ l'ensemble des ensembles de révision, *i.e.* les ensembles de joueurs qui modifient leur stratégie simultanément avec une probabilité positive. Des exemples classiques de classes de processus de révision sont :

- l'apprentissage *asynchrone* qui correspond à $\mathcal{R}_\mu = \{\{u\} | u \in \mathcal{U}\}$: un seul joueur change de stratégie à chaque fois. Cela assure la convergence de l'apprentissage par meilleure réponse ou par la règle fictive dans les jeux de potentiel (voir les sections suivantes).
- l'apprentissage *indépendant* qui correspond à $\mathcal{R}_\mu = \mathcal{P}(\mathcal{U})$. C'est le cas où le temps est discrétisé, les joueurs sont synchronisés, et changent de stratégie à chaque étape avec une probabilité comprise strictement entre 0 et 1, et indépendemment des autres joueurs.
- l'apprentissage *instantané* qui correspond à $\mathcal{R}_\mu = \{\mathcal{U}\}$. Tous les joueurs mettent à jour leur stratégie en même temps.

D'autres exemples pourraient inclure des corrélations entre les joueurs, par exemple si une notion de voisinage existe entre les joueurs, ou bien si le nombre de joueurs révisant leur stratégie est borné, etc...

Incertitude sur les gains

Dans certaines situations, et particulièrement en informatique, le gain d'un joueur est une quantité physique qui, contrairement à des gains monétaires, est soumise à une incertitude liée à sa mesure. C'est le cas, par exemple, lorsque l'on considère un jeu de routage dans lequel la fonction de gain des joueurs est le délai ou le débit. Ces incertitudes, que

13. Néanmoins, cela suppose implicitement que le délai de changement d'action et le temps nécessaire à l'observation des gains sont nuls. Si des délais interviennent, deux révisions qui n'ont pas commencé au même instant peuvent se superposer. Le délai induit une possible simultanéité dans les révisions.

l'on modélisera par des variables aléatoires, peuvent avoir un impact sur la convergence du processus d'apprentissage. De plus, dans certains cas, les incertitudes permettent de sélectionner un équilibre spécifique avec une grande probabilité lorsque plusieurs équilibres sont atteignables dans le modèle non perturbé.

L'introduction de bruit sur les gains du jeu a comme conséquence de modifier la fonction de meilleure réponse des joueurs. Pour voir cela, prenons le cas d'un jeu fini $(\mathcal{U}, \mathcal{S}, c)$. Supposons que les gains d'un joueur soient soumis à des perturbations : au vecteur de gains $(c_u(a, s_{-u}))_{a\in\mathcal{S}_u}$ du joueur u vient s'ajouter un vecteur aléatoire (une erreur de mesure par exemple) $(E_u(a))_{a\in\mathcal{S}_u}(s_{-u})$ [14] dont la valeur est tirée et connue *à chaque fois* que le joueur révise sa stratégie. La correspondance de meilleure réponse devient alors $\mathrm{BR}_u(s_{-u}) = \operatorname*{argmax}_{a\in\mathcal{S}_u} c_u(a, s_{-u}) + E_u(a)$.

Si le joueur connaît la réalisation des variables aléatoires $E_u(a)$, alors choisir une meilleure réponse est un choix déterministe pour lui. Par contre, la probabilité notée $P_{u,a}(s_{-u})$ que l'action a soit choisie avant que les valeurs aléatoires ne soient tirées vaut :

$$P_{u,a}(s_{-u}) \stackrel{\text{def}}{=} \mathbb{P}[c_u(a, s_{-u}) + E_u(a) \geq c_u(b, s_{-u}) + E_u(b), \forall b \in \mathcal{S}_u]. \tag{3.8}$$

Clairement $P_u(s_{-u}) \in \Delta(\mathcal{S}_u)$. On appelle la fonction P_u une *fonction de choix*, qui est donc une stratégie mixte. Il est intéressant de noter qu'ici, la stratégie mixte ne correspond pas à un tirage aléatoire du joueur comme avant, mais à un gain aléatoire qui précède la décision du joueur.

De façon informelle, on peut remarquer que si les bruits sont i.i.d., de moyenne nulle, et avec une variabilité qui est grande par rapport aux gains du jeu, alors la fonction de choix est proche de la loi uniforme. A contrario, si la variabilité est petite devant les fonctions de gain, alors la fonction de choix est proche de la fonction de meilleure réponse dans le jeu sans perturbations.

Le théorème qui suit affirme que la fonction de choix (3.8) peut également être obtenue par une fonction de pénalisation sur l'espace des stratégies mixtes. Même si ce résultat ne correspond à aucune situation pratique connue, il peut être utilisé à des fins de simulation.

On dit qu'une fonction $V_u : \Delta(\mathcal{S}_u) \to \mathbb{R}$ est une *fonction de pénalité déterministe* si elle est strictement convexe, de classe $\mathcal{C}^1$, et telle que $\|\nabla V_u(y)\|$ tend vers l'infini quand y tend vers le bord de $\Delta(\mathcal{S}_u)$.

Théorème 3.21 (*théorème 2.1 dans [HS02]*)

Si pour tout $u \in \mathcal{U}$, $E_u = (E_u(a))_{a\in\mathcal{S}_u}$ est un vecteur aléatoire dont les composantes sont indépendantes et admettent une densité bornée et strictement positive sur $\mathbb{R}$, alors il existe une fonction de pénalité déterministe $V_u : \Delta(\mathcal{S}_u) \to \mathbb{R}$ telle que :

$$P_u(s_{-u}) = \operatorname*{argmax}_{y\in\mathrm{int}(\Delta(\mathcal{S}_u))} \Big(\sum_{a\in\mathcal{S}_u} y_a c_u(a, s_{-u}) - V_u(y) \Big),$$

où $P_u(s_{-u})$ est la fonction de choix (3.8).

14. Pour alléger les notations, on omettra par la suite la dépendance du vecteur aléatoire en s_{-u}.

La démonstration de ce théorème se fait en trois étapes. D'abord il faut montrer qu'il existe une fonction $V^* : \mathbb{R}^{n_u} \to \mathbb{R}$ qui est une fonction de potentiel pour P_u donné par (3.8), *i.e.* telle que $\nabla V^* = P_u$. Ensuite on observe que V^* vérifie les hypothèses suffisantes pour que, en posant $V(x) = \max_{c \in \mathbb{R}^{n_u}} \big(\sum_{a \in \mathcal{S}_u} c_a x_a - V^*(c) \big)$, les hypothèses du théorème 26.5 de [Roc97] soient satisfaites, ce qui implique que $V^*(c_u) = \max_{y \in \Delta(\mathcal{S}_u)} \big(\sum_{a \in \mathcal{S}_u} y_a c_u(a, s_{-u}) - V(y) \big)$ et que $(\nabla V)^{-1} = \nabla V^*$. Enfin, cela permet de montrer que P_u réalise le maximum de $\max_{y \in \Delta(\mathcal{S}_u)} \big(\sum_{a \in \mathcal{S}_u} y_a c_u(a, s_{-u}) - V(y) \big)$, et termine la démonstration.

L'exemple le plus connu de fonction de choix qui peut être obtenue à partir d'un modèle de gain avec incertitude est ce qu'on appelle la fonction de choix logit, et correspond en fait à une *distribution de Gibbs*, c'est-à-dire :

$$\forall u \in \mathcal{U}, \forall a \in \mathcal{S}_u, P_{u,a}(s_{-u}) = \frac{\exp(\beta^{-1} c_u(a, s_{-u}))}{\sum_{b \in \mathcal{S}_u} \exp(\beta^{-1} c_u(b, s_{-u}))},$$

où $\beta > 0$ est un paramètre qui s'interprète comme le niveau de bruit. Lorsque β est proche de zéro, la distribution est concentrée sur les états qui maximisent le gain. Quand β tend vers l'infini, la distribution de Gibbs tend vers la distribution uniforme.

La distribution de Gibbs peut être obtenue par le choix de perturbations aléatoires indépendantes et distribuées selon la loi de Gumbel [15], dont la fonction de répartition est donnée par $F(x) = \exp\big(- \exp(-\beta^{-1} x - \gamma) \big)$, où γ est la constante d'Euler (afin que la moyenne soit nulle). Notons $f(x) = F'(x)$ la fonction de densité de la distribution de Gumbel. Notons pour alléger les notations $c_u(a) = c_u(a, s_{-u})$, et omettons la dépendance en u. Alors :

$$P_a = \mathbb{P}[c(a) + E(a) \geq c(b) + E(b), \forall b] = \mathbb{P}[E(a) \geq E(b) + \delta(b), \forall b],$$

avec $\delta(b) = c(b) - c(a)$. Notons également $\alpha(b) = \exp(-\beta^{-1} \delta(b))$. Alors :

$$\begin{aligned} P_a &= \int_{-\infty}^{\infty} f(x_a) \int_{-\infty}^{x_a + \delta_1} f(x_1) \dots \int_{-\infty}^{x_a + \delta_{a-1}} f(x_{a-1}) \int_{-\infty}^{x_a + \delta_{a+1}} f(x_{a+1}) \dots \int_{-\infty}^{x_a + \delta_n} f(x_n) dx_n \dots dx_{a+1} dx_{a-1} \dots dx_1 dx_a \\ &= \int_{-\infty}^{\infty} F'(x) \prod_{b \neq a} F(x)^{\alpha(b)} dx \\ &= \int_{-\infty}^{\infty} F'(x) F(x)^{\sum_{b \neq a} \alpha(b)} dx. \end{aligned}$$

Une intégration par partie donne alors $P_a = \dfrac{1}{1 + \sum_{b \neq a} \alpha(b)}$, ce qui correspond bien à la distribution de Gibbs.

15. Notons que nous ne connaissons pas de systèmes réels où apparaît ce type de distribution. La distribution de Gumbel semble tout de même être liée (d'après Wikipédia) au maximum de variables exponentielles ce qui pourrait modéliser le maximum de temps d'attente d'un ensemble de files d'attente.

La distribution de Gibbs peut également être obtenue par une fonction de pénalité déterministe d'après le théorème précédent. Le choix de la fonction d'entropie $V_u(y) = \beta \sum_{a \in \mathcal{S}_u} y_a \log(y_a)$ donne le résultat voulu. Cela se vérifie par l'intermédiaire des conditions nécessaires et suffisantes d'optimalité de Karush Kuhn Tucker.

Dans les sections qui suivent, nous détaillons deux modèles classiques d'apprentissage, qui sont l'algorithme de meilleure réponse et l'algorithme du jeu fictif. Nous analysons leur convergence dans les jeux de potentiel finis ainsi que la robustesse du résultat au bruit et au processus de révision.

3.4 L'algorithme de meilleure réponse

L'algorithme de meilleure réponse (l'algorithme 1) consiste en une succession de déviations unilatérales de meilleures réponses. Le processus de révision des stratégies est *asynchrone*.

Algorithme 1: Algorithme de meilleure réponse

initialisation;
Chaque joueur $u \in \mathcal{U}$ choisit une action uniformément dans $\mathcal{S}_u$;
répéter
  Choisir le joueur u uniformément sur $\mathcal{U}$;
  Le joueur u choisit une action uniformément dans $\mathrm{BR}_u(s_{-u})$;
jusqu'à *l'infini*;

Il faut noter que l'algorithme de meilleure réponse est un algorithme aléatoire car :

- le choix initial du profil d'actions est aléatoire (uniforme) sur $\mathcal{S}$,
- le choix du joueur u qui révise sa stratégie est aléatoire (uniforme) sur $\mathcal{S}_u$,
- le choix de l'action de meilleure réponse est aléatoire (uniforme) sur $\mathrm{BR}_u(s_{-u})$ [16].

La suite aléatoire notée $(S_n)_n$ des états induite par l'algorithme est une chaîne de Markov sur l'espace d'état $\mathcal{S}$. Comme $\mathcal{S}$ est fini il existe des états par lesquels la chaîne de Markov passe une infinité de fois et qui sont donc des *états limites*. Dans la suite, nous caractérisons ces états limites, qui correspondent aux résultats de l'apprentissage par l'algorithme de meilleure réponse.

16. On suppose l'uniformité dans toutes les décisions aléatoires. En fait, toute mesure de probabilité dont le support contient tout l'ensemble des choix suffit pour que les résultats de convergence qui suivent soient valides.

Caractérisation des états limites de l'algorithme de meilleure réponse

Notons $N_s \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} \mathbf{1}_{S_n=s}$ le nombre de passages par l'état s. On note $\mathbb{P}_s$ l'unique mesure de probabilité [17] sur $\mathcal{S}^{\mathbb{N}}$ telle que la suite $(S_n)_n$ des coordonnées sous cette mesure de probabilité vérifie les probabilités de transition induites par l'algorithme de meilleure réponse, et $\mathbb{P}_s[S_0 = s] = 1$ (dont l'existence et l'unicité sont assurées par le théorème 13.3.3 dans [LG06]).

Définition 3.22 (*Convergence de l'algorithme de meilleure réponse*)

On dit que l'algorithme de meilleure réponse converge vers l'ensemble $\mathcal{E} \subseteq \mathcal{S}$ si :

$$\mathbb{P}_s[N_s = +\infty] = \begin{cases} 1 & \text{si } s \in \mathcal{E}, \\ 0 & \text{sinon.} \end{cases}$$

Remarquons que, même si un état est dans l'ensemble de convergence, il se peut que la chaîne de Markov ne passe jamais par cet état (cela n'est pas contradictoire avec la définition, par exemple si ce n'est pas l'état initial de la chaîne de Markov). On peut juste affirmer que si un état n'est pas dans l'ensemble de convergence, alors le nombre de passages par cet état est presque sûrement fini.

Par exemple, nous verrons au corollaire 3.26 que, dans les jeux de potentiel, l'algorithme de meilleure réponse converge presque sûrement vers un ensemble d'équilibres de Nash. Cela n'est bien entendu pas vrai pour n'importe quel jeu, même si celui-ci admet un équilibre de Nash, comme l'illustre l'exemple suivant :

Exemple : Soit le jeu à deux joueurs donné par la matrice de gains suivant :

Gains

$(3,1)$	$(0,0)$	$(0,0)$
$(0,0)$	$(1,3)$	$(2,1)$
$(0,0)$	$(4,1)$	$(1,2)$

États

(a,a)	(a,b)	(b,c)
(b,a)	(b,b)	(b,c)
(c,a)	(c,b)	(c,c)

Partant de l'état (c,c) en bas à droite, la trajectoire de l'algorithme de meilleure réponse ne sort jamais de l'ensemble des états (b,b), (b,c), (c,b) et (c,c). Or aucun de ces états n'est un équilibre de Nash. Ici, les seuls états qui ne sont pas dans l'ensemble de convergence sont ceux qui procurent le gain $(0,0)$.

Afin de caractériser l'ensemble limite de l'algorithme de meilleure réponse, il est naturel d'introduire le *graphe de meilleure réponse*.

17. Par la suite, l'usage de cette mesure de probabilité sera implicite, notamment dans le calcul d'espérances, ainsi que dans l'expression "presque sûrement".

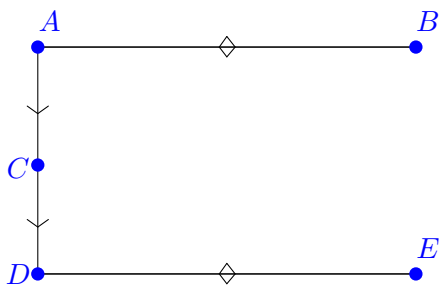
Définition 3.23 (*Graphe de meilleure réponse*)

Soit $(\mathcal{U}, \mathcal{S}, c)$ un jeu. On appelle *graphe de meilleure réponse* du jeu le graphe orienté dont l'ensemble des sommets est $\mathcal{S}$ et il existe un arc du sommet s vers le sommet (a, s_{-u}) si $a \in \mathrm{BR}_u(s_{-u})$.

Ce graphe n'est en général pas fortement connexe. Un *chemin* du graphe est une suite de sommets reliés par un arc du graphe. Si s et t sont deux sommets du graphe, on note $s \to t$ le fait qu'il existe un chemin de s vers t, et $s \leftrightarrow t$ si $s \to t$ et $t \to s$. Il est immédiat de vérifier que la relation "$\leftrightarrow$" est une relation d'équivalence sur les sommets du graphe. Celle-ci partage l'ensemble des sommets en *classes d'équivalence*. Une classe d'équivalence est donc un sous-graphe fortement connexe.

Une classe d'équivalence est dite *récurrente* s'il n'y a aucun arc dirigé vers l'extérieur de la classe, et un sommet est récurrent s'il appartient à une classe récurrente. Cela implique que, si un sommet n'est pas récurrent, alors il existe un chemin qui le relie à un sommet récurrent.

Exemple : Considérons le graphe suivant :



Les classes d'équivalence du graphe sont $\{A, B\}$, $\{C\}$ et $\{D, E\}$. Seule la classe $\{D, E\}$ est récurrente.

Dans la suite, nous ne ferons pas la distinction entre un sommet du graphe de meilleure réponse et le profil d'action (pour le jeu), ou l'état (pour la chaîne de Markov) correspondant.

Le résultat suivant est intuitif et la preuve élémentaire.

Théorème 3.24

L'algorithme de meilleure réponse converge vers l'ensemble des sommets récurrents du graphe de meilleure réponse.

De plus, le temps d'atteinte d'un sommet récurrent est presque sûrement fini.

Le théorème implique :

Corollaire 3.25

Soit $(\mathcal{U}, \mathcal{S}, c)$ un jeu. Si les sommets récurrents du graphe de meilleure réponse sont des équilibres de Nash, alors l'algorithme de meilleure réponse converge presque sûrement vers un ensemble d'équilibres de Nash en un temps presque sûrement fini.

Dans les jeux de potentiel, la fonction de potentiel est strictement croissante lorsqu'un joueur choisit une action qui améliore strictement son gain. Il est clair que, si un état n'est

pas un équilibre de Nash, alors il ne peut pas être un sommet récurrent du graphe de meilleure réponse, si bien que :

Corollaire 3.26

Si $(\mathcal{U}, \mathcal{S}, c)$ est un jeu de potentiel, alors l'algorithme de meilleure réponse converge presque sûrement vers un ensemble d'équilibres de Nash.

Relaxation de la fonction de potentiel, et analyse de la convergence pour l'algorithme de meilleure réponse

Le théorème 3.24 caractérise les profils d'actions qui sont sélectionnés par l'algorithme de meilleure réponse à partir des sommets récurrents du graphe de meilleure réponse. Dans le cas des jeux de potentiel, cela implique que l'algorithme converge presque sûrement vers un ensemble d'équilibres de Nash. Néanmoins, les jeux de potentiel ne sont pas les seuls jeux tels que les sommets récurrents du graphe de meilleure réponse sont des équilibres de Nash.

En relaxant la contrainte d'égalité qui caractérise la fonction de potentiel de différentes manières, on obtient de nouvelles classes de jeux qui partagent certaines propriétés des jeux de potentiel. Nous discutons ici de leurs propriétés de convergence pour l'algorithme de meilleure réponse.

Plusieurs extensions des jeux de potentiel[18] ont été introduites dans [MS96b], parmi lesquelles :

Définition 3.27 (*Extensions de la notion de potentiel*)

Le jeu $(\mathcal{U}, \mathcal{S}, c)$ est un jeu de potentiel

- *pondéré* si $\exists (\omega_u)_{u \in \mathcal{U}}$ positifs et $\exists P : \mathcal{S} \to \mathbb{R}$ tels que :

$$\forall u \in \mathcal{U}, \forall s_{-u} \in \mathcal{S}_{-u}, \forall a, b \in \mathcal{S}_u, c_u(a, s_{-u}) - c_u(b, s_{-u}) = \omega_u(P(a, s_{-u}) - P(b, s_{-u})).$$

- *ordinal* si $\exists P : \mathcal{S} \to \mathbb{R}$ telle que :

$$\forall u \in \mathcal{U}, \forall s_{-u} \in \mathcal{S}_{-u}, \forall a, b \in \mathcal{S}_u, c_u(a, s_{-u}) - c_u(b, s_{-u}) < 0 \Leftrightarrow P(a, s_{-u}) - P(b, s_{-u}) < 0.$$

- *généralisé* si $\exists P : \mathcal{S} \to \mathbb{R}$ telle que :

$$\forall u \in \mathcal{U}, \forall s_{-u} \in \mathcal{S}_{-u}, \forall a, b \in \mathcal{S}_u, c_u(a, s_{-u}) - c_u(b, s_{-u}) < 0 \Rightarrow P(a, s_{-u}) - P(b, s_{-u}) < 0.$$

On constate immédiatement à partir des définitions que les jeux de potentiel pondérés sont des jeux de potentiel ordinaux, qui sont eux-mêmes des jeux de potentiel généralisés.

On dira que l'action d'un joueur est une meilleure réponse *stricte* si c'est une meilleure réponse qui améliore strictement son gain actuel. Il est clair que, si l'algorithme de meilleure réponse est restreint aux strictes meilleures réponses, alors les jeux de potentiel généralisés

18. Pour distinguer la fonction de potentiel classique (la seule qui correspond au sens physique du terme potentiel) des autres formes de potentiel, on la désigne par "potentiel exact".

convergent vers un équilibre de Nash en un nombre fini d'itérations (en supposant que l'algorithme s'arrête à partir du moment où aucun joueur ne peux améliorer son gain unilatéralement). Cela tient au fait que les jeux de potentiel généralisés possèdent la propriété des chemins d'amélioration finis (définition 3.10). On a même :

Proposition 3.28 (*lemme 2.5 dans [MS96b]*)

Un jeu possède la propriété des chemins d'amélioration finis (définition 3.10) si et seulement si c'est un jeu de potentiel généralisé.

La propriété des chemins d'amélioration finis caractérise les jeux de potentiel généralisés. Cependant, cette propriété n'est pas suffisante pour que l'algorithme de meilleure réponse converge vers un équilibre de Nash, comme le prouve l'exemple suivant :

Exemple : Soit un jeu à deux joueurs, chacun ayant deux actions a et b. La matrice des gains est la suivante :

Gains

$(1,0)$	$(0,1)$
$(1,1)$	$(1,0)$

Potentiel

0	1
3	2

Chemins d'amélioration

$$\begin{array}{ccc} (a,a) & \rightarrow & (a,b) \\ \downarrow & & \downarrow \\ (b,a) & \leftarrow & (b,b) \end{array}$$

Chemins de meilleure réponse

$$\begin{array}{ccc} (a,a) & \rightarrow & (a,b) \\ \uparrow & & \downarrow \\ (b,a) & \leftarrow & (b,b) \end{array}$$

On peut vérifier que la fonction de potentiel indiquée est bien un potentiel généralisé. De ce fait, le jeu a la propriété des chemins d'amélioration finis, et le graphe construit à partir des déviations unilatérales qui améliorent strictement le gain du joueur est sans cycle. Par contre, le graphe de meilleure réponse (pas nécessairement stricte) possède un cycle : en effet, le gain du joueur qui choisit les lignes pour le profil d'action (a, a) est égal à celui du profil d'action (b, a). De ce fait, l'algorithme de meilleure réponse passe indéfiniment, avec probabilité 1, par des états qui ne sont pas des équilibres de Nash.

Cet exemple repose sur le fait que l'équilibre de Nash du jeu n'est pas un équilibre strict. Pour remédier à l'absence de convergence, on peut restreindre l'algorithme de meilleure réponse aux changements d'action qui améliorent *strictement* le gain des joueurs. Cependant, cela ne résout pas le problème de convergence dès lors que, comme cela est souvent le cas en pratique, le jeu est soumis à une perturbation aléatoire sur les fonctions de gain. En effet, soumis à des perturbations aléatoires aussi faibles soient-elles, deux actions qui, en moyenne, procurent le même gain, seront alternativement meilleures l'une par rapport à l'autre.

Il s'avère donc que la propriété des chemins d'amélioration finis n'est pas équivalente au fait de converger vers des équilibres de Nash dans l'algorithme de meilleure réponse. Cela n'est pas surprenant car, d'une part, cette propriété ne porte que sur les améliorations strictes, alors que l'algorithme de meilleure réponse admet des déviations qui n'améliorent pas strictement le gain. Et d'autre part, l'algorithme ne porte que sur les meilleures réponses, tandis que la propriété des chemins d'amélioration finis est valable pour n'importe quelle amélioration qui n'est pas forcément une meilleure réponse.

Le potentiel tel que défini dans la classe de jeux suivante tient compte de ces remarques :

Définition 3.29 (*Jeu de potentiel de meilleure réponse [Voo00]*)

$(\mathcal{U}, \mathcal{S}, c)$ est un jeu de potentiel de meilleure réponse si $\exists P : \mathcal{S} \to \mathbb{R}$ telle que :

$$\forall u \in \mathcal{U}, \forall s_{-u} \in \mathcal{S}_{-u}, \mathrm{BR}_u(s_{-u}) = \underset{a \in \mathcal{S}_u}{\mathrm{argmax}} P(a, s_{-u}).$$

Une caractérisation de ces jeux repose sur les *cycles de meilleures réponses* qui sont définis comme une suite d'états $(s^0, \ldots, s^m)$ telle que :

- $s^0 = s^m$.
- chaque transition (s^i, s^{i+1}) est le résultat de la déviation unilatérale de l'un des joueurs, et cette déviation est une meilleure réponse,
- il existe une transition qui est une meilleure réponse stricte (qui améliore strictement le gain du joueur qui dévie),

Alors :

Théorème 3.30 (*théorème 3.2 dans [Voo00]*)

Un jeu fini est un jeu de potentiel de meilleure réponse si et seulement s'il ne possède pas de cycle de meilleures réponses.

Il s'ensuit :

Corollaire 3.31

Dans tout jeu de potentiel de meilleure réponse, l'algorithme de meilleure réponse converge vers un ensemble d'équilibres de Nash.

Démonstration : Il suffit de vérifier que les états récurrents du graphe de meilleures réponses d'un jeu de potentiel des meilleures réponses sont des équilibres de Nash.

Supposons qu'il existe un état récurrent s qui ne soit pas un équilibre de Nash. Alors il existe $u \in \mathcal{U}$ et $a \in \mathrm{BR}_u(s_{-u})$ tels que $c_u(a, s_{-u}) > c_u(s)$. Comme s est récurrent, il existe un cycle de meilleures réponses $(s, (a, s_{-u}), \ldots, s)$ qui comporte une meilleure réponse stricte. Par le théorème 3.30, cela contredit le fait que c'est un jeu de potentiel de meilleure réponse. ■

L'existence d'un potentiel de meilleure réponse fournit une condition suffisante pour que l'algorithme de meilleure réponse converge vers un ensemble d'équilibres de Nash. Cette condition n'est pas nécessaire car il existe des jeux qui ne sont pas des jeux de potentiel de meilleure réponse et qui vérifie la propriété de convergence, comme dans l'exemple suivant.

Exemple : Considérons le jeu à deux joueurs suivant :

Gains

$(3,3)$	$(2,0)$	$(0,0)$
$(0,0)$	$(1,2)$	$(2,1)$
$(0,0)$	$(2,1)$	$(1,2)$

Ce jeu possède un cycle de meilleures réponses constitué des quatre états du carré en bas à droite, ce n'est donc pas un jeu de potentiel de meilleure réponse. On peut cependant sortir du cycle en passant par l'état $(2, 0)$. Le seul état récurrent est donc l'équilibre de Nash $(3, 3)$.

La classe des jeux de potentiel de meilleure réponse ne contient pas tous les jeux de potentiel généralisés, comme le jeu de l'exemple précédent. Par contre, elle contient la classe des jeux de potentiel ordinaux, ce qui se déduit directement des définitions. Enfin, il existe des jeux de potentiel de meilleure réponse qui ne sont pas des jeux de potentiel généralisés. Cela n'est pas surprenant du fait que, même si les jeux de potentiel de meilleure réponse ne contiennent pas de cycles de meilleures réponses, ils peuvent avoir des chemins d'amélioration infinis, comme dans le jeu suivant :

Exemple 4.3 de [Voo00] : Soit le jeu à deux joueurs suivant :

Gains

$(2, 2)$	$(1, 0)$	$(0, 1)$
$(0, 0)$	$(0, 1)$	$(1, 0)$

Potentiel

4	3	0
0	2	1

Chemin d'amélioration

$$\begin{array}{ccccc} (a,a) & & (a,b) & \rightarrow & (a,c) \\ & & \uparrow & & \downarrow \\ (b,a) & & (b,b) & \leftarrow & (b,c) \end{array}$$

On peut vérifier que le potentiel indiqué est effectivement un potentiel de meilleure réponse pour le jeu. Celui-ci possède un cycle dans le graphe des chemins d'amélioration, ce n'est donc pas un jeu de potentiel généralisé.

Les relations entre les différentes classes de jeux de potentiel sont résumées à la figure 3.4.

Pour résumer, la plus grande classe de jeux de potentiel pour laquelle l'algorithme de meilleure réponse converge presque sûrement vers un ensemble d'équilibres de Nash est la classe des jeux de potentiel de meilleure réponse. Cependant, il n'existe pas de caractérisation complète, par l'intermédiaire d'une fonction de potentiel, des jeux qui vérifient cette propriété.

Robustesse de l'algorithme de meilleure réponse

D'après le corollaire 3.26, l'algorithme de meilleure réponse converge vers un ensemble d'équilibres de Nash dans les jeux de potentiel. Dans l'exemple suivant, nous montrons que cela n'est plus vrai dès lors que le processus de révision n'est pas asynchrone :

Exemple : Considérons le jeu de potentiel à deux joueurs suivant :

Gains

$(1, 1)$	$(0, 0)$
$(0, 0)$	$(1, 1)$

Si les joueurs changent leur stratégie simultanément (processus de révision instantané), et que le gain initial est $(0, 0)$, alors les joueurs ne choisissent jamais l'un des équilibres de Nash.

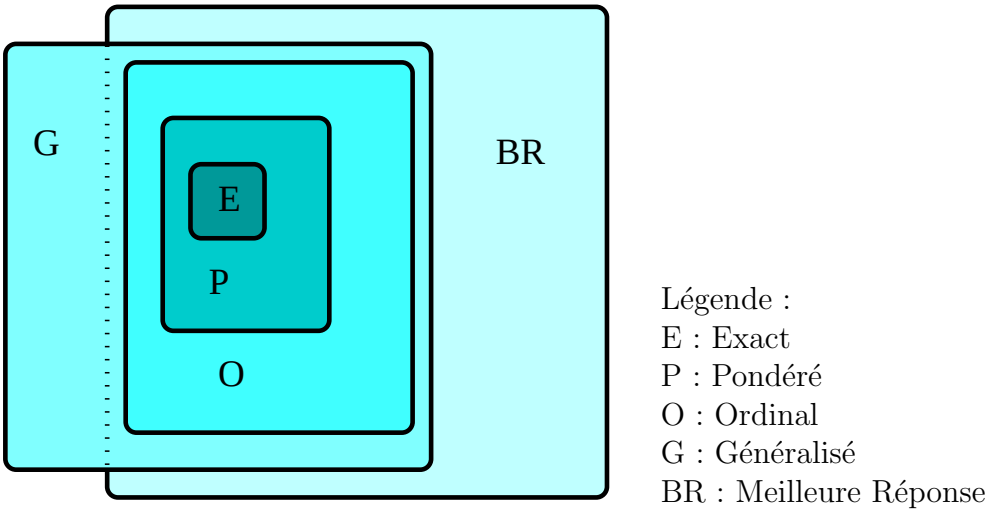


Figure 3.4 – Relations entre les différentes classes de potentiel.

De plus, s'il y a des perturbations aléatoires ajoutées aux gains de la matrice, les correspondances de meilleure réponse sont modifiées à chaque fois, et l'algorithme de meilleure réponse visite une infinité de fois tous les profils d'actions du jeu, quel que soit le processus de révision.

3.5 L'apprentissage par la règle du jeu fictif

Nous terminons ce chapitre par la description brève d'un deuxième exemple de modèle d'apprentissage. Celui-ci repose également sur un processus de meilleure réponse.

Même si nous n'utilisons pas ce modèle par la suite car il ne repose pas sur des hypothèses satisfaites en pratique, il reste intéressant car c'est le premier exemple de modèle d'apprentissage qui évolue dans l'espace des stratégies mixtes du jeu, et la méthode d'analyse est la même que celle que nous utiliserons dans le chapitre 5. Celle-ci repose sur la théorie des approximations stochastiques.

Le modèle d'apprentissage par la règle du jeu fictif suppose que tous les joueurs observent les actions jouées par l'ensemble des joueurs (ou de façon équivalente les profils d'actions). La règle d'apprentissage consiste alors à jouer une meilleure réponse à la distribution empirique des actions qui ont été jouées par les autres joueurs. Cette distribution empirique constitue pour le joueur ce que l'on nomme une *croyance*. Si le joueur suppose que les autres joueurs choisissent leur action en fonction d'une stratégie mixte fixée à l'avance, la croyance est sensée converger vers la stratégie en question.

Il s'agit donc d'un algorithme de meilleure réponse qui se base non pas sur les gains instantanés, mais sur les croyances. Comparativement à l'algorithme de meilleure réponse précédent, il nécessite, afin d'être implémenté, l'observation de toute la succession des profils d'actions.

Ce modèle d'apprentissage a été proposé la première fois dans [Bro51]. Son nom "règle

du jeu fictif" vient du fait qu'un joueur qui a une connaissance complète du jeu peut le jouer fictivement (autrement dit simuler) selon cette règle afin de calculer un équilibre de Nash du jeu (si le processus converge).

On note $y^u(n) \in \Delta(\mathcal{S}_{-u})$ la croyance du joueur u à l'étape n. Il s'agit d'un vecteur qui, à chaque joueur v différent de u, associe un vecteur de probabilité $y^u_v(n) \in \Delta(\mathcal{S}_v)$ correspondant à la distribution empirique (moyenne de Césaro) des actions jouées par v. Si $S_v(n)$ dénote l'action jouée par le joueur v à l'étape n, alors $y^u_v(n) = \frac{1}{n}\sum_{i=1}^{n} S_v(i)$.

L'action jouée par le joueur u à l'étape $n+1$ est une meilleure réponse à la croyance, *i.e.* $S_u(n+1) \in \mathrm{BR}_u(y^u(n))$, meilleure réponse que l'on peut toujours supposer être en stratégies pures. Cette action est observée par tous les joueurs, ce qui leur permet de mettre à jour leur croyance. Cela est résumé à l'algorithme 2.

Algorithme 2: Algorithme du jeu fictif

initialisation;
$n \leftarrow 0$;
Chaque joueur $u \in \mathcal{U}$ initialise sa croyance, par exemple $y^u(0) = 0$;
répéter
  Tous les joueurs choisissent simultanément une stratégie pure $S_u(n+1) \in \mathrm{BR}_u(y^u(n))$;
  Tous les joueurs mettent à jour leur croyance :
  $\forall v \neq u, y^u_v(n+1) = \frac{1}{n+1}\sum_{i=1}^{n+1} S_v(i)$;
  $n \leftarrow n+1$;
jusqu'à *l'infini*;

Plusieurs articles ont montré la convergence de l'algorithme du jeu fictif dans des classes de jeu particulières. Le premier résultat porte sur les jeux à deux joueurs à somme nulle [Rob51]. L'article [Ber07] passe en revue les résultats existants et donne les principales références. En ce qui concerne les jeux de potentiel, on a :

Théorème 3.32 (*théorème A dans [MS96a]*)

Les croyances de l'algorithme du jeu fictif convergent vers un ensemble d'équilibres de Nash dans les jeux de potentiel de meilleure réponse (définition 3.29), et donc dans les jeux de potentiel.

Comme nous le verrons au chapitre 5, ce résultat est robuste aux deux critères que sont le bruit sur les gains et le processus de révision [19]. Cela tient au fait que l'algorithme évolue dans un espace continu, et que les bruits ont tendance à se compenser au fil des itérations.

19. En particulier, la version originale de l'algorithme donnée dans [Bro51] est asynchrone, et partage les mêmes propriétés de convergence.

Notons enfin que l'algorithme du jeu fictif n'est pas utilisable dans la pratique lorsque le nombre de joueurs est grand, car il nécessite l'observation de toutes les actions des autres joueurs, ce qui est rédhibitoire.

CHAPITRE 4

LE MODÈLE STOCHASTIQUE DE MEILLEURE RÉPONSE

Résumé du chapitre

Dans le chapitre précédent, nous avons caractérisé les résultats de l'apprentissage par meilleure réponse *déterministe*. On s'intéresse maintenant à la version *stochastique* qui prend en compte un certain type de fluctuation aléatoire des gains du jeu. Cela amène les joueurs à ne pas choisir uniquement des meilleures réponses lors de la révision de leur stratégie. Nous montrons que cela est équivalent au modèle dans lequel leur choix est tiré selon une distribution de Gibbs. Le paramètre de la distribution de Gibbs peut alors s'interpréter comme le niveau de bruit sur les gains.

Dans la version stochastique de l'algorithme de meilleure réponse, le processus aléatoire induit est une chaîne de Markov qui, quel que soit le jeu, visite infiniment tous les états. Aussi, et même dans le cas d'un jeu de potentiel, le processus ne converge pas vers un ensemble d'équilibres de Nash. Néanmoins, lorsque le niveau de bruit tend à être nul, l'algorithme stochastique "ressemble" à l'algorithme déterministe. Nous montrons alors que les profils d'actions dont la distribution limite quand le bruit tend vers zéro, que nous appelons des *états stochastiquement stables*, sont inclus dans l'ensemble de convergence de l'algorithme déterministe. De plus, dans les jeux de potentiel, et si le processus de révision est asynchrone, les états stochastiquement stables sont ceux qui maximisent le potentiel. L'ajout de bruit permet donc d'avoir une caractérisation plus fine des états limites.

L'organisation du chapitre est la suivante. Dans un premier temps, nous montrons que, pour le processus asynchrone, les seuls états stochastiquement stables sont ceux qui maximisent le potentiel. Cela fournit un algorithme d'optimisation dont l'implémentation dans un problème de routage dans les réseaux de mobiles ad hoc fait l'objet de la section 4.2. Enfin, dans la section 4.3, nous analysons la robustesse de ce résultat par rapport au processus de révision des stratégies utilisé dans l'algorithme stochastique de meilleure réponse.

4.1 Algorithme stochastique de meilleure réponse asynchrone

Dans ce chapitre, nous étudions l'algorithme stochastique de meilleure réponse ainsi que ses extensions. L'algorithme 3 en est la version asynchrone, c'est-à-dire qu'à chaque itération de l'algorithme un et un seul joueur change de stratégie.

Nous nous plaçons dans le cadre d'un jeu fini $(\mathcal{U}, \mathcal{S}, c)$, les notations étant conservées par rapport au chapitre précédent. Contrairement au modèle de meilleure réponse de la section 3.4, les joueurs ne choisissent pas uniquement des meilleures réponses : les actions sont choisies aléatoirement selon une distribution de Gibbs qui dépend d'un paramètre $\eta > 0$. Quand η est proche de 0, la distribution est quasiment nulle sur les états qui ne sont pas dans l'ensemble de meilleure réponse, et quand η est grand, la distribution tend vers la distribution uniforme[1].

