



HAL
open science

Modéliser des systèmes complexes : quelques pistes pour relever le défi

Guillaume Deffuant

► To cite this version:

Guillaume Deffuant. Modéliser des systèmes complexes : quelques pistes pour relever le défi. Sciences de l'environnement. Habilitation à Diriger des Recherches, Ecole doctorale SPI de l'Université Blaise Pascal de Clermont-Ferrand, soutenue le 3 mars 2004, 2004. tel-02583104

HAL Id: tel-02583104

<https://hal.inrae.fr/tel-02583104>

Submitted on 14 May 2020

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

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

Mémoire d'Habilitation à Diriger des Recherches

**Modéliser des systèmes complexes : quelques pistes
pour relever le défi**

Guillaume DEFFUANT

Soutenance le 3 Mars 2004

Jury :

M. J. Ferber	Professeur à l'université de Montpellier	Rapporteur
M. A. Quilliot	Professeur à l'université de Clermont-Ferrand	Rapporteur
M. J.P. Nadal	Directeur de Recherche au CNRS	Rapporteur
M. P. Bernard	Professeur à l'université de Clermont-Ferrand	Examineur
M. G. Kahn	Directeur de Recherche à l'INRIA	Examineur
M. P. Mahey	Professeur à l'université de Clermont-Ferrand	Examineur
M. J. Petitot	Directeur de Recherche au CNRS	Examineur

Table des Matières

1	Introduction	7
2	Un regard sur le paysage	11
2.1	Pour une alternative scientifique et idéologique au paradigme du traitement de l'information.....	11
2.2	Allier géométrie et statistiques pour mettre le « traitement de l'information » sur ses pieds.....	16
2.3	Constituer un monde « pour soi », la boucle sensori-motrice	18
2.4	Conscience de soi et empathie	20
2.5	Conclusion	25
3	Des hyperplans séparateurs dans des espaces déployés : les Support Vector Machines ..	27
3.1	La convergence de la stratégie de décomposition (<i>chunking</i>).....	28
3.2	La minimisation récursive de facettes.....	30
3.3	Application aux <i>svm</i>	34
3.4	Diminuer la sensibilité des <i>svm</i> aux points aberrants.....	44
3.5	Conclusion de ce chapitre	49
4	Les réseaux géométriques dynamiques	51
4.1	Vers des organismes mathématiques ?.....	51
4.2	Le dilemme de l'apprentissage	52
4.3	Les fibres adaptatives : des composants de dimension 1	53
4.4	Les membranes de perceptrons	56
4.5	Conclusion de ce chapitre	65
5	Modèles géométriques dynamiques pour la diffusion d'opinions : comparer le modèle individu-centré à son approximation agrégée	67
5.1	Le modèle de confiance limitée (BC)	68
5.2	Le modèle d'accord relatif et la propagation d'opinions extrêmes	72
5.3	Conclusion de ce chapitre	84
6	Projet Européen IMAGES : un modèle individu-centré de propagation de l'innovation .	87
6.1	Le modèle individu-centré de diffusion de l'innovation.....	88
6.2	Analyse du comportement global du modèle.....	95
6.3	Discussion - conclusion	100
7	Conclusion, perspectives	103
7.1	Le projet d'équipe : les modèles individus-centrés comme source d'expérimentation en écologie et sciences sociales.....	103
7.2	Mon projet de recherche personnel :	107
8	Références	109
9	Annexe : le logiciel SICOPTER d'aide à la formulation de problèmes.....	115

Remerciements

J'ai accumulé beaucoup de dettes intellectuelles auprès de nombreuses personnes depuis mes débuts dans la recherche, au point que je ne sais trop comment en organiser l'inventaire. La chronologie est finalement le critère le plus simple.

Au départ, pendant mon doctorat, Paul Bourguine et Francisco Varela, ont eu une influence décisive. Paul m'a communiqué une parcelle de son enthousiasme formidable, de sa passion toujours en éveil pour la création scientifique. L'atmosphère d'ouverture et de créativité intenses qu'il réussissait à insuffler à l'équipe intelligence artificielle du Cemagref reste pour moi une référence. Sans diminuer en aucune manière son talent d'animateur hors du commun, cette atmosphère devait aussi beaucoup à la réunion de personnalités fortes et imaginatives comme Isabelle Alvarez, Thierry Fuhs et Laurent Bochereau qui donnait à l'équipe une qualité exceptionnelle.

Les premières lectures des travaux de Francisco Varela ont été une sorte de révélation pour moi. L'investigation scientifique de territoires nouveaux, aux fondements du vivant et de la cognition s'ouvrait, devenait possible, tout en évitant l'idéologie du traitement de l'information, que j'ai toujours jugée inacceptable. Les discussions récurrentes avec Francisco durant ma thèse, avec la rigueur de l'argumentation qu'elles exigeaient, restent des moments parmi les plus formateurs de ma carrière. J'aimerais lui rendre un profond hommage.

Mais à la réflexion, le véritable point de départ est mon stage de fin d'études de l'X, deux ans avant le début de ma thèse. Les graines qui germeront deux ans plus tard ont notamment été semées lors du colloque de Cerisy sur les sciences cognitives en 1987, auquel ce stage m'a donné la chance d'assister. J'aimerais témoigner ma reconnaissance à Daniel Andler et Joëlle Proust, pour cette expérience inoubliable et déterminante. J'adresse par la même occasion un salut amical et presque fraternel à Eric Davalo et Christophe Tollu, mes compagnons dans cette aventure. Elle a d'ailleurs suscité des lectures et discussions communes pendant une dizaine d'année. Christophe Tollu m'a alors fait découvrir C. Castoriadis et H. Arendt, qui sont maintenant des éléments clefs de ma colonne vertébrale intellectuelle. Je ne saurai jamais assez le remercier pour cela.