Algorithme 3: Algorithme stochastique de meilleure réponse asynchrone

initialisation;
Chaque joueur $u \in \mathcal{U}$ choisit aléatoirement une action $s_u \in \mathcal{S}_u$;
répéter
  Choisir le joueur u uniformément parmi l'ensemble des joueurs;
  Le joueur u choisit l'action a avec probabilité

$$\frac{\exp\left(\eta^{-1} c_u(a, s_{-u})\right)}{\sum_{b \in \mathcal{S}_u} \exp\left(\eta^{-1} c_u(b, s_{-u})\right)}$$

jusqu'à *l'infini*;

Notons tout d'abord que l'algorithme stochastique de meilleure réponse induit un processus aléatoire sur l'ensemble des stratégies $\mathcal{S}$ du jeu. Dans la suite, nous ne ferons pas de distinction entre l'ensemble des stratégies du jeu et l'ensemble des états du processus aléatoire. Notons $(S^\eta(t))$ ce processus aléatoire. Il s'agit d'une chaîne de Markov qui est homogène si le paramètre η ne dépend pas du temps, ce que nous supposons ici. Alors, la probabilité de transition de l'état s à l'état s' est donnée par :

$$p^\eta_{s,s'} = \begin{cases} \dfrac{1}{U} \dfrac{\exp\left(\eta^{-1} c_u(a, s_{-u})\right)}{\sum_{b \in \mathcal{S}_u} \exp\left(\eta^{-1} c_u(b, s_{-u})\right)} & \text{si } s' = (a, s_{-u}) \text{ où } a \in \mathcal{S}_u, \\ 0 & \text{sinon.} \end{cases} \tag{4.1}$$

1. Dans la suite, afin de simplifier les notations, on supposera que le paramètre η est le même pour tous les joueurs. Le cas contraire n'implique aucune complexité supplémentaire.

4.1.1 Interprétations de l'algorithme stochastique de meilleure réponse à partir de modèles d'apprentissage

Rappelons que l'on note $\Delta(\mathcal{S}_u)$ l'extension mixte de $\mathcal{S}_u$. L'algorithme stochastique de meilleure réponse 3 repose sur une fonction de choix $P_u : \mathcal{S}_{-u} \to \Delta(\mathcal{S}_u)$ pour chacun des joueurs donnée par :

$$\forall a \in \mathcal{S}_u, P_{u,a}(s_{-u}) = \frac{\exp\left(\eta^{-1} c_u(a, s_{-u})\right)}{\sum_{b \in \mathcal{S}_u} \exp\left(\eta^{-1} c_u(b, s_{-u})\right)}. \tag{4.2}$$

Cette fonction de choix est très répandue dans les modèles économiques du fait qu'elle dérive de plusieurs variantes du modèle de meilleure réponse. Elle peut être obtenue en ajoutant du bruit sur les gains des joueurs, mais également en ajoutant une pénalité aux gains des joueurs qui dépend de leur stratégie (mixte), et enfin en supposant une rationalité limité des joueurs. Nous détaillons maintenant ces trois modèles.

Meilleure réponse avec incertitude

Comme nous l'avons vu dans la section 3.3.2, la fonction de choix (4.2) peut être obtenue à partir du modèle d'utilité avec bruit aléatoire (voir à ce sujet [ADPT92], chapitre 9), où le bruit suit une distribution de Gumbel. Un tel bruit aléatoire peut être engendré par des incertitudes sur les gains qui proviennent de mesures physiques.

Ce modèle suppose que, avant le choix d'une meilleure réponse, les gains des joueurs sont soumis à un bruit aléatoire qui s'ajoute au gain initial. Si $E_u(a)$ est la variable aléatoire réelle modélisant le bruit, la probabilité de choisir l'action a est donnée par la fonction de choix suivante :

$$\mathbb{P}[c_u(a, s_{-u}) + E_u(a) \geq c_u(b, s_{-u}) + E_u(b), \forall b \neq a]. \tag{4.3}$$

Si $E_u(a)$ suit une distribution de Gumbel de paramètre η pour tout $a \in \mathcal{S}_u$, alors la distribution (4.3) est la fonction de choix (4.2).

Meilleure réponse avec pénalisation sur l'espace des stratégies

Comme cela a été montré dans [HS02], les fonctions de choix peuvent également dériver d'un modèle de meilleur réponse dans l'extension mixte $\Delta(\mathcal{S}_u)$ sur les gains du jeu fini auxquels on a ajouté une fonction de pénalité $V^\eta : \Delta(\mathcal{S}_u) \to \mathbb{R}$. Dans ce cas, la meilleure réponse est la solution (unique si l'on suppose V^η strictement convexe, et de dérivée infinie sur les bords du domaine de définition) de $\max_{y \in \Delta(\mathcal{S}_u)} \left(\sum_{a \in \mathcal{S}_u} y_a c_u(a, s_{-u}) - V^\eta(y) \right)$. La fonction de choix (4.2) est obtenue en prenant la fonction d'entropie comme fonction de pénalité, *i.e.* $V^\eta(y) = \eta \sum_{a \in \mathcal{S}_u} y_a \log(y_a)$.

Dans le cas du routage avec des flux divisibles, les stratégies mixtes peuvent être vues comme une manière de partager un flux entre plusieurs routes. Dans ce cas, la pénalité

s'interprète comme une taxe imposée aux joueurs afin de les inciter à ne pas router leur flux sur une seule route. La raison à cela est de garantir une certaine diversité dans l'utilisation des ressources du réseau, et une adaptabilité plus grande aux variations lentes des paramètres.

Rationalité limitée

Enfin, la fonction de choix découle du modèle de rationalité limitée des joueurs [Blu93, CFT97] dans lequel on suppose que les joueurs ne connaissent pas les gains, mais qu'ils en ont uniquement une perception. Cette perception est modélisée par un poids $\omega(c_u(a, s_{-u}))$ attribué à chaque action $a \in \mathcal{S}_u$, où $\omega : \mathbb{R} \to (0, +\infty)$ est une fonction croissante et différentiable.

Ils choisissent finalement une réponse selon la distribution :

$$\frac{\omega(c_u(a, s_{-u}))}{\sum_{b \in \mathcal{S}_u} \omega(c_u(b, s_{-u}))}, \tag{4.4}$$

si bien que les actions qui maximisent leur gain ont la plus grande probabilité d'être sélectionnée [2].

Le choix de ω sous le forme $\omega(y) = \exp(\eta^{-1} y)$ est le seul pour lequel le modèle de rationalité limitée et le modèle avec incertitudes coïncident (si bien que l'on peut voir la rationalité limité comme une conséquence des incertitudes aléatoires sur les gains). En effet, on a :

Proposition 4.1 (*Proposition 2.3 dans [HS02]*)

Supposons qu'une distribution P_u s'écrive à la fois sous la forme (4.4) et la forme (4.3). Alors P_u est la fonction de choix (4.2) pour un certain paramètre $\eta > 0$.

Suivant l'interprétation de la fonction de choix (4.2) que l'on choisit, le paramètre η peut être vu soit comme un niveau de bruit (sa variance), soit comme le poids de la fonction de pénalité, soit comme le degré de "limitation de la rationalité" des joueurs.

4.1.2 Analyse de l'algorithme dans les jeux de potentiel

L'un des objectifs de l'analyse de l'algorithme est d'étudier les propriétés du processus aléatoire $(S^\eta(t))$ quand η tend vers 0, notamment comment celui-ci se compare au modèle limite, c'est-à-dire le modèle classique de meilleure réponse. Le cadre des jeux de potentiel permet d'obtenir une caractérisation fine du comportement asymptotique du processus. Les résultats que nous obtenons lorsque l'algorithme est *asynchrone*, et que nous présentons dans cette section, sont classiques. Nous étudierons les autres processus de révision à la section 4.3.

2. Notons également qu'il existe également un modèle de rationalité limitée qui repose sur des erreurs [KMR93], dans lequel les joueurs choisissent une action qui n'est pas une meilleure réponse avec une certaine probabilité indépendante des gains. Les erreurs correspondent aux choix d'actions qui ne sont pas des meilleures réponses.

Dans cette section, on suppose que $(\mathcal{U}, \mathcal{S}, c)$ est un jeu de potentiel. Rappelons que cela implique qu'il existe une fonction de potentiel $P : \mathcal{S} \to \mathbb{R}$ telle que $\forall u \in \mathcal{U}, \forall s_{-u} \in \mathcal{S}_{-u}, \forall a, b \in \mathcal{S}_u$,

$$c_u(a, s_{-u}) - c_u(b, s_{-u}) = P(a, s_{-u}) - P(b, s_{-u}).$$

Dans ce cas, la probabilité de transition de l'état s à l'état (a, s_{-u}) quand le joueur u met à jour sa stratégie dans l'algorithme 3 se réécrit :

$$p^{\eta}_{s,(a,s_{-u})} = \frac{1}{U} \frac{\exp\left(\eta^{-1} P(a, s_{-u})\right)}{\sum_{b \in \mathcal{S}_u} \exp\left(\eta^{-1} P(b, s_{-u})\right)}. \tag{4.5}$$

En effet, la fonction de choix est donnée par une distribution de Gibbs qui ne dépend que des différences relatives entre les valeurs des gains du joueur, et non des valeurs absolues.

Notons que la chaîne de Markov $(S^{\eta}(t))$ sur l'espace d'état $\mathcal{S}$ induite par l'algorithme 3 est une chaîne *irréductible*, car tout état peut être atteint avec une probabilité non nulle à partir de n'importe quel autre état (même en au plus U transitions, où U est le nombre de joueurs), et *apériodique* car la probabilité, à chaque itération, de ne pas changer d'état est positive. Par conséquent, il existe une unique probabilité stationnaire $\pi^{\eta} = (\pi^{\eta}(s))_{s \in \mathcal{S}}$ et la loi forte des grands nombres s'applique : presque sûrement, $\pi^{\eta}(s) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}_{S^{\eta}(t)=s}$.

Comme la chaîne est irréductible, on a en plus que $\pi^{\eta}(s) > 0$ pour tout état s. Cela signifie que, lorsque le paramètre η est fixé, le processus visite chaque état une infinité de fois. Cela constitue une différence fondamentale par rapport à l'algorithme déterministe de meilleure réponse qui converge presque sûrement en un temps fini vers un ensemble d'équilibres de Nash dont il ne sort plus jamais par la suite.

États stochastiquement stables

Lorsque η tend vers 0, l'algorithme 3 est proche de l'algorithme déterministe de meilleure réponse. Même si la chaîne de Markov $(S^{\eta}(t))$ demeure irréductible et passe par tous les états une infinité de fois, la proportion de temps passé dans certains états tend vers 0 alors qu'elle reste positive pour d'autres états. On peut donc classer les états selon que la probabilité $\pi^{\eta}(s)$ quand η tend vers 0 est nulle ou positive.

Définition 4.2 (*Etat stochastiquement stable*)

Soit $S^{\eta}(t)$ un processus aléatoire sur $\mathcal{S}$, ergodique pour toute valeur du paramètre $\eta > 0$, et de distribution stationnaire π^{η}. On dit que l'état $s \in \mathcal{S}$ est *stochastiquement stable* si :

$$\liminf_{\eta \to 0} \pi^{\eta}(s) > 0.$$

Dans le cas où η représente un niveau de bruit sur les gains du jeu, un état est stochastiquement stable s'il est visité une infinité de fois par l'algorithme stochastique de meilleure réponse quand le bruit tend vers zéro. Comme nous allons le voir, les états stochastiquement stables sont un sous ensemble des états limites pour l'algorithme de meilleure réponse sans

bruit : le bruit permet donc de sélectionner plus finement les états limites. Dans le cas des jeux de potentiel, les états stochastiquement stables sont ceux qui maximisent *globalement* le potentiel.

Notons $\mathcal{S}^*$ l'ensemble des états qui maximisent la fonction de potentiel, *i.e.* $\mathcal{S}^* \stackrel{\text{def}}{=} \underset{s\in\mathcal{S}}{\operatorname{argmax}} P(s)$.

Théorème 4.3 (*Reformulation de résultats dans [Blu97]*)

Soit $(S^\eta(t))$ le processus induit par l'algorithme 3. Si le jeu admet une fonction de potentiel, $s \in \mathcal{S}$ est stochastiquement stable si et seulement si $s \in \mathcal{S}^*$.

La démonstration repose sur le fait que la chaîne de Markov est réversible, c'est à dire que pour tout état s et s' on a $\pi^\eta(s)p^\eta_{s,s'} = \pi^\eta(s')p^\eta_{s',s}$. Dans ce cas, on vérifie aisément que la distribution stationnaire vaut :

$$\pi^\eta(s) = \frac{1}{K}\exp(\eta^{-1}P(s)),$$

où $K = \sum_{s'\in\mathcal{S}} \exp(\eta^{-1}P(s'))$. Cela donne le résultat.

Lien avec l'algorithme d'échantillonage de Gibbs

L'algorithme 3 peut également être vu comme un algorithme d'échantillonage de Gibbs (qui fait partie de la classe des algorithmes de Monte Carlo reposant sur les chaînes de Markov, voir par exemple le chapitre 3 de [LPW09]), où l'on génère un état qui a une probabilité élevée d'être optimal à partir des distributions marginales que sont les fonctions de choix de chaque joueur.

Le théorème 4.3 assure donc que lorsque η est proche de zéro, la chaîne de Markov est dans un état qui maximise le potentiel avec une très grande probabilité. Une méthode pour converger presque sûrement vers l'optimum global consiste alors à faire décroître le paramètre η par pallier, en attendant suffisamment longtemps à chaque pallier que le processus ait convergé vers la distribution stationnaire (ce qui est de plus en plus long au fur à mesure que la température décroît).

Une autre possibilité consiste à diminuer la température à chaque itération, mais suffisamment lentement pour ne pas que le processus soit "bloqué" dans un optimum local. Le paramètre η devient donc une fonction du temps $\eta(t)$. En particulier, le choix d'une décroissance pas plus rapide que $\frac{\Delta(P)}{\log(t)}$, où $t \geq 2$ est l'itération de l'algorithme et $\Delta(P)$ est la différence de potentiel maximale entre deux états, est connu pour converger presque sûrement vers un optimum global du potentiel. Il s'agit d'une version particulière de la méthode de recuit simulé par échantillonage de Gibbs. Le processus est alors une chaîne de Markov non-homogène dont on peut montrer qu'elle est fortement ergodique (voir [Bre99], page 314) et converge bien vers la distribution limite. Notons qu'il existe d'autres façons d'échantillonner, notamment par la méthode de Metropolis dont la convergence a été prouvée dans [Haj88].

La méthode de recuit simulé est très intéressante si l'objectif est de maximiser la fonction de potentiel, car elle garantit avec probabilité un de converger vers un optimum global. La contrepartie est que la convergence est très lente quand η est petit (voir page 118 dans [BEK05]). De plus, lorsque η est trop petit, des problèmes numériques apparaissent : certaines probabilités sont arrondies à zéro ce qui créé plusieurs classes de récurrence (*i.e.* la chaîne de Markov n'est plus irréductible).

La proposition suivante dit que la probabilité d'être dans un état optimal est décroissante avec le paramètre η. Il est donc intéressant en pratique de prendre η le plus petit possible mais de manière à éviter les problèmes numériques que nous venons de mentionner.

Proposition 4.4

Soit $\pi^\eta(S^*)$ la probabilité stationnaire d'être dans un état optimal. Alors $\pi^\eta(S^*)$ est strictement décroissante en fonction de η.

Démonstration : Soit $M = \max_{s\in\mathcal{S}} P(s)$. Alors

$$\pi^\eta(S^*) = \sum_{s\in S^*} \frac{\exp(\eta^{-1}M)}{\sum_{s'\in\mathcal{S}} \exp(\eta^{-1}P(s))} = |S^*| \frac{1}{\sum_{s\in\mathcal{S}} \exp(\eta^{-1}(P(s)-M))}.$$

Comme $\forall s \in \mathcal{S}, P(s) - M \leq 0$, alors $\exp(P(s) - M) \leq 1$. Ainsi, $\exp(\eta^{-1}(P(s) - M)) = (\exp(P(s) - M))^{\eta^{-1}}$ est décroissant en η^{-1}, et son inverse est décroissant en η. ■

Exemple : Considérons le jeu de potentiel à deux joueurs suivant :

Gains

$(1,3)$	$(0,1)$
$(0,2)$	$(3,4)$

Potentiel

2	0
1	3

Espace d'état

(a,a)	(a,b)
(b,a)	(b,b)

Ce jeu comporte les deux équilibres de Nash (a,a) et (b,b), qui ont pour potentiel respectivement la valeur 2 et 3. La figure 4.1 montre la distribution stationnaire en fonction du paramètre η. Quand il tend vers zéro, la distribution se concentre sur l'état (b,b) qui maximise le potentiel.

4.2 Utilisation de l'algorithme stochastique de meilleure réponse pour l'optimisation du routage dans les réseaux ad hoc de mobiles

Les réseaux ad hoc de mobiles sont constitués de transmetteurs mobiles répartis sur un espace géographique qui ont la capacité de s'auto-configurer en vue de former un réseau sans topologie préalable (voir par exemple [GLJ09]). Ces réseaux sont utilisés dans différents contextes : le rétablissement des communications lors de catastrophes naturelles sur des

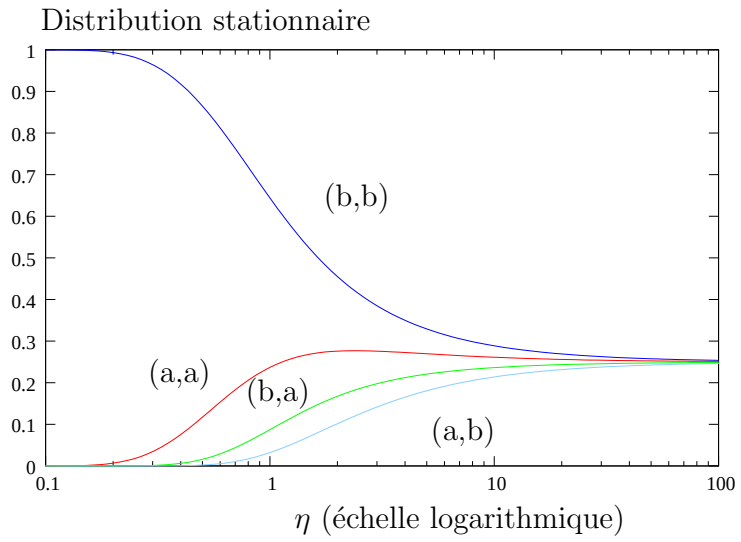


Figure 4.1 – Distribution stationnaire de l'algorithme 3 dans le jeu de potentiel en fonction du paramètre η (qui peut être vu comme un niveau de bruit).

zones larges, assurer la liaison entre différentes entités d'un groupe mobile comme un groupe de voitures (ces réseaux portent le nom de VANET)... Sans infrastructure préalable, ces réseaux sont rapides et peu coûteux à déployer.

La particularité de ces réseaux réside dans une topologie qui varie au cours du temps en raison de la mobilité de ses noeuds. Par conséquent, le problème majeur est de pouvoir maintenir les communications en assurant l'existence de routes entre le mobile source et le mobile destinataire. Dans cette optique, plusieurs protocoles de routage ont été proposés. Les protocoles les plus répandus sont AODV [CBR04] (pour Ad-hoc On-demand Distance Vector), DSR (Dynamic Source Routing), DSDV (Destination Sequenced Distance Vector), et TORA (Temporally Ordered Routing Algorithm). Ces quatre protocoles, qui utilisent des méthodes de routage diverses (proactives ou réactives), utilisent la même métrique pour évaluer la qualité du routage obtenu : le nombre de noeuds relai. Plus petite est la distance entre les mobiles communicants en terme de nombre de mobiles relais, meilleur est le routage. Par conséquent, ils ne tiennent compte ni de l'interaction entre les différentes communications qui peut se manifester par de l'engorgement au niveau de certains noeuds, ni du type de demande (application supportée par la communication), ni de la qualité des liens.

Nous proposons dans cette section une manière de prendre en compte de manière efficace ces paramètres dans une version modifiée du protocole AODV, qui repose sur l'algorithme stochastique de meilleure réponse.

4.2.1 Modélisation du réseau ad hoc

On considère un réseau ad hoc de mobiles que l'on représentera par un graphe orienté, dont l'ensemble des sommets, noté $\mathcal{N}$, est l'ensemble des terminaux mobiles qui peuvent

communiquer par un protocole radio prédéfini (par exemple 802.11). Il existe un arc reliant un sommet n au sommet m si n peut transmettre des informations à m. Comme il s'agit de communications radio, les arcs ne sont pas nécessairement symétriques (la capacité de réception dépend de la qualité de l'antenne, alors que la capacité d'émission dépend essentiellement de la puissance et de la capacité de la batterie).

Sur ce graphe, un ensemble $\mathcal{U}$ de communications sont établies entre un sommet source et un sommet destination. On suppose que ces flux ont un débit d'émission qui est stationnaire, par exemple des paquets sont émis selon un processus de Poisson d'intensité constante. Chaque communication appartient à une classe de priorité, de 1 à C, la classe la plus petite étant la plus prioritaire. Ces classes permettent de prendre en compte les contraintes de qualité de service des différentes application. A chaque classe de priorité est associé un poids qui représente l'importance relative des priorités les unes par rapport aux autres. Le flux u a un poids noté ω_u dépendant de la classe qui lui est associé.

Pour chaque flux u, on note $\mathcal{S}_u$ l'ensemble des routes simples joignant la source à la destination du flux sur le graphe de communication. Une route est donc une suite de terminaux mobiles qui connectent la source au mobile destinataire. Le choix d'un de ces chemins par le flux u est noté s_u. Le profil d'actions $s = (s_u)_{u\in\mathcal{U}}$ est appelé un *routage* des flux, et on note $\mathcal{S}$ l'ensemble des routages possibles, *i.e.* $\mathcal{S} = \underset{u\in\mathcal{U}}{\times} \mathcal{S}_u$.

Notons[3] par $\ell_n(s)$ l'état d'un sommet n du graphe sous le routage s : c'est un vecteur binaire $(\ell_n^u)_{u\in\mathcal{U}}(s)$ tel que $\ell_n^u(s) = 1$ si le flux u passe par le sommet n, *i.e.* $n \in s_u$, et 0 sinon. Lorsqu'aucune confusion n'est possible, on omettra de noter la dépendance de l'état des noeuds en fonction de s. Enfin, le délai moyen en un sommet est noté $d_n^u(\ell_n)$, et le délai total sur un chemin $p \in \mathcal{S}_u$, quand le routage est s, noté $D_p^u(s)$. On fait les hypothèses suivantes :

- Le délai total d'un flux sur un chemin est dû exclusivement aux délais sur les sommets traversés (le délai de transmission entre deux sommet est négligé).
- Les délais sont additifs sur les chemins, c'est-à-dire $D_p^u(s) = \sum_{n\in p} d_n^u(\ell_n)$.
- Le délai en un sommet dépend de la priorité des flux. Les flux les plus prioritaires ont des délais inférieurs aux flux ayant une priorité inférieure. Dans la section numérique, nous utilisons un modèle de file d'attente avec préemption par les flux prioritaires.

Le problème consiste alors à trouver un routage des flux sur le réseau afin de minimiser le délai global, où le délai global[4] vaut $F(s) = \sum_{u\in\mathcal{U}} \omega_u D_{s_u}^u(s) = \sum_{u\in\mathcal{U}} \sum_{n\in s_u} \omega_u d_n^u(\ell_n)$. Le problème d'optimisation s'écrit simplement :

$$\min_{s\in\mathcal{S}} F(s).$$

Ce problème peut être résolu en explorant toutes les possibilités de routage, mais cette approche par force brute centralisée est inefficace en pratique en raison de la taille exponentielle de l'ensemble des routages. Une autre alternative consiste à exploiter les

3. Nous reprenons les notations générales pour les problèmes d'allocation de ressources que nous avons utilisées à la section 2.3.3.

4. Il s'agit en fait du délai global pondéré par les priorités des flux.

éventuelles structures des fonctions de délais, par exemple la convexité comme dans [GL07]. Néanmoins, ces méthodes reposent sur un contrôleur centralisé ce qui les rend inadaptées au routage dans les réseaux ad hoc, qui sont par essence des structures sans entité centrale.

Notre approche consiste à placer les prises de décision au niveau des flux. Ces décisions doivent tenir compte du délai du flux, ainsi que de son impact sur les autres flux. D'après la section 2.3.3, si l'on définit la fonction de coût par arc suivante :

$$\delta_n^u(\ell_n) \stackrel{\text{def}}{=} \omega_u d_n^u(\ell_n) - \sum_{v \neq u | n \in s_v} \Big(\omega_v d_n^v(\ell_n - e_n^u) - \omega_v d_n^v(\ell_n) \Big), \tag{4.6}$$

où e_n^u est le vecteur de taille $|\mathcal{U}|$ qui vaut 1 en la composante u et zéro ailleurs, et si l'on pose $\Delta_p^u(s) \stackrel{\text{def}}{=} \sum_{n \in p} \delta_n^u(\ell_n)$, alors le jeu définit par le triplet $(\mathcal{U}, \mathcal{S}, \Delta)$ est un jeu de potentiel dont F est une fonction de potentiel.

Par conséquent, l'algorithme de meilleure (en version asynchrone) réponse converge vers un maximum local du potentiel. L'un des avantages de cet algorithme est le fait qu'il peut être implémenté de manière complètement décentralisée et que, bien qu'on puisse exhiber des exemples pour lesquels il converge en un temps exponentiel, il est en pratique très rapide. Mais il ne converge pas vers un optimum global. En raison de la capacité finie des réseaux, on ne peut pas borner la "dérivée" de la fonction de délai, et par conséquent on ne peut pas avoir de garantie sur la dégradation maximale entre le pire maximum local et l'optimum local comme le montre l'exemple suivant.

Prix de l'anarchie non borné : Considérons le réseau de la figure 4.2. Supposons que deux communications de même priorité aient lieu entre la source A et la destination B, et entre la source C et la destination D, au débit d'émission de $1kb/s$. Notons par u et v les flux correspondants.

Un chemin est décrit par l'ensemble des noeuds qui le compose. Ici, $\mathcal{S}_u = (U, V), (X, Y)$ et $\mathcal{S}_v = ((U, Y), (V, X))$. Supposons que les délais soient ceux d'une file d'attente $M/M/1$, c'est-à-dire que le délai pour franchir un noeud vaut $\dfrac{\lambda}{\mu - \lambda}$ (λ est le débit arrivant dans la file et μ le taux de service), et que les taux de service soient $2 + 2\varepsilon$, $2 + \varepsilon$, 3, $2 + \varepsilon$ (en kb/s) pour, respectivement, les noeuds X, Y, U et V. La matrice du jeu de potentiel associé, donc tenant compte de l'impact des des joueurs, est :

	$(U \quad , \quad Y)$	$(V \quad , \quad X)$
(U, V)	$\left(3.5 + \dfrac{1}{1+\varepsilon} \quad , \quad 3.5 + \dfrac{1}{1+\varepsilon}\right)$	$\left(0.5 - \dfrac{1}{1+\varepsilon} + \dfrac{4}{\varepsilon} \quad , \quad \dfrac{1}{1+2\varepsilon} - \dfrac{1}{1+\varepsilon} + \dfrac{4}{\varepsilon}\right)$
(X, Y)	$\left(\dfrac{1}{1+2\varepsilon} - \dfrac{1}{1+\varepsilon} + \dfrac{4}{\varepsilon} \quad , \quad 0.5 - \dfrac{1}{1+\varepsilon} + \dfrac{4}{\varepsilon}\right)$	$\left(\dfrac{1}{1+\varepsilon} - \dfrac{1}{1+2\varepsilon} + \dfrac{2}{\varepsilon} \quad , \quad \dfrac{1}{1+\varepsilon} - \dfrac{1}{1+2\varepsilon} + \dfrac{2}{\varepsilon}\right)$

On voit qu'il y a deux équilibres de Nash qui sont $\text{NE}_1 = ((U, V), (U, Y))$ et $\text{NE}_2 = ((X, Y), (V, X))$, dont les délais globaux sont respectivement $4 + \dfrac{2}{1+\varepsilon}$ et $\dfrac{2}{\varepsilon} + \dfrac{2}{1+\varepsilon}$, et dont le rapport tend vers zéro quand ε tend vers 0. Le rapport entre les deux performances n'est donc pas borné.

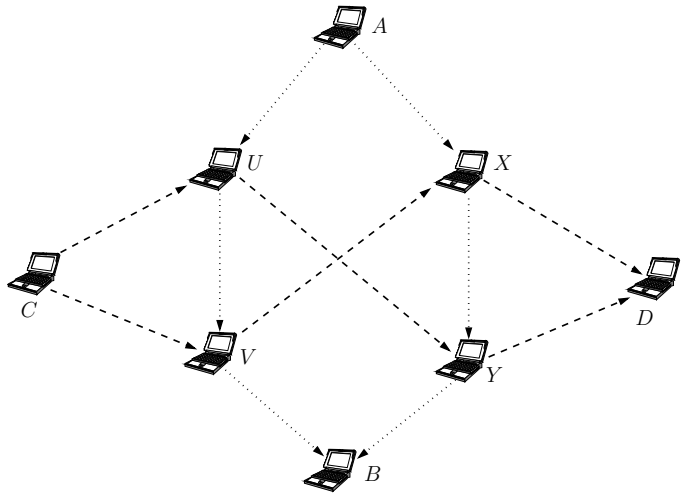


Figure 4.2 – Exemple de réseau dans lequel le ratio des performances entre le pire équilibre et l'optimum global est non borné.

4.2.2 Implémentation de l'algorithme stochastique de meilleure réponse

Dans cette section, nous expliquons comment, en pratique, les calculs et les échanges d'informations nécessaires à l'exécution de l'algorithme stochastique de meilleure réponse sont opérés.

Il faut d'abord noter que la fonction de coût (4.6) $\delta_n^u(\ell_n)$ doit pouvoir être calculée en chaque noeud. Ce calcul nécessite de connaître le délai des flux autre que u sans ce dernier, ce qui peut être rédhibitoire. Une option consiste à tester puis interrompre le flux sur chacune des routes possibles. Cela suppose, et nous ferons cette hypothèse, que le temps nécessaire pour obtenir une bonne estimation des délais moyens de chaque flux est très inférieur au temps de convergence de l'algorithme. Il est également nécessaire que le temps soit universel, en ce sens que les horloges des mobiles sont synchronisées. Cela est essentiel pour le calcul des délais à partir du temps marqué sur les paquets qui circulent dans le réseau.

Sous ces hypothèses, l'algorithme théorique peut être transformé en un programme implémenté dans les réseaux de manière distribuée. Une partie du programme (le calcul des fonctions de coût) est exécutée par les noeuds du réseau (les mobiles), et une autre partie (la meilleure réponse stochastique) par chaque flux, ou plus précisément par la source émettrice du flux.

Chaque paquet venant d'un flux donné est marqué des informations suivantes : sa source, son chemin (la suite des sommets) et sa date d'émission. Les paquets du flux suivent alors le chemin indiqué. A chaque sommet, de courts messages d'information, dont nous détaillons l'utilité après, sont envoyés à la source. On suppose que ces messages ont un impact négligeable sur les performances du réseau.

Programme côté réseau

Le rôle des sommets intermédiaires qui transmettent les flux est de calculer les coûts (4.6) et de les transmettre à l'émetteur du flux. Nous expliquons comment l'on peut procéder. Chaque sommet peut calculer le coût du flux u en suivant la suite d'actions suivante :

1. Mesurer le délai de traversée du sommet de chaque flux (ce sont des moyennes effectuées sur des fenêtres de temps suffisamment larges).
2. Une fois que le flux u cesse d'utiliser le sommet, mesurer le délai des flux restant.
3. A partir de ces mesures, calculer la fonction de coût associée au flux u.
4. Envoyer cette valeur à la source du flux (bien souvent, cela se fait en attachant cette information au prochain paquet transitant vers la source).

Programme côté flux

Ici, l'algorithme est exécuté en boucle infinie de façon nécessairement asynchrone. Pour cela, on introduit un temps d'attente aléatoire entre deux boucles de l'algorithme qui suit une loi exponentielle dont le paramètre est supérieur au délai maximal sur le réseau. Les étapes de l'algorithme sont :

1. A la fin du temps d'attente, le flux interrompt sa communication sur le chemin courant.
2. Il reçoit les messages provenant de tous les sommets sur le chemin courant. Il calcule la somme des fonctions de coût.
3. Le flux choisit sa nouvelle route suivant la distribution donnée dans l'algorithme stochastique de meilleure réponse.
4. Le flux génère un temps d'attente aléatoire distribué selon une loi exponentielle, dont le paramètre dépend du délai d'une itération de l'algorithme.

Estimation du délai par des mesures de charge

Afin d'éviter la procédure complexe de mesure des délais et des fonctions de coût, il est possible d'utiliser plutôt des estimations de la charge des sommets.

Les sommets mesurent le débit de chaque flux qui l'utilisent (le nombre de paquets arrivant par unité de temps) plutôt que le délai de chaque paquet. Le délai moyen est ensuite estimé en utilisant des formules aussi réalistes que possible reliant le délai aux taux moyens, comme les formules (4.7). De façon similaire, la fonction de coût d'un flux est estimée à partir des formules sans avoir besoin d'interrompre le flux.

Cette approche possède plusieurs avantages comparativement à l'approche qui repose sur les mesures. D'une part, cela ne nécessite pas que les sommets utilisent un temps universel (pour le calcul des délais). D'autre part, les fonctions de coût peuvent être calculées sans interrompre les flux.

Le principal inconvénient de cette approche tient à la perte de qualité de l'estimation du délai. Cependant, si on limite le nombre de chemins possibles pour chaque flux, le choix

d'une meilleure réponse ne nécessite pas un calcul très précis à partir du moment où les différences de délais sur les routes est grande. Si les différences ne sont pas significatives, il est nécessaire de recourir à des mesures afin d'éviter une erreur sur le choix du chemin le plus court.

4.2.3 Étude numérique

Dans cette section, des simulations de l'algorithme de routage sont proposées sur un modèle très simplifié de réseau ad hoc. Le but de ces simulations est d'une part d'évaluer le temps de convergence et la qualité de la solution comparativement aux performances que l'on peut s'attendre à obtenir avec le protocole AODV [5], et d'autre part d'étudier l'impact de certains paramètres tel que le nombre de chemins par flux. Les simulations ont été implémentées en utilisant le logiciel Maple. Nous commençons pas présenter le détail des simulations, en particulier le choix des paramètres.

La topologie du réseau : le réseau est représenté par un graphe orienté (les communications sans fil ne sont pas nécessairement symétriques) à $n = 20$ sommets, qui est généré aléatoirement de la manière suivante : chaque paire de sommets est connectée avec une probabilité p, et tant que le graphe généré n'est pas fortement connexe, c'est-à-dire tant qu'il existe deux sommets qui ne sont pas connectés, on retire un nouveau graphe. Il s'agit d'une méthode d'échantillonage avec rejet, la loi du tirage est donc uniforme sur l'ensemble des graphes fortement connexes si $p = 0.5$. Enfin, les taux de service des arcs du graphe valent tous 10 (l'unité est arbitraire).

Les flux : le nombre de flux est 25. Un flux est caractérisé par deux sommets : le sommet source et le sommet destination que l'on choisit différents. Deux flux différents peuvent avoir la même paire source destination, mais les applications portées par les flux peuvent être différentes, notamment leur priorité. Les sources et les destinations de chaque flux sont uniformément choisies dans l'ensemble des sommets. Les flux sont également caractérisés par un niveau de priorité et un taux d'émission des données au niveau de la source. Nous considérons deux niveaux de priorité : un flux de niveau 1 est prioritaire sur un flux de niveau 2. La priorité de chaque flux est tirée avec probabilité 0.5. Les flux de priorité 1 (resp. 2) ont un taux d'émission λ (resp. 5λ). Les flux de priorité 1 correspondent à des communications portant des applications ayant des contraintes de délai, tels des applications streaming ou voix.

Délais : le délai sur chaque arc est modélisé par un délai moyen dans une file d'attente M/M/1 avec deux classes de priorités dans le cas *préemptif*. Cela signifie qu'un paquet

5. Il faut noter que ces simulations se placent dans un cadre idéal très éloigné de la réalité. Des travaux ont montré que les performances attendues de AODV peuvent être très loin de celles que l'on observe par la mise en pratique de ce protocole [CJWK02].

en service interrompt immédiatement son service lorsqu'un paquet plus prioritaire que lui arrive dans la file d'attente.

Notons par λ_1 et λ_2 les taux d'arrivées des flux de chacune des classes de priorité dans une file d'attente, et l'on suppose que les arrivées entre ces flux sont indépendantes. Soit μ le taux de service de la file d'attente. Alors le délai moyen d_1 (resp. d_2) pour les paquets les plus (resp. les moins) prioritaires est donné par :

$$\begin{aligned} d_1 &= \frac{\lambda_1}{\mu - \lambda_1}, \\ d_2 &= \frac{1}{\lambda_2}\left(\frac{\lambda^2}{\mu - \lambda} - \frac{\lambda_1^2}{\mu - \lambda_1}\right), \end{aligned} \tag{4.7}$$

où $\lambda = \lambda_1 + \lambda_2$. Le fait d'être dans un système préemptif donne le délai pour la classe la plus prioritaire : tout se passe comme si le flux prioritaire était seul. Ensuite, par la loi de Little, on a $N_i = \lambda_i d_i$, avec N_i le nombre moyen de paquets de la classe i dans la file d'attente. Comme les flux sont indépendants, le processus d'arrivées du flux global (la somme des deux classes) est un processus ponctuel de Poisson, donc le temps de séjour moyen pour un paquet est $d = \dfrac{\lambda}{\mu - \lambda}$. En utilisant la loi de Little sur le flux global, on obtient $N = \lambda d$. Finalement, $d_2 = \dfrac{1}{\lambda_2} N_2 = \dfrac{1}{\lambda_2}(N - N_1)$. Cela donne le résultat pour d_2. Ce calcul peut être étendu en itérant ce schéma au cas où il y aurait plus de deux classes de priorité.

Ensemble de chemins : On fixe a priori la cardinalité de l'ensemble des chemins pour chaque flux (la cardinalité est un paramètre de l'algorithme). Le premier chemin est celui obtenu par le protocole AODV, c'est-à-dire que c'est celui qui minimise le nombre de relais entre la source et la destination du flux. Ce calcul de plus court chemin est effectué par le protocole de manière décentralisée. Le second chemin est ensuite calculé en utilisant une nouvelle fois le protocole AODV, mais sur un graphe modifié, dans lequel on a retiré le premier arc du premier chemin [6]. Le troisième chemin est obtenu de la même manière, mais en retirant les premiers arcs des chemins précédents, et ainsi de suite. Cette façon de choisir les chemins possède l'avantage d'être facilement implémentable, puisque c'est le sommet source qui modifie les chemins en changeant l'arc initial. D'autres alternatives pourraient néanmoins être testées, par exemple en modifiant un arc intermédiaire.

Courbes et commentaires

La figure 4.3 montre une trajectoire de l'algorithme sur une instance aléatoire choisie avec les paramètres décrits précédemment. Le délai global est en ordonnée, l'abscisse étant le nombre d'itérations, en échelle logarithmique. La courbe est globalement décroissante, et

6. Si aucun chemin ne connecte la source à la destination dans le graphe modifié, AODV ne renvoie aucun chemin supplémentaire.

les petits ressauts correspondent à la visite d'états qui permettent de sortir de minimums locaux. On voit que la convergence vers une bande de valeurs de délai très étroite (moins de 1% de variation par rapport à la valeur minimale que l'on obtient sur toute la trajectoire) a lieu en quelques dizaine d'itérations.