Progressivement, je suis ensuite devenu encadrant moi-même, ce qui occasionne autant de dettes intellectuelles, même si elles sont un peu différentes. Merci à Etienne Monneret et Thierry Fuhs pour la confiance, l'énergie et le talent qu'il ont bien voulu consacrer à notre collaboration. Thierry Fuhs a notamment permis d'obtenir des résultats théoriques sur lesquels je butais vainement depuis plusieurs années. Les thèses de Nicolas Meuleau, Rémi Munos et Olivier Higelin n'étaient pas directement inspirées par mes propres recherches, elles m'ont donc permis une ouverture vers de nouveaux horizons. Un grand merci à mes deux derniers thésards, Lambert Rousseau et Frédéric Amblard qui viennent de soutenir, pour tout ce qu'ils m'ont apporté aux plans scientifique et humain. Ils ont permis d'établir parfaitement ce lien si particulier entre thésard et encadrant, fait d'estime réciproque et de respect du rôle de chacun. Enfin je voudrais dire tout le plaisir que j'ai d'encadrer la thèse de Sophie Martin, qui s'annonce tout à fait passionnante.

J'aimerais remercier les membres du LISC pour la confiance qu'ils me témoignent depuis plusieurs années. Sans cette confiance, rien n'est possible. Au delà de cette confiance, l'énergie, l'opiniâtreté et l'efficacité de Sylvie Huet ont été une aide irremplaçable. De même, je mesure bien la chance d'avoir un collaborateur tel que Stéphan Bernard, capable de résoudre tant de problèmes insurmontables pour moi. Merci à Isabelle Alvarez, avec laquelle je collabore depuis plus de 12 ans maintenant. Sa clarté, sa précision, son imagination (et son

amour de la géométrie) sont extrêmement précieux. François Goreaud a choisi le corps du GREF puis le LISC alors que les voies les plus prestigieuses de l'état s'ouvraient à lui. Ses qualités sont un apport incomparable à l'équipe. Thierry Faure, recruté plus récemment, a immédiatement su s'y intégrer, apportant sa culture en physique et en informatique.

Je ne peux pas nommer tous les participants au projet IMAGES, mais je leur adresse à tous mes remerciements les plus chaleureux. Cette collaboration a été particulièrement structurante pour moi. Elle s'est prolongée par des travaux communs avec Gérard Weisbuch, qui a bien voulu nous transmettre un peu de son expérience et de son savoir-faire. J'aimerais lui témoigner ma gratitude.

Enfin, j'adresse mes remerciements les plus profonds aux membres du jury. Je sais combien ils sont sollicités et je sais mesurer l'honneur qu'ils me font en acceptant de me consacrer leur temps.

Il est d'usage d'avoir un mot pour ses proches pour terminer ce genre d'exercice. Une pudeur un peu étrange me retient de me conformer à cette règle. Mais, si les personnes concernées lisent ces lignes, qu'elles se souviennent que l'intensité des sentiments suggérée entre les lignes peut être plus forte que celle d'une expression plus directe mais forcément maladroitement ou convenue dans le présent contexte.

1 Introduction

L'écriture d'un tel document oblige à figer une réflexion qui est encore en devenir, ce qui laisse toujours une forme d'insatisfaction. C'est d'ailleurs cette insatisfaction qui m'a fait reporter plusieurs fois l'achèvement de ce travail. J'imagine que ce sentiment touche tous ceux qui sont confrontés à ce genre d'exercice. Mais il est probable que certaines particularités de mon travail m'y exposent davantage.

En effet, ce travail se situe à la fois dans des disciplines bien constituées et sûres de leur identité comme les mathématiques appliquées et l'informatique, et dans d'autres qui sont beaucoup plus nouvelles, mouvantes et immatures comme les sciences de la cognition, l'intelligence artificielle. Ces nouveaux champs de recherche sont des continents nouveaux dont l'exploration ne fait que commencer. Ils sont extrêmement attirants pour le jeune chercheur en mal d'aventures, et qui n'a pas idée des dangers qui le guettent.

L'un des pièges les plus redoutables est le sable mouvant des questions fondamentales qui parsème ces contrées. En sciences de la cognition ou en intelligence artificielle, ce risque est permanent. Qu'est-ce que l'information ? Qu'est-ce la connaissance ? Ces questions ne sont nullement résolues. Plus grave, elles glissent vite vers d'autres questions encore plus angoissantes : Qu'est-ce que l'objectivité, la subjectivité, la conscience ? Et voilà que notre jeune chercheur en informatique se met à ânonner laborieusement Kant ou Spinoza.

Cependant, qui évitera soigneusement ces régions marécageuses a toutes chances de se perdre dans des déserts, s'épuisant à poursuivre les mirages qui les peuplent. Combien d'effets de mode, d'annonce, d'engouements passagers agitent régulièrement ces communautés ? Généralement, un mot bien choisi et appliqué à une vague métaphore suffit. La recherche, à l'ère du marketing, prospère sur des slogans : réseau neuronal, algorithme génétique, traitement de l'information, vie artificielle, multi-agents... Et voilà notre jeune chercheur, qui, croyant participer à la révolution mondiale et ses lendemains qui chantent, se réveille membre de l'église de scientologie ou de la secte des Raëliens.

Au seuil de la présentation de ce bilan d'une quinzaine d'années d'activité dans ces domaines, il me faut bien l'annoncer d'emblée : Je suis tombé dans les deux pièges. Oui, je me suis passionné pour les réseaux connexionnistes, les algorithmes génétiques, les modèles multi-agents, l'apprentissage par renforcement, les fractals, le chaos, la complexité et bien d'autres choses encore. Oui, j'ai passé de longues heures avec les classiques de la tradition philosophique, les moins classiques aussi d'ailleurs. C'est aussi la pleine conscience, déjà ancienne, de cette dépense d'énergie excessive au regard des critères académiques qui suscite mon insatisfaction. Je plaide donc doublement coupable. Mais j'aimerais faire valoir quelques circonstances atténuantes.

D'abord, avouons-le, l'insatisfaction est mêlée à un grand plaisir à parcourir ces contrées inhospitalières. Les dangers qu'elles recèlent donnent son goût âpre à l'aventure intellectuelle.

On ne pratique pas pendant des années des auteurs comme Castoriadis, Dupuy ou Varela sans être contaminé par une soif de fondements, une volonté de dépasser le morcellement des disciplines scientifiques comme seul horizon de sens. Et puis, pouvais-je vraiment faire autrement ? C'était bien la règle du jeu que de se lancer à corps perdu dans ces explorations.

Par ailleurs, je pense avoir rapidement élaboré des stratégies pour combattre les pièges. Je ne me suis pas contenté de m'agiter en tous sens dans les marécages. J'ai tenté, comme d'autres avant moi, d'utiliser des matériaux solides importés des contrées civilisées, pour planter des piliers et couler des fondations. Certains mélanges entre ces matériaux, comme les statistiques et la géométrie, sont encore peu utilisés, alors qu'ils me semblent particulièrement efficaces sur toute une partie du terrain. Cette même méthode m'a permis de repérer plus rapidement les mirages, en donnant des points de repère solides.

D'autre part, cette réflexion sur les fondements a toujours été menée en parallèle à des recherches sur des questions concrètes, parfois très appliquées. La plupart ne sont pas mentionnées dans ce document, car je tenais à lui conserver une certaine unité, alors que ces travaux appliqués s'éparpillent dans de nombreuses directions (guidage d'engins, propagation du risque de contamination dans les abattoirs, classification de zones humides...). Cette préoccupation pour les applications concrètes a souvent été un antidote utile aux vertiges des questions fondamentales. Elle m'a permis aussi de tester le véritable potentiel des nouveautés récurrentes. La plupart des travaux plus spécialisés que je présente portent tout de même la marque de la double exigence de cadres fondamentaux solides et de potentiel d'application concrète.

Enfin, ces errances et ces explorations ont bénéficié de multiples aides, collaborations, discussions avec d'autres chercheurs, avec les doctorants. La discussion critique bienveillante, la controverse argumentée m'ont protégé de bien des pièges.

A la réflexion, mon insatisfaction est peut-être finalement davantage due à l'ampleur de la tâche qui reste à accomplir qu'aux erreurs ou aux difficultés rencontrées.

J'ai donc choisi de commencer ce document par un regard rapide sur le paysage des questions fondamentales, même si cette réflexion est loin d'être suffisamment mûre pour constituer un exposé scientifique convaincant. Il m'a semblé impossible de la passer sous silence car elle explique beaucoup des choix et des orientations du reste de mon travail. Après être revenu rapidement sur le problème de l'information, je propose de distinguer trois niveaux de cognition et quelques pistes pour les modéliser et les comprendre. En particulier, au premier niveau, je plaide pour l'utilisation de concepts issus des statistiques et de la géométrie.

Le reste du document présente des travaux plus techniques, que l'on peut rattacher principalement au premier niveau de cognition défini au chapitre 2.

Le chapitre 3 est consacré à un premier modèle mêlant statistiques et géométrie : les « support vector machines » (*svm*). Ces modèles, dont le principe est très simple, sont très efficaces notamment en discrimination statistique. Je propose un algorithme d'optimisation particulier de ces machines, qui permet en particulier de moduler le poids accordé aux points aberrants et ainsi améliorer la robustesse de la méthode.

Le chapitre 4 présente un ensemble plus large de modèles : les réseaux géométriques dynamiques. Définis à partir de composants géométriques simples, soumis à des dynamiques individuelles, ces modèles construisent une forme globale plus ou moins stable. Nous proposons plusieurs variantes de ces modèles, dont les propriétés sont intéressantes pour résoudre des problèmes statistiques.

Le chapitre 5 présente une autre famille de réseaux géométriques dynamiques, dont l'objectif est de simuler certaines dynamiques d'évolution d'opinions. Nous étudions le comportement global de populations de ces composants dans différentes conditions. Nous identifions des comportements typiques, qui trouvent des interprétations sociales pertinentes.

Le chapitre 6 décrit un modèle individu-centré de diffusion de l'innovation, qui s'appuie sur le précédent. Ce modèle est issu du travail d'équipe mené dans le cadre du projet Européen IMAGES pour modéliser la diffusion de mesures agri-environnementales dans les populations d'agriculteurs. Il comprend, outre les dynamiques d'influence sociale, des dynamiques de propagation d'informations. Nous montrons en explorant l'espace des paramètres que le modèle rend compte de différentes situations typiques de diffusion de l'innovation.

Enfin, le dernier chapitre dresse un bilan partiel de ces travaux et présente mes orientations de travail pour l'avenir, dans le cadre de celles de mon équipe.

L'annexe présente le logiciel SICOPTER d'aide à la décision de groupe en gestion du territoire, issu de la thèse de L. Rousseau et du projet Européen FIRMA.

2 Un regard sur le paysage

La mode est actuellement au langage de « l'information ». On nous dit que le vivant cueille des informations dans la nature et les traite de diverses façons. Ce langage doit être impitoyablement condamné. On a jamais vu des informations pousser dans les champs au printemps ou à l'automne. Le vivant crée de l'information pour lui. Rien n'est information que pour un soi – qui peut transformer, ou ne pas transformer, le X du choc extérieur en information. Les ondes radio n'offrent aucune information aux vivants terrestres, et le théorème de Weirstrass-Stone aucune information à ma boulangère (..).

C. Castoriadis. Les carrefours du labyrinthe V. p. 260.

Ce chapitre expose brièvement le cadre général dans lequel s'inscrivent les travaux plus spécialisés présentés dans les chapitres suivants, et aborde des questions générales sur l'information et la cognition. Cet intérêt pour les questions fondamentales remonte au début de ma thèse. Il est au départ stimulé par une forte défiance vis à vis de l'idéologie ambiante du « traitement de l'information », partagée par mon directeur de thèse, F. Varela. Puis ma réflexion sur ces questions prend progressivement un caractère plus constructif, proposant des outils et des cadres conceptuels alternatifs. Je remets tous les ans cette réflexion sur le métier dans le cadre du séminaire sur les sciences de la cognition que je donne depuis 8 ans à l'ENSAE.