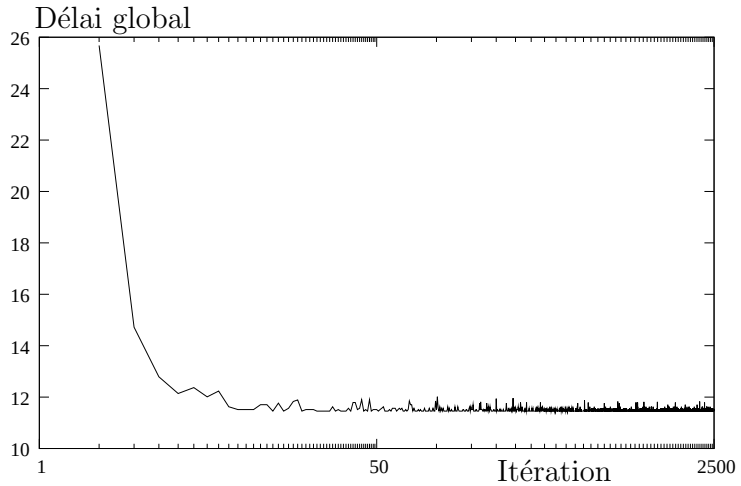


Figure 4.3 – Convergence d'une trajectoire de l'algorithme de meilleure réponse stochastique avec une échelle logarithmique en abscisse.

La figure 4.4 montre le gain que l'on peut obtenir en ajoutant à l'ensemble des chemins possibles un, deux ou trois chemins au chemin choisi par le protocole AODV. Lorsque les taux d'émission de chaque flux augmentent (l'axe des abscisses), et donc que le trafic sur le réseau augmente, le gain croît considérablement lorsque l'on ajoute un chemin alternatif. Alors que l'ajout d'un chemin apporte un bénéfice conséquent, ajouter une seconde route n'a que peu d'impact et une troisième alternative est inutile.

Enfin, la figure 4.5 montre le temps de convergence nécessaire à l'algorithme pour atteindre un routage dont le délai global ne dépasse pas de plus de 1% le délai minimal obtenu en 5000 itérations. Moins de 50 itérations sont suffisantes dans la plupart des cas, ce qui est favorable à une implémentation de ce type d'algorithme dans les réseaux ad hoc. La moyenne ici est de $18,8$ itérations, sur quelques centaines d'instances.

4.3 Robustesse de la dynamique stochastique de meilleure réponse aux processus de révision des stratégies

Dans les sections précédentes, nous avons fait l'hypothèse que les prises de décision sont parfaitement asynchrones, ou en d'autres termes qu'il n'y a jamais deux joueurs qui modifient leur stratégie en même temps. Cela n'est pas réaliste dans les grands systèmes distribués, à moins de disposer d'un contrôleur central qui coordonne les prises de décisions des mobiles. Mais bien souvent, un tel contrôleur n'existe pas, comme c'est le cas dans les

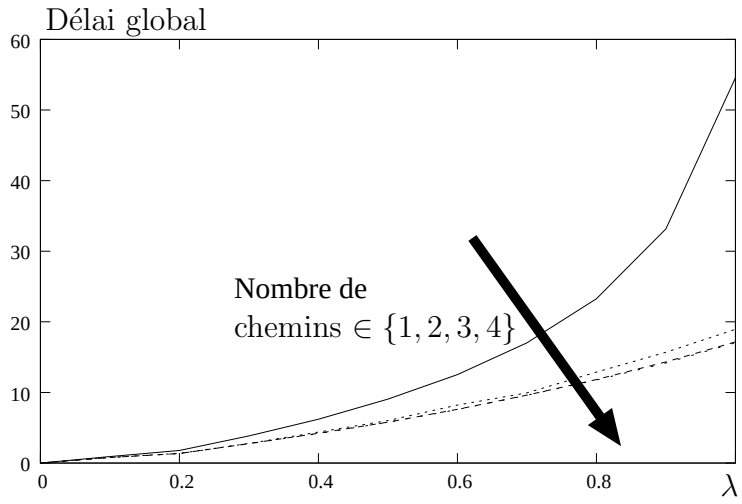


Figure 4.4 – Délai global en fonction de la quantité de flux dans le réseau, pour une topologie fixée, avec $1, 2, 3$ et 4 chemins possibles. Les intervalles de confiance à 95% sont trop petits pour apparaître. L'unité de λ est arbitraire.

réseaux ad hoc, les réseaux pair à pair, ou bien les systèmes hétérogènes comme les réseaux sans fil avec plusieurs technologies.

Afin de résoudre ce problème de synchronisation sans contrôleur central, la méthode que nous avons utilisée (et inspirée de [CB10]) consiste à imposer à chaque mobile d'attendre un temps aléatoire tiré selon une loi exponentielle entre chaque révision. Ce faisant, la probabilité que deux mobiles agissent simultanément est nulle. De plus, cela suppose que les prises de décision n'induisent aucun délai ce qui n'est pas non plus réaliste.

Pour ces raisons pratiques, nous analysons maintenant le comportement de l'algorithme 3 pour des processus de révisions plus généraux que le processus asynchrone. Nous allons en particulier montrer que, dans ce cas, l'algorithme stochastique de meilleure réponse ne converge plus vers un maximum global du potentiel. Les actions des joueurs sont prises de la même manière que précédemment. Notamment, on suppose que lorsque plusieurs joueurs révisent leur stratégie simultanément, ils ne forment pas de coalition, c'est-à-dire que leur décision ne repose que sur leurs gains *individuels*.

Rappelons qu'un *processus de révision* est une mesure de probabilité μ sur l'ensemble $\mathcal{P}(\mathcal{U})$ des sous-ensembles de joueurs telle que :

$$\forall u \in \mathcal{U}, \exists \mathcal{V} \subseteq \mathcal{U} \text{ tel que } u \in \mathcal{V} \text{ et } \mu(\mathcal{V}) > 0,$$

où $\mu(\mathcal{V})$ est la probabilité que tous les joueurs dans $\mathcal{V}$ modifient leur stratégie simultanément. L'algorithme 4 généralise l'algorithme de meilleure réponse stochastique pour des processus de révision généraux.

Le processus $(S^\eta(t))$ induit par l'algorithme, où $S^\eta(t)$ est l'état (stratégie) du jeu à la $t^{\text{ème}}$ itération, est une chaîne de Markov homogène sur $\mathcal{S}$. Soient s et s' deux états et $\mathcal{V}$ un sous-ensemble de joueurs tel que $s_u = s'_u$ si $u \notin \mathcal{V}$. Alors la probabilité de transition de s

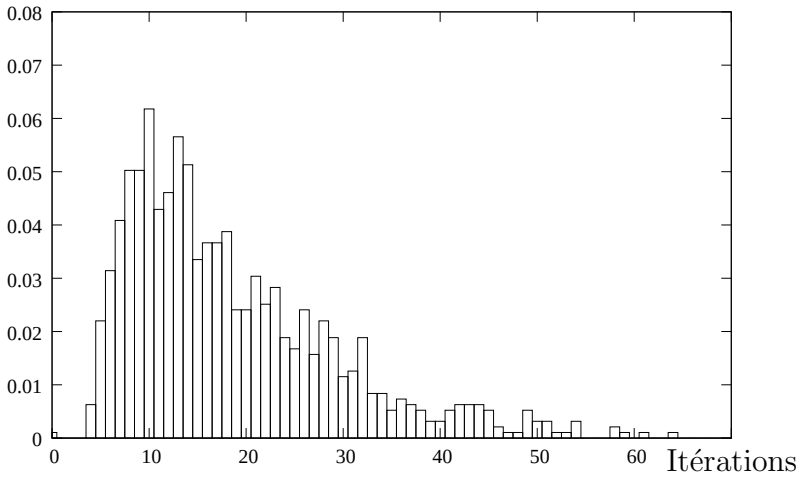


Figure 4.5 – Distribution du nombre maximal d'itérations par mobile pour atteindre un routage ayant un délai supérieur d'au plus 1% au délai du routage optimal.

Algorithme 4: Algorithme stochastique de meilleure réponse sous le processus de révision μ

initialisation;
Chaque joueur $u \in \mathcal{U}$ choisit une action s_u uniformément dans $\mathcal{S}_u$;
répéter
Choisir un ensemble de joueurs $\mathcal{V} \in \mathcal{P}(\mathcal{U})$ avec probabilité $\mu(\mathcal{V})$;
Chaque joueur u dans $\mathcal{V}$ choisit l'action a avec probabilité

$$\frac{\exp\left(\eta^{-1} c_u(a, s_{-u})\right)}{\sum_{b \in \mathcal{S}_u} \exp\left(\eta^{-1} c_u(b, s_{-u})\right)}$$

jusqu'à *l'infini*;

à s' est maintenant donnée par :

$$p^{\eta}_{s,s'} = \sum_{\mathcal{Z} \supseteq \mathcal{V}} \mu(\mathcal{Z}) \prod_{u \in \mathcal{Z}} \frac{\exp\left(\eta^{-1} c_u(s'_u, s_{-u})\right)}{\sum_{b \in \mathcal{S}_u} \exp\left(\eta^{-1} c_u(b, s_{-u})\right)}. \tag{4.8}$$

Comme pour le cas asynchrone, la chaîne de Markov est irréductible et apériodique, si bien qu'il existe une distribution stationnaire π^{η}. Rappelons qu'un état s est *stochastiquement stable* si $\liminf_{\eta \to 0} \pi^{\eta}(s) > 0$.

4.3.1 Caractérisation des états stochastiquement stables pour les processus de révision généraux

Dans cette section, nous montrons que les états stochastiquement stables de l'algorithme stochastique de meilleure réponse sont inclus dans l'ensemble des sommets récurrents du graphe de meilleure réponse associé au processus de révision. Ce résultat est vrai dans tout jeu. Le cas des jeux de potentiels est abordé dans les sections suivantes.

Le théorème 4.3 repose sur le fait que la chaîne de Markov associée à l'algorithme est réversible ce qui nous permet de calculer explicitement la distribution stationnaire. Néanmoins, cela n'est plus le cas dès lors que le processus de révision n'est pas asynchrone. Nous utilisons alors une formule explicite de la distribution stationnaire qui repose sur les arbres couvrants de l'ensemble des stratégies $\mathcal{S}$.

Formulation de la distribution stationnaire par les arbres couvrants

Étant donné un jeu $(\mathcal{U}, \mathcal{S}, c)$ et un profil d'actions $s \in \mathcal{S}$, un *s-arbre* sur l'ensemble des sommets $\mathcal{S}$ est un graphe orienté tel qu'il existe un unique chemin reliant tout sommet à la racine s. Il s'agit donc d'un arbre couvrant de l'ensemble des sommets $\mathcal{S}$ ayant s comme racine.

Pour un processus de révision μ donné, on dira qu'un s-arbre A est *faisable* pour le processus de révision μ si pour toute arête $(a, b) \in A$, la transition de a à b est faisable, *i.e.* s'il existe $\mathcal{V}$ un sous-ensemble de joueurs tel que $b_{-\mathcal{V}} = a_{-\mathcal{V}}$ (notation compacte de $\forall u \notin \mathcal{V}, b_u = a_u$) et $\mu(\mathcal{V}) > 0$. Nous notons $\mathcal{A}(s)$ l'ensemble des s-arbres faisables pour le processus de révision courant.

De manière générale, nous dirons qu'un graphe sur $\mathcal{S}$ est faisable pour le processus de révision[7] μ si chaque arc du graphe est faisable sous μ.

Afin de simplifier les notations, on suppose que les processus de révisions sont tels qu'il existe un plus petit ensemble de joueur qui réalise chaque transition faisable. Néanmoins les théorèmes de cette section restent vrais dans le cas général. La fonction qui associe à chaque transition, c'est-à-dire chaque arc du graphe, le plus petit ensemble de joueurs qui réalise cette transition est appelée *fonction de révision du graphe*.

Définition 4.5 (*Fonction de révision d'un graphe*)

Soit μ un processus de révision et $G = (\mathcal{S}, \mathcal{E})$ un graphe orienté faisable pour le processus de révision μ. La fonction de révision du graphe associée à μ est une fonction $R : \mathcal{E} \to \mathcal{P}(\mathcal{U})$ sur l'ensemble des arcs du graphe telle que, pour tout $(a, b) \in \mathcal{E}$:

- $\mu(R(a, b)) > 0$,
- si $\mathcal{V} \subseteq \mathcal{P}(\mathcal{U})$ est tel que $\mu(\mathcal{V}) > 0$ et $b_{-\mathcal{V}} = a_{-\mathcal{V}}$, alors $R(a, b) \subseteq \mathcal{V}$.

Dans le cas de l'apprentissage asynchrone, un arc (s, t) est faisable si s et t diffèrent en au plus une composante. Ainsi, la fonction de révision de tout graphe faisable est donnée par $R((a, s_{-u}), (b, s_{-u})) = \{u\}$.

7. Ou plus brièvement faisable, quand il n'y a pas de confusion possible sur le processus de révision en question.

Le théorème 4.9 qui suit caractérise les états stochastiquement stables. Il repose sur la formule explicite de la distribution stationnaire d'une chaîne de Markov à partir des s-arbres donnée dans la proposition suivante (qui n'est pas spécifique aux chaînes de Markov obtenues par l'algorithme stochastique de meilleure réponse).

Proposition 4.6 (*Lemme 3.1 du chapitre 6 dans [FW98]*)

Soit une chaîne de Markov ergodique définie sur l'espace d'état $\mathcal{S}$ fini, et de probabilités de transition $p_{a,b}$ pour tout a et b dans $\mathcal{S}$. Soit π la distribution stationnaire et $\mathcal{A}(s)$ l'ensemble des s-arbres. Alors, pour tout $s \in \mathcal{S}$:

$$\pi(s) \propto \sum_{A \in \mathcal{A}(s)} \prod_{(a,b) \in A} p_{a,b}. \tag{4.9}$$

Nous proposons une démonstration de cette proposition qui est faite de manière compacte dans [FW98]. De plus, les techniques utilisées dans cette démonstration (essentiellement les constructions d'arbres) seront réutilisées dans des preuves ultérieures.

Notons $\mathrm{Succ}(s)$ et $\mathrm{Pred}(s)$ l'ensemble des sommets successeurs et prédécesseurs de s, *i.e.* $t \in \mathrm{Succ}(s)$ si $p_{s,t} > 0$ et $t \in \mathrm{Pred}(s)$ si $p_{t,s} > 0$. L'opérateur $\cup$ désigne l'ajout d'un arc au graphe, et $\setminus$ la suppression d'un arc. Le lemme suivant est un résultat sur les graphes qui est au fondement de la proposition :

Lemme 4.7

Soit $r \in \mathrm{Succ}(s)$ fixé. Alors l'ensemble des s-arbres vaut :

$$\mathcal{A}(s) = \bigcup_{t \in \mathrm{Pred}(s)} \bigcup_{A \in \mathcal{A}(t) | (s,r) \in A} \Big(A \backslash (s,r) \cup (t,s) \Big),$$

et par conséquent :

$$\sum_{A \in \mathcal{A}(s)} \prod_{(a,b) \in A} p_{a,b} = \sum_{t \in \mathrm{Pred}(s)} p_{t,s} \sum_{A \in \mathcal{A}(t) | (s,r) \in A} \prod_{(a,b) \in A | a \neq s} p_{a,b}. \tag{4.10}$$

Démonstration : On dira qu'un sommet s est en amont (resp. en aval) d'un sommet t dans un arbre s'il existe un chemin de s à t (resp. de t à s).

Soit $A = A_t \backslash (s,r) \cup (t,s)$, où A_t est un t-arbre qui contient l'arc (s,r). Montrons que c'est un s-arbre. Comme A a le même nombre d'arcs que tout arbre sur $\mathcal{S}$, il suffit de montrer qu'on peut relier tout sommet z à s. Si z est en aval de s dans A_t, alors z est relié à t qui est lui même relié à s par l'arc (t,s). Sinon, z est en aval de s et est donc directement relié à s.

A contrario, soit A_s un s-arbre. Montrons qu'il existe $t \in \mathrm{Pred}(s)$ tel que $A_t = A_s \backslash (t,s) \cup (s,r)$ est un t-arbre. En effet, choisissons $t \in \mathrm{Pred}(s)$ tel que t est un prédécesseur de s dans l'arbre A_s et t est en aval de r. Ce sommet existe et A_t est bien un t-arbre (voir la figure 4.6) : en effet, si un sommet est en amont de t dans l'arbre A_s alors il est connecté à t dans A_t, et s'il est en aval, il est connecté à s qui est lui-même connecté à t.

■

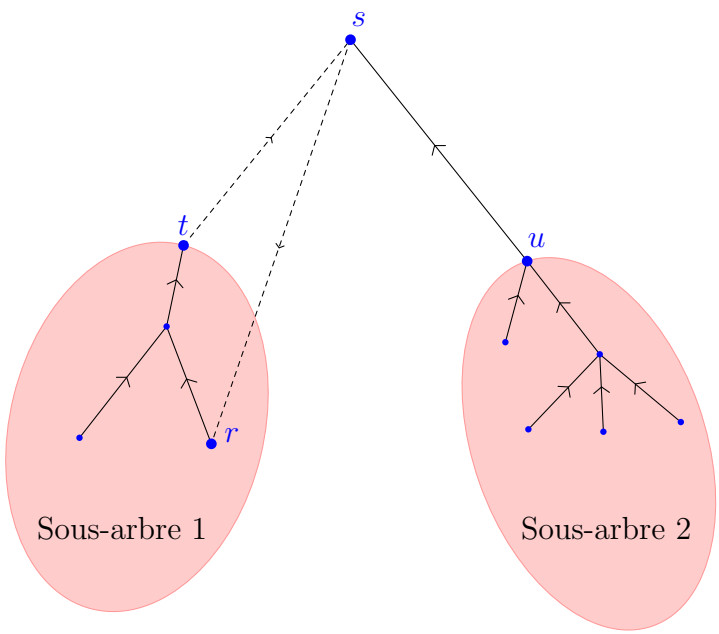


Figure 4.6 – Exemple de construction d'un t-arbre qui contient l'arête (s, r) à partir d'un s-arbre, où t est le sommet racine du sous-graphe qui contient r. Il suffit d'enlever l'arête (t, s) et d'ajouter l'arête (s, r).

Démonstration (Proposition 4.6) : Il s'agit maintenant de montrer que la distribution π donnée par (4.9) vérifie l'équation de balance locale $\sum_{t \in \mathrm{Pred}(s)} \pi(t) p_{t,s} = \pi(s)$ pour tout $s \in \mathcal{S}$. En effet (pour alléger les formules, nous omettons le coefficient de proportionnalité dans l'expression de π) :

$$\begin{aligned}
\sum_{t \in \mathrm{Pred}(s)} \pi(t) p_{t,s} &= \sum_{t \in \mathrm{Pred}(s)} p_{t,s} \sum_{A \in \mathcal{A}(t)} \prod_{(a,b) \in A} p_{a,b} \\
&= \sum_{t \in \mathrm{Pred}(s)} p_{t,s} \sum_{r \in \mathrm{Succ}(s)} p_{s,r} \sum_{A \in \mathcal{A}(t) | (s,r) \in A} \prod_{(a,b) \in A | a \neq s} p_{a,b},
\end{aligned}$$

où la dernière égalité vient de la partition de l'ensemble des t-arbres selon le successeur (unique) de s. En fin de compte :

$$\begin{aligned}
\sum_{t \in \mathrm{Pred}(s)} \pi(t) p_{t,s} &= \sum_{r \in \mathrm{Succ}(s)} p_{s,r} \sum_{t \in \mathrm{Pred}(s)} p_{t,s} \sum_{A \in \mathcal{A}(t) | (s,r) \in A} \prod_{(a,b) \in A | a \neq s} p_{a,b} \\
&= \sum_{r \in \mathrm{Succ}(s)} p_{s,r} \sum_{A \in \mathcal{A}(s)} \prod_{(a,b) \in A} p_{a,b} \\
&= \sum_{A \in \mathcal{A}(s)} \prod_{(a,b) \in A} p_{a,b} \\
&= \pi(s),
\end{aligned}$$

où la deuxième égalité vient du lemme précédent, et la troisième égalité vient du fait que $\sum_{r \in \mathrm{Succ}(s)} p_{s,r} = 1$. ■

Caractérisation des états stochastiquement stables

La proposition 4.6 nous donne une formule explicite de la distribution stationnaire pour l'algorithme 4, même dans le cas où le processus n'est pas réversible. Nous expliquons maintenant comment cette proposition est utilisé pour la détermination des états stochastiquement stables, en supposant, afin de simplifier les calculs, que le processus de révision est asynchrone. Dans ce cas, les arcs de l'arbre sont formés par une paire de sommets qui diffèrent en exactement une composante. Par la proposition 4.6 on a :

$$\begin{aligned}
\pi^\eta(s) &\propto \sum_{A \in \mathcal{A}(s)} \prod_{(r,(a,r_{-u})) \in A} p^\eta_{(r,(a,r_{-u}))} \\
&\propto \sum_{A \in \mathcal{A}(s)} \prod_{(r,(a,r_{-u})) \in A} \mu(\{u\}) \frac{\exp\big(\eta^{-1} c_u(a, r_{-u})\big)}{\sum_{b \in \mathcal{S}_u} \exp\big(\eta^{-1} c_u(b, r_{-u})\big)} \\
&\propto \sum_{A \in \mathcal{A}(s)} K(A) f_\eta(A) \exp\big(-\eta^{-1} C(A)\big).
\end{aligned}$$

où :

- $K(A) = \prod_{(r,(a,r_{-u})) \in A} \mu(\{u\}) > 0$ ne dépend pas de η,
- $C(A) = \sum_{(r,(a,r_{-u})) \in A} d_u(b, r_{-u})$, et $d_u(a, r_{-u}) = \max_{\alpha \in \mathcal{S}_u} c_u(\alpha, r_{-u}) - c_u(a, r_{-u}) \geq 0$ ne dépend pas non plus de η,
- $f_\eta(A) = \prod_{(r,(a,r_{-u})) \in A} f_\eta(r_{-u})$, et $f_\eta(r_{-u}) = \dfrac{1}{\sum_{b \in \mathcal{S}_u} \exp\big(-\eta^{-1} d_u(b, r_{-u})\big)} > 0$, tend vers une constante positive quand η tend vers zéro.

Finalement, quand $\eta \to 0$, seuls les s-arbres qui minimisent $C(A)$ interviennent à la limite. Soit A^*_s un tel arbre pour le sommet s. Il vient alors que l'état s est stochastiquement stable si $C(A^*_s) \leq C(A^*_t)$ pour tout $t \in \mathcal{S}$. Plus explicitement, s est stochastiquement stable si il réalise le minimum de :

$$\min_{s \in \mathcal{S}} \min_{A \in \mathcal{A}(s)} \sum_{(r,(a,r_{-u})) \in A} \max_{\alpha \in \mathcal{S}_u} c_u(\alpha, r_{-u}) - c_u(a, r_{-u}).$$

Dans le cas où le jeu possède une fonction de potentiel P, $\max_{\alpha \in \mathcal{S}_u} c_u(\alpha, r_{-u}) - c_u(a, r_{-u}) = \max_{\alpha \in \mathcal{S}_u} P(\alpha, r_{-u}) - P(a, r_{-u})$. Alors un sommet est stochastiquement stable s'il minimise :

$$\min_{s \in \mathcal{S}} \min_{A \in \mathcal{A}(s)} \sum_{(r,(a,r_{-u})) \in A} \max_{\alpha \in \mathcal{S}_u} P(\alpha, r_{-u}) - P(a, r_{-u}).$$

Cela implique que c'est un maximum global du potentiel. En effet, supposons que le sommet t ne soit pas un maximum global, et soit A_t un t-arbre tel que $C(A_t) = \min_{A \in \mathcal{A}(t)} C(A)$. Soit s un maximum global du potentiel. On construit le s-arbre A_s à partir de A_t en inversant le sens des arcs du chemin qui lie s à t. Il s'agit bien d'un s-arbre et on peut vérifier qu'il satisfait $C(A_s) < C(A_t)$. Par conséquent t n'est pas stochastiquement stable. Donc, seuls les sommets qui maximisent le potentiel sont stochastiquement stables. On retrouve ainsi le résultat du théorème 4.3 sans utiliser la réversibilité de la chaîne de Markov sous-jacente.

Remarquons que la stabilité des états fait intervenir la valeur $\max_{\alpha \in \mathcal{S}_u} c_u(\alpha, r_{-u}) - c_u(a, r_{-u})$ de chaque arc faisable pour le processus de révision. Cette grandeur positive représente le "coût" pour le joueur u de ne pas jouer une meilleure réponse.

Dans le cas d'un processus de révision asynchrone, l'ensemble des révisions faisables pour passer de l'état s à l'état (a, s_{-u}) est unique : il s'agit de l'ensemble $\{u\}$. Pour des processus de révision généraux, il n'y a pas unicité, et on utilise alors le plus petit ensemble de révision donné par la fonction de révision (définition 4.5). Cela nous permet de définir le coût d'un graphe :

Définition 4.8 (*Coût d'un graphe*)

Soit $G = (\mathcal{S}, \mathcal{E})$ un graphe faisable pour un processus de révision μ. Le coût du graphe est le nombre positif suivant :

$$\mathrm{Cout}(G) = \sum_{(a,b) \in \mathcal{E}} \sum_{u \in R(a,b)} \max_{\alpha \in \mathcal{S}_u} c_u(\alpha, a_{-u}) - c_u(b_u, a_{-u}), \tag{4.11}$$

où $R : \mathcal{E} \to \mathcal{P}(\mathcal{U})$ est la fonction de révision du graphe associée à μ.

En particulier, le coût d'une arête (a, b) faisable pour un processus de révision vaut $\sum_{u \in R(a,b)} \max_{\alpha \in \mathcal{S}_u} c_u(\alpha, a_{-u}) - c_u(b_u, a_{-u})$. Ce coût ne dépend pas de l'état d'arrivée b, mais de l'état (b_u, a_{-u}) qui est la valeur observée par le joueur u avant sa décision. On constate alors qu'une arête a un coût nul si et seulement si elle correspond à une meilleure réponse unilatérale des joueurs qui dévient.

Le calcul que l'on a fait dans le cas d'un processus de révision asynchrone se généralise directement, ce qui donne le théorème suivant :

Théorème 4.9 (*Théorème 1 dans [AFN10]*)

Étant donné un processus de révision, l'ensemble des états stochastiquement stables pour l'algorithme stochastique de meilleure réponse 4 est l'ensemble des états qui réalisent le minimum de :

$$\min_{s \in \mathcal{S}} \min_{A \in \mathcal{A}(s)} \mathrm{Cout}(A). \tag{4.12}$$

Il est intéressant de noter que la stabilité stochastique ne dépend que du coût des arbres couvrants. L'ensemble des arbres couvrants est lui-même défini par le processus de révision. Cependant, cet ensemble dépend *uniquement* du support du processus de révision, c'est-

à-dire des ensembles de joueurs qui ont une probabilité positive de réviser leur stratégie.

Ce dernier résultat est du même ordre que celui du modèle d'apprentissage avec erreur dans [KMR93, You93] : dans ce modèle, les joueurs ont une probabilité fixe de commettre une erreur, c'est-à-dire de ne pas jouer une meilleure réponse, indépendemment des gains de chaque action. On montre alors qu'un état est stable si un arbre dont il est la racine minimise le nombre d'erreurs. Contrairement à notre modèle, le nombre d'erreurs est une quantité discrète.

Calcul des états stochastiquement stables dans un jeu de potentiel : Considérons le jeu de potentiel à deux joueurs suivant :

Gains

$(1,1)$	$(0,0)$
$(0,0)$	$(1,1)$

Potentiel

1	0
0	1

Espace d'état

$$\begin{tikzcd} (a,a) \arrow[r] \arrow[d] & (a,b) \\ (b,a) \arrow[r] \arrow[ur, leftrightarrow] & (b,b) \arrow[u] \end{tikzcd}$$

Supposons un processus d'apprentissage indépendant (les joueurs révisent leur stratégies seuls ou à deux). Les flèches indiquent les transitions qui ont un coût positif (qui ne sont pas des meilleures réponses). Celui-ci vaut 2 sur la diagonale (car c'est la somme des coûts de chaque joueur) et 1 ailleurs. On voit alors que $\min_{A\in\mathcal{A}(a,a)} \mathrm{Cout}(A) = \min_{A\in\mathcal{A}(b,b)} \mathrm{Cout}(A) = 1$ et $\min_{A\in\mathcal{A}(a,b)} \mathrm{Cout}(A) = \min_{A\in\mathcal{A}(b,a)} \mathrm{Cout}(A) = 2$, donc (a,a) et (b,b) sont stochastiquement stables. Considérons maintenant un processus de révision instantané, où les deux joueurs modifient leur stratégie à chaque fois. Cette fois, la transition de (a,b) vers (a,a) a un coût qui n'est pas nul : en effet, pour réaliser cette transition, il a fallu que l'un des joueurs ne joue pas une meilleure réponse. Dans ce cas $\min_{A\in\mathcal{A}(a,a)} \mathrm{Cout}(A) = \min_{A\in\mathcal{A}(b,b)} \mathrm{Cout}(A) = \min_{A\in\mathcal{A}(a,b)} \mathrm{Cout}(A) = \min_{A\in\mathcal{A}(b,a)} \mathrm{Cout}(A) = 2$, donc tous les états sont stochastiquement stables, en particulier ceux qui ne sont pas des équilibres de Nash.

4.3.2 Convergence vers les équilibres de Nash

L'exemple précédent montre que, même dans les jeux de potentiel, les états stochastiquement stables ne sont pas nécessairement des équilibres de Nash. Pour garantir cela, des hypothèses supplémentaires sont nécessaires. Ces hypothèses portent sur le graphe de meilleure réponse associé au processus de révision.

Rappelons que le *graphe de meilleure réponse* (défini au chapitre précédent pour un processus de révision asynchrone) est un graphe orienté sur $\mathcal{S}$ tel qu'il existe un arc de s à (a, s_{-u}) si $a \in \mathrm{BR}_u(s_{-u})$. De la même manière, on définit un graphe de meilleure réponse pour n'importe quel processus de révision μ comme un graphe orienté sur $\mathcal{S}$ tel que pour tout ensemble de joueurs $\mathcal{V} \subseteq \mathcal{P}(\mathcal{U})$, il y a un arc entre s et $(a_{\mathcal{V}}, s_{-\mathcal{V}})$ si pour tout $u \in \mathcal{V}$, $a_u \in \mathrm{BR}_u(s_{-u})$ et $\mu(\mathcal{V}) > 0$. Il faut noter qu'il s'agit de meilleures réponses *unilatérales*. En particulier, le fait qu'il existe un arc de s à $(a_{\mathcal{V}}, s_{-\mathcal{V}})$ n'implique pas que chaque joueur

qui dévie ait un gain supérieur dans l'état $(a_{\mathcal{V}}, s_{-\mathcal{V}})$ à celui qu'il a dans l'état s. Même, l'état d'arrivée peut être pire pour tous les joueurs qui dévient.

Comme dans la section 3.4, un sommet (ou état) est récurrent s'il appartient à une classe récurrente du graphe de meilleure réponse, c'est-à-dire une classe d'équivalence pour la relation de connexité forte entre sommets tel qu'il n'y a aucun arc dirigé vers l'extérieur de la classe. En particulier, pour tout sommet qui n'est pas récurrent, il existe un chemin dans le graphe qui le connecte à un sommet récurrent. Alors, les états stochastiquement stables sont des états récurrents :

Théorème 4.10

Si $s \in \mathcal{S}$ est stochastiquement stable pour l'algorithme stochastique de meilleure réponse 4, avec le processus de révision des stratégies μ, alors s est un sommet récurrent du graphe de meilleure réponse associé à μ.

Démonstration : Soit s un sommet qui n'est pas récurrent dans le graphe de meilleure réponse, et soit A_s un s-arbre de coût minimal. D'après le théorème 4.9, il suffit de montrer qu'il existe un sommet t et un t-arbre A_t de coût strictement inférieur à celui de A_s.

Comme s n'est pas récurrent, il existe un sommet récurrent r et un chemin $C_{s,r}$ de s à r dans le graphe de meilleure réponse. Le coût du chemin de s à r est donc nul.

Il existe également un chemin de r à s dans A_s. Ce chemin est de coût strictement positif. En effet, si le coût est nul, cela signifie qu'il existe un chemin entre r et s dans le graphe de meilleure réponse, donc que s et r font partie de la même classe d'équivalence, et en particulier que s est récurrent ce qui est une contradiction. Il existe donc un arc $(t, t+)$ sur le chemin dans A_s entre r et s de coût strictement positif.

Considérons maintenant le graphe G constitué de l'union des arcs de A_s et des arcs de $C_{s,r}$ à laquelle on a retiré l'arc $(t, t+)$. Alors :

- par construction $\text{Cout}(G) < \text{Cout}(A_s)$,
- G contient un t-arbre. En effet, les sommets en amont de t dans A_s sont connectés à t, en particulier le sommet r. Ensuite, on remarque que $(t, t+) \notin C_{s,r}$, car $(t, t+)$ ne correspond pas à une meilleure réponse. Par conséquent s est connecté à r, et donc s est connecté à t. Finalement, les sommets qui sont en aval de t dans A_s sont connectés à s, et donc à t.

Notons A_t le t-arbre contenu dans G. On a alors $\text{Cout}(A_t) \leq \text{Cout}(G) < \text{Cout}(A_s)$. ■

Le théorème affirme donc que les états stochastiquement stables de l'algorithme *stochastique* de meilleure réponse sont dans l'ensemble des états vers lesquels l'algorithme *déterministe* de meilleure réponse converge (au sens de la définition 3.22).

Remarquons que la réciproque du théorème n'est pas vraie. Il peut exister des sommets récurrents du graphe de meilleure réponse qui ne sont pas stochastiquement stables : c'est le cas par exemple des équilibres de Nash, dans les jeux de potentiel et sous un processus de révision asynchrone, qui ne maximisent pas le potentiel (d'après le théorème 4.3).

Corollaire 4.11

Si les sommets récurrents du graphe de meilleure réponse sont des équilibres de Nash, alors les états stochastiquement stables de l'algorithme stochastique de meilleure réponse sont des équilibres de Nash.

Il est donc intéressant de pouvoir caractériser les sommets récurrents d'un graphe de meilleure réponse, en particulier sous quelles conditions ces sommets sont des équilibres de Nash. Comme nous l'avons vu au chapitre précédent, lorsque le processus de révision est asynchrone, les *jeux de potentiel de meilleure réponse* vérifient cette propriété. Mais cela n'est en général pas vrai, même pour les jeux de potentiel exacts, dès que le processus de révision n'est plus asynchrone :

Exemple : Le jeu suivant est un jeu de potentiel. Le graphe de meilleure réponse montre que, pour un processus de révision qui autorise les déviations des deux joueurs en même temps, tous les états sont récurrents.

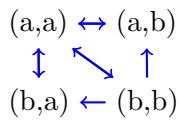
Gains

$(2,2)$	$(1,2)$
$(2,1)$	$(0,0)$

Potentiel

1	1
1	0

Graphe de meilleure réponse



Cet exemple tient au fait que les équilibres de Nash du jeu ne sont pas stricts. En effet, considérons un processus de révision μ tel que pour tout $u \in \mathcal{U}$, $\mu(\{u\}) > 0$. Notons $\mathcal{R}_{async}$ (resp. $\mathcal{R}_\mu$) l'ensemble des états récurrents du graphe de meilleure réponse sous le processus de révision asynchrone (resp. μ). Une condition suffisante pour que $\mathcal{R}_{async} = \mathcal{R}_\mu$ est qu'il n'y ait pas d'arcs sortant de l'ensemble $\mathcal{R}_{async}$ dans le graphe de meilleure réponse associé à μ. Cela est en particulier vérifié si $\mathcal{R}_{async}$ est un ensemble d'équilibres de Nash stricts.

Corollaire 4.12

Soit un processus de révision μ tel que pour tout $u \in \mathcal{U}$, $\mu(\{u\}) > 0$. Si les sommets récurrents du graphe de meilleure réponse associé au processus de révision asynchrone sont des équilibres de Nash stricts, alors les états stochastiquement stables de l'algorithme stochastique de meilleure réponse associé à μ sont des équilibres de Nash stricts.

Les hypothèses du corollaire sont vérifiées pour un jeu de potentiel de meilleure réponse qui n'a que des équilibres stricts.

4.3.3 Contre-exemples sur la sélection des équilibres optimaux dans les jeux de potentiel

On sait maintenant que les états stochastiquement stables sont nécessairement des états récurrents du graphe de meilleure réponse. Dans le cas des jeux de potentiel, peut-on affirmer, comme c'est le cas pour le théorème 4.3, que ces états maximisent le potentiel ?

Nous allons montrer par des contre-exemples que cela n'est pas vrai pour les processus de révision indépendants [8], même si on a un jeu de potentiel exact (premier exemple). Et cela n'est pas vrai non plus pour des jeux de potentiel pondérés, même si le processus de révision est asynchrone (deuxième exemple).

Cela montre que l'algorithme stochastique de meilleure réponse n'est pas robuste au processus de révision. Aucune garantie de performance dans les systèmes distribués ne peut être obtenue dès lors que l'on ne maîtrise pas le processus de révision.

Processus de révision général dans les jeux de potentiel : Considérons le potentiel d'un jeu à 3 joueurs (le troisième joueur choisissant la matrice) suivant [9] (extrait de [AFN10]) :

Potentiel Matrice 1

10	6	0
6	0	0
0	0	9

Potentiel Matrice 2

0	0	0
0	1	1
0	1	1

Si l'on considère un processus de révision asynchrone, seul l'état dont le potentiel vaut 10 est stochastiquement stable. Si l'on considère un processus indépendant, on peut montrer que le coût minimal d'un arbre de racine l'état de potentiel 10 est 9 alors que celui de l'état de potentiel 9 vaut 8. Par conséquent, seul le deuxième état est stochastiquement stable. Le chemin pour aller de l'état 10 à l'état 9 qui donne le coût de l'arbre égal à 8 est, par exemple, la première diagonale de la matrice 1 [10].

Jeu de potentiel pondéré avec processus de révision asynchrone : Considérons le jeu de potentiel pondéré à deux joueurs suivant [11] :

Gains

$(2,2)$	$(0,0)$
$(0,0)$	$(10,1)$

Potentiel

2	-6
0	4

Espace d'état

(a,a)	(a,b)
(b,a)	(b,b)

8. C'est-à-dire où chaque joueur choisit de réviser sa stratégie à chaque itération avec une probabilité indépendante des autres. Il s'agit du processus de révision le plus naturel pour modéliser un système distribué sans contrôleur centralisé.

9. Un jeu qui possède ce potentiel est, par exemple, le jeu où les gains des joueurs sont identiques, et donnés par la fonction de potentiel.

10. Sur ce petit exemple, il est facile de trouver pour chaque sommet s, le s-arbre de coût minimal. Dès que la taille du jeu augmente, le nombre d'arbres couvrants explose. On peut alors utiliser le critère proposé à la proposition 3 de [AFN10]. Intuitivement, ce résultat dit qu'un sommet est asymptotiquement stable si le coût pour sortir de son bassin d'attraction est plus faible que le coût d'y entrer. Ce critère a l'avantage d'être local et a donc une complexité inférieure à celui de la recherche exhaustive de tous les arbres couvrants.

11. On peut vérifier que ce n'est pas un jeu de potentiel *exact* car la somme des différences de gain sur les chemins $(b,a) \to (b,b)$ et $(b,a) \to (a,a) \to (a,b) \to (b,b)$ devraient alors être égales.

Ce jeu comporte les deux équilibres de Nash (a, a) et (b, b), qui ont pour potentiel respectivement la valeur 2 et 4. Le calcul de la distribution stationnaire en fonction de η pour le processus de révision asynchrone montre que seul l'équilibre (a, a) est stochastiquement stable alors qu'il ne maximise pas le potentiel. La distribution est tracée à la figure 4.7. Il est intéressant de constater la non monotonie de la probabilité de choisir le maximum global du potentiel.