Je reviens tout d'abord sur les difficultés posées par le « traitement de l'information », puis je retrace les trois jalons principaux d'une vision alternative : l'apport de concepts venant des statistiques et de la géométrie, le rôle particulier de l'action dans une boucle « sensori-motrice », la spécificité des dynamiques engendrées par un opérateur d'empathie, comme approche privilégiée pour comprendre la complexité humaine. Les deux derniers paragraphes reprennent partiellement certaines publications (Deffuant 1997, 1998, 2004). Je développe peu les arguments, car mon objectif est d'informer le lecteur du contexte de mon travail, plutôt que de tenter de le convaincre d'adopter mes positions.

2.1 Pour une alternative scientifique et idéologique au paradigme du traitement de l'information

L'hégémonie du paradigme du traitement de l'information m'apparaît au mieux comme un aveuglement idéologique collectif, au pire comme une vaste malhonnêteté intellectuelle. Bien que très minoritaire, cette position n'est pas originale (elle est très marquée par l'influence de Varela et Castoriadis notamment). Cependant, les enjeux idéologiques de ces débats scientifiques me semblent souvent dangereusement sous-estimés, voire ignorés, et ils

motivent une grande partie de mon activité de recherche. J'essaie de me limiter à l'essentiel, en renvoyant à des argumentaires souvent développés ailleurs.

2.1.1 La « théorie de l'information » ne parle pas d'information

Dans son livre « la redécouverte de l'esprit » (Searle 1992), Searle considère les spécialistes modernes de la philosophie de l'esprit comme atteints collectivement d'une névrose obsessionnelle qui les ramène constamment dans la même impasse. Selon lui, la cause de cette névrose est le choix initial de développer une théorie de l'esprit qui exclut complètement la conscience. Cette exclusion initiale impose des contorsions intellectuelles de plus en plus désespérées pour éviter la notion de conscience dans la description du fonctionnement de l'esprit. Mais la conscience est selon lui un phénomène tellement central et incontournable de la cognition, que toute tentative de le contourner est vouée à l'échec. On assiste donc toujours au même scénario : les théories de la cognition tentent de se définir en faisant abstraction totale de la conscience, mais il s'avère ensuite que cette option initiale les mène à des contradictions, ce qui suscite des modifications subtiles et sophistiquées de la théorie, menant à d'autres contradictions, et cela indéfiniment. Et Searle de se désoler de ce vaste gaspillage d'intelligence et de talents.

Un syndrome très similaire semble identifiable dans l'activité de recherche autour de la théorie de l'information en sciences de la cognition¹. En effet, en choisissant dès l'origine de faire abstraction du sens des signes qu'elle étudie, la théorie de l'information se condamne inéluctablement à s'écarter radicalement de l'acception commune du terme information. Dans le langage courant en effet, une information désigne avant tout le sens porté par les signes (ou les formes), plutôt que les signes eux-mêmes. L'information, c'est le sens. La théorie de l'information fait donc le choix de laisser à part ce qui est au cœur du concept auquel elle s'intéresse. On assiste donc à des contorsions conceptuelles un peu similaires à celles dont Searle se plaint à propos de la conscience dans la philosophie de l'esprit². Il s'agit de parler de l'information, sans parler du sens tout en en parlant quand même. Ainsi, par exemple le terme de « quantité d'information » est bien caractéristique de cette tendance. Comme si quantifier un objet autorisait à le dénaturer totalement ! Le fait de compter des pommes de terre les transforme-t-il en carottes ? Ce caractère insatisfaisant apparaît d'ailleurs rapidement puisqu'une communauté de chercheurs essaie de quantifier la pertinence des informations. Mais on cherche à nouveau à le faire en évacuant a priori l'idée de sens et en se fondant sur la théorie de l'information. Sperber le fait en déployant des trésors d'intelligence et d'ingéniosité. Mais comment ne pas voir que cette technicité ne fait que cacher une vaine fuite en avant, dont la cause vient d'une position de principe évacuant le sens à l'origine ? Dans le langage courant, le sens est tout autant constitutif de la pertinence qu'il l'est de l'information. Il est donc à craindre que tous ces magnifiques talents soient déployés en pure perte.

Dans cette perspective, la dénomination « théorie de l'information » apparaît comme usurpée, et confine même à l'imposture intellectuelle. En effet, cette théorie ne peut en aucun cas se targuer d'avoir donné un contenu scientifique précis à un concept courant, comme par exemple les travaux sur les systèmes formels et les machines de Turing ont permis de préciser la notion de calculabilité. La dénomination « théorie de la communication », utilisée par Shannon n'apparaît pas plus heureuse. Il s'agit en fait d'une théorie permettant d'optimiser la transmission de signaux discrets, en excluant également a priori la question du sens des

¹ Notre jugement sur l'application de la théorie de l'information en physique est totalement différent (voir plus loin).

² Ce parallèle n'est pas du tout fortuit. Le syndrome de la théorie de l'information et celui de la philosophie de l'esprit contemporaine puisent évidemment leurs causes aux mêmes sources idéologiques.

signes. Or, dans l'acception courante, la communication humaine concerne avant tout le sens, les signes n'étant qu'un moyen. Quelle désolation de voir de doctes consultants en communication se référer pompeusement à Shannon pour donner un vernis scientifique à un discours qui porte principalement sur ce qui est exclu de cette théorie ! Constaté que ces dérives se perpétuent, et probablement s'amplifient depuis des décennies est peut-être encore plus déprimant que le tableau que fait Searle de la philosophie de l'esprit contemporaine. Car il s'agit d'un phénomène à la fois plus ample et plus profond, le syndrome de la philosophie de l'esprit n'en étant qu'une manifestation parmi d'autres.

En effet, la théorie de l'information a permis l'émergence du paradigme du « traitement de l'information », qui s'étend naturellement aux sciences de l'homme. Ainsi elle a quitté les problèmes de transmission télégraphique, de statistiques ou de physique dans lesquels elle pouvait encore trouver une interprétation précise, qui limitait les effets de l'inadéquation à son objet. De plus, le paradigme du traitement de l'information présente dans notre perspective des difficultés supplémentaires.