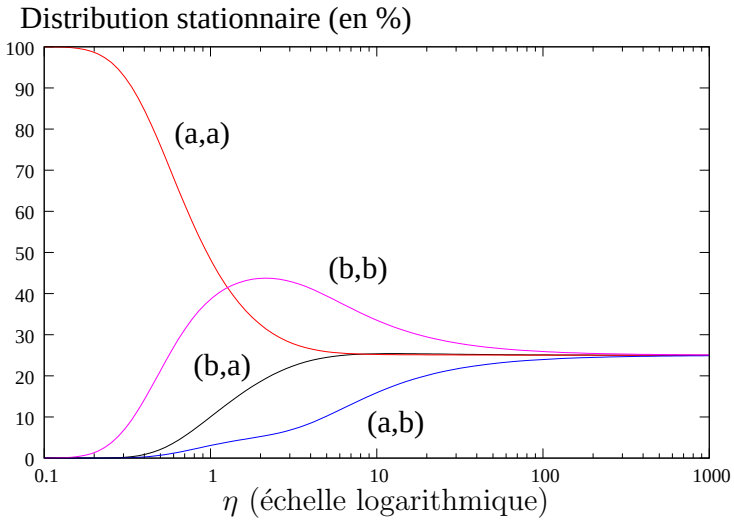


Figure 4.7 – Distribution stationnaire de l'algorithme 4 sous le processus de révision asynchrone dans le jeu de potentiel pondéré en fonction du paramètre η.

CHAPITRE 5

EXTENSION MIXTE DU MODÈLE DE MEILLEURE RÉPONSE

Résumé du chapitre

Ce chapitre constitue la contribution principale de la thèse.

Dans les jeux de potentiel finis, le processus d'apprentissage par meilleures réponses converge vers un ensemble d'équilibres de Nash. Néanmoins, ce résultat est remis en cause dès lors que les gains du jeu sont soumis à des incertitudes aléatoires, ou que le processus de révision des stratégies n'est pas asynchrone. Dans ce chapitre, nous proposons une extension de l'algorithme de meilleure réponse aux stratégies mixtes. Intuitivement, les dynamiques qui évoluent de façon continue sont robustes aux incertitudes et au processus de révision, tous les phénomènes aléatoires tendant à se compenser après un temps suffisamment long.

Une dynamique de meilleure réponse dans l'extension mixte est une équation différentielle telle qu'en tout point, chaque joueur suit une direction de plus grande pente de sa fonction de gain. Il s'agit bien de l'analogie directe du cas discret, à la différence près que la direction de plus grande pente dépend, dans le cas continu, de la métrique employée.

Dans la première section du chapitre, nous donnons des conditions suffisantes sur ces métriques pour que les solutions de l'équation différentielle existent. Nous analysons ensuite les principales propriétés de ces dynamiques, en particulier dans les jeux de potentiel.

Le problème de l'implémentation des dynamiques de meilleure réponse est abordé dans la deuxième section. Notre solution repose sur les approximations stochastiques. Nous étudions ensuite les propriétés de convergence de cette implémentation.

Enfin, dans la dernière section, nous proposons un algorithme distribué, qui repose sur l'implémentation des dynamiques de meilleure réponse, pour résoudre le problème de l'association optimale de mobiles à des points d'accès sans fil. Par des simulations, nous comparons plusieurs heuristiques pour le choix des pas de l'algorithme, et nous montrons le gain de performance que l'on obtient par rapport à des protocoles actuellement en usage.

5.1 Dynamique de meilleure réponse dans l'extension mixte des jeux finis

Dans ce chapitre, nous nous plaçons dans l'extension mixte d'un jeu fini. La dynamique de meilleure réponse est un système dynamique continu qui modélise le comportement de joueurs qui suivent à chaque instant la direction de plus grande pente de leur fonction de gain. Pour assurer l'existence de solutions à ce système dynamique (dont les trajectoires sont contraintes à rester dans un espace compact) nous introduisons des métriques particulières. Cela fait, nous donnons les propriétés des dynamiques de meilleure réponse correspondant à ces métriques. Comme dans les chapitres précédents, l'existence d'une fonction de potentiel permet d'obtenir des résultats de convergence forts.

Un cas particulier de dynamique de meilleure réponse est la célèbre *dynamique de réplication*. Il s'avère que de nombreux résultats qui s'appliquent à cette dynamique s'étendent à toutes les dynamiques de meilleure réponse.

5.1.1 Construction d'une métrique qui garantit l'existence de solutions

Rappelons que la direction de plus grande pente d'une fonction différentiable est relative à un produit scalaire. La donnée d'un produit scalaire définit une métrique et donc un choix particulier de direction de plus grande pente. La dynamique qui suit les plus grandes pentes dépend donc du produit scalaire.

Dans certains cas, les trajectoires sont contraintes à rester dans un ensemble compact donné (par exemple l'ensemble des stratégies mixtes). Toutes les dynamiques qui suivent les plus grandes pentes ne satisfont pas cette contrainte. Dans cette section, nous donnons des conditions suffisantes sur les métriques qui permettent de garantir l'existence de solutions.

Cette section est assez technique, et principalement inspirée de [ABB05]. Elle se place dans un cadre plus général que celui des jeux.

Dynamique définie par les directions de plus grande pente

Soit un domaine $\mathcal{X} = \overline{\mathcal{O}} \cap \mathcal{H}$ où :
- $\mathcal{O}$ est un ouvert convexe et non vide de $\mathbb{R}^n$ et $\overline{\mathcal{O}}$ est son adhérence,
- $\mathcal{H}$ est un sous-espace affine de $\mathbb{R}^n$. Celui-ci peut s'écrire $\{x \in \mathbb{R}^n | Ax = b\}$, où A est une matrice $m \times n$, avec $m \leq n$, de rang plein, et b un vecteur de taille m.

Par exemple, l'ensemble des stratégies mixtes d'un joueur d'un jeu fini peut s'écrire de cette manière, si l'on pose $\mathcal{O} = \{x \in \mathbb{R}^n | x > 0\}$, A la matrice $1 \times n$ qui vaut 1 partout, et $b = 1$.

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction de classe $\mathcal{C}^1$. La direction de plus grande pente est donnée par le gradient de f. Dans le cas du produit scalaire canonique, le gradient est égal au vecteur des dérivées partielles, ce qui donne un moyen pratique de le calculer.

L'objectif est de définir un système dynamique qui :

– suit les directions de plus grande pente de f pour une métrique donnée,
– et tel que les trajectoires sont incluses dans $\mathcal{X}$. Cela est crucial si l'on prend $\mathcal{X}$ comme étant un espace de probabilité : cela n'aurait aucun sens d'avoir, par exemple, des composantes négatives.

Tout d'abord, il est possible que la direction de plus grande pente ne soit pas incluse dans $\mathcal{H}$ [1], et que, par conséquent, une dynamique qui suit cette direction sorte de $\mathcal{H}$. C'est le cas si l'on prend, par exemple, $\mathcal{O} = \mathbb{R}^2$, $\mathcal{H} = \{(x, y) \in \mathbb{R}^2 | y = 0\}$, et $f(x, y) = y$, alors $\nabla f(x, y) = (0, 1)$ qui est orthogonal à $\mathcal{H}$. Pour cette raison, on définit, comme à la section 3.2.2, le gradient de la fonction f restreinte au sous-espace affine $\mathcal{H}$ par :

$$\nabla f_{|\mathcal{H}}(x) = \mathrm{Proj}_{\mathcal{H}_0}(\nabla f(x)), \tag{5.1}$$

où $\mathrm{Proj}_{\mathcal{H}_0} : \mathbb{R}^n \to \mathcal{H}_0$ est la projection orthogonale, pour le produit scalaire courant, sur l'espace vectoriel $\mathcal{H}_0 \stackrel{\text{def}}{=} \mathrm{Ker}(A) = \{x | Ax = 0\}$ [2].

On s'intéresse naturellement au système dynamique donné par l'équation différentielle :

$$\dot{x}(t) = \nabla f_{|\mathcal{H}}(x(t)),$$

avec $x(0) \in \mathcal{X}$. En raison de la projection sur $\mathcal{H}_0$, les trajectoires restent incluses dans $\mathcal{H}$. Cependant, rien ne les interdit de sortir de l'ensemble $\mathcal{O}$, et donc de $\mathcal{X}$.

Cela nous amène à considérer des produits scalaires qui dépendent continûment de x. Soit $G : \mathcal{X} \to \mathbb{S}^n_{++}$ une fonction continue, où $\mathbb{S}^n_{++}$ désigne l'ensemble des matrices définies positives. On définit le produit scalaire $\langle \cdot, \cdot \rangle^G_x$ par [3] :

$$\forall u, v \in \mathcal{R}^n, \langle u, v \rangle^G_x \stackrel{\text{def}}{=} \langle G(x)u, v \rangle,$$

où $\langle \cdot, \cdot \rangle$ est le produit scalaire canonique.

Calculons le gradient, que l'on notera $\nabla_G f(x)$, associé à ce produit scalaire particulier. Par définition du gradient, on a, pour tout $u \in \mathcal{R}^n$: $\langle \nabla f(x), u \rangle = \langle \nabla_G f(x), u \rangle^G_x$. Cela implique que pour tout $u \in \mathcal{R}^n$, $\langle \nabla f(x), u \rangle = \langle G(x) \nabla_G f(x), u \rangle$, et donc que $\nabla f(x) = G(x) \nabla_G f(x)$. Comme $G(x)$ est définie positive, c'est une matrice inversible, d'où :

$$\nabla_G f(x) = G(x)^{-1} \nabla f(x).$$

Finalement, si l'on restreint f au sous-espace affine $\mathcal{H}$, on considère l'équation différentielle ordinaire indépendante du temps :

$$\begin{cases} \dot{x} = \nabla_G f_{|\mathcal{H}}(x), \\ x(0) \in \mathcal{X}. \end{cases} \tag{5.2}$$

1. Ou plus précisément dans le plan tangent à $\mathcal{H}$.
2. On projette en fait sur l'espace tangent à $\mathcal{H}$ qui s'avère être $\mathcal{H}_0$.
3. Les produits scalaires obtenus de cette façon définissent une métrique riemannienne sur $\mathcal{X}$. La géométrie riemannienne est en dehors du cadre de cette thèse et des connaissances de l'auteur. Dans notre cadre très simple, une métrique sera un "produit scalaire qui dépend de x", ce qui peut toujours se définir sans avoir recours à la théorie générale. L'usage de ces métriques est d'utilisation courante en optimisation convexe pour les méthodes de points intérieurs (voir [BGLS97]). Notons également que la classique méthode de Newton pour l'optimisation peut s'interpréter comme une descente de gradient pour une métrique particulière (voir [Iou07]).

Notons que, comme à l'équation (5.1), $\nabla_G f_{|\mathcal{H}}(x) = \mathrm{Proj}_{\mathcal{H}_0}(\nabla_G f(x))$, où la projection est relative au produit scalaire $\langle .\,,.\rangle_x^G$.

Nous donnons maintenant des conditions suffisantes sur la fonction G pour que les trajectoires du système dynamique (5.2) restent dans $\mathcal{X}$.

Existence des solutions

On dira qu'il existe une solution au système dynamique s'il existe une trajectoire $(x(t))$ satisfaisant le système (5.2) telle que $\forall t \in [0, \infty), x(t) \in \mathcal{X}$.

Le théorème suivant donne des conditions sur la fonction $G : \mathcal{R}^n \to \mathbb{S}^n_{++}$ définissant le produit scalaire, pour que les solutions du système dynamique existent. Une des conditions est que G soit égale à la matrice hessienne $\nabla^2 g(x)$ d'une fonction g qui est de type Legendre.

Définition 5.1 (*Fonctions de type Legendre (chapitre 26 dans [Roc97])*)

Une fonction $g : \mathcal{O} \to \mathbb{R}$ est de type Legendre si elle satisfait les conditions suivantes :
- g est différentiable,
- $\|\nabla g(x)\|$ tend vers l'infini quand x tend vers le bord de $\mathcal{O}$,
- g est strictement convexe (si g est de classe $\mathcal{C}^2$, alors $\nabla^2 g(x)$ est définie positive).

Une fonction qui vérifie la deuxième condition peut se voir comme une "barrière" que la dynamique ne peut pas franchir. Cela nous permet d'introduire le théorème fondamental de cette section :

Théorème 5.2 (*Théorème 4.1 dans [ABB05]*)

Supposons que la fonction $g : \mathcal{O} \to \mathbb{R}$ est de classe $\mathcal{C}^2$, et de type Legendre. Supposons également que l'ensemble $\mathcal{X}$ est borné, donc compact. Alors il existe une solution au système dynamique (5.2), où le produit scalaire $\langle .\,,.\rangle_x^G$ est donné par la fonction $G : \mathcal{X} \to \mathbb{S}^n_{++}$ telle que $G(x) = \nabla^2 g(x)$.

Calculons la dérivée de f le long d'une trajectoire du système dynamique (5.2) :

$$\begin{aligned}
\frac{d}{dt} f(x(t)) &= \langle \nabla_G f(x(t)), \dot{x}(t)\rangle_x^G \\
&= \langle \nabla_G f(x(t)), \nabla_G f_{|\mathcal{H}}(x(t))\rangle_x^G \\
&= \langle \nabla_G f_{|\mathcal{H}}(x(t)), \nabla_G f_{|\mathcal{H}}(x(t))\rangle_x^G \\
&= \left(\|\nabla_G f_{|\mathcal{H}}(x(t))\|_x^G \right)^2,
\end{aligned}$$

où la troisième égalité vient du fait que $\nabla_G f_{|\mathcal{H}}(x(t)) \in \mathcal{H}$, donc que le produit scalaire est inchangé si on projette $\nabla_G f(x(t))$ sur $\mathcal{H}$. On appelle *point stationnaires* du système dynamique (5.2), les points tels que $\nabla_G f_{|\mathcal{H}}(x) = 0$. Comme conséquence des calculs précédents, on a :

Proposition 5.3

En dehors des points stationnaires, la fonction f est strictement croissante le long des trajectoires du système dynamique (5.2).

En particulier, si l'on suppose que $\mathcal{X}$ est compact, alors f admet une borne supérieure, et sa valeur converge sur toute trajectoire de la dynamique (5.2).

Calcul de l'expression du système dynamique lorsque le domaine est décrit par des contraintes explicites

Nous donnons une manière de construire une fonction G qui vérifie les hypothèses du théorème fondamental d'existence de solutions dans le cas où l'ensemble $\mathcal{O}$ est défini par des inégalités.

Supposons que l'ensemble ouvert et convexe $\mathcal{O}$ soit décrit par un ensemble de I contraintes $\mathcal{O} = \{x \in \mathbb{R}^n | \forall i = 1 \dots I, c_i(x) > 0\}$, où les fonctions $c_i : \mathbb{R}^n \to \mathbb{R}$ sont des fonctions affines. Soit $h : (0, \infty) \to \mathbb{R}$ une fonction qui vérifie les hypothèses suivantes :

Hypothèse 5.4

- h est de classe $\mathcal{C}^2$,
- $\lim_{s \to 0^+} h'(s) = -\infty$, et
- pour tout $s \in (0, \infty)$, $h''(s) > 0$.

Par exemple, ces hypothèses sont satisfaites par les fonctions $-\log(s)$, $\frac{1}{s}$, $s\log(s) - s$, et $-\frac{1}{\alpha}s^\alpha$ pour $\alpha \in (0, 1)$.

La proposition 4.10 dans [ABB05] implique que la fonction $g : \mathcal{O} \to \mathbb{R}$ définie par :

$$g(x) = \sum_{i=1}^{I} h(c_i(x)), \tag{5.3}$$

satisfait les hypothèses du théorème 5.2. Par le calcul, on obtient :

$$\nabla^2 g(x) = \sum_{i=1}^{I} \Big(h''(c_i(x)) \nabla c_i(x) \nabla c_i(x)^T + h'(c_i(x)) \nabla^2 c_i(x) \Big). \tag{5.4}$$

Posons $G(x) = \nabla^2 g(x)$. Si le sous-espace affine est donné par $\mathcal{H} = \{x | Ax = b\}$, alors, on a plus explicitement [4] :

$$\nabla_G f_{|\mathcal{H}}(x) = G(x)^{-1} \Big(I - A^T (AG(x)^{-1} A^T)^{-1} AG(x)^{-1} \Big) \nabla f(x). \tag{5.5}$$

5.1.2 Dynamique de meilleure réponse

On se place maintenant dans le cadre de l'extension mixte des jeux finis. Supposons que chaque joueur a une fonction de gain et peut, individuellement, suivre la direction de

4. Remarquons que l'expression se simplifie si A est inversible. Cependant, si A est inversible, le domaine est réduit à un point.

plus grande pente de son gain comme dans (5.2). Le produit des trajectoires de tous les joueurs constitue une trajectoire de meilleure réponse.

Une telle trajectoire modélise un processus d'apprentissage naturel, à savoir, comme dans l'algorithme de meilleure réponse dans le cas discret, des joueurs qui suivent la direction de plus grande pente de leur fonction de gain. Contrairement au cas discret, la plus grande pente n'est pas définie de manière unique : elle dépend de la métrique considérée. Quelle est la métrique qui modélise le mieux le comportement d'un joueur ? Cette question semble absurde du fait que les joueurs n'ont aucune conscience d'une quelconque géométrie du jeu. La justification d'un tel modèle d'apprentissage apparaît de façon naturelle dans l'approche évolutionnaire des jeux, dans laquelle les stratégies mixtes représentent des proportions d'individus d'une population infinie (ici de plusieurs populations infinies) qui choisissent une action donnée.

Néanmoins, ce problème de modélisation peut être mis de côté dès lors que l'on suppose que le mécanisme d'apprentissage est intégré dans le comportement des joueurs, par exemple si les joueurs sont des machines qui sont codées pour agir de cette manière.

Formulation de la dynamique de meilleure réponse

Considérons l'extension mixte $(\mathcal{U}, \Delta(\mathcal{S}), f)$ d'un jeu fini. Rappelons que l'on note n_u le cardinal de l'ensemble des actions du joueur u. Pour reprendre les notations de la section précédente, on pose $\mathcal{O}_u = \{x_u \in \mathbb{R}^{n_u} | x_u > 0\}$, et $\mathcal{H}_u = \{x_u \in \mathbb{R}^{n_u} | \sum_{a \in \mathcal{S}_u} x_{u,a} = 1\}$. Pour tout $u \in \mathcal{U}$, on a $\Delta(\mathcal{S}_u) = \overline{\mathcal{O}_u} \cap \mathcal{H}_u$. La dynamique de meilleure réponse est définie de la façon suivante :

Définition 5.5 (*Dynamique de meilleure réponse*)

Soit $(\mathcal{U}, \Delta(\mathcal{S}), f)$ l'extension mixte d'un jeu fini. Pour tout $u \in \mathcal{U}$, une fonction $g_u : \mathcal{O}_u \to \mathbb{R}$ est donnée qui vérifie les hypothèses du théorème 5.2, et on pose $G_u = \nabla^2 g_u$. La dynamique de meilleure réponse est la solution du système d'équations différentielles :

$$\forall u \in \mathcal{U}, \dot{x}_u = \nabla_{G_u} f_{|\mathcal{H}_u}(x), \tag{5.6}$$

et $x(0) \in \mathrm{int}(\Delta(\mathcal{S}))$.

Il s'agit d'un système d'équation différentielles qui ne rentre pas de façon immédiate dans le cadre de la section précédente. Contrairement à la proposition 5.3, les fonctions de gain f_u de chaque joueur ne sont pas nécessairement monotones le long des trajectoires de la dynamique et le théorème 5.2 ne s'applique pas en l'état. Il reste tout de même valide comme nous allons le montrer, en adaptant de façon directe la preuve dans [ABB05] pour montrer que les trajectoires ne sortent pas de $\Delta(\mathcal{S}) = \underset{u \in \mathcal{U}}{\times} \Delta(\mathcal{S}_u)$:

Démonstration : On procède par l'absurde. Supposons qu'il existe un joueur u telle que la trajectoire $x_u(t)$ sorte de l'espace de ses stratégies mixtes $\Delta(\mathcal{S}_u)$. Cela implique qu'il existe un temps $T > 0$ fini, et une suite croissante $(t_n)_n$ d'instants plus petits que T telle que $x_u(t_n) \to \bar{x}_u$, où $\bar{x}_u$ appartient au bord de l'espace des stratégies.

Notons que, comme la fonction de gain $f_u : \underset{u\in\mathcal{U}}{\times} \mathbb{R}^{n_u} \to \mathbb{R}$ est affine en chacune des variables, son gradient pour le produit scalaire canonique est borné : $\|\nabla f_{|\mathcal{H}_u}(x)\| \leq M$.

Par intégration on obtient :

$$\begin{aligned}
\nabla g_u(x_u(t_n)) &= \nabla g_u(x_u(0)) + \int_0^{t_n} \frac{d}{dt}\nabla g_u(x_u(s))ds \\
&= \nabla g_u(x_u(0)) + \int_0^{t_n} G_u(x_u(s))\dot{x}_u(s)ds \\
&= \nabla g_u(x_u(0)) + \int_0^{t_n} G_u(x_u(s))G_u(x_u(s))^{-1}\nabla f_{|\mathcal{H}_u}(x(s))ds \\
&= \nabla g_u(x_u(0)) + \int_0^{t_n} \nabla f_{|\mathcal{H}_u}(x(s))ds
\end{aligned}$$

Donc :

$$\begin{aligned}
\|\nabla g_u(x_u(t_n))\| &= \|\nabla g_u(x_u(0)) + \int_0^{t_n} \nabla f_{|\mathcal{H}_u}(x(s))ds\| \\
&\leq \|\nabla g_u(x_u(0))\| + \int_0^{t_n} \|\nabla f_{|\mathcal{H}_u}(x(s))\|ds \\
&\leq \|\nabla g_u(x_u(0))\| + TM.
\end{aligned}$$

Comme on a supposé T fini, et comme le gradient est continu, il s'ensuit que $\limsup_n \|\nabla g_u(x_u(t_n))\| < \infty$, ce qui est une contradiction avec le fait que g_u est de type Legendre. ■

Remarque sur la terminologie "dynamique de meilleure réponse"

Il faut noter que la dénomination "dynamique de meilleure réponse" n'est pas standard, cette appellation faisant habituellement référence à la dynamique définie dans [GM91] par l'inclusion différentielle :

$$\dot{x}_u \in \mathrm{BR}_u(x) - x_u.$$

Dans cette dynamique, les joueurs se dirigent en "ligne droite" vers la meilleure stratégie, direction qui n'est pas forcément celle qui augmente leur gain le plus rapidement (aller dans la direction d'un sommet ne fait pas nécessairement monter le plus rapidement). Néanmoins, il a été montré que, dans le cas particulier de dynamique de meilleure réponse qui est la dynamique de réplication, les deux dynamiques sont fortement corrélées par le fait que la trajectoire moyenne (au sens de Césaro) de la dynamique de réplication est une solution perturbée de l'inclusion différentielle. Ce résultat a été découvert dans [HSV08]. Il n'est pas clair que cela soit vrai pour toutes les dynamiques de meilleure réponse.

Dans le reste du chapitre, nous choisissons néanmoins la terminologie "dynamique de meilleure réponse" pour la dynamique (5.6) par analogie au modèle de meilleure réponse dans les jeux finis : ici, chaque joueur choisit non pas une action, mais une direction qui optimale par rapport à une certaine métrique.

Sous-classe de dynamiques de meilleure réponse

Nous utilisons la formulation explicite donnée par (5.5) pour la dynamique de meilleure réponse.

Ici, le domaine est l'espace des stratégies mixtes, et $\mathcal{O}_u$ est décrit par un ensemble d'inégalités : $\mathcal{O}_u = \{x_u \in \mathbb{R}^{n_u} | \forall a \in \mathcal{S}_u, x_{u,a} > 0\}$. Par conséquent, on peut construire des fonctions g_u vérifiant les hypothèses du théorème 5.2 à partir de l'équation (5.3), c'est-à-dire $g_u(x_u) = \sum_{a \in \mathcal{S}_u} h(x_{u,a})$ avec $h : (0, \infty) \to \mathbb{R}$ une fonction [5] vérifiant les hypothèses 5.4. Dans ce cas, la matrice hessienne de g_u, notée G_u, est tout simplement la matrice diagonale : $G_u(x_u) = \mathrm{Diag}(h''(x_{u,a}))$. Posons $k(x_{u,a}) = h''(x_{u,a})^{-1}$ ($h''(x_{u,a}) > 0$ par hypothèse), alors $G_u(x_u)^{-1} = \mathrm{Diag}(k(x_{u,a}))$. On note $\|k(x_u)\| = \sum_{a \in \mathcal{S}_u} k(x_{u,a})$.

Rappelons que $\nabla f_u(x) = (f_{u,a}(x))_{a \in \mathcal{S}_u}$, où $f_{u,a}(x)$ est le gain espéré par le joueur u quand il choisit l'action a, si bien que, d'après les équation (5.4) et (5.5), la dynamique de meilleure réponse s'écrit, par composantes :

$$\dot{x}_{u,a} = k(x_{u,a}) \left(f_{u,a}(x) - \sum_{b \in \mathcal{S}_u} \frac{k(x_{u,b})}{\|k(x_u)\|} f_{u,b}(x) \right). \tag{5.7}$$

En particulier :
- si $h(s) = -\log(s)$, alors $k(s) = s^2$,
- si $h(s) = \frac{1}{s}$, alors $k(s) = \frac{1}{2} s^3$,
- si $h(s) = s \log(s) - s$, alors [6] $k(s) = s$,
- si $h(s) = -\frac{1}{\alpha} s^\alpha$, avec $\alpha \in (0, 1)$ alors $k(s) = \frac{1}{1-\alpha} s^{2-\alpha}$.

On peut noter que, au vu des hypothèses sur la fonction h, $\lim_{s \to 0} k(s) = 0$.

Analysons la formule explicite de la dynamique. D'abord, il y a le facteur $k(x_{u,a})$: comme $k(x_{u,a}) \to 0$ quand $x_{u,a} \to 0$, cette fonction a un rôle de "barrière" empêchant que $x_{u,a}$ ne devienne négatif. Ensuite, le vecteur $\left(\frac{k(x_{u,b})}{\|k(x_u)\|} \right)_{b \in \mathcal{S}_u}$ définit un vecteur de probabilité sur $\mathcal{S}_u$. Par conséquent, $\sum_{b \in \mathcal{S}_u} \frac{k(x_{u,b})}{\|k(x_u)\|} f_{u,b}(x)$ s'interprète comme le gain espéré pour cette probabilité. Comme $\lim_{s \to 0} k(s) = 0$, le gain donné par l'action b a un impact d'autant plus faible dans cette moyenne que $x_{u,b}$ est petit. Enfin, $\left(f_{u,a}(x) - \sum_{b \in \mathcal{S}_u} \frac{k(x_{u,b})}{\|k(x_u)\|} f_{u,b}(x) \right)$ est la différence entre le gain espéré en jouant l'action a, et la moyenne. Au final, la probabilité de jouer une action croît, à un facteur positif près, comme cette différence.

5. On utilise la même fonction h pour tous les joueurs, mais cela n'est pas nécessaire.
6. Le produit scalaire obtenu par ce choix de fonction est connu sous le nom de produit scalaire de Shahshahani. Ce choix donne la dynamique de réplication (voir par exemple [Hof85]).

Un exemple de la théorie des jeux évolutionnaires : la dynamique de réplication

Prenons $h(s) = s\log(s) - s$. Alors $k(s) = s$, et de plus, $\|k(x_u)\| = \sum_{a\in\mathcal{S}_u} k(x_{u,a}) = \sum_{a\in\mathcal{S}_u} x_{u,a} = 1$. Dans ce cas, la dynamique s'écrit simplement :

$$\dot{x}_{u,a} = x_{u,a}\Big(f_{u,a}(x) - \sum_{b\in\mathcal{S}_u} x_{u,b} f_{u,b}(x)\Big).$$

Cette dynamique s'appelle *dynamique de réplication*. A l'instar du modèle de proies et prédateurs de Lotka-Volterra, la dynamique de réplication a été très étudiée en biologie pour décrire les dynamiques de population [TJ78]. Cette dynamique est au fondement de la théorie des jeux évolutionnaires, une branche de la théorie des jeux, qui étudie l'évolution macroscopique de populations infinies dans lesquelles les individus interagissent à l'échelle microscopique.

Les jeux évolutionnaires en eux-mêmes ne sont pas l'objet de ce chapitre puisque nous supposons un nombre fini de joueurs. Néanmoins, les dynamiques qui y sont étudiées peuvent être utilisées comme modèles d'apprentissage dans les jeux finis. Bien que cet apprentissage soit difficile à justifier dans le cas de joueurs humains (agents économiques par exemple), celui-ci peut être "codé" dans les machines impliquées dans le jeu.

A titre d'illustration de la dynamique de réplication, considérons le jeu suivant :

Gains

$(0,0)$	$(5,2)$
$(1,6)$	$(1,3)$

(5.8)

On identifie l'ensemble des stratégies avec $[0,1]\times[0,1]$. Le profil des trajectoires de la dynamique est tracé à la figure 5.1. On observe que les trajectoires convergent vers les équilibres de Nash, c'est-à-dire vers $(0,0)$, $(1,1)$ et $(0.6,0.2)$. L'équilibre mixte est un équilibre instable, car les trajectoires qui passent dans son voisinage s'éloignent en direction des équilibres purs. Comme nous allons le voir, ceci est une propriété générale des dynamiques de meilleures réponses.

5.1.3 Propriétés de la dynamique de meilleure réponse

Le théorème "folk" de la théorie des jeux évolutionnaires établit des correspondances entre les ensembles stables de la dynamique de réplication [7] et les équilibres de Nash du jeu. Nous allons montrer que ces résultats restent vrais pour toute dynamique de meilleure réponse décrite par l'équation (5.7). En étendant par continuité la fonction k de la dynamique par $k(0) = 0$, on obtient une dynamique définie sur $\Delta(\mathcal{S})$, et pas uniquement sur son intérieur.

7. Ce théorème a été étendu à des classes de dynamiques plus générales [Wei95].

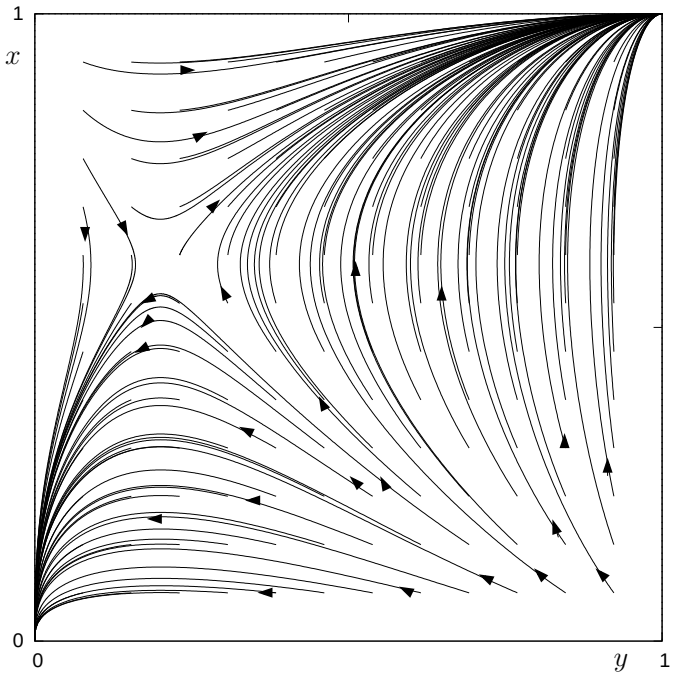


Figure 5.1 – Trajectoires de la dynamique de réplication dans un jeu à deux joueurs. x est la probabilité de choisir la ligne du haut, et y la probabilité de choisir la colonne de droite (pour conserver la même disposition que la matrice du jeu).

Rappels sur les équations différentielles ordinaires

Afin d'étudier les propriétés des dynamiques, quelques rappels de base sur les équations différentielles ordinaires sont nécessaires. Pour plus de détails, on peut se référer au chapitre 9 de [HS74], et également, dans le cas spécifique des dynamiques de population, à [HS98].

En premier lieu, on sait que les solutions du système dynamique existent, mais nous sommes incapable, en général, d'obtenir une formule analytique de la solution. Bien que les solutions peuvent être approchées en utilisant des méthodes classiques d'intégration numérique, nous nous intéressons plutôt à leur comportement asymptotique. De ce fait, nous nous contenterons ici de considérations qualitatives sur les points limites des trajectoires.

Soit $f : \mathbb{R}^n \to \mathbb{R}^n$ et $\dot{x} = f(x)$ une équation différentielle ordinaire ne dépendant pas du temps dont les solutions $x(t), t \geq 0$ existent, sont déterminées de manière unique par la condition initiale $x(0) = x_0$, et restent dans un domaine compact $\mathcal{X}$ de $\mathbb{R}^n$. L'ensemble *ω-limite* d'une solution x est l'ensemble des points d'accumulation de la trajectoire quand $t \to \infty$:

$$\omega(x) \stackrel{\text{def}}{=} \{y \in \mathcal{X} | \exists (t_n)_n \text{ une suite telle que } t_n \to \infty \text{ et } x(t_n) \to y\}.$$

En général, l'ensemble ω-limite est vide. Cependant, lorsque les trajectoires sont bornées,

par exemple lorsqu'elles ne sortent pas de l'ensemble compact $\mathcal{X}$, toute suite $x(t_n)$ admet des points d'accumulation dans $\mathcal{X}$. Alors, on peut vérifier que $\omega(x)$ est un ensemble :

- fermé, donc compact,
- invariant, et
- connexe.

Un ensemble est *invariant* si toute trajectoire passant par un point de cet ensemble est incluse dans l'ensemble. Par exemple, $\mathcal{X}$ est invariant, et les points stationnaires sont des ensembles invariants. Rappelons que x est *stationnaire* pour la dynamique si $f(x) = 0$. Comme nous le verrons, les faces du polytope $\Delta(\mathcal{S})$ sont des ensembles invariants de la dynamique (5.7).

Les ensembles invariants contiennent donc les points de convergence des solutions de la dynamique. Cela donne finalement peu d'information puisque $\mathcal{X}$ est lui-même invariant. Nous allons utiliser un autre critère pour discriminer ces ensembles invariants qui est la *stabilité*. Intuitivement, un ensemble est stable si en partant près de cet ensemble, la trajectoire reste dans un voisinage de l'ensemble. Plus précisément, soit $\mathcal{A}$ un ensemble dans $\mathcal{X}$, $\mathcal{B}$ est un voisinage de $\mathcal{A}$ *dans* $\mathcal{X}$ s'il existe B voisinage de $\mathcal{A}$ dans $\mathbb{R}^n$ tel que $\mathcal{B} = B \cap \mathcal{X}$ [8]. Alors :

Définition 5.6 (*Stabilité*)

Étant donnée une dynamique, et un ensemble $\mathcal{A}$ invariant pour la dynamique, on dira que $\mathcal{A}$ est :

- un *attracteur* s'il existe un voisinage $\mathcal{B}$ tel que toute trajectoire avec $x(0) \in \mathcal{B}$ est telle que $\omega(x) \subseteq \mathcal{A}$.
- *stable au sens de Lyapunov* si, pour tout voisinage $\mathcal{B}$, il existe un voisinage $\mathcal{B}' \subseteq \mathcal{B}$ tel que toutes les trajectoires avec $x(0) \in \mathcal{B}'$ demeurent dans $\mathcal{B}$.
- *asymptotiquement stable* s'il est stable au sens de Lyapunov et que c'est un attracteur.

Notons qu'un attracteur n'est pas nécessairement stable au sens de Lyapunov, et inversement. Enfin, un ensemble est *instable* s'il n'est pas stable au sens de Lyapunov. Cela signifie qu'il existe un voisinage $\mathcal{B}$ tel que pour tout voisinage $\mathcal{B}' \subseteq \mathcal{B}$, il existe une solution ayant une condition initiale dans $\mathcal{B}'$ qui n'est pas entièrement incluse dans $\mathcal{B}$.

Une façon efficace de montrer qu'un ensemble est asymptotiquement stable est de trouver une *fonction de Lyapunov* pour la dynamique, c'est-à-dire une fonction qui est monotone le long des trajectoires. Plus précisément, $L : \mathcal{X} \to \mathbb{R}$ est une fonction de Lyapunov, si pour toute trajectoire $(x(t))$ de l'équation différentielle, $t \mapsto L(x(t))$ est strictement monotone en dehors des points stationnaires de la dynamique. Si une telle fonction existe, alors tout ensemble ω-limite est constitué de points stationnaires. Enfin, si $\bar{x}$ est un maximum local de la fonction de Lyapunov et que pour tout x dans un voisinage de $\bar{x}$, $\dfrac{d}{dt}L(x(t)) > 0$, alors $\bar{x}$ est asymptotiquement stable. Il est important de noter que, bien que les fonctions de Lyapunov donnent beaucoup d'information sur la dynamique, il n'existe pas de méthode générale pour déterminer si une telle fonction existe, et pour la trouver le cas échéant.

8. À partir de maintenant, la notion de voisinage fera implicitement référence au voisinage dans $\mathcal{X}$.

Une autre manière d'étudier la stabilité d'un point stationnaire $\bar{x}$, est d'étudier la linéarisation du système dynamique[9]. Un critère de stabilité est donné par les valeurs propres de la matrice Jacobienne de f en $\bar{x}$:

- s'il existe une valeur propre dont la partie réelle est strictement positive, alors $\bar{x}$ est instable,
- si toutes les valeurs propres ont des parties réelles strictement négatives, alors $\bar{x}$ est asymptotiquement stable.

Quand les valeurs propres ont des parties réelles négatives, et certaines nulles, on ne peut rien affirmer directement. Une étude au second ordre est parfois déterminante.

Enfin, il existe une troisième méthode, qui s'avérera efficace, pour montrer qu'un ensemble invariant à l'intérieur du domaine n'est pas un attracteur. Elle repose sur l'étude du flot global induit par la dynamique : il suffit de montrer que la dynamique préserve le volume[10]. En effet, soit $\mathcal{E} \subseteq \text{int}(\mathcal{X})$ un ensemble invariant. Supposons que la dynamique préserve le volume, c'est-à-dire que si $\mathcal{A} \subseteq \text{int}(\mathcal{X})$ est un ensemble mesurable de volume $\text{vol}(\mathcal{A})$ (on note $\mathcal{A}(t)$ l'ensemble des points de chaque trajectoire au temps t dont la condition initiale est dans $\mathcal{A}$) alors pour tout t, $\text{vol}(\mathcal{A}(t)) = \text{vol}(\mathcal{A})$. Par l'absurde, supposons que $\mathcal{E}$ est un attracteur. Alors il existe un voisinage $\mathcal{A}$ mesurable et de volume strictement supérieur à celui de $\mathcal{E}$ tel que toutes les solutions partant de A convergent vers $\mathcal{E}$. Par le théorème de convergence dominée (dominé par le volume fini du compact $\mathcal{X}$), $\text{vol}(A(t)) \to \text{vol}(\mathcal{E})$ ce qui contredit le fait que le volume est conservé.

Intuitivement, une dynamique qui préserve le volume peut s'apparenter à un fluide incompressible. Si le fluide est contraint à rester dans un volume borné, celui-ci peut, par exemple, être en rotation, ou être immobile. Remarquons que si le fluide est immobile, cela signifie que la dynamique est stationnaire en tous points. Et si tous les points sont stationnaires, aucun n'est attracteur.