2.1.2 Le traitement de l'information marche sur la tête

Nous ne pouvons pas nous attarder sur l'analyse du succès fulgurant de l'idée de « traitement de l'information ». Signalons simplement, en suivant H. Gardner (Gardner 1985), une conjonction de travaux en linguistique (Chomsky), psychologie (Miller) et informatique (Simon) à la fin des années 50, qui lui a donné une crédibilité apparente. Ces travaux sont incontestablement d'une grande qualité et d'une grande créativité. L'engouement pour les nouvelles pistes de recherche qu'ils ouvraient est parfaitement compréhensible. En simplifiant, on peut dire que la théorie de l'information s'est associée au « linguistic turn », lui-même associé aux progrès décisifs de la logique formelle au 20^{ème} siècle, pour produire cette machine de guerre conceptuelle redoutable qu'est le « traitement de l'information ». Cette association nous semble reposer sur l'ambiguïté de la théorie de l'information, qui légitime l'idée d'un calcul sur des éléments de sens, et la logique formelle qui identifie calcul et langage. Ainsi, le « traitement de l'information » sensorielle et le « traitement de l'information » symbolique langagière peuvent-ils faire appel aux mêmes mécanismes dans ce cadre. Toute activité cognitive est finalement linguistique. Ce pari est tout à fait explicite dans le livre de Fodor « le langage de la pensée » (Fodor 1975), et il imprègne encore en profondeur des travaux plus récents très influents (Pinker 2000).

Cependant, outre la référence douteuse à la théorie de l'information, ce paradigme nous semble source de graves confusions. Pour résumer, d'une part il étend excessivement le rôle du langage et rend ainsi incompréhensibles les activités dont le langage est absent, et d'autre part il mutile irrémédiablement le langage en faisant abstraction de sa nature sociale. Mentionnons quelques jalons de cette critique.

Le connexionnisme nous semble être la voix la plus forte du premier versant de cette critique. Il postule l'existence de mécanismes parallèles, non syntaxiques et non symboliques, qui se veulent plus proches d'une inspiration biologique. C'est d'ailleurs le manque de réalisme biologique des théories du traitement de l'information pure et dure qui est l'un de leurs arguments les plus forts. Il reste que, selon nous, le connexionnisme n'a fait que la moitié du chemin, car il n'a pas remis en cause l'idée de « traitement de l'information ». Il en résulte une certaine confusion qui souligne le manque d'un cadre conceptuel plus approprié.

L'autre versant de la critique attaque l'absence de composante sociale du traitement de l'information qui ramène l'activité cognitive à des calculs de conséquences, essentiellement individuels. Cette vision entre en opposition avec certaines théories du langage notamment la théorie des « actes de langage », qui, au contraire pose la relation à l'autre comme centrale

dans l'activité linguistique (Austin 1962) (pour un développement plus complet de cette critique, voir (Winograd et Florès 1986)).

L'échec relatif du premier courant de l'intelligence artificielle, qui a dû constamment revoir ses ambitions à la baisse, fait écho à cette double critique. En effet, encore aujourd'hui, des tâches usuelles comme comprendre une langue naturelle ou conduire une automobile en ville semblent très largement hors de portée des systèmes artificiels actuels. Les recherches en intelligence artificielle ont depuis longtemps consacré l'inadéquation de l'orthodoxie inspirée du logico-empirisme en privilégiant souvent des assemblages de solutions pragmatiques (cadres, objets, scénarios, connaissances théoriques du domaine) dans les systèmes à base de connaissances.

A ce stade, il pourrait apparaître que nous nous acharnons à tirer sur une ambulance, voire sur un corbillard. Ce serait négliger le puissant courant idéologique que le traitement de l'information à la fois traduit et renforce. Cette collusion idéologique explique à notre avis la persistance et même à certains égards le renforcement du paradigme du traitement de l'information, malgré ses faiblesses et ses contradictions bien identifiées. Qui ambitionne de combattre ce phénomène a tout intérêt à bien en prendre la mesure.

2.1.3 La toute puissance de l'idéologie sous-jacente est néfaste

H. Arendt (Arendt 1961) disait : « ce qu'il y a de fâcheux dans les théories modernes du comportement, ce n'est pas qu'elles sont fausses, c'est qu'elles peuvent devenir vraies, c'est qu'elles sont, en fait, la meilleure mise en concept possible de certaines tendances évidentes de l'époque moderne ». Cette phrase visait plus particulièrement le behaviorisme et les théories utilitaristes en économie, mais elle pourrait tout autant être appliquée au paradigme du traitement de l'information, qui leur est fortement apparenté. Peu importe que ce paradigme soit incohérent et faux, car il traduit et renforce des partis pris idéologiques modernes, qui ont presque acquis la force de l'évidence. Essayons de mieux les identifier.

Dans « le jardin imparfait » (Todorov 1998), Todorov distingue l'humanisme de deux autres familles de pensée moderne : le scientisme et l'individualisme³. Malgré la lourdeur un peu désagréable de ces mots en « isme », cette classification nous semble très éclairante sur les enjeux idéologiques qui nous intéressent.

En effet, le paradigme du traitement de l'information nous semble aménager une alliance entre scientisme et individualisme, qui renforce ces deux familles et leur assure une position hégémonique au détriment de l'humanisme. En retour, le traitement de l'information bénéficie du double soutien inconditionnel de ces familles idéologiques, qui lui savent gré des services rendus. Cette alliance s'appuie selon nous sur deux points fondamentaux : la possibilité de déduire des lois globales de mécanismes individuels de traitement de l'information, et la déduction des valeurs à partir des faits.