Enfin, pour montrer que le volume est préservé, on peut utiliser le théorème de Liouville qui affirme que cela est vrai si la trace de la matrice Jacobienne de la dynamique est nulle, ce qui se traduit ici par : $\forall x \in \mathcal{X}, \sum_n \frac{\partial f_n}{\partial x_n}(x) = 0$.

Ensembles invariants de la dynamique de meilleure réponse

Nous revenons maintenant à la dynamique de meilleure réponse (5.7). Elle est définie sur le domaine $\Delta(\mathcal{S})$ des stratégies mixtes du jeu. Ce domaine est un polytope de dimension $\sum_{u \in \mathcal{U}} n_u$, où n_u est la cardinalité de l'ensemble d'actions du joueur u. Chaque face de ce polytope est déterminée de manière unique par un ensemble d'indices (u, a), avec $u \in \mathcal{U}$ et $a \in \mathcal{S}_u$, noté $\mathcal{I}$, de la façon suivante :

$$F_{\mathcal{I}} = \{x \in \Delta(\mathcal{S}) | x_{u,a} = 0, \forall (u, a) \in \mathcal{I}\}. \tag{5.9}$$

9. Les systèmes dynamiques linéaires sont complètement résolus. Le lien entre la stabilité d'un point stationnaire et la stabilité dans le système linéarisé est donné par le théorème de Hartman–Grobman.

10. Cette propriété a été avancée dans la remarque page 133 de [EA83].

En particulier, $F_\emptyset = \Delta(\mathcal{S})$, et les sommets du polytope qui correspondent aux stratégies pures sont $F_\mathcal{I}$ avec $\mathcal{I}$ qui contient, pour chaque joueur u, toutes les actions dans $\mathcal{S}_u$ sauf une.

Comme nous avons posé $k(0) = 0$ dans la dynamique de meilleure réponse, le résultat suivant est évident :

Proposition 5.7

Les faces du polytope $\Delta(\mathcal{S})$ sont invariantes par la dynamique de meilleure réponse (5.7).

En particulier, les sommets de $\Delta(\mathcal{S})$, c'est-à-dire les stratégies pures du jeu, sont des faces de dimension 0 (*i.e.* des points) et sont nécessairement des points stationnaires. Néanmoins, ce ne sont pas les seuls points stationnaires car, comme nous allons le voir, tous les équilibres de Nash le sont.

Ensembles stables de la dynamique de meilleure réponse

Étant donné une face F de $\Delta(\mathcal{S})$, il existe un ensemble d'indices $\mathcal{I}$ tel que $F_\mathcal{I} = F$, où $F_\mathcal{I}$ est défini par (5.9). On défini l'intérieur d'une face par :

$$\mathrm{int}(F_\mathcal{I}) \stackrel{\text{def}}{=} \{x \in F | 0 < x_{u,a} < 1, \forall (u,a) \notin \mathcal{I}\}. \tag{5.10}$$

De manière équivalente, l'intérieur d'une face est l'ensemble des points de la face qui ne sont pas inclus dans une face plus petite.

Rappelons que x est un équilibre de Nash du jeu $(\mathcal{U}, \Delta(\mathcal{S}), f)$ si pour tout $u \in \mathcal{U}, \forall a \in \mathcal{S}_u, f_{u,a}(x) < \max_{b \in \mathcal{S}_u} f_{u,b}(x) \Rightarrow x_{u,a} = 0$. L'équivalent du "folk" théorème de la théorie des jeux évolutionnaires est le suivant :

Théorème 5.8 (*Extension du théorème 7.2.1 dans [HS98]*)

Considérons la dynamique de meilleure réponse (5.7) associée au jeu $(\mathcal{U}, \Delta(\mathcal{S}), f)$.

1. Si $\bar{x} \in \Delta(\mathcal{S})$ est un équilibre de Nash du jeu, alors c'est un point stationnaire de la dynamique.
2. Si $\bar{x}$ est l'ensemble ω-limite d'une trajectoire $(x(t))$ dans $\mathrm{int}(\Delta(\mathcal{S}))$, alors c'est un équilibre de Nash.
3. Si $\bar{x}$ est stable au sens de Lyapunov, alors c'est un équilibre de Nash.
4. Si $\bar{x}$ est un équilibre de Nash strict, alors il est asymptotiquement stable.

Démonstration : La preuve s'adapte directement de la preuve du théorème 7.2.1 dans [HS98] excepté le dernier point.

1. Si $\bar{x} \in \Delta(\mathcal{S})$ est un équilibre de Nash, alors, pour tout $u \in \mathcal{U}$, il existe un réel K tel que pour tout $a \in \mathcal{S}_u$ tel que $\bar{x}_{u,a} > 0$, $f_{u,a}(\bar{x}) = K$. Donc

$$\dot{x}_{u,a} = f_{u,a}(\bar{x}) - \frac{1}{\|k(\bar{x}_u)\|} \sum_{b \in \mathcal{S}_u} k(\bar{x}_{u,b}) f_{u,b}(\bar{x}) = K - K \frac{\sum_{b \in \mathcal{S}_u} k(\bar{x}_{u,b})}{\|k(\bar{x}_u)\|} = 0.$$

2. Comme n'importe quel sommet qui n'est pas un équilibre de Nash est stationnaire et donc ensemble ω-limite, il est nécessaire de supposer que $x(t) \in \text{int}(\Delta(\mathcal{S}))$.

   Soit $x(t)$ dans $\text{int}(\Delta(\mathcal{S}))$ dont on suppose qu'il converge vers $\bar{x}$ qui n'est pas un équilibre de Nash. Alors il existe $(u, a) \in \mathcal{U} \times \mathcal{S}_u$ tels que $f_{u,a}(\bar{x}) - \sum_{b \in \mathcal{S}_u} \frac{k(\bar{x}_{u,b})}{\|k(\bar{x}_u)\|} f_{u,b}(\bar{x}) > \varepsilon > 0$, et donc tels que $\dot{x}_{u,a}(t) > k(x_{u,a}(t))\varepsilon$ pour t suffisamment grand (par continuité des espérances de gain). Comme k est strictement positive à l'intérieur du domaine, et que $x(t)$ converge vers un point stationnaire, cela implique que $k(\bar{x}_{u,a}) = 0$, ce qui est une contradiction car k ne s'annule qu'en 0.

3. Supposons que $\bar{x}$ soit un point stationnaire qui n'est pas un équilibre de Nash. Il existe donc (u, a) avec $\bar{x}_{u,a} < 1$ tels que $f_{u,a}(\bar{x}) - \sum_{b \in \mathcal{S}_u} \frac{k(\bar{x}_{u,b})}{\|k(\bar{x}_u)\|} f_{u,b}(\bar{x}) > \varepsilon > 0$. Comme $\bar{x}$ est stationnaire, on a nécessairement $\bar{x}_{u,a} = 0$. Il existe donc un voisinage V de $\bar{x}$ suffisamment petit sur lequel $\dot{x}_{u,a} > 0$ (si $x_{u,a} > 0$), et donc, toute trajectoire partant de V sort finalement de V, ce qui montre que $\bar{x}$ n'est pas stable au sens de Lyapunov.

4. Soit $\bar{x}$ un équilibre de Nash strict, qui est donc en stratégies pures. Il existe un voisinage V tel que pour tout $x \in V$, pour chaque joueur u, $\forall b \neq a$, $f_{u,a}(x) > f_{u,b}(x)$, où a est l'action telle que $\bar{x}_{u,a} = 1$. Cela implique que, dans V, $\dot{x}_{u,a} > 0$ et $\dot{x}_{u,b} < 0$. Il suffit alors de remarquer que la fonction qui à x associe sa distance (en norme 1) à $\bar{x}$ est strictement décroissante le long des trajectoires, c'est donc localement une fonction de Lyapunov, d'où le résultat.

■

Le deuxième point de la proposition affirme qu'aucune trajectoire dans l'intérieur du domaine ne converge vers des points stationnaires qui ne sont pas des équilibres de Nash. Le troisième point affirme de plus que ces points sont instables. Finalement, les seuls points potentiellement asymptotiquement stables sont les équilibres de Nash. Ce résultat peut encore être raffiné : comme nous allons le montrer, les équilibres de Nash qui ne sont pas stricts ne sont pas asymptotiquement stables, ou plus précisément, pas des attracteurs.

Théorème 5.9 (*Propriété des équilibres asymptotiquement stables*)

Soit $\bar{x}$ un équilibre de Nash du jeu $(\mathcal{U}, \Delta(\mathcal{S}), f)$. Si $\bar{x}$ est un point asymptotiquement stable pour la dynamique de meilleure réponse (5.7), alors c'est un équilibre de Nash en stratégies pures.

Soit $\mathcal{I}$ un ensemble d'indices et $F_{\mathcal{I}}$ la face associée par (5.9). Dans cette face, pour chaque joueur, il existe au moins une action, que l'on notera o, telle que $x_{u,o} > 0$. Le théorème repose sur le résultat suivant :

Lemme 5.10 (*Préservation du volume*)

Considérons le changement de variable $y_{u,a} = H_{u,a}(x_u)$ défini par (rappelons que h est la fonction qui définit le produit scalaire) :

$$\forall (u,a) \notin \mathcal{I}, \forall x \in \mathrm{int}(F_{\mathcal{I}}), H_{u,a}(x_u) = h'(x_{u,a}) - h'(x_{u,o}).$$

Alors, le changement de variable H est un homéomorphisme de $\mathrm{int}(F_{\mathcal{I}})$ dans $\mathbb{R}^n$, où n est la dimension de la face, et la dynamique obtenue par changement de variable conserve le volume.

Démonstration (Lemme) : Il s'agit de la généralisation du résultat dans [Hof96] qui porte sur la dynamique de réplication.

Comme $h''(s) > 0$ pour tout $s \in (0,\infty)$, $\lim_{s\to 0} h'(s) = -\infty$, alors h' est un homéomorphisme de $(0,\infty)$ dans $(-\infty, M)$, où $M = \lim_{s\to\infty} h'(s)$. Notons ℓ la fonction réciproque de h', qui est strictement croissante.

Montrons que H est un homéomorphisme dans le cas où il n'y a qu'un seul joueur (on omet u), c'est-à-dire que pour tout $y \in \mathbb{R}^n$, il existe un unique x tel que $H(x) = y$. Soit $K \in \mathbb{R}$. Pour tout $a \neq o$, on pose $x_a(K) = \ell(y_a + K)$, et $x_o(K) = \ell(K)$. On a bien $x_a(K) > 0$, et $\sum_a x_a(K)$ est strictement croissant continu et prend ses valeurs entre 0 et $+\infty$. Par continuité, il existe un unique K tel que $x(K) \in \mathrm{int}(F_{\mathcal{I}})$, ce qui donne la fonction réciproque. Sa continuité découle de celle de ℓ.

On montre maintenant que le volume est préservé par la dynamique, après le changement de variable. Rappelons que $k(s) = h''(s)^{-1}$, alors :

$$\begin{aligned}
\dot{y}_{u,a} &= h''(x_{u,a})\dot{x}_{u,a} - h''(x_{u,o})\dot{x}_{u,o} \\
&= \frac{k(x_{u,a})}{k(x_{u,a})}(f_{u,a}(x) - \sum_{b\in\mathcal{S}_u} \frac{k(x_{u,b})}{\|k(x_u)\|} f_{u,b}(x)) - \frac{k(x_{u,o})}{k(x_{u,o})}(f_{u,o}(x) - \sum_{b\in\mathcal{S}_u} \frac{k(x_{u,b})}{\|k(x_u)\|} f_{u,b}(x)) \\
&= f_{u,a}(x) - f_{u,o}(x)
\end{aligned}$$

Comme $f_{u,a}(x)$ et $f_{u,o}(x)$ sont indépendants de x_u, $\frac{\partial f_{u,a}}{\partial y_{u,a}}(H^{-1}(y)) = 0$, et donc $\frac{\dot{y}_{u,a}}{\partial y_{u,a}} = 0$. Finalement, la trace de la matrice jacobienne de la dynamique est nulle ce qui implique, par le théorème de Liouville, que le volume est constant. ■

Démonstration (Théorème 5.9) : D'après le lemme précédent, la dynamique de meilleure réponse après changement de variable préserve le volume. Par conséquent, il n'y a pas d'attracteurs dans $\mathbb{R}^n$, et puisque le changement de variable est un homéomorphisme, cela implique qu'il n'y a pas non plus d'attracteurs dans $\mathrm{int}(F_{\mathcal{I}})$.

Si un équilibre de Nash n'est pas pur, il est inclus dans l'intérieur d'une face. Par conséquent, il ne peut pas être un attracteur. ■

A la figure 5.2, on a tracé la même dynamique (c'est-à-dire pour le même jeu) qu'à la figure 5.1, mais après le changement de variable $y_{u,a} = \ln(\frac{x_{u,a}}{x_{u,o}})$. L'équilibre mixte qui était

$(0.6, 0.2)$ vaut maintenant $(\ln(1.5), -\ln(4))$. Les trajectoires évoluent dans $\mathbb{R}^2$ et le volume est conservé.

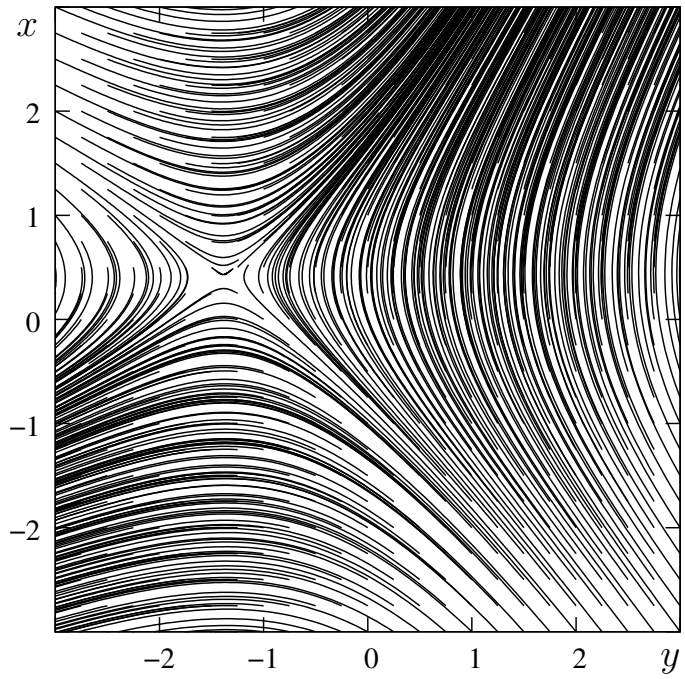


Figure 5.2 – Trajectoires de la dynamique de réplication avec le changement de variable du lemme 5.10 dans le jeu (5.8). La dynamique préserve le volume : on a l'image d'un fluide incompressible qui s'écoule.

Un point asymptotiquement stable est un équilibre de Nash pur du jeu. Si cet équilibre $\bar{x}$ n'est pas strict, alors il existe une action de l'un des joueurs qui lui procure un gain identique et forme un état $\hat{x}$. Si l'on considère la dynamique sur l'arête joignant $\bar{x}$ à $\hat{x}$, tous les points de l'arête sont stationnaires. Par conséquent, aucun point n'est asymptotiquement stable, d'où :

Corollaire 5.11

Si $\bar{x}$ est un point asymptotiquement stable pour la dynamique de meilleure réponse, alors c'est un équilibre de Nash strict.

Bien que les équilibres qui ne sont pas des stratégies pures ne soient pas des attracteurs, cela ne signifie pas qu'aucune trajectoire, en dehors de la trajectoire stationnaire, ne converge vers ces points (voir l'équilibre de Nash en stratégies mixtes à la figure 5.1). Comme nous le verrons, par l'ajout de bruit aléatoire dans la discrétisation de la dynamique, ces équilibres (sous certaines hypothèses) ont cependant une probabilité nulle d'être choisis asymptotiquement.

Notons enfin que les équilibres dans l'intérieur du domaine peuvent cependant être stables au sens de Lyapunov, comme dans le jeu suivant, dans lequel il n'y pas d'équilibre

de Nash en stratégies pures (voir le profil des trajectoires à la figure 5.3). L'unique équilibre de Nash $(2/3, 1/2)$ est stable au sens de Lyapunov puisque les trajectoires forment des cycles de plus en plus petits autour de l'équilibre.

Gains

$(2,0)$	$(1,1)$
$(0,2)$	$(3,0)$

(5.11)

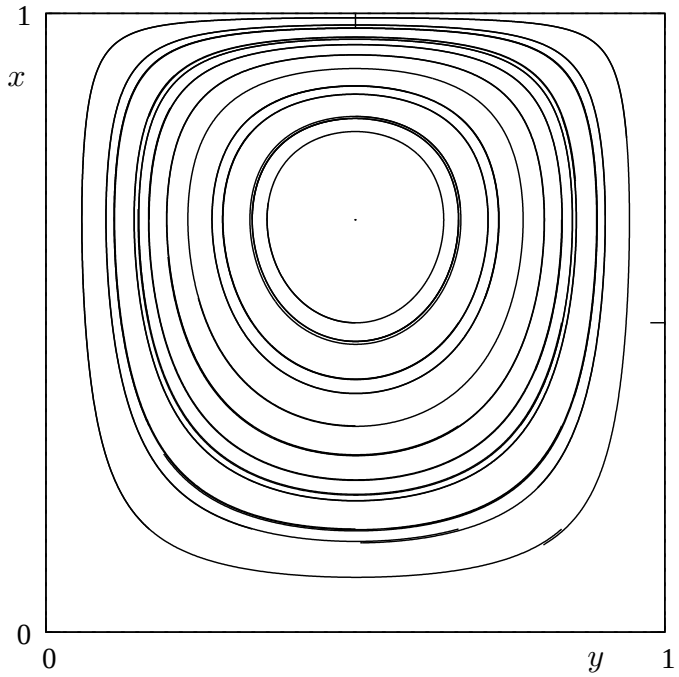


Figure 5.3 – Trajectoires de la dynamique de réplication dans le jeu (5.11) sans équilibre de Nash en stratégie pure : les trajectoires sont des orbites.

5.1.4 Convergence dans les jeux de potentiel

Rappelons que l'extension mixte d'un jeu fini est un jeu de potentiel s'il existe une fonction $F : \underset{u \in \mathcal{U}}{\times} \mathbb{R}^{n_u} \to \mathbb{R}$ de classe $\mathcal{C}^1$ telle que :

$$\forall x \in \Delta(\mathcal{S}), \forall u \in \mathcal{U}, \nabla F_{|\mathcal{H}_u}(x) = \nabla f_{|\mathcal{H}_u}(x).$$

Dans ce cas, la dynamique de meilleure réponse (5.7) s'exprime en fonction de la fonction de potentiel par :

$$\forall u \in \mathcal{U}, \dot{x}_u = \nabla_{G_u} F_{|\mathcal{H}_u}(x).$$

Si l'on note $G(x)$ la matrice diagonale par bloc, où les blocs sont les matrices $G_u(x)$, cela se réécrit de façon compacte :

$$\dot{x} = \nabla_G F_{|\mathcal{H}}(x). \tag{5.12}$$

Il est immédiat de voir que la fonction G définie de cette manière vérifie les conditions du théorème 5.2 où l'on prend $\mathcal{O} = \underset{u\in\mathcal{U}}{\times} \mathcal{O}_u$ et $\mathcal{H} = \underset{u\in\mathcal{U}}{\times} \mathcal{H}_u$. En particulier, par la proposition 5.3, la fonction de potentiel F est strictement croissante le long des trajectoires, en dehors des points stationnaires. Comme F est bornée sur $\Delta(\mathcal{S})$, sa valeur converge pour toute trajectoire de la dynamique de meilleure réponse (à l'image du potentiel dans l'algorithme de meilleure réponse des jeux finis). De plus, F est une fonction de Lyapunov et donc, si un point appartient à un ensemble ω-limite d'une trajectoire, alors c'est un point stationnaire (cela n'est pas vrai en général, par exemple dans le jeu (5.11) dans lequel chaque point appartient à un cycle ω-limite).

Un ensemble $\mathcal{E} \subseteq \Delta(\mathcal{S})$ est appelé un *strict maximum local* de F si :

- $\mathcal{E}$ est connexe,
- tout point de $\mathcal{E}$ est un maximum local pour F (pas nécessairement strict),
- il existe un voisinage $\mathcal{V}$ de $\mathcal{E}$ [11] tel que $F(x) > F(y)$ pour tout $y \in \mathcal{V}\backslash\mathcal{E}$ et tout $x \in \mathcal{E}$.

Il est clair que F est constante sur $\mathcal{E}$, et comme F est continue, $\mathcal{E}$ est compact. Ici, la fonction de potentiel F est multi-affine, ce qui implique que $\mathcal{E}$ est une face de $\Delta(\mathcal{S})$. En effet, si F est une fonction de $(x_1 \dots x_n)$, et si l'on fixe toutes les variables sauf une, disons x_1, la fonction $x_1 \in [0,1] \mapsto F(x_1 \dots x_n)$ est affine. Par conséquent, elle est maximale soit pour $x_1 = 0$, soit pour $x_1 = 1$, soit pour tout x_1 dans $[0,1]$; en résumé elle n'est jamais maximale *uniquement* pour une valeur strictement comprise entre 0 et 1. Comme F est une fonction de Lyapunov sur le voisinage de la face $\mathcal{E}$, on a directement :

Proposition 5.12

Les ensembles asymptotiquement stables de la dynamique de meilleure réponse dans un jeu de potentiel sont $\Delta(\mathcal{S})$ et ses faces qui sont des stricts maximums locaux de la fonction de potentiel.

On retrouve ainsi (mais par des arguments différents) le résultat du corollaire 5.11 qui affirme que les seuls points asymptotiquement stables sont les équilibres de Nash stricts. En effet, d'après la proposition précédente, un point est asymptotiquement stable si c'est une face, donc si c'est une stratégie pure, et s'il maximise localement, et strictement, le potentiel. Ce dernier point implique que le point est un équilibre de Nash strict.

La fonction de potentiel garantit que tout point d'un ensemble asymptotiquement stable est un équilibre de Nash et qu'un tel ensemble existe. Finalement, si la dynamique de meilleure réponse converge vers un ensemble asymptotiquement stable, alors elle converge vers un ensemble d'équilibres de Nash. Ce résultat est proche de celui qu'on obtient dans le cas discret, où l'algorithme de meilleure réponse converge presque sûrement vers un ensemble d'équilibres de Nash. Cependant, il existe des trajectoires de la dynamique qui ne convergent pas vers un ensemble asymptotiquement stable.

11. Notons que ce voisinage peut être $\mathcal{E}$ lui-même si $\mathcal{E} = \Delta(\mathcal{S})$. Ce cas advient si la fonction de potentiel est constante sur $\Delta(\mathcal{S})$, ce qui traduit le fait que les gains du jeu sont constants pour chaque joueur.

Exemple : Considérons le jeu de potentiel suivant dont les trajectoires pour la dynamique de réplication sont représentées à la figure 5.4 :

Gains

$(1,1)$	$(2,2)$
$(1,1)$	$(0,0)$

(5.13)

Le profil d'actions correspondant aux gains $(2,2)$ est un équilibre de Nash strict et donc un strict maximum local qui est asymptotiquement stable pour la dynamique de meilleure réponse. La face correspondant à $y = 0$ a un potentiel constant et contient un ensemble d'équilibres de Nash ($x \in [0, 0.5]$) et des trajectoires convergent vers cet ensemble. Cependant cet ensemble n'est pas un strict maximum local et n'est donc pas asymptotiquement stable.

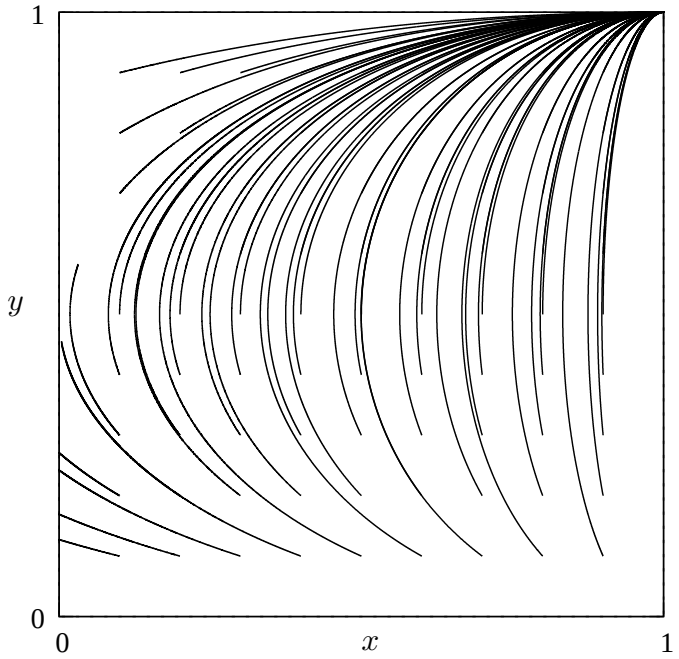


Figure 5.4 – Trajectoires de la dynamique de réplication dans le jeu de potentiel. Le seul point asymptotiquement stable est le sommet $(1,1)$.

Par la suite, nous aurons besoin d'une notion d'instabilité plus forte que l'instabilité au sens de Lyapunov. En effet, bien que les points à l'intérieur du domaine ne soient pas des attracteurs, il n'existe pas nécessairement une direction d'instabilité, ou autrement dit, ils ne sont pas *linéairement instables*. Un point est linéairement instable si la matrice jacobienne de la dynamique en ce point admet une valeur propre de partie réelle strictement positive.

Pour la dynamique de meilleure réponse, on voit que, en un équilibre de Nash mixte, $\dfrac{\partial \dot{x}_{u,a}}{\partial x_{u,a}} = 0$ pour tout (u,a) tel que $0 < x_{u,a} < 1$. Cela implique que la trace de la matrice

jacobienne en ce point est nulle, et donc également la somme des valeurs propres. Au final, un équilibre de Nash dans l'intérieur du domaine n'est éventuellement pas linéairement instable si toutes les valeurs propres ont une partie réelle nulle. Cela arrive si la dynamique est identiquement nulle, mais dans ce cas, tous les points sont stationnaires. Cela arrive également dans le jeu à 3 joueurs dont la fonction de potentiel est donnée par :

Potentiel

1	0
0	1

0	1
1	0

La stratégie mixte $(1/2, 1/2, 1/2)$ est un équilibre de Nash. En ce point, la matrice jacobienne de la dynamique est nulle.

5.2 Implémentation de la dynamique de meilleure réponse

Supposons que l'on souhaite mettre au point un algorithme de calcul effectif de la dynamique de meilleure réponse [12]. Deux problèmes se posent :

1. d'une part, le système dynamique défini par la dynamique de meilleure réponse évolue continûment dans le temps,
2. et d'autre part, les fonctions d'évolutions des stratégies de chaque joueur dépendent de $(f_{u,a}(x))_{u,a}$, les *espérances* de gain sur chaque action. Les espérances ne sont pas connues, seules les réalisations de la variable aléatoire associée (le gain) le sont.

Une solution classique pour le premier point consiste à utiliser un schéma d'approximation itératif comme la méthode d'Euler, ou des méthodes au deuxième ordre, comme Runge-Kutta. La méthode d'Euler peut facilement être implémentée de manière décentralisée chez chaque joueur. Pour la méthode de Runge-Kutta, cela semble plus difficile car elle suppose que chaque joueur connaisse l'état d'arrivée avant d'effectuer une itération, état d'arrivée qui dépend des décisions des autres joueurs. Nous nous baserons donc sur une approximation au premier ordre, dans la veine de la méthode d'Euler.

Le deuxième point concentre en fait la principale difficulté. Si l'on connaît les stratégies de chaque joueur, ainsi que le jeu, on est capable de calculer les espérances. Mais en pratique, aucune de ces informations n'est connue des joueurs dans son intégralité. Cependant, étant donnée une stratégie $x \in \Delta(\mathcal{S})$, si les joueurs choisissent simultanément, et indépendamment, des actions selon cette stratégie, le joueur u peut estimer $f_{u,a}(x)$ en prenant la moyenne d'échantillons $c_u(a, S_{-u})$ où S_{-u} est distribuée selon x_{-u}. Une fois que tous les joueurs ont une bonne estimation, une itération du schéma d'Euler peut être

12. Notons que cette démarche est à contresens de la vision "évolutionnaire" de la dynamique de réplication, qui, à partir d'un modèle d'interactions dans une population finie mais grande, aboutit à la dynamique continue par passage à la limite. Ici, on part de la dynamique, et on cherche une procédure implémentable dont le comportement moyen correspond à la dynamique.

exécutée à partir de cette estimation, et l'on recommence cette procédure à partir de la nouvelle stratégie.

Cette dernière méthode semble implémentable en pratique mais nécessite la synchronisation des joueurs afin qu'ils attendent un temps suffisamment long pour que tous aient une bonne estimation. La méthode que nous proposons est plus dynamique et n'est pas soumise à cette contrainte : elle consiste à effectuer le schéma d'Euler, non pas à partir de moyennes empiriques, mais à partir d'un échantillon unique à chaque fois. L'idée est donc de faire l'estimation et les itérations de façon simultanée [13]. L'analyse de cette méthode repose sur la théorie des approximations stochastiques d'équations différentielles.

Nous commençons par présenter les résultats généraux concernant les approximations stochastiques. Ensuite nous formulons l'algorithme d'approximation stochastique des dynamiques de meilleure réponse. Enfin, nous analysons la convergence de cet algorithme.

5.2.1 Résultats généraux sur les approximation stochastiques

Dans cette section, nous donnons les résultats fondamentaux sur les approximations stochastiques que nous utiliserons dans le cas de la dynamique de meilleure réponse. La présentation s'inspire de [Ben99] et [BEK05].

Soit le système dynamique dans $\mathbb{R}^n$ donné par :

$$\begin{cases} \dot{x} = h(x) \\ x(0) = x_0, \end{cases} \tag{5.14}$$

où $h : \mathbb{R}^n \to \mathbb{R}^n$ est une fonction de classe $\mathcal{C}^1$, que l'on suppose satisfaire les hypothèses suffisantes (par exemple celle du théorème de Cauchy Lipschitz : h est localement lipschitzienne) pour que les solutions existent et soient uniques.

Une méthode classique d'approximation de la solution de ce système dynamique est la méthode d'Euler. Celle-ci consiste à discrétiser la dynamique par des itérations de la forme $x(n+1) = x(n) + \gamma(n+1)h(x(n))$, où $\gamma(n+1)$ est une suite de pas (réels) positifs.

Dans certaines situations, comme pour la dynamique de meilleure réponse, la fonction h est difficilement calculable. Par contre, pour tout x, on sait générer des échantillons d'une variable aléatoire $Z(x)$ intégrable dont la loi dépend de x telle que $\mathbb{E}[Z(x)] = h(x)$. Intuitivement, l'approximation stochastique consiste à remplacer la fonction h par une réalisation de la variable aléatoire dans la méthode d'Euler, ce qui donne l'algorithme aléatoire $X(n+1) = X(n) + \gamma(n+1)Z(X(n))$. Un pas de l'approximation stochastique est, *en moyenne*, un pas de la méthode d'Euler.

13. Cette idée est à rapprocher des méthodes de recuit simulé [Haj88], dans lesquelles la température décroît à chaque pas de la la chaîne de Markov, alors que le processus n'a pas convergé vers son régime stationnaire.

Écriture de l'approximation stochastique de la dynamique déterministe

On appelle approximation stochastique de la dynamique (5.14) l'algorithme aléatoire décrit par l'équation de récurrence[14] :

$$\begin{cases} X(n+1) = X(n) + \gamma(n+1)\Big(h(X(n)) + U(n+1)\Big) \\ X(0) = x_0, \end{cases} \tag{5.15}$$

où la suite $(U(n))$ est une suite de différences de martingales adaptée à la filtration canonique $\{\mathcal{F}_n\}$. Cela signifie que pour tout n, $U(n)$ est intégrable et $\mathbb{E}[U(n+1)|\mathcal{F}_n] = 0$.

Un point central de l'étude des approximations stochastiques consiste à comparer la dynamique déterministe (5.14) et son approximation stochastique (5.15). Cette comparaison porte essentiellement sur le comportement asymptotique des trajectoires : on montre par exemple que les ensembles limites de la trajectoire stochastique sont inclus dans des ensembles invariants de la dynamique. Le premier exemple d'algorithme qui s'écrit comme l'approximation stochastique d'une équation différentielle est l'algorithme de Robbins-Monro introduit en 1951 afin de calculer les zéros de l'espérance d'une variable aléatoire réelle.

Le choix des pas de l'approximation stochastique est crucial pour la convergence du processus. En particulier, le choix de pas constants n'assure qu'une convergence faible du processus. Des résultats dans ce cas sont fournis dans [KHY97]. Ici, nous considérons des pas *décroissants*, ce qui donne une convergence presque sûre (détaillé dans la suite). On s'intéresse à deux classes de pas décroissants spécifiées par les hypothèses 5.13 et 5.14 :

Hypothèse 5.13

- $\sum_n \gamma(n) = +\infty$,
- $\sum_n \gamma(n)^2 < +\infty$.

Hypothèse 5.14

- $\sum_n \gamma(n)^2 = +\infty$,
- pour tout $\eta > 0$, $\sum_n \exp(-\frac{\eta}{\gamma(n)}) < +\infty$.

On désignera par *pas décroissant rapidement* toute suite de pas qui vérifie les hypothèses 5.13, et par *pas décroissant lentement* si elle vérifie les hypothèses 5.14. Les pas décroissant rapidement (resp. lentement) sont typiquement $\gamma(n) = \frac{1}{n^\alpha}$ avec $1 \geq \alpha > 0.5$ (resp. $0 < \alpha \leq 0.5$).

La plupart des résultats sur les approximations stochastiques requièrent que le processus $(X(n))$ soit presque sûrement borné, condition qui est difficile à vérifier en général (qui

14. Dans la formulation de l'approximation stochastique, on sépare l'espérance de $Z(X)$ qui est $h(X)$ du bruit autour de l'espérance, que l'on note $U \stackrel{\text{def}}{=} Z(X) - h(X)$.

porte le nom de critère de stabilité). Or, on verra dans la proposition 5.20, que l'approximation stochastique de la dynamique de réplication est bornée de façon déterministe. Par conséquent, on fait l'hypothèse suivante sur le processus :

Hypothèse 5.15

Soit $\mathcal{X}$ un compact de $\mathbb{R}^n$. Toute trajectoire de l'approximation stochastique (5.15) avec $X(0) \in \mathcal{X}$ est incluse dans $\mathcal{X}$.

Sous cette hypothèse, et si les pas sont décroissants lentement ou rapidement, alors les propositions 4.1 et 4.4 dans [Ben99] affirment que l'interpolation linéaire de l'approximation stochastique notée $(X(t))$ est presque sûrement une *pseudo trajectoire asymptotique* du système dynamique. Intuitivement, cela signifie que, sur une fenêtre de temps finie, la distance entre la trajectoire déterministe et la trajectoire stochastique tend vers zéro. Formellement cela s'écrit :

$$\forall T > 0, \lim_{t\to\infty} \sup_{0\leq s\leq T} d(X(t+s), x_t(s)) = 0,$$

où $d(,)$ est la fonction distance, et $(x_t(s))$ est la trajectoire déterministe qui vérifie $x_t(0) = X(t)$.

Les propriétés des ensembles limites de l'approximation stochastique que nous énonçons dans la suite sont une conséquence de cette propriété trajectorielle.

Ensembles limites de l'approximation stochastique

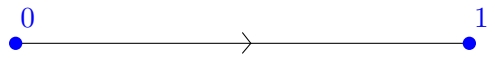
Notons $\mathcal{A}(X)$ l'ensemble des points d'accumulation de la trajectoire $(X(n))$ de l'approximation stochastique (5.15), c'est-à-dire $\mathcal{A}(X) = \{y \in \mathbb{R}^n | \exists$ une sous-suite (n_k) telle que $X(n_k) \to y\}$. Le théorème suivant énonce des propriétés sur $\mathcal{A}(X)$.

Théorème 5.16 (*Proposition 5.3 et théorème 5.7 dans [Ben99]*)

Soit $(X(n))$ une trajectoire de l'approximation stochastique (5.15) qui satisfait l'hypothèse 5.15, et l'hypothèse 5.13 ou bien 5.14. Alors l'ensemble $\mathcal{A}(X)$ est presque sûrement inclus dans un ensemble $\mathcal{E}$ qui :

- est compact,
- est connexe,
- est invariant pour la dynamique (5.14),
- ne contient aucun attracteur propre, *i.e.* $\mathcal{E}$ ne contient strictement aucun ensemble qui soit attracteur pour la dynamique restreinte à $\mathcal{E}$.

Exemple : Prenons l'exemple d'une dynamique définie sur le segment $[0, 1]$, qui est stationnaire en 0 et en 1 et strictement positive entre les deux.



L'ensemble des points d'accumulation est inclus dans un ensemble connexe et invariant, qui peut donc être soit $\{0\}$, soit $\{1\}$, soit $[0, 1]$. Mais $[0, 1]$ contient un attracteur propre qui est $\{1\}$. Finalement la trajectoire de l'approximation stochastique converge presque sûrement vers l'une des deux extrémités.

Sur l'exemple, si la condition initiale de l'approximation stochastique est à l'intérieur du segment, il semble que la probabilité de converger vers $\{0\}$ est plus faible que celle de converger vers $\{1\}$ car $\{0\}$ est un point instable pour la dynamique. Cependant, le théorème ne donne pas d'informations à ce sujet. Il s'avère que, en général, les points qui sont instables pour la dynamique peuvent être, avec une probabilité non nulle, des points d'accumulation de l'approximation stochastique.

Le théorème suivant permet dans certains cas de résoudre ce problème :

Théorème 5.17 (*Théorème 6.10 dans [Ben99]*)

Soit $\mathcal{E}$ un attracteur de la dynamique ayant un bassin d'attraction $V(\mathcal{E})$ et soit $\mathcal{K} \subset V(\mathcal{E})$ un compact. Si une trajectoire de l'approximation stochastique passe une infinité de fois dans $\mathcal{K}$, alors l'ensemble de ses points d'accumulation est inclus dans $\mathcal{E}$.

Dans l'exemple précédent, le bassin d'attraction de $\{1\}$ est $(0, 1]$. Si l'approximation visite une infinité de fois un compact inclus dans $(0, 1]$, cela implique qu'elle ne converge pas vers $\{0\}$, et donc qu'elle converge vers $\{1\}$. Ici, le théorème 5.17 est une tautologie.

Le cas des dynamiques qui admettent une fonction de Lyapunov

Rappelons que $L : \mathbb{R}^n \to \mathbb{R}$ est une fonction de Lyapunov pour la dynamique (5.14) si la dérivée de L est strictement monotone (disons positive) en tout point x non stationnaire d'une trajectoire : $\langle \nabla L(x), h(x) \rangle > 0$. Alors :

Théorème 5.18 (*Proposition 6.4 dans [Ben99]*)

Supposons les mêmes hypothèses qu'au théorème 5.16, qu'il existe une fonction de Lyapunov L pour la dynamique, et que l'ensemble des valeurs prises par la fonction de Lyapunov sur l'ensemble des points stationnaires $\{y \in \mathbb{R}^n | h(y) = 0\}$ est d'intérieur vide. Alors $\mathcal{A}(X)$ est inclus dans l'ensemble des points stationnaires.

Dans ce cas, l'approximation stochastique converge presque sûrement vers un ensemble connexe de points stationnaires sur lequel la fonction de Lyapunov est constante.

L'hypothèse "$\{y \in \mathbb{R}^n | h(y) = 0\}$ est d'intérieur vide" est satisfaite, si, par exemple, la fonction de Lyapunov est constante sur chaque composante connexe de l'ensemble des points stationnaires, et que l'ensemble des composantes est discret.

5.2.2 Approximation stochastique de la dynamique de meilleure réponse

Rappelons que la dynamique de meilleure réponse s'écrit pour chaque composante :

$$\dot{x}_{u,a} = k(x_{u,a}) \left(f_{u,a}(x) - \sum_{b \in \mathcal{S}_u} \frac{k(x_{u,b})}{\|k(x_u)\|} f_{u,b}(x) \right).$$

La fonction $f_{u,a}(x)$ est l'espérance de gain du joueur u lorsqu'il choisit l'action a sous la stratégie x, *i.e.* $f_{u,a}(x) = \mathbb{E}[c_u(a, S_{-u})]$ où S_{-u} est une variable aléatoire de loi x_{-u}.

L'hypothèse suivante est primordiale pour l'invariance de l'espace des stratégies $\Delta(\mathcal{S})$ par l'approximation stochastique :

Hypothèse 5.19

Les gains du jeu sont strictement positifs : $\forall u \in \mathcal{U}, \forall s \in \mathcal{S}, c_u(s) > 0$.

Cela implique que les espérances de gain sont strictement positives : $\forall (u,a), f_{u,a}(x) > 0$. Notons que l'ajout d'une constante aux gains du jeu permet d'assurer des gains positifs et ne modifie ni les équilibres de Nash, ni le profil des trajectoires de la dynamique de meilleure réponse.

Formulation de l'algorithme d'approximation stochastique

La construction de l'approximation stochastique de la dynamique de meilleure réponse donne lieu à l'algorithme suivant :

1. $X(0) = x_0 \in \mathrm{int}(\Delta(\mathcal{S}))$,
2. à l'instant n, les joueurs choisissent simultanément un profil d'action $S(n)$ dont la loi (qui est le produit des stratégies indépendantes de chaque joueur) est donnée par :

$$\mathbb{P}[S(n) = s] = \prod_{u \in \mathcal{U}} X_{u,s_u}(n),$$

3. simultanément, les joueurs mettent à jour leur stratégie selon la règle suivante :

$$X_{u,a}(n+1) = X_{u,a}(n) + \gamma(n+1) Z_{u,a}(n+1), \tag{5.16}$$

où $(\gamma(n))_n$ est une suite de pas positifs, et $Z(n+1)$ est la mise à jour aléatoire donnée par (on allège les notations en écrivant S et X à la place de $S(n+1)$ et $X(n)$) :

$$Z_{u,a}(n+1) = c_u(S) \frac{k(X_{u,a})}{X_{u,S_u}} \left(\mathbf{1}_{S_u = a} - \frac{k(X_{u,S_u})}{\|k(X_u)\|} \right).$$

Quelques remarques immédiates : tout d'abord, le calcul de la mise à jour de la stratégie repose uniquement, pour chaque joueur, sur l'observation de son gain courant $c_u(S)$. Aucune information sur les gains et les stratégies des autres joueurs n'est nécessaire.

Lorsqu'une action a été choisie à l'instant n, la probabilité de la choisir à l'instant $n+1$ augmente (car les gains sont positifs), et ce, d'autant plus que le gain est important, alors que celle des autres actions diminue.

Enfin, le processus est markovien, dans ce sens où les probabilités de transition du processus $(X(n))$ à l'instant n ne dépendent que de l'état à l'instant n.

Vérifions qu'il s'agit bien d'une approximation stochastique de la dynamique de meilleure réponse. Pour voir cela, on calcule la moyenne de la mise à jour, c'est-à-dire de $Z_{u,a}$ (pour

plus de lisibilité, on omet u là où aucune confusion n'est possible) :

$$\begin{aligned}
\mathbb{E}[Z_a(n+1)|X(n)] &= \sum_{b\in\mathcal{S}} \mathbb{P}[S=b]\mathbb{E}[c_u(b,S_{-u})|X(n)]\frac{k(X_a)}{X_b}\left(\mathbf{1}_{b=a} - \frac{k(X_b)}{\|k(X)\|}\right) \\
&= X_a f_a(X)\frac{k(X_a)}{X_a}(1-\frac{k(X_a)}{\|k(X)\|}) - \sum_{b\neq a} X_b f_b(X)\frac{k(X_a)}{X_b}\frac{k(X_b)}{\|k(X)\|} \\
&= k(X_a)\left(f_a(X) - \sum_b f_b(X)\frac{k(X_b)}{\|k(X)\|}\right).
\end{aligned}$$

Pour que l'approximation stochastique soit implémentable, il faut s'assurer que $X_u(n)$ reste bien un vecteur de probabilité. Il est clair que ses composantes somment à un. Il reste à garantir que $X_{u,a}(n)$ reste positif. Cela est le cas si les pas de l'algorithme vérifient les hypothèses de la proposition suivante :

Proposition 5.20

Supposons que l'hypothèse 5.19 est satisfaite et que la condition initiale $X(0)$ est dans l'intérieur du domaine $\Delta(\mathcal{S})$. Supposons également que $\frac{s}{k(s)}$ est minoré par $\nu > 0$ sur $(0,1)$, et que, pour tout n, les pas vérifient :

$$\gamma(n) < \frac{K}{M}\nu^2,$$

où K est la valeur minimale de $\|k(x)\|$ et M est le gain maximal du jeu. Alors les trajectoires de l'approximation stochastique (5.16) demeurent dans l'intérieur de $\Delta(\mathcal{S})$.

Démonstration : Il suffit de vérifier par induction que chaque composante de $(X(n))$ reste positive. En effet, si $X_{u,a}(n) > 0$, alors :

$$\begin{aligned}
\gamma(n+1)c_u(S)\frac{k(X_{u,a})}{X_{u,S_u}}\frac{k(X_{u,S_u})}{\|k(X_u)\|} &< \frac{K}{M}\nu^2 M X_{u,a}\frac{k(X_{u,a})}{X_{u,a}}\frac{k(X_{u,S_u})}{X_{u,S_u}}\frac{1}{\|k(X_u)\|} \\
&< X_{u,a}(n).
\end{aligned}$$

Comme on a supposé que les gains sont positifs :

$$X_{u,a}(n+1) \geq X_{u,a}(n) - \gamma(n+1)c_u(S)\frac{k(X_{u,a})}{X_{u,S_u}}\frac{k(X_{u,S_u})}{\|k(X_u)\|} > 0.$$

■

Notons que $\frac{s}{k(s)}$ est minoré par $\nu > 0$ sur $(0,1)$ si $\lim_{s\to 0}\frac{s}{k(s)} > 0$. Cette hypothèse n'est pas vérifiée par toutes les fonctions satisfaisant les hypothèses 5.4, comme, par exemple, $h(s) = -\frac{1}{\alpha}x^\alpha$ si $0 < \alpha < 1$. Mais elle l'est dans le cas de la dynamique de réplication.

Approximation stochastique de la dynamique de réplication

Dans la suite, nous nous bornons à l'étude de l'approximation stochastique de la dynamique de réplication. Celle-ci s'écrit plus simplement :

$$X_{u,a}(n+1) = X_{u,a}(n) + \gamma(n+1)c_u(S(n+1))\Big(\mathbf{1}_{S_u(n+1)=a} - X_{u,a}(n)\Big), \tag{5.17}$$

et le processus est assuré de rester dans l'espace des stratégies si $\gamma(n) < \dfrac{1}{M}$.

Exemple : Nous montrons sur cet exemple une trajectoire de l'approximation stochastique de réplication sur le jeu suivant :

Gains

$(0,0)$	$(5,2)$
$(1,6)$	$(1,3)$

La dynamique de réplication pour ce jeu s'écrit (x est la probabilité du choisir la ligne du haut, et y la probabilité de choisir la colonne de droite) :

$$\begin{cases} \dot{x} = x(1-x)(5y-1) \\ \dot{y} = y(1-y)(5x-3) \end{cases}$$

A la figure 5.5, on a superposé la trajectoire de la dynamique déterministe et une trajectoire du processus aléatoire en partant de la condition initiale $x = y = 0.5$. Ici, même si le processus aléatoire s'éloigne de la trajectoire déterministe, celui-ci converge finalement vers le même point. Mais cela n'est pas vrai en général : une approximation stochastique ne converge pas nécessairement vers le même ensemble limite que la trajectoire déterministe qui a la même condition initiale.

Notons que l'approximation stochastique (5.17) de la dynamique de réplication est un algorithme qui existe déjà dans la littérature des automates et de l'ordonnancement (voir par exemple [SPT94, BBBC08]). Cependant, ces articles ne considèrent que le cas où le pas est constant.

Approximation stochastique d'autres dynamiques

Une question naturelle est de savoir s'il est possible d'écrire l'approximation stochastique de n'importe quelle dynamique, de la même manière que pour la dynamique de meilleure réponse (5.16).

Supposons que l'on ait une dynamique définie par une équation différentielle qui est linéaire par rapport aux espérances de gain ($f_{u,a}$). Plus précisément, supposons que la dynamique s'écrive sous la forme suivante (explicitée pour une coordonnée) :

$$\dot{x}_{u,a} = \sum_b \ell_b(x_u) f_{u,b}(x),$$

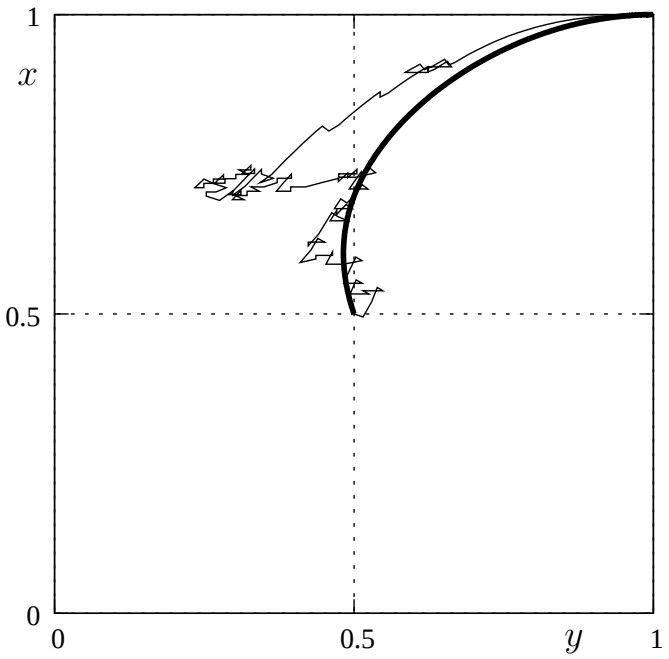


Figure 5.5 – Trajectoires de la dynamique de réplication (en gras) et du processus aléatoire donné par l'approximation stochastique.

avec éventuellement un terme additionnel indépendant de $f_{u,a}$, et où $(\ell_b)_b$ est une famille de fonctions qui vérifient les hypothèses suffisantes pour que la dynamique soit bien définie. Alors, une approximation stochastique de cette dynamique est donnée par :

$$X_{u,a}(n+1) = X_{u,a} + \gamma(n+1) \sum_b \mathbf{1}_{S_u=b} c(b, S_{-u}) \frac{\ell_b(X_u)}{X_{u,b}},$$

i.e. on peut vérifier que $\mathbb{E}[\frac{X_{u,a}(n+1) - X_{u,a}(n)}{\gamma(n+1)} | X(n)] = \sum_b \ell_b(X_u) f_{u,b}(X)$. Cela fonctionne car l'espérance de la somme des gains est la somme des espérances de gains.

Lorsque la dynamique n'est pas linéaire par rapport aux espérances de gain, on ne peut plus écrire d'approximation stochastique de la dynamique.

Montrons cela dans un jeu à deux joueurs et deux actions. Les stratégies des joueurs 1 et 2 sont respectivement x et y. L'espérance de gain du joueur 1 sur l'action a peut s'écrire $f_{1,a}(y) = \alpha y + \beta(1-y)$, où α et β sont les gains du jeu selon que le joueur 2 a choisi l'action a ou b.

Supposons donnée une dynamique qui s'écrit simplement $\dot{x} = \ell(f_{1,a}(y))$, où ℓ est une fonction réelle, et qu'il existe une approximation stochastique de la forme :

$$X(n+1) = X(n) + \gamma(n+1) h(c_1(a, S)),$$

où $\mathbb{E}[h(c_1(a, S_{-1}))] = \ell(f_{1,a}(y))$. Comme $\mathbb{E}[h(c_1(a, S_{-1}))] = yh(\alpha) + (1-y)h(\beta)$, alors on doit avoir $yh(\alpha) + (1-y)h(\beta) = \ell(f_{1,a}(y)) = \ell(y\alpha + (1-y)\beta)$ pour tout $y \in [0,1]$, α et β.

En prenant $y = 1$ on obtient $\ell(\alpha) = h(\alpha)$ pour tout α, donc $\ell = h$. De plus on voit que ces fonctions sont nécessairement linéaires.

Ainsi, il n'est pas possible de construire l'approximation stochastique d'une dynamique dans laquelle les espérances de gain n'interviennent pas de façon linéaire, comme c'est le cas de plusieurs dynamiques des jeux évolutionnaires parmi lesquelles la dynamique de projection [LS08], la dynamique de meilleure réponse (au sens classique) [GM91], la dynamique logit [FL98], et la dynamique de Brown von Neumann Nash [HOR09] (initialement introduite dans un contexte différent des jeux évolutionnaires dans [BVN50]).

Cela justifie donc l'usage de l'approximation stochastique des dynamiques de meilleure réponse (et en particulier la dynamique de réplication) par rapport aux autres dynamiques [15].

5.2.3 Prise en compte du processus de révision et de l'incertitude sur les gains dans l'approximation stochastique

On se place dans le cas de l'approximation stochastique de la dynamique de réplication.

L'approximation stochastique de la dynamique de réplication étant bornée, son ensemble limite est un ensemble connexe et compact. Le théorème 5.18 donne des propriétés supplémentaires sur cet ensemble lorsque le jeu est un jeu de potentiel (car la fonction de potentiel est une fonction de Lyapunov pour la dynamique de meilleure réponse). Nous montrons ici que l'introduction de bruit blanc sur les gains du jeu et l'extension de l'algorithme d'approximation stochastique à des processus de révisions qui ne sont pas synchrones ne modifient pas les propriétés des points limites données par ces théorèmes.

Ce résultat est intuitif car, contrairement au cas discret, chaque pas de l'algorithme modifie très peu la stratégie globale, et, sur une longue période de temps, les bruits aléatoires tendent à se moyenner.

Ajout d'un bruit blanc borné sur les gains du jeu

Nous modélisons les incertitudes sur les gains par un bruit blanc, c'est-à-dire de moyenne nulle. On suppose que ce bruit est indépendant pour chacun des joueurs, et à chaque itération de l'algorithme.

Notons ξ ce bruit de moyenne nulle et intégrable, alors :

$$\begin{aligned}
\mathbb{E}[\frac{X_{u,a}(n+1) - X_{u,a}(n)}{\gamma(n+1)}|X(n)] &= \mathbb{E}[(c_u(S) + \xi)\big(\mathbf{1}_{S_u=a} - X_{u,a}\big)|X(n)] \\
&= \mathbb{E}[c_u(S)\big(\mathbf{1}_{S_u=a} - X_{u,a}\big)|X(n)] + \mathbb{E}[\xi\big(\mathbf{1}_{S_u=a} - X_{u,a}\big)|X(n)] \\
&= \mathbb{E}[c_u(S)\big(\mathbf{1}_{S_u=a} - X_{u,a}\big)|X(n)].
\end{aligned}$$

La dynamique moyenne est donc la même avec ou sans bruit, et donc également le comportement asymptotique.

15. Par conséquent, parmi les dynamiques proposées dans l'optimisation du multihoming dans [SAK07], seule la dynamique de réplication est effectivement implémentable.

Néanmoins, il faut s'assurer que le processus va bien rester dans l'espace des stratégies. Si le bruit est non borné, la mise à jour de la stratégie peut faire que, avec probabilité positive, $X_{u,a}$ sorte des limites (*i.e.* 0 et 1). Dans le cas contraire, si l'on connaît un majorant B du bruit, on garantit que la trajectoire stochastique reste à l'intérieur de l'espace des stratégies en bornant le pas de l'algorithme par $\frac{1}{BM}$ de la même manière qu'à la proposition 5.20.

Quid des processus de révisions non synchrones ?

Rappelons qu'un processus de révision est la donnée d'une probabilité qu'un ensemble de joueurs modifient leur stratégie simultanément. C'est donc une mesure μ sur l'ensemble des sous-ensembles de joueurs $\mathcal{P}(\mathcal{U})$. Notons $U(n)$ l'ensemble aléatoire des joueurs qui révisent leur stratégie à la $n^{\text{ème}}$ itération de l'approximation stochastique. On suppose que le processus de révision ne dépend pas du temps, si bien que pour tout n, la probabilité que $U(n)$ soit égale à l'ensemble (non aléatoire) U, vaut $\mu(U)$. Le processus de révision définit un schéma asynchrone pour l'approximation stochastique qui s'écrit :

$$X_{u,a}(n+1) = X_{u,a} + \gamma(n+1)\mathbf{1}_{u\in U(n)}c_u(S)\big(\mathbf{1}_{S_u=a} - X_{u,a}\big). \tag{5.18}$$

En d'autres termes, si $u \in U(n)$ alors $\mathbf{1}_{u\in U(n)} = 1$ et le joueur u modifie sa stratégie, sinon elle reste inchangée.

Les variables aléatoires $U(n)$ et $S(n)$ étant indépendantes, le calcul de l'espérance de la mise à jour donne :

$$\begin{aligned}
\mathbb{E}[\frac{X_{u,a}(n+1) - X_{u,a}(n)}{\gamma(n+1)}|X(n)] &= \mathbb{E}[\mathbf{1}_{u\in U(n)}c_u(S)\big(\mathbf{1}_{S_u=a} - X_{u,a}\big)|X(n)] \\
&= \sum_{U|u\in U} \mu(U)X_{u,a}\big(f_{u,a}(X) - \sum_{b\in\mathcal{S}_u} X_{u,b}f_{u,b}(X)\big) \\
&= \pi_u X_{u,a}\big(f_{u,a}(X) - \sum_{b\in\mathcal{S}_u} X_{u,b}f_{u,b}(X)\big),
\end{aligned}$$

où $\pi_u \stackrel{\text{def}}{=} \sum_{U|u\in U} \mu(U)$ est la probabilité que le joueur u révise sa stratégie.

Au final, le processus de révision modifie juste la vitesse relative de la dynamique de chaque joueur [16]. En particulier, s'il existe une fonction de potentiel, celle-ci reste une fonction de Lyapunov pour la dynamique.

5.2.4 Convergence de l'approximation stochastique dans les jeux de potentiel

Dans cette section, nous analysons l'approximation stochastique (5.17) de la dynamique de réplication dans le cas des *jeux de potentiel*. Comme la fonction de potentiel F est une

16. Il s'agit d'un cas particulier du résultat principal dans [Bor06].

fonction de Lyapunov pour la dynamique, qu'elle est constante sur les composantes connexes de l'ensemble des points stationnaires de la dynamique, et que le processus aléatoire est borné, alors le théorème 5.18 s'applique :

Corollaire 5.21

Si les pas de l'approximation stochastique décroissent en vérifiant l'hypothèse 5.13 ou 5.14, alors l'approximation stochastique de la dynamique de réplication converge vers un ensemble connexe de points stationnaires de la dynamique de réplication.

Les équilibres de Nash sont des points stationnaires et sont donc potentiellement des points limites. Mais ce ne sont pas les seuls points stationnaires puisque toute stratégie pure est un point stationnaire. Comme nous allons le voir, ce résultat ne peut pas être beaucoup affiné. On va montrer que :

1. les points stationnaires isolés linéairement instables qui ne sont pas purs ne sont pas des points limites (proposition 5.22),

2. si l'approximation stochastique converge vers une face de points stationnaires, que le processus est localement une sous-martingale, et que les pas convergent lentement, alors elle converge presque sûrement vers un point pur (proposition 5.25),

3. si les pas convergent lentement ou sont constants, alors le processus converge avec probabilité positive vers un point qui n'est pas un équilibre de Nash.

Dans le cas des pas décroissant rapidement, savoir si le processus converge presque sûrement vers un équilibre de Nash reste un problème ouvert.

Évitement des stratégies non pures

On démontre le premier point dans le cas d'un point stationnaire qui est dans l'intérieur de $\Delta(\mathcal{S})$. Dans le cas de points stationnaires qui ne sont ni purs ni dans l'intérieur de $\Delta(\mathcal{S})$, le résultat découle avec la même démonstration en se restreignant aux composantes non pures.

Proposition 5.22

Soit $\bar{x}$ un point stationnaire de la dynamique de réplication qui est isolé, dans l'intérieur de $\Delta(\mathcal{S})$, et linéairement instable. On suppose que la suite des pas $(\gamma(n))$ satisfait les hypothèses 5.13 ou 5.14. Alors, presque sûrement, $\bar{x}$ n'est pas un point d'accumulation de l'approximation stochastique (5.17).

Nous utiliserons le résultat suivant pour la démonstration de la proposition :

Lemme 5.23

Soit $x = (x_a)_{a\in\mathcal{A}}$ un vecteur de $\mathbb{R}^n$ dans l'intérieur du simplexe, *i.e.* $x_a \geq 0$ pour tout $a \in \mathcal{A}$ et $\sum_{a\in\mathcal{A}} x_a = 1$. Soit $\theta = (\theta_a)_{a\in\mathcal{A}}$ un vecteur unitaire pour la norme 1 tel que $\sum_{a\in\mathcal{A}} \theta_a = 0$ (θ est dans le plan tangent du simplexe). Notons $s^+ = \max(s, 0)$. Alors il existe $\nu > 0$ indépendant de θ tel que :

$$\sum_{b\in\mathcal{A}} \big(\theta_b - \sum_{a\in\mathcal{A}} x_a\theta_a\big)^+ \geq \nu.$$

Démonstration (Lemme) : Comme x est dans l'intérieur du simplexe, il existe $\varepsilon >$ tel que $x_a \geq \varepsilon$ pour tout $a \in \mathcal{A}$.

Notons $\mathcal{A}^+$ (resp. $\mathcal{A}^-$) l'ensemble des $a \in \mathcal{A}$ tels que θ_a est strictement positif (resp. négatif). Désignons par M l'indice tel que $\theta_M \geq \theta_a$ pour tout $a \in \mathcal{A}$. Comme θ est unitaire, on a $\sum_{a\in\mathcal{A}^+} \theta_a - \sum_{b\in\mathcal{A}^-} \theta_b = 1$. De plus, θ est dans le plan tangent du simplexe donc $\sum_{a\in\mathcal{A}^+} \theta_a = -\sum_{b\in\mathcal{A}^-} \theta_b$. Cela implique que $-\sum_{b\in\mathcal{A}^-} \theta_b = 0.5$, et donc que $-\sum_{b\in\mathcal{A}^-} x_b\theta_b \geq 0.5\varepsilon$.

Au final, $\theta_M - \sum_{a\in\mathcal{A}} x_a\theta_a = \theta_M - \sum_{a\in\mathcal{A}^+} x_a\theta_a - \sum_{b\in\mathcal{A}^-} x_b\theta_b \geq \theta_M - \sum_{a\in\mathcal{A}^+} x_a\theta_M + 0.5\varepsilon \geq 0.5\varepsilon$. Le résultat est obtenu pour $\nu = 0.5\varepsilon$, qui ne dépend pas de θ. ■

Démonstration (Proposition 5.22) : Il suffit de vérifier que le théorème 1 dans [Pem90] s'applique ici. Intuitivement, l'argument central est que le processus aléatoire en $\bar{x}$ a une composante aléatoire strictement positive dans la direction d'instabilité de la dynamique.

Notons que le théorème 1 dans [Pem90] suppose des pas décroissant rapidement, mais il s'étend au cas des pas lent, car l'hypothèse des pas rapides sert à s'assurer que le processus converge presque sûrement ce qui est le cas également avec les pas lents (cela a été découvert plusieurs années après la publication de [Pem90], voir [BS00]).

Les hypothèses *i*, *ii* et *iv* du théorème sont évidentes. Il reste à montrer le point *iii*. Celui-ci nécessite que la partie aléatoire du processus stochastique au point stationnaire $\bar{x}$ soit strictement positive dans toutes les directions (et en particulier, dans une direction d'instabilité, par exemple celle du vecteur propre associé à la valeur propre positive). Pour l'approximation stochastique (5.17), la partie aléatoire est $\xi_{u,a}(n+1) = c_u(S(n+1))\big(\mathbf{1}_{S_u(n+1)=a} - x_{u,a}(n)\big)$. Formellement, il suffit de montrer que pour tout vecteur unitaire (nous prendrons la norme 1) $\theta = (\theta_{u,a})_{u\in\mathcal{U},a\in\mathcal{S}_u}$ tel que $\sum_{a\in\mathcal{S}_u} \theta_{u,a} = 0$, on a $\mathbb{E}[\big(\sum_{u,a} \theta_{u,a}\xi_{u,a}(n+1)\big)^+|\mathcal{F}_n] \geq \eta > 0$, où $s^+ = \max(s, 0)$ et ξ est considéré en $\bar{x}$, et ceci indépendemment de η. Pour alléger (un peu) les notations et sans perte de généralité, on montre l'inégalité pour un joueur quelconque. Ainsi on omet l'indice u, et également l'itération n. Alors :

$$\begin{aligned}\mathbb{E}[\big(\sum_a \theta_a \xi_a\big)^+|\mathcal{F}] &= \mathbb{E}[\big(c(S)\sum_a \theta_a(\mathbf{1}_{S=a} - \bar{x}_a)\big)^+|\mathcal{F}] \\ &= \sum_b \bar{x}_b f_b(\bar{x})\big(\theta_b - \sum_a \bar{x}_a\theta_a\big)^+\end{aligned}$$

Comme $\bar{x}$ est dans l'intérieur du domaine alors $\bar{x}_b > 0$ pour tout b. De plus $f_b(\bar{x}) > 0$ car on a supposé les gains strictement positifs. Donc $\bar{x}_b f_b(\bar{x}) \geq \varepsilon$ pour tout b.

Par le lemme 5.23, pour tout θ, $\sum_b \big(\theta_b - \sum_a \bar{x}_a\theta_a\big)^+ \geq \nu > 0$, et on obtient le résultat en prenant $\eta = \varepsilon\nu$. ■

Cas où le potentiel est constant et lien avec les urnes de Polya

Supposons que le potentiel est constant sur l'ensemble du domaine $\Delta(\mathcal{S})$. Cela signifie que le gain de chaque joueur est constant (ce qui peut arriver plus vraisemblablement sur une sous-face du domaine). On le suppose égal à 1. Alors, l'approximation stochastique est un processus indépendant pour chaque joueur et s'écrit :

$$X_a(n+1) = X_a(n) + \gamma(n+1)\Big(\mathbf{1}_{S(n+1)=a} - X_a(n)\Big). \tag{5.19}$$

Proposition 5.24

Le processus (5.19) converge presque sûrement pour toute suite de pas $(\gamma(n))$. Si les pas décroissent lentement, le point limite est pur.

Démonstration : Le processus $(X_a(n))_n$ donné par (5.19) est une martingale car :

$$[X_a(n+1)|\mathcal{F}_n] = X_a(n) + \gamma(n+1)\Big(X_a(n)(1 - X_a(n)) - \sum_{b\neq a} X_b(n)X_a(n)\Big) = X_a(n).$$

Comme le processus est borné, il converge presque sûrement.

Pour tout n, on vérifie par récurrence que :

$$X_a(n) = X_a(0) + \sum_{i=1}^n \gamma(i)Z_a(i),$$

où $Z_a(i) = \mathbf{1}_{S(i)=a} - X_a(i)$ qui vérifie $\mathbb{E}[Z_a(i+1)|\mathcal{F}_i] = 0$. Alors :

$$\begin{aligned}\mathbb{E}[X_a(n)^2] &= \mathbb{E}[\big(X_a(0) + \sum_{i=1}^n \gamma(i)Z_a(i)\big)^2] \\ &\geq \mathbb{E}[\big(\sum_{i=1}^n \gamma(i)Z_a(i)\big)^2] \\ &= \sum_{i=1}^n \gamma(i)^2\mathbb{E}[Z_a(i)^2].\end{aligned}$$

La dernière égalité vient du fait que pour tout i, $\mathbb{E}[Z_a(i)Z_a(i+1)] = \mathbb{E}[Z_a(i)\mathbb{E}[Z_a(i+1)|\mathcal{F}_i]] = 0$, et donc, de proche en proche, pour tout i et j différents,

$$\mathbb{E}[Z_a(i)Z_a(j)] = 0. \tag{5.20}$$

Comme le processus est borné, alors pour tout n et pour tout a, $\mathbb{E}[X_a(n)^2]$ est borné (et même compris entre 0 et 1). Par conséquent, $\sum_{i=1}^{\infty} \gamma(i)^2 \mathbb{E}[Z_a(i)^2] < \infty$. Dans le cas où les pas décroissent lentement, $\sum_n \gamma(n)^2 = \infty$, ce qui implique qu'il existe une sous-suite $(n_i)_i$ telle que $\mathbb{E}[Z_a(n_i)^2] \to 0$ et donc que $X_a(n_i)(1 - X_a(n_i)) \to 0$. Cette dernière relation implique que $X_a(n_i)$ converge soit vers 0, soit vers 1. Or $X_a(n)$ est presque sûrement convergent, donc il converge vers la même limite que la sous-suite.

Finalement, $X(n)$ converge vers un point pur. ■

Comme $(X(n))$ est une martingale dans ce cas, il tend *en loi* vers une variable $X(\infty)$ qui a la même distribution que $X(0)$: $\mathbb{P}[X_a(\infty) = 1] = X_a(0)$.

Dans le cas classique d'un pas décroissant rapidement comme $1/n$, il est connu [BEK05] que le processus converge en loi vers une variable aléatoire qui admet une densité strictement positive sur tout le domaine, et donc pas uniquement les points purs.

Il est intéressant de noter l'analogie de l'approximation stochastique (5.19) avec le modèle des urnes de Polya [17]. Supposons qu'une urne contienne des boules de différentes couleurs, où l'ensemble des couleurs est $\mathcal{A}$. A chaque étape, on tire une boule au hasard dans l'urne que l'on remet dans l'urne avec une boule de la même couleur. Il y a donc une boule supplémentaire à chaque étape. Notons $X_a(n)$ la proportion de boules de couleur a à l'étape n. On peut alors vérifier que l'évolution de $(X_a(n))$ est donnée par l'équation (5.19) pour $\gamma(n) = 1/n$.

Dans le cas où les gains sont constants pour un joueur, le processus d'évolution de ses stratégies est donc celui d'une urne de Polya *généralisée* (c'est-à-dire avec des pas quelconques).

La figure 5.6 montre une simulation de l'évolution d'une urne qui contient des boules de deux couleurs différentes, avec, initialement, autant de boules de chaque couleur. Pour observer le phénomène de convergence vers les points purs quand le pas décroît lentement, on interdit au processus de s'approcher à moins de ε de 0 ou de 1 (c'est-à-dire que l'on prend le maximum entre ε et la valeur du processus, et le minimum avec $1 - \varepsilon$). Ainsi, la trajectoire des pas qui décroissent rapidement converge vers une valeur intermédiaire, alors que la trajectoire des pas qui décroissent lentement ne converge pas à cause de la restriction, mais s'éloigne et revient à chaque fois vers 0 ou vers 1.

17. En fait, la dynamique de réplication peut en général être vue comme une équation différentielle limite d'une urne de Polya généralisée, voir [Sch01].

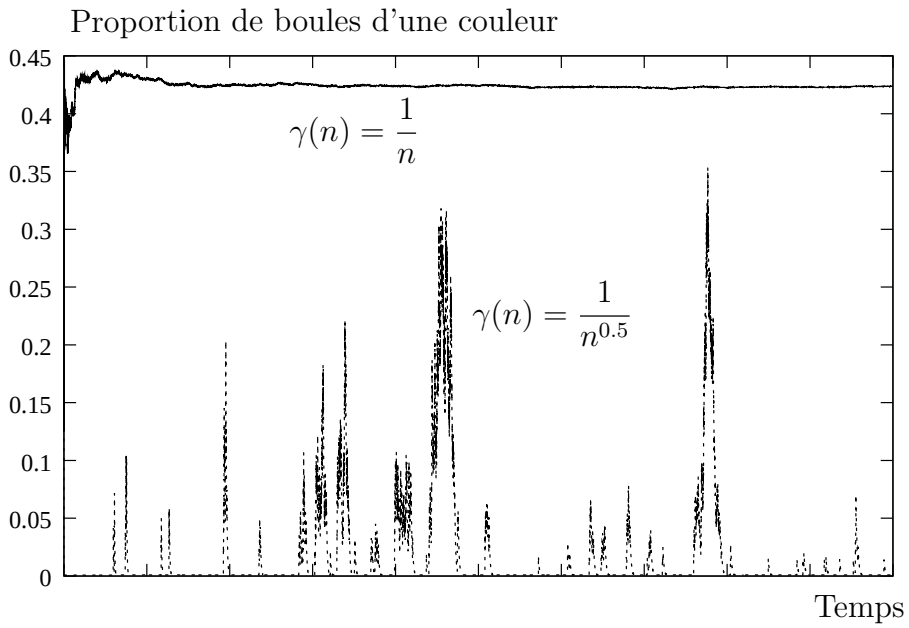


Figure 5.6 – Proportion de boules d'une couleur donnée dans un modèle d'urne de Polya généralisée décrit par l'équation (5.19) avec deux couleurs. Pour $\gamma(n) = \frac{1}{n}$, on observe la convergence vers une valeur mixte (à peu près 0.43). Pour $\gamma(n) = \frac{1}{n^{0.5}}$, comme on interdit la convergence vers les bords 0 et 1, le processus ne converge jamais.

Convergence vers un point pur d'une face de points stationnaires

Le corollaire 5.21 ne permet pas d'exclure que, lorsqu'une face de $\Delta(\mathcal{S})$ est constituée de points stationnaires, alors l'approximation stochastique n'admet pas de points limites qui ne sont pas purs dans cette face. Cela est pourtant vrai si $\Delta(\mathcal{S})$ tout entier est constitué de points stationnaires, *i.e.* le potentiel est constant, et que les pas décroissent lentement.

En fait, la proposition 5.24 se généralise dans le cas où l'approximation stochastique est une sous ou une surmartingale au voisinage d'une telle face.

Proposition 5.25

Soit $\mathcal{F}$ une face de $\Delta(\mathcal{S})$ constituée de points stationnaires de la dynamique de réplication. Supposons qu'il existe un voisinage $\mathcal{V}$ de $\mathcal{F}$ tel que pour tout $u \in \mathcal{U}$, pour tout $a \in \mathcal{S}_u$, $\dot{x}_{u,a}$ est de signe constant (au sens large) sur tout $\mathcal{V}$, et que les pas de l'approximation stochastique (5.17) décroissent lentement. Alors les points de la face qui ne sont pas des points purs ne sont pas des points d'accumulation de l'approximation stochastique (presque sûrement).

Démonstration : Par le corollaire 5.21, l'approximation stochastique converge presque sûrement vers un ensemble connexe de points stationnaires. Cela implique que, étant donnée une trajectoire de l'approximation stochastique, presque sûrement ses points d'accumula-

tion sont soit dans $\mathcal{F}$ soit dans l'ensemble complémentaire.

On peut donc se restreindre aux trajectoires qui convergent vers $\mathcal{F}$.

L'hypothèse "$\dot{x}_{u,a}$ est de signe constant sur un voisinage $\mathcal{V}$ de $\mathcal{F}$" implique que l'approximation stochastique est une sous ou une surmartingale. Donc, pour tout $u \in \mathcal{U}$ et pour tout $a \in \mathcal{S}_u$, $X_{u,a}(n)$ converge presque sûrement vers un point.

On utilise alors le même argument que dans la preuve de la proposition 5.24, c'est-à-dire que l'on calcule $\mathbb{E}[X_{u,a}(n)^2]$ qui est borné, pour montrer que $X_{u,a}(n)$ converge soit vers 0 soit vers 1. La seule différence est dans l'équation (5.20), où il faut remplacer l'égalité par un signe supérieur ou égal, ce qui ne contredit pas la preuve. ■

Convergence avec probabilité positive vers un point qui n'est pas un équilibre de Nash

Pour certains pas qui décroissent lentement ou sont constants, l'approximation stochastique converge avec une probabilité non nulle vers un point qui n'est pas un équilibre de Nash.

Nous montrons cela sur un exemple : considérons le jeu à un seul joueur avec deux actions possibles telles que la première donne un gain 1 et la deuxième un gain 2. Notons x la probabilité de choisir la deuxième action. Le seul équilibre de Nash consiste à choisir la deuxième action, *i.e.* $x = 1$.

L'approximation stochastique de la dynamique de réplication de ce jeu s'écrit :

$$X(n+1) = X(n) + \gamma(n+1) \begin{cases} 2(1 - X(n)) & \text{avec probabilité } X(n) \\ (-X(n)) & \text{avec probabilité } 1 - X(n) \end{cases} \tag{5.21}$$

Il est immédiat de constater que $(X(n))$ est une sous-martingale qui converge donc soit vers 0 soit vers 1. Notons $X(\infty)$ sa limite. Si la suite des pas vérifie :

$$\sum_n \prod_{i=1}^{n} (1 - \gamma(i)) < \infty, \tag{5.22}$$

alors $\mathbb{P}[X(\infty) = 0] > 0$.

En effet, $X(n)$ tend vers 0 si, par exemple, à chaque itération, la première action est choisie, ce qui se produit avec la probabilité $1 - X(n)$. Donc $\mathbb{P}[X(\infty) = 0] > \prod_n (1 - X(n))$, avec $X(n) = X(0) \prod_{i=1}^{n} (1 - \gamma(n))$. Un résultat classique affirme que $\prod_n (1 - X(n))$ est strictement positif si et seulement si $\sum_n X(n) < +\infty$, ce qui équivaut à la condition (5.22).

Si les pas sont constants (plus petit que 1), alors (5.22) est clairement vérifiée. Si $\gamma(n) = \dfrac{1}{n^\alpha}$ avec $0 < \alpha < 1$, alors on peut montrer par récurrence que $\prod_{i=1}^{n} (1 - \gamma(i)) \leq \gamma(n)$, et donc que (5.22) est vérifiée.

Enfin, dans le cas où $\gamma(n) = 1/n$, la proposition 5 dans [LPT04] affirme que $\mathbb{P}[X(\infty) = 0] = 0$. Mais ce résultat ne se généralise pas directement pour des jeux plus généraux.

Discussion sur la vitesse de décroissance des pas

Nous avons étudié deux types de pas pour l'approximation stochastique : les pas qui décroissent rapidement et ceux qui décroissent lentement. Le choix de ces pas permet d'avoir une convergence presque sûre de l'approximation stochastique alors qu'avec des pas constants, la convergence est faible (en loi uniquement).

En passant de la dynamique à l'approximation stochastique, on perd la propriété de convergence vers un équilibre de Nash dans le cas de pas constants ou qui décroissent lentement. Il n'est pour l'instant pas prouvé qu'il n'existe pas un choix particulier de pas pour l'approximation stochastique, par exemple $\gamma(n) = 1/n$, et ce même dans le cas d'un joueur unique, pour lequel le processus converge vers un équilibre de Nash avec probabilité 1. Notons également qu'il est possible que, si le pas initial $\gamma(0)$ tend vers 0, alors la probabilité de ne pas aller vers un équilibre de Nash tende vers 0.