En effet, le scientisme privilégie des lois globales, économiques ou biologiques au détriment de la liberté individuelle, qui lui apparaît comme une illusion. Il s'oppose en cela à l'individualisme, qui considère au contraire la liberté individuelle comme première. Certains modèles économiques et biologiques permettent de lever cette contradiction, en associant liberté, égoïsme individuel et bénéfices collectifs. Nous ne pouvons revenir sur l'histoire passionnante de cette idée typiquement moderne (voir (Dumont 1983; Renaut 1989)), dont la matrice serait la monadologie de Leibniz. Les mécanismes apparentés au traitement de l'information donnent une crédibilité renouvelée à ce schéma. En effet, l'individu de

³ Nous laissons de côté la famille des conservateurs identifiée comme faisant partie des modernes par Todorov, mais qui, finalement, s'oppose à la modernité.

l'économie libérale optimise son intérêt individuel, et ce faisant optimise l'utilité collective (la « main invisible » d'A. Smith). De même, l'individu biologique cherche à optimiser ses chances d'avoir une descendance et participe ainsi au grand jeu de l'évolution, qui permet la création d'espèces plus adaptées. De plus, avec l'idée de « traitement de l'information » génétique, le paradigme s'inscrit au cœur d'une des avancées scientifiques majeures du 20^{ième} siècle, et permet à un autre niveau la jonction entre dynamiques individuelles et collectives. Le traitement de l'information individuelle s'accorde aux principes de l'individualisme, et les lois globales engendrées par ces comportements individuels sont caractéristiques du scientisme.

L'autre fondement de l'alliance entre les deux familles est moins souvent souligné, me semble-t-il, mais il est peut-être encore plus profond. L'une des caractéristiques du scientisme que de penser que les valeurs peuvent se déduire des faits. La démarche scientifique doit permettre à l'homme de se construire des valeurs rationnelles, conformes à la vérité. Or, la théorie de l'information amalgame les faits et les valeurs en une même quantité, et semble ainsi rendre cette distinction non pertinente. Certes, la distinction réapparaît dans les théories utilitaristes, mais le calcul de la valeur, monétaire le plus souvent, d'un fait est un traitement de l'information comme un autre, exécuté au niveau strictement individuel, et la fonction d'utilité est au mieux la somme des intérêts individuels d'un groupe. Le statut des valeurs n'est donc pas très différent de celui des faits. Le paradigme du traitement de l'information contribue donc au discrédit d'une distinction fondamentale entre faits et valeurs, tant au niveau individuel que collectif, et permet ainsi une nouvelle jonction entre scientisme et individualisme.

La position hégémonique de l'alliance entre individualisme et scientisme, lui permet de prétendre couvrir l'ensemble du champ, de l'individuel au collectif, et reléguer la troisième famille, l'humanisme, au rang de pauvres vieillards indigents perdus dans les chimères d'un autre temps. Mais les exemples de paradigmes triomphants, qui s'écroulent tout-à-coup comme des châteaux de cartes, ne manquent pas. Il est permis d'espérer que le « paradigme informationnel » du 20^{ième} siècle figure en bonne place avec le behaviorisme dans le bestiaire des chimères conceptuelles qu'établiront les historiens des sciences de la fin du 21^{ième}, et que cela bouleverse la fortune des familles idéologiques. Contribuer, même de la manière la plus minime, à un tel renversement est une des motivations les plus profondes de mon activité.

Je juge l'hégémonie actuelle des familles individualistes et scientistes invraisemblablement excessive et potentiellement dangereuse. Selon moi, sa vision de l'homme est mutilée, atrophiée, en particulier car elle ignore la découverte fondamentale des humanistes qui est que le « je » et le « tu » naissent en même temps et sont indissociables, et remet en cause cet acquis formidable de la modernité qu'est la distinction entre faits et valeurs. Elle maintient l'intersubjectivité hors du champ scientifique, et dans un monde dominé par la pensée scientifique produit des hommes étrangers à eux-mêmes.

Le renversement que j'appelle de mes vœux ne peut advenir que sous l'influence d'un cadre conceptuel rigoureux, aux fondements scientifiques solides qui parvienne à traduire la pensée humaniste. Car, dépourvue de cet appui, son combat contre les autres familles, qui bénéficient de l'allure formelle et quantitative du traitement de l'information, sera toujours celui du pot de terre contre le pot de fer. Je m'attache donc depuis longtemps à esquisser les contours possibles d'un tel cadre.

2.2 Allier géométrie et statistiques pour mettre le « traitement de l'information » sur ses pieds

2.2.1 Les statistiques distinguent l'estimateur et les données

Crier « le sens, le sens ! » ou « l'intersubjectivité, l'intersubjectivité ! » en sautant comme un cabri, est bien dérisoire. Cette attitude comporte en outre le risque de fournir des arguments en faveur d'une dérive nihiliste et anti-scientifique, que nous nous empressons de condamner avec la dernière énergie. Il faut reconnaître au paradigme du traitement de l'information la volonté de produire des descriptions précises et rigoureuses des phénomènes cognitifs, et d'être le seul actuellement à pouvoir prétendre le faire avec une certaine cohérence. D'ailleurs cette allure scientifique est une des explications de son succès, notamment car elle impressionnait les financeurs, ce que les chercheurs ont exploité parfois abusivement (Segal 2003).

Le défi est donc de contribuer à l'émergence d'une alternative crédible. La première étape est selon nous de mettre le « traitement de l'information » sur ses pieds. Nous entendons par là introduire un cadre conceptuel mathématique qui rend compte de la cognition non linguistique. Sur ce point, il nous semble qu'une alternative possible est actuellement en gestation. Nous tentons d'en identifier certains contours, ce qui n'est pas toujours aisé, car les composantes de cette alternative sont à notre avis un peu dispersées.

Tout d'abord, comme Castoriadis et beaucoup d'autres le soulignent, il est indispensable de considérer le couple système / environnement. Se focaliser uniquement sur l'un ou sur l'autre est voué à l'échec. Une information n'existe pas en elle-même. Ne revenons pas sur cette évidence. Il s'agit maintenant de l'exprimer plus formellement.

Signalons que l'utilisation de l'information de Fisher en physique intègre bien ce couple en considérant l'interaction entre le phénomène mesuré et le dispositif physique utilisé pour faire la mesure. En cela, la démarche proposée nous semble moins critiquable que celle de Shannon, sans remettre en cause la critique de fond exprimée au paragraphe précédent⁴.