On a également montré que, sous certaines hypothèses, l'approximation stochastique avec des pas qui décroissent lentement va vers des points purs, ce qui n'est pas forcément le cas pour les pas rapides, pour lesquels le processus peut converger vers une face de points stationnaires. Or, tous les points de la face ont le même potentiel, et il est facile de passer d'une stratégie mixte à une stratégie pure qui donne les mêmes gains en choisissant arbitrairement une action pour chaque joueur qui est dans le support de la stratégie mixte.

Discussion sur la vitesse de convergence de l'algorithme

La fonction de potentiel est une fonction de Lyapunov de la dynamique de meilleure réponse, ce qui permet d'avoir des bornes sur le temps de convergence de la dynamique déterministe. En particulier, en utilisant une métrique particulière (la distance de Kullback-Leibler), et sous des hypothèses fortes sur la nature du jeu, les trajectoires de la dynamique de réplication convergent exponentiellement rapidement vers un équilibre de Nash (voir le théorème 8 dans [MBML11]).

Même dans ce dernier cas, en ce qui concerne l'approximation stochastique, aucune borne *uniforme*[18] ne peut être obtenue sur l'espérance du nombre d'itérations nécessaires pour que l'algorithme converge. En effet, considérons l'exemple simple d'un joueur unique qui a deux actions possibles, et que l'approximation stochastique soit donnée par (5.21). On sait que, si les pas sont $\gamma(n) = 1/n$, alors l'approximation stochastique converge presque sûrement vers 1 quelle que soit la condition initiale. Or on peut toujours choisir la condition initiale suffisamment proche de 0 de manière à ce que le nombre d'itérations nécessaires pour entrer dans un voisinage donné de 1 soit plus grand (il s'agit d'une borne déterministe) qu'une valeur donnée. Et par conséquent, l'espérance est n'est pas uniformément bornée sur $(0, 1)$.

18. Qui ne dépend pas de la condition initiale

5.3 Application au problème d'association de mobiles à des réseaux hétérogènes

Dans cette section, nous proposons une application où l'utilisation de l'approximation stochastique de la dynamique de meilleure réponse s'avère être une solution algorithmique efficace. Il s'agit du problème de l'association de terminaux mobiles à des antennes radio de différentes technologies.

5.3.1 Présentation du problème général

Généralités sur les réseaux sans fil

Actuellement, les communication sans fil sont dominées par six technologies majeures : GSM, UMTS, HSDPA, Wifi, WiMAX et LTE. Chacune de ces technologies a ses propres avantages et inconvénients et aucune n'est amenée à prendre le pas sur les autres. D'autre part, les équipements radio tendent à satisfaire plusieurs standards, ce qui leur permet de se connecter à plusieurs antennes de technologies diverses. Cela offre la possibilité de basculer d'une technologie à une autre[19], et permet d'améliorer la qualité de service du mobile. Ce basculement est appelé *handover vertical*, par analogie au handover horizontal qui désigne le changement de cellule d'une même technologie dû à la mobilité. La décision d'initier un handover vertical peut soit être prise par le mobile lui-même (ce qui inclut le choix de l'usager), soit par le réseau, ce qui nécessite des communications entre les antennes concernées. La norme 802.21 est actuellement développée pour gérer ces communications.

Jusqu'à présent, les protocoles utilisant le handover vertical reposent sur des choix *statiques*. Par exemple, le protocole UMA donne priorité à la connexion via le Wifi par rapport à toute autre connexion, dès qu'un point d'accès Wifi est disponible. Cela ne tient nullement compte de la congestion et de la qualité de service requise par les mobiles. Par exemple, dans le protocole Wifi, aucune garantie de délai n'est considérée pour les communications temps réel, alors que cela est intégré dans les réseaux cellulaires (UMTS et LTE). De plus, un mobile qui se connecte via un point d'accès Wifi dans de mauvaises conditions radio aura non seulement un débit faible, mais réduira aussi sérieusement le débit des autres mobiles attachés à ce point d'accès (voir [HRBSD03]).

Formulation du problème d'association

On considère un ensemble $\mathcal{U}$ de mobiles pouvant se connecter à un réseau via un ensemble de points d'accès sans fil. Ces points d'accès peuvent être de technologies différentes.

19. Cela permet également de maintenir plusieurs communications en parallèle, technique appelée "multihoming". Le multihoming entre différentes technologies est complexe en raison de l'hétérogénéité des protocoles, et induit beaucoup de trafic additionnel (par exemple il faut gérer le fait que les paquets n'arrivent pas dans l'ordre dans lequel ils ont été envoyés). Pour ces raisons, le handover vertical simple lui est généralement préféré. Notons également que pour le handover vertical, plusieurs connexions peuvent être maintenues en même temps pour permettre de faire du handover "soft" plutôt que "hard", rendant la procédure de basculement rapide et transparente du point de vue du mobile.

L'ensemble des points d'accès de chaque mobile dépend bien évidemment de sa localisation géographique, mais également du terminal et de l'abonnement à l'opérateur du réseau. On note $\mathcal{S}_u$ l'ensemble des points d'accès auxquels le mobile u peut se connecter. Une action du mobile est le choix d'un point d'accès $a \in \mathcal{S}_u$. La figure 5.7 montre un exemple de configuration du réseau, avec plusieurs cellules [20] Wifi locales qui sont recouvertes par une unique et large cellule de technologie WiMAX.

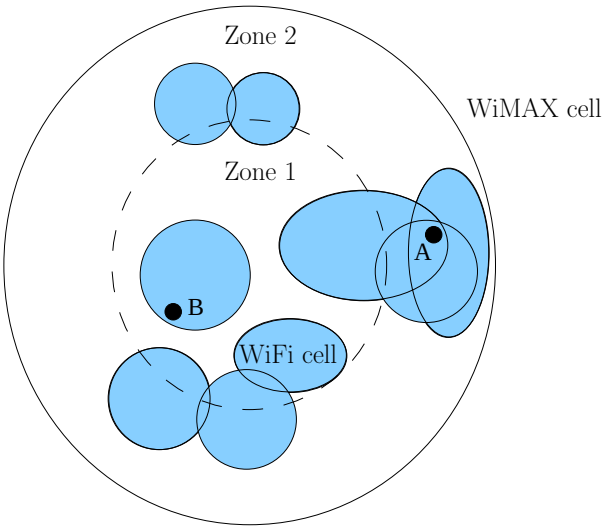


Figure 5.7 – Un système constitué de cellules de technologies sans fil hétérogènes : une large cellule MAN (Metropolitan Area Network, par exemple WiMAX) et un ensemble de petite cellules LAN (Local Area Networks, par exemple WiFi). Lorsque le mobile B (dans la zone 1) est proche de l'antenne WiMAX, il peut utiliser un codage plus efficace, et donc avoir un débit utile plus important, que le mobile A (dans la zone 2) : par exemple le codage QAM 16 plutôt que le codage QPSK. Les zones sont représentées en lignes pointillées, par opposition aux cellules en ligne pleines.

Ici, le gain du jeu est le *débit* de connexion du mobile. Par débit, on fait référence à la quantité d'informations utiles (les donnés liées aux protocoles réseau mises à part) transitant entre le point d'accès et le mobile par unité de temps (aussi appelé "goodput" dans la littérature anglophone). Le choix de ce gain se justifie pour des applications longues telles que le téléchargement de données, mais pas pour des applications temps réel (voix, streaming vidéo) pour lesquelles le délai est prépondérant [21].

Le débit d'un mobile sur un point d'accès dépend à la fois de l'état de charge du point d'accès, donc des autres mobiles connectés, et également des paramètres propres au mobile : ceux-ci incluent la position géographique (degré d'interférences et taux d'atténuation), ainsi que les paramètres de la carte sans fil (parmi lesquels on a le type de codage, et la version

20. Le terme cellule désigne la couverture géographique d'un point d'accès.
21. En fait d'autres critères pourraient être pris en compte en théorie. L'implémentation que nous proposons est néanmoins spécifique au débit des mobiles.

de TCP). Cela implique que les mobiles connectés à un même point d'accès n'ont pas nécessairement le même débit. Enfin, l'état de charge, ou la *charge d'une cellule* est un vecteur, que nous notons $\ell_a = (\ell_a^u)_{u\in\mathcal{U}}$, tel que $\ell_a^u(s) = 1$ si $s_u = a$ et 0 sinon.

Nous supposons ici que le débit d'un mobile dépend *uniquement* du vecteur de charge du point d'accès auquel il est connecté. En particulier, il ne dépend pas de l'activité des cellules voisines. Cela se justifie si les cellules utilisent des spectres de fréquences différentes, par exemple lorsqu'il s'agit de technologies différentes. Le débit du mobile u connecté au point d'accès a s'écrit $d_a^u(\ell_a)$. Finalement, la fonction de gain du joueur u est $c_u(s) = d_{s_u}^u(\ell_{s_u}(s))$.

Laisser les usagers choisir leur point d'accès peut amener à une utilisation peu efficace des ressources, ainsi qu'à de perpétuels basculements de technologie d'une partie d'entre eux. De la même manière, les choix fixés par priorité de technologie n'utilisent pas au mieux la capacité de toutes les cellules. La construction d'une procédure d'association doit donc permettre d'améliorer les performances, et une bonne solution doit être non seulement *Pareto optimale*, mais également *α-équitable* (le paramètre α autorise de la flexibilité dans le choix entre l'efficacité globale et l'équité entre les mobiles). Le problème d'optimisation associé s'écrit donc :

$$\max_{s\in\mathcal{S}} \sum_{u\in\mathcal{U}} c_u^\alpha(s) \text{ où } c_u^\alpha(s) = \frac{c_u(s)^{1-\alpha}}{1-\alpha}. \tag{5.23}$$

La procédure doit en plus être complètement distribuée. Cela signifie que les seules communications possibles sont entre les mobiles et le point d'accès auquel ils sont attachés.

Notons que, sans hypothèses sur les fonctions de débit (par exemple la convexité), ce problème d'optimisation ne peut être résolu sans avoir au préalable calculé la fonction $\sum_{u\in\mathcal{U}} c_u^\alpha(s)$ en chaque état de $\mathcal{S}$. Le nombre d'états croissant exponentiellement avec le nombre de mobiles, il est illusoire d'avoir un mécanisme général qui converge rapidement vers un maximum global. Au mieux, on peut espérer converger vers un maximum local en un temps raisonnable.

5.3.2 Implémentation de l'algorithme d'association des mobiles aux cellules

Nous détaillons ici la façon dont l'approximation stochastique de la dynamique de meilleure réponse peut en pratique être implémentée pour la gestion de l'association des mobiles aux cellules.

Nous considérons l'extension mixte du jeu fini $(\mathcal{U}, \mathcal{S}, c^\alpha)$: on note $x_{u,a}$ la probabilité pour le mobile u de se connecter au point d'accès a. La fonction à optimiser dans (5.23) s'écrit, sur l'espace des stratégies mixtes :

$$F(x) \stackrel{\text{def}}{=} \sum_{u\in\mathcal{U}} \mathbb{E}[c_u^\alpha(S)], \tag{5.24}$$

où S est le profil d'action aléatoire sous la stratégie x. Notons que F est multi-affine et donc maximale en des stratégies pures.

Mécanisme d'incitation par fonction de potentiel

Nous utilisons les résultats de la section 2.3.2 afin de construire un mécanisme d'incitation de manière à ce que F soit une fonction de potentiel du jeu modifié. Pour cela, on ajoute une fonction de pénalité aux gains du jeu. Cette pénalité prend en compte l'impact de la décision de chaque joueur sur les autres. Il est pratique d'appeler *utilité* la nouvelle fonction de gain, notée r^α, qui est définie par :

$$r_u^\alpha(a, s_{-u}) \stackrel{\text{def}}{=} c_u^\alpha(s) - \sum_{i \neq u | s_i = a} \Big(c_i^\alpha(\bar{a}, s_{-u}) - c_i^\alpha(s) \Big), \tag{5.25}$$

où $\bar{a}$ désigne toute action différente de a. Le terme $c_i^\alpha(\bar{a}, s_{-u}) - c_i^\alpha(s)$ mesure l'impact du joueur u sur le gain du joueur i, qui est positif uniquement si $s_i = a$.

Pour garantir que l'approximation stochastique reste bien dans l'espace des stratégies mixtes, il faut s'assurer que les utilités que nous avons définies sont *strictement positives* (voir la proposition 5.20). Pour cela, il suffit d'ajouter une constante positive assez grande K_α (qui dépend du paramètre d'équité α) aux fonctions de gain (*i.e* de débit), si bien que la nouvelle utilité vaut $r_u^\alpha + K_\alpha$: la même constante est ajoutée aux fonctions d'utilité.

La capacité de chaque cellule étant finie, le nombre de mobiles connectés à un point d'accès est fini et les utilités sont donc bornées. On assure alors la stricte positivité en ajoutant $K_\alpha \geq \max(0, \max_{\ell,u,a} -r_u^\alpha(\ell_a))$. Notons que le fait d'ajouter une constante aux utilités ne modifie pas le fait que la fonction (5.24) est une fonction de potentiel du jeu $(\mathcal{U}, \mathcal{S}, r^\alpha + K_\alpha)$.

Supposons que le débit de chaque mobile soit borné par $C_{\min} > 0$ et $C_{\max}$ quelle que soit la cellule et la charge, et que l'on puisse majorer le nombre de mobiles connectés en même temps à un même point d'accès par $N_{\max}$. Le choix suivant :

$$K_\alpha \stackrel{\text{def}}{=} N_{\max}\big(G_\alpha(C_{\max}) - G_\alpha(C_{\min})\big), \tag{5.26}$$

satisfait $K_\alpha \geq \max(0, \max_{\ell,u,a} -r_u^\alpha(\ell_a))$, et assure donc la positivité des utilités. La valeur K_α ainsi définie peut être vue comme une constante universelle puisqu'elle ne dépend pas du système considéré [22], et on la suppose connue des mobiles.

Formulation de l'algorithme

L'approximation stochastique (5.17) appliquée au jeu dont les gains sont les utilités (5.25) plus la constante définie par (5.26) définit un algorithme qui comporte deux parties : d'un côté, les mobiles mettent à jour, à chaque itération, leur stratégie (leur vecteur x_u), et d'un autre côté chaque point d'accès calcule l'utilité de chaque mobile qui lui est attaché, et lui envoie cette information.

22. Néanmoins, la valeur de K_α dépend du choix des unités utilisées pour calculer le débit. Par précaution, il vaut mieux calculer K_α à partir des unités les plus petites.

Les mobiles ont une stratégie initiale uniforme sur l'ensemble de leurs actions. A chaque itération (typiquement à chaque envoi d'un nombre fixe de paquets), une partie des mobiles, qui dépend du processus de révision, exécute la boucle de l'algorithme 5 de manière indépendante.

Algorithme 5: Algorithme d'association des mobiles aux cellules pour le mobile u

initialisation;
Le mobile $u \in \mathcal{U}$ choisit une stratégie $x_u(0)$ uniforme sur $\mathcal{S}_u$;
répéter
- Choisir un point d'accès selon la stratégie $x_u(n)$;
- Recevoir l'utilité $r_u^\alpha(S(n))$;
- Mettre à jour la stratégie $x_u(n+1)$ selon :

$$x_{u,a}(n+1) \leftarrow x_{u,a}(n) + \gamma(n+1)\big(r_u^\alpha(S(n)) + K_\alpha\big)\Big(\mathbf{1}_{S_u=a} - x_{u,a}(n)\Big) \tag{5.27}$$

jusqu'à *l'infini*;

D'un autre côté, les points d'accès des cellules mesurent les débits descendants vers les mobiles. Ils calculent ensuite les utilités. Cela est résumé à l'algorithme 6.

Algorithme 6: Algorithme d'association des mobiles aux cellules pour le point d'accès a

répéter
- **pour chaque** *mobile u attaché au point d'accès a ($S_u(n) = a$)* **faire**
  - Mesurer le débit $c_u(S(n))$;
  - Calculer $c_u^\alpha(S(n))) = \dfrac{c_u(S(n))^{1-\alpha}}{1-\alpha}$;
- **pour chaque** *mobile u attaché au point d'accès a* **faire**
  - Calculer l'utilité $r_u^\alpha(S(n))$ selon la formule (5.25);
  - Envoyer la valeur $r_u^\alpha(S(n))$ au mobile u;

jusqu'à *l'infini*;

Finalement, cet algorithme ne nécessite aucune communication excepté entre chaque mobile et son point d'accès. Cela permet de libérer les ressources sur les autres points d'accès. Cet algorithme distribué repose sur des mesures en temps réel, et il est donc également auto-adaptatif, c'est-à-dire qu'il tient compte de l'évolution lente des paramètres du système.

Plusieurs choix pour le pas de l'algorithme et pour un critère d'arrêt sont proposés dans les heuristiques détaillées à la section 5.3.3.

Exemple : Considérons une exécution typique de l'algorithme dans un système constitué de 10 mobiles et de 10 point d'accès. La figure 5.8 montre l'évolution de la stratégie d'un mobile. Au début, la distribution est uniforme. Avec le temps, toutes les probabilités sauf une convergent vers zéro, donc vers une stratégie pure.

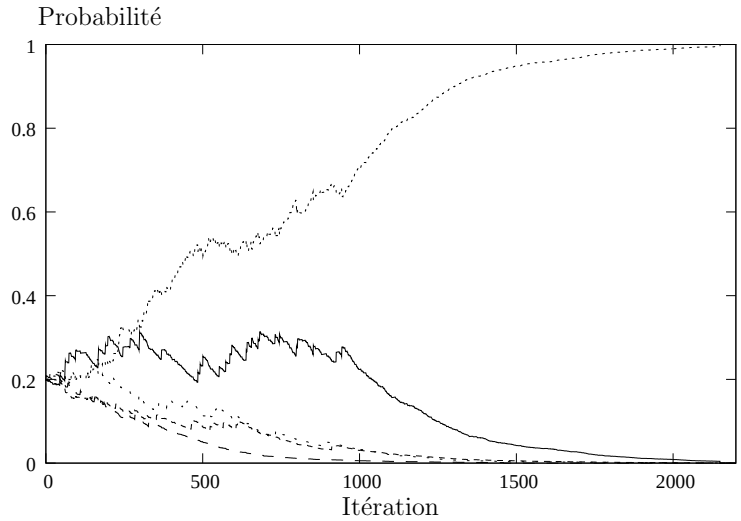


Figure 5.8 – Convergence de la stratégie d'un joueur pour chacun des 5 choix possibles de point d'accès.

Détails d'implémentation

La principale difficulté d'implémentation réside dans le calcul de la fonction d'utilité des mobiles à partir de la mesure des débits. En effet, cette fonction repose sur la connaissance de l'impact de la décision d'un mobile sur les autres mobiles connectés sur le même point d'accès. Alors que le débit peut être mesuré en ligne, le débit qu'un mobile aurait si un autre mobile n'était pas là nécessite de mémoriser certaines valeurs.

Tout d'abord, voici la description de la mesure du débit. Rappelons que le débit fait référence au débit utile descendant vers un mobile. Celui-ci est, par définition, la limite du débit moyen. Étant donnée une fenêtre de temps, on mesure la quantité de paquets qui ont été transmis durant cette période, ce qui suppose qu'il y a toujours des paquets en attente d'être transmis. Dans le cas contraire, le débit peut être estimé à partir des paramètres de connexions qui dépendent de la technologie utilisée : ces paramètres sont, entre autres, la modulation utilisée et les ressources affectées au mobile (puissance d'émission, codes, fréquences, temps).

Pour calculer les utilités (5.25), il suffit que, à chaque itération de l'algorithme, le mobile interrompe sa connexion puis se reconnecte, éventuellement à un nouveau point d'accès. Au moment de la reconnexion, le point d'accès à gardé en mémoire les débits des mobiles avant que celui-ci ne se reconnecte. Il est donc en mesure de calculer l'utilité du mobile.

5.3.3 Simulation de l'algorithme

Cette section est consacrée à des simulations de l'algorithme 5 et 6 afin d'étudier plusieurs heuristiques possibles portant sur le choix des pas et du critère d'arrêt, et également afin de mesurer le gain par rapport à des solutions classiques. Nous commençons par décrire le scénario sur lequel les simulations reposent.

Scénario des simulations

Nous considérons une configuration simple d'un opérateur fournissant à des abonnés un service de communication sans fil disponible soit par technologie WiMAX, donc par des cellules larges (quelques kilomètres de couverture), soit par technologie Wifi (quelques dizaines de mètres de couverture).

Pour chacune des simulations, une topologie est choisie aléatoirement, comme à la figure 5.7, selon trois paramètres : le nombre de mobiles, le nombre de points d'accès Wifi, et le nombre de choix possibles par mobile. Plus précisément, pour chaque mobile :

- Le premier choix correspond à la cellule WiMAX et il est situé aléatoirement dans l'une des huit zones possibles qui définissent la modulation et donc le débit instantané reçu par le mobile.
- Les autres choix sont tirés uniformément parmi l'ensemble des points d'accès Wifi. On ne considère pas de zones pour le Wifi : le débit dépend uniquement du nombre de mobiles connectés.

Débit des sessions TCP en Wifi et WiMAX

Avoir une formule de débit pour des terminaux connectés par interface radio est extrêmement difficile en raison, entre autres, de la complexité du système physique : contrairement aux liaisons filaires, où le médium physique est isolé de l'extérieur et dont les performances fluctuent peu, la qualité des interfaces radio change constamment en raison des variations de l'environnement, celui de l'air, et des obstacles physiques. Par conséquent, les formules closes dans la littérature ont été obtenues en faisant des hypothèses fortes et ne sont pertinentes que si l'on considère des débits moyens sur des fenêtres de temps suffisamment larges. Les formules reposent essentiellement sur une approximation fluide du trafic.

De surcroît, le débit utile d'une connexion dépend des protocoles de communication. Il y a principalement deux protocoles qui ont un impact important : le premier est le protocole de la couche physique qui dépend de la technologie sans fil utilisée, le deuxième étant le protocole de transport. Dans nos simulations, on considère le cas de flux TCP pour lesquels de bonnes approximations existent. De plus, l'utilisation de flux UDP, ou d'un mélange de flux TCP et UDP n'a pas de réel impact sur les performances de l'algorithme : le choix du protocole UDP ou TCP peut se modéliser par l'ajout d'une zone supplémentaire dans chaque cellule.

Nous utilisons les résultats dans [MKA06] pour établir les équations de débit dans le

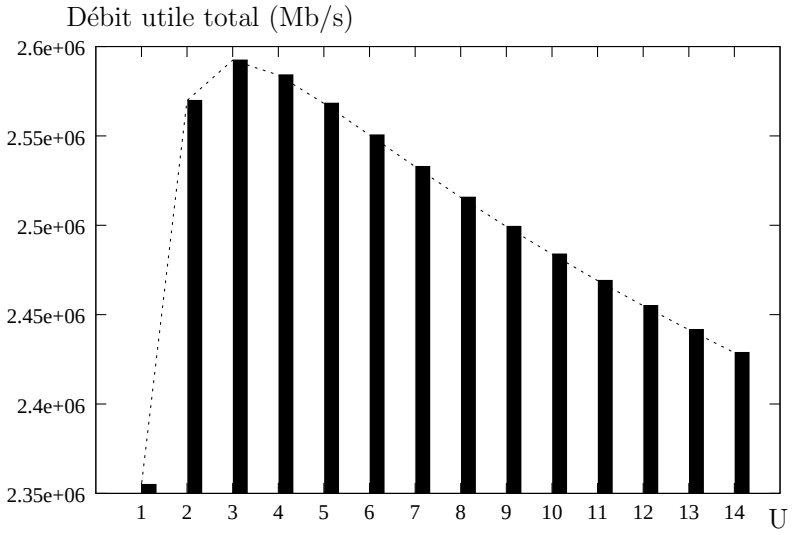


Figure 5.9 – Débit utile global d'une cellule Wifi en fonction de sa charge (en Mbit/s). La valeur maximale est obtenue avec trois mobiles.

cas du Wifi. Étant donné le vecteur de charge ℓ sur un point d'accès Wifi, et $L = \sum_{u \in \mathcal{U}} \ell_u$ le nombre de mobiles connectés, le débit est donné par :

$$c_u(\ell) = \frac{\mathrm{L_{TCP}}}{L\big(\mathrm{T_{DATA}} + \mathrm{T_{ACK}} + 2\mathrm{T_{TBO}}(L) + 2\mathrm{T_W}(L)\big)}, \tag{5.28}$$

où $\mathrm{L_{TCP}} = 8000$bit est la taille d'un paquet TCP, $\mathrm{T_{ACK}}$ est le temps de transmission d'un accusé de réception en TCP (approximativement 1 ms), $\mathrm{T_{DATA}}$ le temps de transmission d'un paquet de données en TCP (de l'ordre de 2 ms). Ensuite, $\mathrm{T_{TBO}}(.)$ et $\mathrm{T_W}(.)$ sont respectivement le temps moyen total perdu à cause des collisions de paquets et des backoffs. Ces valeurs dépendent de la probabilité de collision de chaque paquet et peuvent être calculées à partir de la charge de la cellule. La probabilité de collision peut être calculée numériquement en résolvant une équation de point fixe donnée dans [MKA06]. La figure 5.9 montre le débit global d'une cellule Wifi obtenue par ces calculs en fonction de la charge.

Dans la spécification du protocole 802.11, une seule communication est active à chaque instant, et l'accès au canal de communication est géré par le protocole CSMA-CA. Au contraire, pour la technologie WiMAX, plusieurs communications sont actives simultanément grâce au multiplexage par la méthode OFDMA. Ainsi, à chaque mobile est alloué un ensemble de porteuses, chacune de ces porteuses transmettant à un certain débit dépendant de la modulation choisie et du codage, qui dépendent essentiellement des conditions radio entre l'émetteur et le récepteur. On considère ici un partage équitable des porteuses (voir [TC06]), ce qui signifie que les porteuses d'un point d'accès WiMAX sont partagées de façon égale entre les mobiles connectés. Par conséquent, le débit utile reçu par chaque mobile dans une zone (*i.e.* pour un codage donné) est approximativement le débit qu'il aurait s'il était le seul mobile connecté divisé par le nombre de mobiles actuellement connectés.

Pour un mobile connecté tout seul dans une cellule WiMAX, on choisit les valeurs mesurées expérimentalement dans [YDW07] pour la norme IEEE 802.16d qui comporte huit zones de codage :

Modulation	QAM64 3/4	QAM64 2/3	QAM16 3/4	QAM16 1/2
TCP débit utile	9.58	8.88	6.80	4.50
Modulation	QPSK 3/4	QPSK 1/2	BPSK 3/4	BPSK 1/2
TCP débit utile	3.37	2.21	1.65	1.08

Heuristiques pour le choix du pas de l'algorithme

Nous proposons maintenant plusieurs heuristiques simples pour le choix du pas de l'algorithme d'approximation stochastique. Bien que le pas doive être suffisamment petit pour assurer la convergence avec une grande probabilité vers un maximum local, des grandes valeurs sont préférables afin de diminuer le temps de convergence de l'algorithme. Par conséquent, il faut trouver un bon compromis entre ces deux objectifs opposés.

Chaque heuristique comporte deux composantes : une règle de calcul du pas de l'algorithme, et un critère d'arrêt.

Avec le temps, l'algorithme tend vers une stratégie pure, et la probabilité de choisir chacune des actions tend soit vers zéro soit vers un. Afin d'arrêter l'algorithme, on fixe des seuils δ_m et δ_M tels que pour tout $u \in \mathcal{U}$:

- si $x_{u,a}(n) < \delta_m$, alors $x_{u,a}(n+1) = 0$ et $x_{u,b}(n+1) = \dfrac{x_{u,b}(n+1)}{\sum_{c \in \mathcal{S}_u | c \neq a} x_{u,c}(n)}$ pour tout $b \in \mathcal{S}_u$, $b \neq a$,
- si $x_{u,a}(n) > 1 - \delta_M$, alors $x_{u,a}(n+1) = 1$ et pour tout $b \in \mathcal{S}_u$ et $b \neq a$, $x_{u,b}(n+1) = 0$.

Dans les tests que nous avons effectués, on a fixé $\delta_m = 0.05$ et $\delta_M = 0.3$.

Plusieurs règles de calcul des pas ($\gamma(n)$) de l'algorithme ont été envisagées :

- *Pas constant (CSS)* : les pas sont fixés au début de l'algorithme et constants au cours du temps : $\forall n, \gamma(n) = \gamma$. Pour des valeurs basse (CSS_L), typiquement $\gamma = 0.01$, l'algorithme converge dans la plupart des cas vers l'optimum global, mais au détriment d'un temps de convergence important. Pour des valeurs élevées (CSS_H), typiquement $\gamma = 1$, la performance est bien moins bonne. Des valeurs intermédiaires (CSS_M), typiquement $\gamma = 0.1$, peuvent constituer un compromis.
- *Mise à jour constante (CUS)* : à chaque itération, chaque joueur calcule la taille de pas maximale de manière à ce que la distance (en norme infinie) entre la stratégie avant et après l'itération est bornée par une constante γ fixée à l'avance (fixée à 0.1 dans les tests). En bornant la taille des mises à jour de chaque mobile, cette règle garantie des changements réguliers de stratégies et on peut donc s'attendre à ce qu'elle "suive" bien l'évolution de la trajectoire déterministe.
- *Pas décroissant (DSS)* : L'idée sous-jacente de cette règle est de faire quelques itérations avec des pas larges pour converger rapidement vers des zones d'attractions de maximum locaux, et ensuite d'utiliser des pas plus petits pour bien converger vers l'optimum en question. On considère deux variantes. Dans la première variante, les

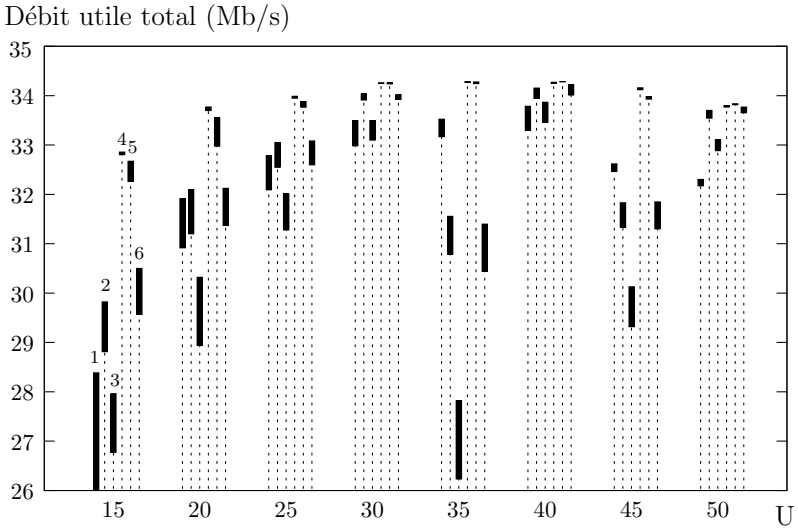


Figure 5.10 – Performances moyennes pour les différentes heuristiques ($CUS, DSSSA, DSSCSS, CSS_L, CSS_M$ et CSS_H resp.) avec différents nombre de mobiles (avec des intervalles de confiance de 95%).

pas décroissent par cycle ($DSSSA$) : dans les tests, $\gamma(n) = 3/(n \bmod 10)$. Dans la deuxième variante ($DSSCSS$), on considère un pas décroissant rapidement suivi ensuite par un pas constant large : dans les tests $\gamma(n) = 4/n$ si $n < 120$, puis $\gamma(n) = 3$ ensuite. L'idée ici étant que, dans la première phase, les stratégies de quelques mobiles vont se stabiliser, et que dans la deuxième phase, on assure une convergence rapide des autres mobiles.

La figure 5.10 montre les performances, c'est-à-dire le débit utile global obtenu par chacune des heuristiques en fonction du nombre total U de mobiles. Pour chaque valeur de U, toutes les heuristiques ont été testées sur la même topologie qui a été préalablement tirée au hasard suivant la procédure décrite précédemment.

L'heuristique CSS_L utilisant des petits pas donne les meilleurs performances. On a vérifié en testant toutes les combinaisons jusqu'à $U = 20$ que cette heuristique atteint la plupart du temps l'optimum global.

Toutes les heuristiques exceptée $DSSCSS$ dont les performances peuvent être très mauvaises, ne dégradent pas les performances de plus de 10% comparée à l'heuristique CSS_L. On peut également noter que le débit global du système est borné par $10 * 2.6 + 9.58 Mbit/s$, où 2.6 est le débit maximal sur un point d'accès Wifi et 9.58 sur un point d'accès WiMAX. On voit alors que les meilleures heuristiques sont à moins de 5% de cette borne. On peut noter également que les performances obtenues avec des pas constants intermédiaires CSS_M sont très proches des valeurs avec un pas petit, et que l'heuristique CUS se comporte de mieux en mieux par rapport à CSS_L à mesure que le nombre de mobiles croît.

Le nombre d'itérations avant convergence est très variable selon les heuristiques, même en représentant les valeurs sur une échelle logarithmique. Les résultats sont montrés à la

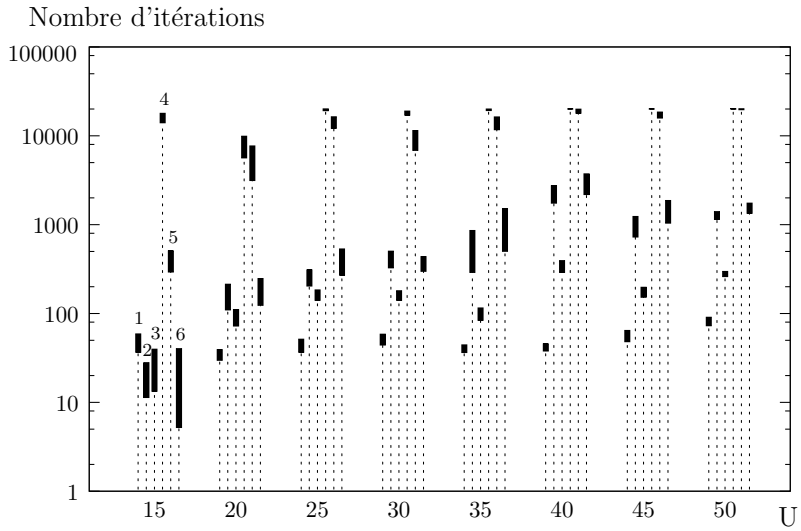


Figure 5.11 – Nombre moyen d'itérations avant la convergence des heuristiques ($CUS, DSSSA, DSSCSS, CSS_L, CSS_M$ et CSS_H resp.) avec différents nombre de mobiles (avec des intervalles de confiance de 95%).

figure 5.11. L'heuristique CUS est clairement la meilleure ici, avec un nombre d'itérations moyen inférieur à 80 dans tous les cas. Le nombre d'itérations nécessaires pour la convergence de l'heuristique CSS_L dépasse le nombre d'itérations limite fixé ici à 20000.

Lorsque le nombre de mobiles est grand, CUS semble être le meilleur compromis avec une convergence rapide et des performances raisonnables par rapport aux autres heuristiques. Lorsque le nombre de mobiles est faible, le choix de pas constants moyens CSS_M est également intéressant.

Alors que des pas petits semblent donner des performances quasiment optimales, toutes les heuristiques, excepté CUS, nécessitent plusieurs centaines d'itérations avant de converger dans des scénarios faisant intervenir plus d'une dizaine de mobiles et de points d'accès. Le nombre d'itérations de CUS ne dépasse jamais 100, et l'association atteinte donne de bonnes performances. De plus, la figure 5.12 montre le nombre moyen de handovers effectués par mobile. On voit qu'il reste faible dans tous les cas. Nous utilisons donc cette heuristique dans les tests qui suivent.

Impact du paramètre d'équité

Considérons le scénario suivant : un ensemble de 20 mobiles, chacun ayant 3 possibilités de connexion parmi 10 points d'accès. Le point d'accès WiMAX est numéroté 0 et ses 8

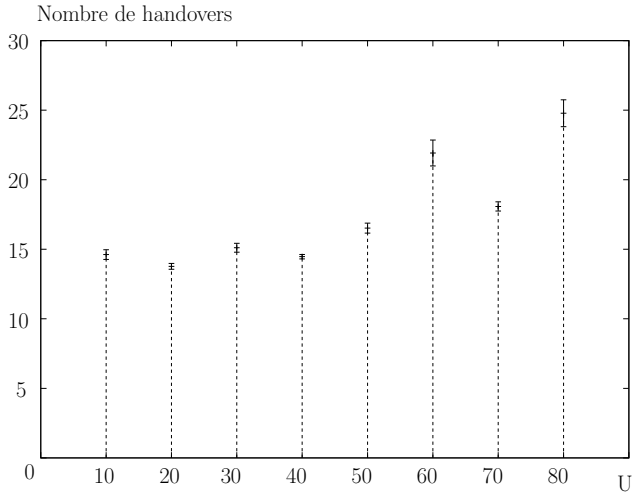


Figure 5.12 – Nombre moyen de handovers par mobile en fonction du nombre de mobiles dans le système pour l'heuristique CUS (avec intervalles de confiance à 95%).

zones sont numérotées de 0 à 7. L'ensemble des choix des mobiles est le suivant :

$$\begin{array}{llll}
\{\{0,1\},\{8\},\{1\}\} & \{\{0,5\},\{6\},\{4\}\} & \{\{0,1\},\{6\},\{9\}\} & \{\{0,2\},\{2\},\{6\}\} \\
\{\{0,3\},\{8\},\{9\}\} & \{\{0,6\},\{4\},\{9\}\} & \{\{0,7\},\{3\},\{6\}\} & \{\{0,4\},\{1\},\{2\}\} \\
\{\{0,6\},\{6\},\{9\}\} & \{\{0,5\},\{3\},\{4\}\} & \{\{0,6\},\{3\},\{1\}\} & \{\{0,7\},\{9\},\{6\}\} \\
\{\{0,3\},\{8\},\{1\}\} & \{\{0,6\},\{4\},\{7\}\} & \{\{0,6\},\{9\},\{5\}\} & \{\{0,0\},\{6\},\{5\}\} \\
\{\{0,5\},\{4\},\{1\}\} & \{\{0,6\},\{6\},\{4\}\} & \{\{0,3\},\{3\},\{4\}\} & \{\{0,3\},\{8\},\{4\}\}.
\end{array}$$

Les associations optimales pour $\alpha = 0$, c'est-à-dire qui maximise la somme des débits, appelée association *efficace*, et pour $\alpha = 2$, appelée association *équitable*, sont respectivement :

$$\begin{aligned}
s_{\text{eff}} &= \{2,1,2,1,1,1,1,2,2,2,1,1,2,2,2,0,2,1,1,1\}, \\
s_{\text{equi}} &= \{0,1,0,1,0,2,1,2,1,1,2,1,1,2,2,2,1,2,0,1\},
\end{aligned}$$

ce qui correspond aux débits :

$$\begin{aligned}
c_{\text{eff}} = {} & 0.824,\ 1.225,\ 0.824,\ 1.225,\ 1.225,\ 1.225,\ 0.824,\ 1.225,\ 0.824,\ 1.225, \\
& 0.824,\ 0.824,\ 0.824,\ 2.245,\ 2.246,\ 9.58,\ \ 0.824,\ 1.225,\ 0.824,\ 1.225. \\
c_{\text{equi}} = {} & 2.22,\ \ 1.225,\ 2.22,\ \ 1.225,\ 1.125,\ 1.225,\ 1.225,\ 1.225,\ 1.225,\ 1.225, \\
& 2.245,\ 1.225,\ 1.225,\ 2.246,\ 1.225,\ 1.225,\ 1.225,\ 1.225,\ 1.125,\ 1.225.
\end{aligned}$$

L'association efficace atteint ici un débit total de 31.29 Mb/s. L'association équitable dégrade ce débit d'un peu moins de 10%, celui-ci valant 28.34 Mb/s. Cependant, si l'on regarde plus attentivement les tables de valeurs, on voit que l'association efficace mène à de grandes disparités entre les mobiles : par exemple, le premier mobile a un débit de 0.8 Mb/s alors que le neuvième a un débit de 9.58 Mb/s. Au contraire, avec l'association

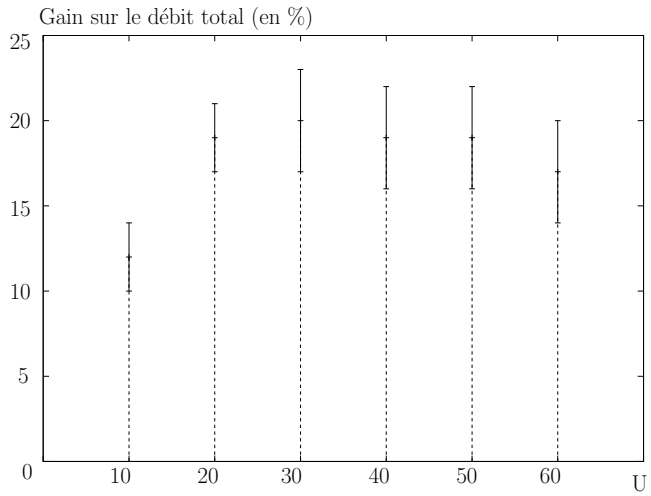


Figure 5.13 – Pourcentage de gain moyen (avec intervalles de confiance à 95%) sur le débit total utile en utilisant notre algorithme (heuristique CUS) par rapport au choix prioritaire de la cellule Wifi pour chaque mobile arrivant dans le système.