Mais il nous semble plus clair encore de considérer le cadre mathématique dont est issu initialement l'information de Fisher : l'estimation statistique, et notamment les travaux sur l'apprentissage artificiel qui s'y rattachent. En effet, ce cadre formalise l'influence de données, distribuées dans un espace quelconque, sur un objet qui peut être modifié par ces données. Des distributions de points dans un espace nous semblent la meilleure manière d'aborder mathématiquement la « diversité sensorielle », les chocs X dont parle Castoriadis. Ainsi, le cadre de l'estimation statistique présente l'avantage d'étudier des cas très simples d'évolution d'un système sous l'influence de « chocs » (l'arrivée des données). Plus profondément, ces systèmes permettent en général de résumer, ou d'extraire une sorte d'ordre de ces données. Le traitement de l'information est parfois qualitativement associé à cette « mise en ordre », ce qui ferait de l'estimation statistique le prototype du traitement de l'information.

Cette opération de mise en ordre, de structuration, issue du couplage entre l'estimateur statistique et les données nous semble effectivement rendre compte d'un aspect fondamental de la cognition, touchant à la fois à la mémoire, la reconnaissance et la capacité d'anticipation. Ces différentes facultés reposent en effet sur une structuration du « divers

⁴ Cette critique ne vise pas, évidemment, la pertinence des travaux en physique issus du principe d'extremum d'information, travaux très impressionnants pour celui qui, comme moi, ne possède qu'une culture générale moyenne en physique. La critique porte sur le choix du terme information qui nous semble impropre.

empirique », que les statistiques nous permettraient de formaliser. Cette formalisation nous apprend notamment que cette structuration dépend à la fois des propriétés de l'estimateur, et des caractéristiques de la distribution à laquelle il est confronté. La distribution des données doit présenter certaines propriétés, elle doit être « organisable », d'une manière compatible avec les propriétés de l'estimateur. Tout cela est philosophiquement confortable.

Il nous semble que cette vue n'est cependant pas entièrement satisfaisante. En effet, la question du sens que nous mettons au centre de notre critique de la théorie de l'information reste difficile à exprimer dans le cadre des statistiques. Le modèle statistique construit-il du sens pour lui-même ? On est plutôt tenté de dire qu'il en construit pour son utilisateur humain, ce qui nous ramène aux mystères de la cognition humaine. En fait le modèle statistique semble atteindre ici ses limites, et il faut appeler d'autres concepts à la rescousse.

2.2.2 Mathématiques des formes, distinction entre saillance et prégnance

Les sciences modernes se sont développées en opérant une distinction entre faits et valeurs, et en se focalisant strictement sur les faits. La théorie de l'information se trouve selon nous prise dans une contradiction liée à cette distinction. En effet, en tant que théorie scientifique, elle se doit d'exclure les valeurs de son discours, c'est pourquoi elle élimine le sens, car sens et valeur ont partie liée. D'ailleurs ce problème de la valeur d'une information est signalé par de nombreux auteurs, notamment Delahaye (Delahaye 1999), qui l'utilise pour mettre en évidence l'inadéquation de la notion, mais aussi Brilloin, qui cherche à établir une équivalence entre les quantités d'information et une valeur monétaire. Mais elle aboutit en fait à une théorie qui, dans le meilleur des cas confond structuration de données (et encore dans un sens extrêmement faible) et attribution de valeur. On reconnaît là une caractéristique de la famille scientifique, dont nous souhaitons profondément nous affranchir.

Le défi est donc de proposer un cadre rigoureux permettant d'appréhender cette distinction entre faits et valeurs. Les concepts de saillance et prégnance, proposées par R. Thom, nous semblent des candidats particulièrement intéressants pour remplir ce rôle (Thom 1988). Nous rapprochons la saillance, fondée sur l'organisation d'un ensemble de singularités, de la structuration de données. Mais dans notre cadre, la forme est extraite statistiquement d'une distribution de points dans un espace. Elle dépend donc à la fois du phénomène considéré et des caractéristiques du système qui l'organise : par exemple l'échelle d'observation ou les capacités de détection influenceront l'organisation des données. Nous rapprochons la prégnance d'une valeur associée à la forme par un processus biologique (reconnaissance innée de proies, de partenaires ou de prédateurs) ou par un processus social (apprentissage social du goût et des convenances par exemple). Les valeurs peuvent donc évoluer au cours du temps, indépendamment de la forme à laquelle elles s'attachent. Par exemple, une bottine de femme peut acquérir une très forte intensité sexuelle pour certaines personnes (exemple de prédilection de Castoriadis), ou encore prononcer le nom d'un dictateur peut déclencher l'effroi des individus qui en subissent le joug.

Conservé à l'esprit ce couple saillance / prégnance nous semble être un repère particulièrement utile dans toutes ces discussions. En effet, la saillance reste par définition neutre de valeur, elle est seulement le résultat d'une activité de structuration de données. Ce terme de saillance serait donc beaucoup plus approprié que celui d'information pour rendre compte de la théorie de l'information, qui ne serait, selon nous, que l'amorce d'un discours sur la saillance. En effet, la « quantité d'information » portée par une distribution de probabilité est élevée lorsque cette distribution est concentrée en un point de l'espace, et elle est au contraire faible lorsqu'elle est diffuse dans un large volume. Ainsi, la quantité d'information d'une distribution gaussienne est-elle l'inverse de son écart-type. N'est-ce pas

là le cas le plus simple de saillance, correspondant à une singularité portée par un point de l'espace ?

Mais comment une valeur, une prégnance, s'associe-t-elle à une saillance ? La difficulté de cette question est inaperçue dans le paradigme du traitement de l'information. Or, elle implique un changement de perspective dans la pratique scientifique habituelle, car, comme le défend notamment Castoriadis, cette attribution de valeur ne peut être comprise qu'à l'intérieur d'un monde « pour soi », créé par le système étudié. Il nous faut mobiliser des concepts différents pour aborder cette nouvelle difficulté.