équitable, tous les mobiles ont un débit supérieur à 1.1 Mb/s. Le paramètre α permet donc un compromis fin entre le débit total maximal et l'équité entre les mobiles.

Afin de comprendre les différences de valeurs observées, comparons la charge sur les neuf points d'accès Wifi :

$$\ell_{\text{eff}}^{wifi} = \{3, 2, 3, 2, 1, 2, 1, 2, 3\}, \quad \ell_{\text{equi}}^{wifi} = \{1, 2, 2, 2, 2, 2, 1, 2, 2\}.$$

D'après la figure 5.9, le débit total maximal d'une cellule Wifi est obtenu lorsque 3 mobiles y sont attachés. Ainsi, l'association efficace essaie autant que possible de regrouper par 3 les mobiles sur les points d'accès Wifi. De la même manière, le débit sur la cellule WiMAX est maximisé si seuls les mobiles de la zone 0 y sont connectés. Dans notre exemple, il n'y a qu'un seul mobile dans la zone 0, qui reçoit donc tout le débit.

D'un autre côté, pour l'association équitable, les charges des cellules Wifi sont beaucoup plus homogènes, et la cellule WiMAX est partagée entre plusieurs mobiles qui ne sont pas nécessairement dans la zone 0.

Gain de l'algorithme par rapport au protocole UMA (Unlicensed Mobile Access)

Nous comparons les performances obtenues par l'utilisation du protocole UMA qui donne la priorité systématique à la connexion via des points d'accès Wifi par rapport aux autres technologies. Il est clair que ce protocole ne peut pas être très efficace lorsque la charge des cellules est importante. La figure 5.13 montre que l'on gagne en moyenne de l'ordre de 20% sur le débit global en utilisant notre approche.

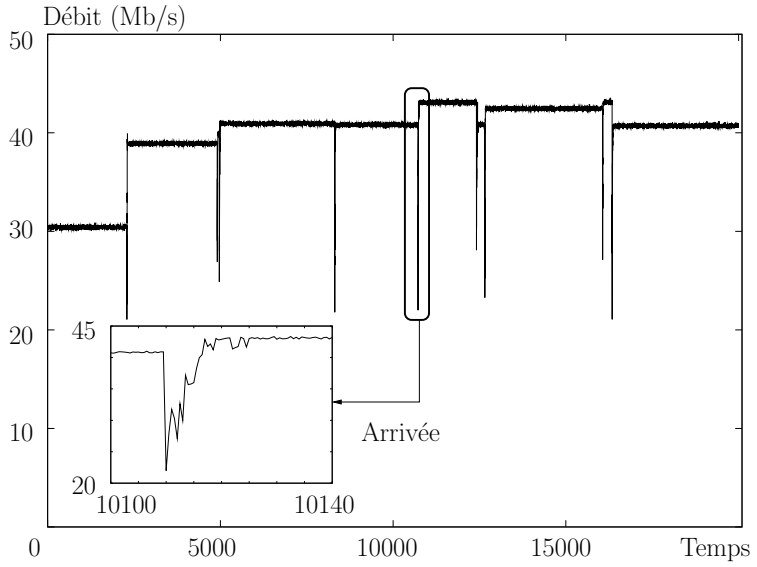


Figure 5.14 – Deux caractéristiques : 1) Adaptation aux arrivées et départs des mobiles : l'heuristique reconverge rapidement après un changement d'état. 2) Stabilité vis à vis du bruit blanc : les débits sont perturbés par un bruit blanc Gaussien de variance 0.45 MB/s.

Adaptation aux perturbations et à la dynamique du système

Les propriétés de l'algorithme sont indépendantes de l'ajout d'un bruit blanc sur les débits mesurés par les mobiles. Nous considérons des perturbations rapides de type fading que nous modélisons par un bruit blanc gaussien.

Lorsque l'ensemble des mobiles n'est plus statique mais varie au cours du temps par des arrivées, des départs et de la mobilité, l'algorithme d'association doit être exécuté à chaque changement.

On simule le comportement de l'algorithme en présence d'arrivées et de départs tout en prenant en compte le bruit blanc. La figure 5.14 montre un échantillon dans lequel les arrivées des mobiles suivent un processus de Poisson, et où chaque mobile reste le temps nécessaire pour télécharger un fichier dont la taille aléatoire suit une loi exponentielle. Une unité de temps correspond sur la figure à l'exécution d'une itération de l'algorithme. Les échelles de temps sont ici de l'ordre d'une arrivée par minute, et de une seconde pour la convergence de l'algorithme.

Prise en compte des trafics "souris et éléphants"

Ici, on considère un trafic global constitué de deux types de trafics appelés *éléphant* et *souris*. Les souris correspondent aux connexions courtes (de l'ordre de la seconde) et les éléphants aux connexions longues (plus d'une minute). Il est connu que, à l'heure actuelle, la proportion de connexions souris est de l'ordre de 90%, mais que le trafic éléphant représente 85% du volume total des communications. Alors que notre algorithme est bien adapté pour

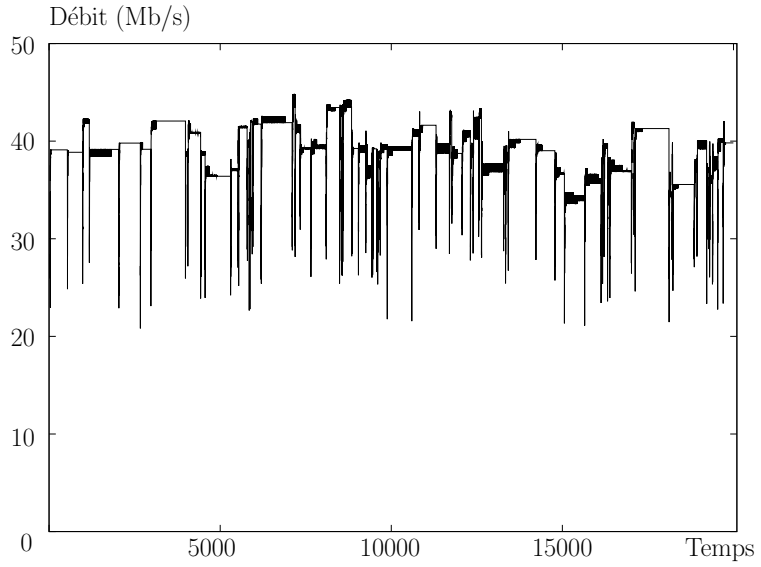


Figure 5.15 – Variation du débit total en fonction du temps lorsque l'algorithme d'association est exécuté pour les connexions souris et éléphant. Le débit moyen est 39.05 Mb/s.

le trafic de type éléphant, puisque le temps de convergence est négligeable devant le temps de connexion, cela n'est plus le cas pour le trafic souris. Aux figures 5.15 et 5.16, on compare deux scénarios : dans le premier, l'algorithme est exécuté à toutes les arrivées, dans le second, seulement pour les éléphants tandis que les souris se connectent automatiquement à l'une des cellules Wifi. Au final, la deuxième méthode permet d'éviter de trop nombreux handovers tout en conservant un débit acceptable.

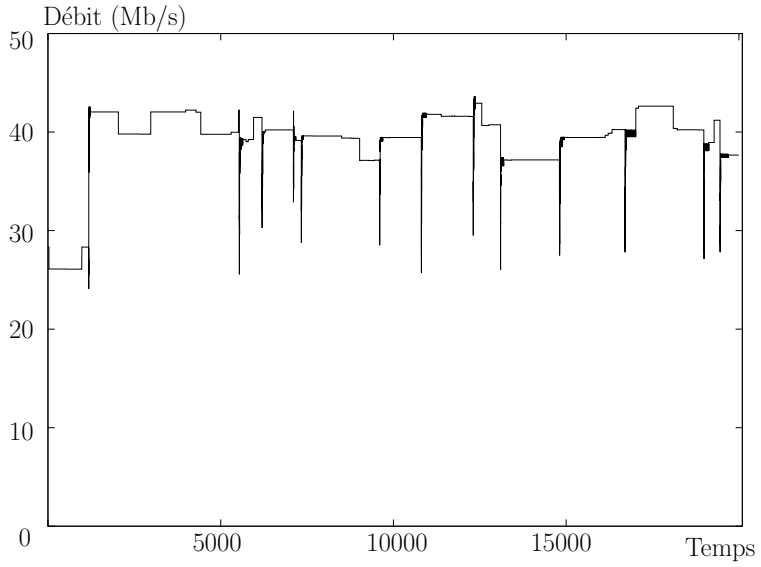


Figure 5.16 – Variation du débit total en fonction du temps lorsque l'algorithme d'association est exécuté uniquement pour les connexions éléphant, les connexions souris étant attachées directement à un point d'accès Wifi. Le débit moyen est 39.19 Mb/s.

CONCLUSION ET EXTENSIONS

L'optimisation des réseaux sans fil intervient à plusieurs niveaux qui vont du dimensionnement du réseau jusqu'à la gestion des ressources des mobiles et des points d'accès. Alors que la gestion des ressources s'opère sur une échelle de temps si brève qu'elle n'autorise pas l'utilisation de stratégies complexes, le routage des communications, que ce soit le choix d'un point d'accès au réseau (le problème du handover), ou bien le choix d'une route dans les réseaux ad hoc, permet, le temps d'une communication, de chercher une solution optimale. Le routage se distingue également des autres optimisations par le fait que les terminaux mobiles, et donc les usagers, sont au cœur des décisions. C'est pourquoi nous avons abordé ce problème d'optimisation sous l'angle de la théorie des jeux.

Le sujet principal de la théorie des jeux est la détermination des résultats d'un jeu donné. Cela suppose de modéliser le comportement des joueurs, ainsi que l'information dont ils disposent pour prendre leurs décisions. Dans les réseaux sans fil, cette information se réduit essentiellement à la qualité de service perçue par les mobiles, et aux informations transmises par le point d'accès au réseau qui a une vision générale des mobiles qui lui sont attachés. Ces informations permettent de créer localement de la coordination (ou de l'incitation) entre les mobiles.

Jusqu'à présent, les mécanismes d'incitation reposaient sur l'instauration de pénalités afin d'obtenir un jeu avec des stratégies dominantes (par exemple le mécanisme VCG). Nous avons montré que, dans le cadre du routage, il n'existe pas un tel mécanisme général qui tienne compte de l'anonymat des joueurs, et qui s'adapte donc à toutes les configurations. Par contre, il existe un mécanisme tel que le jeu obtenu admet une fonction de potentiel.

Les jeux de potentiel fournissent un cadre d'optimisation naturel dans lequel les gains de l'ensemble des joueurs sont agrégés en une fonction unique. Dans ce cas, l'algorithme stochastique de meilleure réponse conduit les joueurs à choisir un profil d'actions qui maximise la fonction de potentiel. Néanmoins, cela n'est plus vrai dès lors que les gains du jeu sont soumis à des fluctuations aléatoires ou si le processus de révision des stratégies des joueurs n'est pas complètement asynchrone. Or, en pratique, ces deux points ne peuvent être garantis : d'une part, les gains du jeu sont obtenus par des mesures physiques et sont donc nécessairement soumis à des imprécisions (sans parler du phénomène de fading

inhérent aux communications sans fil), et d'autre part, il est très complexe et coûteux de garantir l'asynchronisme dans les grands systèmes distribués.

L'algorithme de meilleure réponse évoluant dans l'espace des stratégies mixtes permet d'intégrer ces deux facteurs, c'est-à-dire que ses points limites ne dépendent pas du fait qu'il y ait du bruit sur les gains, ou que le processus soit ou ne soit pas asynchrone. Cependant, les propriétés des points limites ne sont pas aussi fortes que dans l'algorithme discret, ceux-ci pouvant même ne pas être des équilibres de Nash. Cette approche ne permet donc pas d'avoir des garanties théoriques fortes (c'est-à-dire avec probabilité un). Néanmoins, il est important de constater que, même avec l'algorithme discret et asynchrone, la convergence vers un optimum global est un résultat asymptotique qui n'est donc pas atteint en pratique (sans compter les problèmes numériques). En pratique, la convergence vers un optimum global a lieu avec grande probabilité, mais pas avec probabilité un. Finalement, du point de vue de l'implémentation, il semble plus instructif d'avoir des propriétés sur la distribution limite plutôt que des propriétés de convergence presque sûre. Mais cela reste plus difficile à obtenir en théorie. Les simulations permettent alors d'avoir une estimation de la distribution. En particulier, celles que nous avons proposées dans cette thèse montrent le bon comportement de l'algorithme qui évolue dans l'espace des stratégies mixtes.

Nous proposons maintenant cinq extensions des travaux présentés dans cette thèse qui nous paraissent pertinentes aussi bien du point de vue théorique qu'applicatif.

Extensions sur les mécanismes d'incitation

Un mécanisme peut être vu comme la construction d'un jeu par un opérateur qui dispose de différents moyens d'action, parmi lesquels l'établissement de fonctions de pénalités sur les actions prises par les joueurs. Un mécanisme est incitatif si les résultats du jeu correspondent aux états qui maximisent la préférence de l'opérateur.

La notion de résultat d'un jeu sous-entend que le comportement des joueurs (rationalité, modèles d'apprentissage) est donné avec le jeu. Au chapitre 2, nous avons supposé que ce résultat était connudans les jeux de potentiel, et dans les jeux qui admettent des stratégies dominantes. Nous avons montré les possibilités et les limites de la construction d'un mécanisme d'incitation reposant sur chacune de ces classes de jeux.

Une telle étude pourrait être étendue à d'autres classes de jeux. En particulier les jeux dont on connaît le résultat pour certains modèles d'apprentissage, comme les jeux stables [HS09] et les jeux supermodulaires.

Dans toute la thèse, nous avons également supposé que :

- le système était statique (par exemple l'ensemble des joueurs ne change pas au cours du temps),
- et que l'opérateur cherchait à maximiser sa préférence *instantanée*.

Il est cependant naturel, dans le cadre des réseaux sans fil, de considérer la dynamique liée à la mobilité des usagers. Est-il possible de construire un mécanisme d'incitation dans le cas où les préférences de l'opérateur tiendraient compte du futur (un gain moyen par exemple) ? Un tel mécanisme permettrait l'optimisation distribuée de systèmes dynamiques. Cela est actuellement réalisé par des méthodes centralisées (par exemple les méthodes d'op-

timisation des processus markoviens [Put05]) qui sont donc difficilement implémentables dans des grands systèmes distribués.

Extension du modèle de meilleure réponse aux coalitions

L'étude des coalitions dans les réseaux s'est beaucoup développée ces dernières années (voir la thèse [Saa10]). Les communications entre usagers leur permettent en effet de coordonner leurs actions afin d'augmenter conjointement leur qualité de service.

Supposons que l'on étende le modèle de meilleure réponse du chapitre 3 en autorisant des déviations non plus unilatérales, mais par coalitions. Le premier problème à résoudre consiste à définir ce qu'est une meilleure réponse pour la coalition. En effet, il n'y a pas d'ordre total sur les vecteurs de gain. Par exemple, quel est le meilleur gain entre $(3, 1)$ et $(2, 2)$? Une solution naturelle consiste à remplacer la meilleure réponse par une réponse qui est Pareto optimale.

Deux questions se posent alors :

- Peut-on caractériser l'ensemble de convergence de l'algorithme de meilleure réponse (ou plutôt de réponse Pareto optimale) à partir du graphe de meilleures réponses ? Et dans quels jeux cette dynamique converge vers des équilibres ?
- Comment adapter l'algorithme *stochastique* de meilleure réponse (du chapitre 4) dans ce cas ?

Prise en compte du passé dans l'apprentissage par le modèle stochastique de meilleur réponse

Tous les modèles d'apprentissage que nous avons étudiés reposent sur les gains courants des joueurs. Ce sont des modèles markovien.

Considérons l'algorithme stochastique de meilleure réponse du chapitre 4 dans le cas des jeux de potentiel. On a montré (théorème 4.3) que, lorsque le processus de révision est asynchrone, le processus d'apprentissage converge vers un maximum global du potentiel avec grande probabilité. Est-ce que ce résultat reste toujours valable si l'on remplace les gains courants par une moyenne des gains passés ?

Plus formellement, supposons que la fonction de choix (la probabilité que le joueur u choisisse l'action a) de l'algorithme stochastique de meilleure réponse à l'étape n soit modifiée de la manière suivante :

$$\frac{\exp\left(\eta^{-1} C_{u,a}(n)\right)}{\sum_{b \in \mathcal{S}_u} \exp\left(\eta^{-1} C_{u,b}(n)\right)},$$

où $C_{u,a}(n)$ est le gain moyen pour du joueur u sur l'action a. Le gain moyen est défini récursivement par $C_{u,a}(0) = c_u(a, S_{-u}(0))$ et :

$$C_{u,a}(n+1) = \begin{cases} \gamma(n+1) c_u(a, S_{-u}(n+1)) + (1 - \gamma(n+1)) C_{u,a}(n) & \text{si } u \text{ révise sa stratégie,} \\ C_{u,a}(n) & \text{sinon.} \end{cases}$$

Le choix $\gamma(n) = 0$ pour tout n est l'algorithme classique, c'est-à-dire sans mémoire. Si $\gamma(n)$ est constant positif ou bien décroissant, alors le processus est l'approximation stochastique d'une équation différentielle qui possède une fonction de Lyapunov si le jeu est un jeu de potentiel (voir [CMS10] [23]). Cela permet de caractériser les points limites du processus. Néanmoins, cela ne fournit aucune information sur la distribution limite du processus. En particulier, est-ce que la distribution se concentre sur les états de potentiel maximal, quand $\eta \to 0$, comme pour $\gamma(n) = 0$?

Si cela était le cas, on aurait un algorithme qui converge avec une grande probabilité vers un maximum global de la fonction de potentiel, et qui, de plus, serait robuste au processus de révision et aux incertitudes sur les gains du jeu. Des simulations effectuées sur de petits exemples ont montré que, si l'on choisit $\gamma(n)$ constant positif, la distribution tend à se concentrer, quand $\eta \to 0$, sur l'état qui maximise le potentiel. De plus, la convergence est plus rapide [24] que dans le cas $\gamma(n) = 0$.

Autre implémentation des dynamiques continues de meilleure réponse

L'implémentation de la dynamique de meilleure réponse du chapitre 5 repose sur l'approximation stochastique de cette dynamique. Comme nous l'avons montré, les résultats théoriques de convergence, quoique robustes au processus de révision et aux fluctuations aléatoires des gains, apportent peu de garanties. Même dans les jeux de potentiel, l'approximation stochastique converge vers un état qui n'est pas un équilibre de Nash avec probabilité positive.

Rappelons que, si F est le potentiel du jeu, l'approximation stochastique s'écrit sous la forme :

$$X(n+1) = X(n) + \gamma(n+1)\big(\nabla_G F(X(n)) + U(n+1)\big),$$

où $\nabla_G F$ est le gradient par rapport à une métrique particulière définie par la fonction G.

Afin d'obtenir des garanties sur l'état limite, et même de converger presque sûrement vers un état qui maximise globalement le potentiel, une possibilité consiste à s'inspirer des méthodes de recuit simulé dans des espaces continus comme dans [GM90] (bien que les hypothèses de cet article ne soient pas toutes vérifiées dans notre cas). L'idée est d'ajouter à l'approximation stochastique un terme aléatoire exogène, ce qui donne :

$$X(n+1) = X(n) + \gamma(n+1)\big(\nabla_G F(X(n)) + U(n+1)\big) + \beta(n+1)W,$$

où W est un bruit blanc gaussien, et $(\beta(n))$ est une suite de pas pour le bruit exogène qui décroît plus lentement que la suite $(\gamma(n))$.

23. Cela est analysé dans cet article dans le cas où, pour reprendre la terminologie de la section sur les approximations stochastiques, les pas sont décroissants rapidement. Mais les résultats restent vérifiés pour des pas décroissant lentement.

24. Il s'agit ici d'une convergence en loi, et nous utilisons la distance en variation totale pour quantifier la vitesse.

Prise en compte de l'hétérogénéité des joueurs : joueurs atomiques et non-atomiques

Comme nous l'avons mentionné à la fin du chapitre 5, la plupart des communications sur les réseaux sont brèves (appelées communications souris), mais la quantité globale du trafic provient de communications longues (éléphants). Optimiser les communications brèves implique de faire fréquemment appel à un mécanisme d'optimisation, alors que l'état global ne change pas de façon brutale. Il semble donc plus intéressant d'optimiser uniquement les communications longues.

Cette situation peut se modéliser par un jeu dans lequel coexistent deux types de joueurs : des joueurs atomiques (les éléphants) et un continuum de joueurs non-atomiques (les souris) [25]. En supposant que les joueurs se répartissent selon un équilibre de Wardrop, des bornes sur le prix de l'anarchie sont connues [CCSM09]. Et s'il est possible d'optimiser toutes les communications, alors il n'y a aucune dégradation de performance.

Une piste de recherche intéressante est d'établir le prix de l'anarchie dans le cas intermédiaire, dans lequel les joueurs atomiques sont incités (par un opérateur) à agir de manière optimale, sachant que les joueurs non-atomiques se répartissent ensuite sur le réseau selon un équilibre de Wardrop. Cela permettrait de quantifier la dégradation de performance induite par le fait que l'opérateur n'agit que sur les joueurs atomiques, et donc de décider s'il est également intéressant d'intervenir sur les autres joueurs.

25. Notons que ce type de jeu est courant en microéconomie où la taille des acteurs est très variable. Par exemple, la plupart des marchés impliquent à la fois de grosses entreprises et les individus qui, pris séparément, ont un impact négligeable sur les prix.

BIBLIOGRAPHIE

[ABB05] Felipe Alvarez, Jerome Bolte, and Olivier Brahic. Hessian riemannian gradient flows in convex programming. *SIAM journal on control and optimization*, 43 :477–501, 2005.

[ABP10] U. Ayesta, O. Brun, and BJ Prabhu. Price of anarchy in non-cooperative load balancing. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–5. IEEE, 2010.

[ADPT92] S.P. Anderson, A. De Palma, and J.F. Thisse. *Discrete choice theory of product differentiation*. Mit Pr, 1992.

[AFN10] C. Alós-Ferrer and N. Netzer. The logit-response dynamics. *Games and Economic Behavior*, 68(2) :413–427, 2010.

[AJM07] E. Arcaute, R. Johari, and S. Mannor. Network formation : Bilateral contracting and myopic dynamics. *Internet and Network Economics*, pages 191–207, 2007.

[AM93] J. Andreoni and J.H. Miller. Rational cooperation in the finitely repeated prisoner's dilemma : Experimental evidence. *The Economic Journal*, 103(418) :570–585, 1993.

[Aum87] R.J. Aumann. Correlated equilibrium as an expression of bayesian rationality. *Econometrica : Journal of the Econometric Society*, pages 1–18, 1987.

[Aum99] R.J. Aumann. " Acceptable points in general cooperative n-person games. *Topics in mathematical economics and game theory : essays in honor of Robert J. Aumann*, 23 :1, 1999.

[BBBC08] Dominique Barth, Olivier Bournez, Octave Boussaton, and Johanne Cohen. A dynamic approach for load balancing. Technical report, LORIA Research Report, 2008. http ://www.loria.fr/ bournez/load/Soumis-Octave-Fev-2008.pdf.

[BC09] O. Bournez and J. Cohen. Learning Equilibria in Games by Stochastic Distributed Algorithms. *Arxiv preprint arXiv :0907.1916*, 2009.

[BEK05] M. Benaïm and N. El Karoui. *Promenade aléatoire : Chaînes de Markov et simulations ; martingales et stratégies*. Editions Ecole Polytechnique, 2005.

[Ben99] M. Benaïm. Dynamics of stochastic approximation algorithms. *Seminaire de probabilites XXXIII*, pages 1–68, 1999.

[Ber07] U. Berger. Brown's original fictitious play. *Journal of Economic Theory*, 135(1) :572–578, 2007.

[BGLS97] J.F. Bonnans, J.C. Gilbert, C. Lemaréchal, and C. Sagastizábal. *Optimisation Numérique : aspects théoriques et pratiques*. Springer-Verlag, 1997.

[Blu93] L. Blume. The statistical mechanics of strategic interaction. *Games and economic behavior*, 5(3) :387–424, 1993.

[Blu97] L.E. Blume. Population Games. *The economy as an evolving complex system II*, page 425, 1997.

[Bor06] V.S. Borkar. Stochastic approximation with controlled Markov noise. *Systems & control letters*, 55(2) :139–145, 2006.

[Bou04] T. Boulogne. Jeux stratégiques non-atomiques et applications aux réseaux. *Hal. inria*, 2004.

[BR97] M. Balinski and G. Ratier. Of stable marriages and graphs, and strategy and polytopes. *SIAM review*, 39(4) :575–604, 1997.

[Bre99] P. Bremaud. *Markov Chains, Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer, 1999.

[Bro51] G. W. Brown. Iterative solution of games by fictitious play. *"Activity Analysis of Production and Allocation"*, 1951.

[BS00] M. Benaïm and S.J. Schreiber. Ergodic properties of weak asymptotic pseudotrajectories for semiflows. *Journal of Dynamics and Differential Equations*, 12(3) :579–598, 2000.

[BVN50] G.W. Brown and J. Von Neumann. Solutions of games by differential equations. *Contributions to the Theory of Games I*, 24 :73–79, 1950.

[CB10] C. S. Chen and F. Baccelli. Self-optimization in mobile cellular networks : power control and user association. *IEEE International Conference on Communications*, pages 1–6, 2010.

[CBR04] Ian D. Chakeres and Elizabeth M. Belding-Royer. Aodv routing protocol implementation design. In *Proceedings of the 24th International Conference on Distributed Computing Systems*, 2004.

[CCSM09] Roberto Cominetti, José R. Correa, and Nicolás E. Stier-Moses. The impact of oligopolistic competition in networks. *OPERATIONS RESEARCH*, 57 :1421–1437, 2009.

[CFT97] H.C. Chen, J.W. Friedman, and J.F. Thisse. Boundedly rational Nash equilibrium : a probabilistic choice approach. *Games and Economic Behavior*, 18(1) :32–54, 1997.

[CHT11] Pierre Coucheney, Emmanuel Hyon, and Corinne Touati. Admission and Allocation Policies in Heterogeneous Wireless Networks with Handover. In

54th IEEE Global Communications Conference (GLOBECOM 2011), Houston, 2011. (submitted paper).

[CHTG09] P. Coucheney, E. Hyon, C. Touati, and B. Gaujal. Myopic versus clairvoyant admission policies in wireless networks. In *Proceedings of the Fourth International ICST Conference on Performance Evaluation Methodologies and Tools*, pages 1–10. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009.

[CJWK02] K.W. Chin, J. Judge, A. Williams, and R. Kermode. Implementation experience with MANET routing protocols. *ACM SIGCOMM Computer Communication Review*, 32(5) :49–59, 2002.

[CMS10] R. Cominetti, E. Melo, and S. Sorin. A payoff-based learning procedure and its application to traffic games. *Games and Economic Behavior*, 70(1) :71–83, 2010.

[DWV07] D. Dubois, M. Willinger, and P.N. Van. Risque et sélection d'équilibre dans un jeu de coordination : Une analyse expérimentale. *Annales d'Économie et de Statistique*, pages 97–114, 2007.

[EA83] I. Eshel and E. Akin. Cevolutionary instability of mixed Nash solutions. *Journal of mathematical biology*, 18(2) :123–133, 1983.

[FL98] D Fudenberg and D Levine. *Theory of Learning in Games*. MIT Press, 1998.

[FW98] M.I. Freidlin and A.D. Wentzell. *Random perturbations of dynamical systems*. Springer Verlag, 1998.

[GGS03] G. Gottlob, G. Greco, and F. Scarcello. Pure nash equilibria : Hard and easy games. In *Proceedings of the 9th Conference on Theoretical Aspects of Rationality and Knowledge*, pages 215–230. ACM, 2003.

[Gib73] A. Gibbard. Manipulation of voting schemes : a general result. *Econometrica : Journal of the Econometric Society*, pages 587–601, 1973.

[GL07] J. Galtier and A. Laugier. Flow on data network and a positive semidefinite representable delay function. *Journal of Interconnection Networks*, 8 :29–43, 2007.

[GLJ09] Khoriba Ghada, Jie Li, and Yusheng Ji. Cross-layer approach for energy efficient routing in wanets. In *IEEE MASS*, 2009.

[GM90] S.B. Gelfand and S.K. Mitter. Recursive stochastic algorithms for global optimization in ird. In *Proceedings of the 29th IEEE Conference on Decision and Control*, pages 220–221. IEEE, 1990.

[GM91] I. Gilboa and A. Matsui. Social stability and equilibrium. *Econometrica*, 59(3) :859–867, 1991.

[GS62] D. Gale and L.S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1) :9–15, 1962.

[Haj88] Bruce Hajek. Cooling schedules for optimal annealing. *MATHEMATICS OF OPERATIONS RESEARCH*, 13 :311–329, 1988.

[HHKS10] T. Harks, M. Hoefer, M. Klimm, and A. Skopalik. Computing pure Nash and strong equilibria in bottleneck congestion games. *Algorithms–ESA 2010*, pages 29–38, 2010.

[HMC89] S. Hart and A. Mas-Colell. Potential, value, and consistency. *Econometrica : Journal of the Econometric Society*, pages 589–614, 1989.

[Hof85] J. Hofbauer. The selection mutation equation. *Journal of mathematical biology*, 23(1) :41–53, 1985.

[Hof96] Josef Hofbauer. Evolutionary dynamics for bimatrix games : a hamiltonian system ? *Journal of mathematical biology*, 34 :675–688, 1996.

[HOR09] J. Hofbauer, J. Oechssler, and F. Riedel. Brown-von Neumann-Nash dynamics : the continuous strategy case. *Games and Economic Behavior*, 65(2) :406–429, 2009.

[HRBSD03] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda. Performance anomaly of 802.11 b. In *Twenty-Second Annual Joint Conference of the IEEE Computer and Communications.*, volume 2, pages 836–843. IEEE, 2003.

[HS74] M.W. Hirsch and S. Smale. *Differential equations, dynamical systems, and linear algebra*. Academic press, 1974.

[HS98] J. Hofbauer and K. Sigmund. *Evolutionary games and population dynamics*. Cambridge Univ Pr, 1998.

[HS02] J. Hofbauer and W.H. Sandholm. On the global convergence of stochastic fictitious play. *Econometrica*, 70(6) :2265–2294, 2002.

[HS09] J. Hofbauer and W.H. Sandholm. Stable games and their dynamics. *Journal of Economic Theory*, 144(4) :1665–1693, 2009.

[HSV08] J. Hofbauer, S. Sorin, and Y. Viossat. Time average replicator and best reply dynamics. *Mathematics of Operation Research*, 2008.

[Iou07] A. Iouditski. Efficient methods in optimization, 2007.

[Kak41] S. Kakutani. A generalization of Brouwer's fixed point theorem. *Duke Mathematical Journal*, 8(3) :457–459, 1941.

[KHY97] Kushner, J. Harold, and George G. Yin. *Stochastic Approximation Algorithms and Applications*. Springer-Verlag, New-York, 1997.

[KMR93] M. Kandori, G.J. Mailath, and R. Rob. Learning, mutation, and long run equilibria in games. *Econometrica*, 61(1) :29–56, 1993.

[KP98] E. Koutsoupias and C. Papadimitriou. Worst-case equilibria. In *Proc. of STACS*, 1998.

[LG06] J.F. Le Gall. Intégration, probabilités et processus aléatoires. *Ecole Normale Supérieure de Paris*, 2006.

[LPT04] D. Lamberton, G. Pagès, and P. Tarrès. When can the two-armed bandit algorithm be trusted ? *The Annals of Applied Probability*, 14(3) :1424–1454, 2004.

[LPW09] D.A. Levin, Y. Peres, and E.L. Wilmer. *Markov chains and mixing times*. Amer Mathematical Society, 2009.

[LS08] R. Lahkar and W.H. Sandholm. The projection dynamic and the geometry of population games. *Games and Economic Behavior*, 64(2) :565–590, 2008.

[MBML11] P. Mertikopoulos, E.V. Belmega, A.L. Moustakas, and S. Lasaulce. Distributed learning policies for power allocation in multiple access channels. *Arxiv preprint arXiv :1103.3541*, 2011.

[MKA06] Daniele Miorandi, Arzad A. Kherani, and Eitan Altman. A queueing model for HTTP traffic over IEEE 802.11 WLANs. *IEEE Computer Networks*, 50 :63–79, 2006.

[MS96a] D. Monderer and L.S. Shapley. Fictitious play property for games with identical interests. *Journal of Economic Theory*, 68(1) :258–265, 1996.

[MS96b] Dov Monderer and Lloyd S. Shapley. Potential games. *Games and Economic Behavior*, 14 :124–143, 1996.

[MT07] P. Maille and B. Tuffin. Why VCG auctions can hardly be applied to the pricing of inter-domain and ad hoc networks. In *Next Generation Internet Networks, 3rd EuroNGI Conference on*, pages 36–39. IEEE, 2007.

[MVN47] O. Morgenstern and J. Von Neumann. *Theory of games and economic behavior*, volume 3. Princeton University Press, 1947.

[MW98] J. Mo and Jean Walrand. Fair end-to-end window-based congestion control. In *Proc. of SPIE, International Symposium on Voice, Video and Data Communications*, 1998.

[Nas50] J. F. Nash. Equilibrium points in n-person games. *Proceeding of the National Academy of Sciences*, 36 :48–49, 1950.

[Nas51] J. Nash. Non-cooperative games. *The Annals of Mathematics*, 54(2) :286–295, 1951.

[Nis07] N. Nisan. *Algorithmic game theory*. Cambridge Univ Pr, 2007.

[ORS93] A. Orda, R. Rom, and N. Shimkin. Competitive routing in multiuser communication networks. *IEEE/ACM Transactions on Networking (TON)*, 1(5) :510–521, 1993.

[Pem90] R. Pemantle. Nonconvergence to unstable points in urn models and stochastic approximations. *The Annals of Probability*, 18(2) :698–712, 1990.

[Pig52] A.C. Pigou. *The economics of welfare*. Transaction Publishers, 1952.

[Pot06] O. Pottié. Etude des Equilibres de Nash, jeux de Potentiel et Jeux de Congestion. *Hal. inria*, 2006.

[Put05] M. Puterman. *Markov Decision Processes, Discrete Stochastic Dynamic Programming*. Wiley Series in Probability and Statistics, 2005.

[Rob51] J. Robinson. An iterative method of solving a game. *The Annals of Mathematics*, 54(2) :296–301, 1951.

[Roc97] R.T. Rockafellar. *Convex analysis*. Princeton University Press, 1997.

[Ros65] JB Rosen. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica : Journal of the Econometric Society*, pages 520–534, 1965.

[Ros73] R. W. Rosenthal. A class of games possessing pure-strategy Nash equilibria. *Int. J. Game Theory*, 2 :65–67, 1973.

[Saa10] W. Saad. *Coalitional Game Theory for Distributed Cooperation in Next Generation Wireless Networks*. PhD thesis, University of Oslo, 2010.

[SAK07] Srinivas Shakkottai, Eitan Altman, and Anurag Kumar. Multihoming of users to access points in wlans : A population game perspective. *IEEE Journal on Selected Areas in Communication, Special Issue on Non-Cooperative Behavior in Networking*, 25(6) :1207–1215, 2007.

[San01] W. H. Sandholm. Potential games with continuous player sets. *Journal of Economic Theory*, 24 :81–108, 2001.

[San07] W.H. Sandholm. Pigouvian pricing and stochastic evolutionary implementation. *Journal of Economic Theory*, 132(1) :367–382, 2007.

[San10] W.H. Sandholm. *Population games and evolutionary dynamics*. MIT press Cambridge, 2010.

[Sch01] S.J. Schreiber. Urn models, replicator processes, and random genetic drift. *SIAM Journal on Applied Mathematics*, 61(6) :2148–2167, 2001.

[Sha97] LS Shapley. A VALUE FOR n-PERSON GAMES1. *Classics in game theory*, page 69, 1997.

[SPT94] P.S. Sastry, V.V. Phansalkar, and A.L. Thathachar. Decentralized Learning of Nash Equilibria in Multi-person Stochastic Games with Incomplete Information. *IEEE Transactions on System, man, and cybernetics*, 24(5) :769–777, 1994.

[TC06] Chadi Tarhini and Tijani Chahed. System capacity in ofdma-based wimax. In *Proc. of the International Conference on Systems and Networks Communications*, 2006.

[TJ78] P.D. Taylor and L.B. Jonker. Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, 40(1-2) :145–156, 1978.

[TT10] H.H. Tran and B. Tuffin. Inter-domain pricing : challenges and possible approaches. *International Journal of Network Management*, 2010.

[Voc06] B. Vocking. Congestion games : Optimization in competition. In *Proceedings of the 2nd Algorithms and Complexity in Durham Workshop*, pages 9–20. Citeseer, 2006.

[Voo00] M. Voorneveld. Best-response potential games. *Economics Letters*, 66(3) :289–295, 2000.

[War52] J.G. Wardrop. Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers*, 1 :325–378, 1952.

[Wei95] Jorgen W. Weibull. *Evolutionary Game Theory.* MIT Press, 1995.

[YDW07] Faqir Yousaf, Kai Daniel, and Christian Wietfeld. Performance evaluation of ieee 802.16 wimax link with respect to higher layer protocols. In *Proc. of the International Symposium on Wireless Communication Systems*, pages 180–184, 2007.

[You93] H.P. Young. The evolution of conventions. *Econometrica*, 61(1) :57–84, 1993.