2.3 Constituer un monde « pour soi », la boucle sensori-motrice

Dans (Deffuant 1997), je tente de montrer que la notion d'un monde pour soi permet de lever des contradictions apparentes entre les visions de l'autonomie de Castoriadis et Varela. La thèse soutenue est que, en particulier Varela ne distingue pas suffisamment clairement le point de vue scientifique global sur le système et son environnement, du point de vue complètement différent du système sur cet environnement, c'est à dire son monde pour soi. Parfois, en effet, il passe d'un point de vue à l'autre sans le préciser, ce qui rend l'interprétation très difficile. Conserver explicitement à la fois la vision globale de l'estimateur et des données, et celle de l'estimateur lui-même, nous semble donc indispensable.

La théorie de l'information essaie au contraire de faire croire que qu'il est possible d'en faire l'économie, et c'est là l'un de ses effets les plus pernicioeux. En effet, établir cette double vision, dans une démarche scientifique, est un chantier immense, et beaucoup des talents et des efforts déployés dans des directions stériles actuellement, y seraient à mon sens beaucoup mieux employés. Mais l'enjeu fait peur, légitimement sans doute. Car les pratiques scientifiques se sont construites sur l'éviction de la subjectivité, et il s'agit bien de la faire revenir au centre de la scène. Cette question fait bien sûr l'objet de nombreux débats, et depuis longtemps. Je ne peux ici qu'en mentionner très succinctement quelques aspects, qui retiennent plus particulièrement mon attention.

Tout d'abord, quelles sont les dynamiques minimales permettant l'émergence d'un monde « pour-soi » ? Cette question est évidemment très délicate, puisque ce monde est extrêmement difficile à imaginer pour nous. Varela semble considérer que les dynamiques de l'autopoïèse sont une condition suffisante à cette constitution (Varela 1989). En effet, l'environnement prend un sens en fonction de son potentiel bénéfique ou menaçant le maintien de l'unité. On peut se demander, en général, dans quelles conditions un système adaptatif complexe (Gell-Mann 1997), se constitue un monde pour soi. Cette question me semble importante et largement ignorée.

Est-il possible de distinguer différentes catégories de monde pour soi, associées aux dynamiques qui en permettent la création ? J'ai proposé (Deffuant 1997; Deffuant 1998) quelques pistes pour répondre à cette question. Cette catégorisation est en fait très proche de celle de (Edelman 1992), et peut avoir certains points communs avec celle de Damasio (Damasio 1999). Cependant, il me semble que Damasio et Edelman ne prennent pas suffisamment en considération les dynamiques sensori-motrices, dont le rôle est mis en évidence de manière éclatante par les expériences de Bach y Rita (Bach y Rita 1994; Bach y Rita 1995).

Rappelons brièvement ces expériences et leurs résultats pour éclaircir ce point. Bach y Rita utilise un dispositif technique qui relie les pixels d'une caméra numérique à des effecteurs disposés sur la peau d'un sujet (sur le ventre ou dans le dos par exemple). Selon le niveau de gris du pixel, l'effecteur varie l'intensité de l'impulsion sur la peau. Ainsi, la peau sert de

rétine visuelle. Dans notre perspective, le résultat le plus remarquable de ces expériences est le suivant :

- lorsque la caméra reste fixe, les sujets peuvent discriminer des formes, mais ils ressentent les impulsions sur leur peau, et ne distinguent pas vraiment leur cause externe, ces sensations pourraient être causées de l'intérieur comme des démangeaisons.
- En revanche, lorsque les sujets peuvent bouger la caméra, faire des zooms, etc, un changement fondamental se produit : ils perçoivent les objets dans un espace extérieur.

Ce résultat nous semble tout à fait remarquable, et trop peu souvent souligné (et toute l'équipe de l'université technologique de Compiègne qui s'intéresse plus particulièrement à ce point, notamment C. Lenay, n'en a à nos yeux que plus de mérite). En effet, cette expérience montre le passage d'un monde pour soi d'une monde sans espace extérieur à celui d'un être sensori-moteur percevant un monde externe. Il apparaît donc que les boucles sensori-motrices, en apprenant à anticiper les modifications des perceptions induites par les actions, permettent de constituer la sensation d'un espace externe.

Le cadre morphodynamique présenté par J. Petitot (Petitot 1992) nous semble particulièrement adapté pour rendre compte de cette transformation. En effet, il permet de faire apparaître perceptions et actions dans un même espace, et d'en identifier les saillances et les régularités. Ainsi, la création d'un monde extérieur « pour-soi » pourrait-elle se formaliser. La différence entre saillance et prégnance reste d'ailleurs tout aussi pertinente dans ce cadre. En effet, les aveugles ont une certaine réticence à utiliser le dispositif, alors qu'il leur rend des services pratiques immédiats (Lenay 1997). Cette réticence s'expliquerait par une difficulté particulière à associer les prégnances de la perception visuelle normale. Il en résulterait une forme de malaise.

Cependant ce formalisme de la morphodynamique ne nous semble pas élucider complètement le bouclage entre action et anticipation de sensation. En effet, il ne rend pas compte du processus d'apprentissage spontané qui associe les modifications des sensations de la grille aux actions sur la caméra. Le sujet déploie spontanément une exploration de ces sensations en choisissant des actions, et en anticipant les modifications des sensations que ces actions vont provoquer. Comment formaliser ce processus itératif ?

Le domaine de l'apprentissage par renforcement, et la programmation dynamique adaptative nous semblent fournir des concepts et des formalismes prometteurs. En effet, ces formalismes s'attaquent à cette question du choix de l'action en fonction de l'anticipation. Notons en outre qu'ils pratiquent une distinction explicite entre la saillance et la prégnance (qui est la récompense associée à certaines situations).

Il nous semble donc que des pistes sérieuses existent pour construire un discours scientifiquement acceptable sur la constitution de certains types de mondes pour soi. Il est clair que les conclusions de ce discours seront en général indirectes. Par exemple : il est possible d'expliquer la perception d'un monde externe par l'apprentissage de régularités entre des actions sur le monde et leurs effets sur les sensations, car ces régularités peuvent expliquer l'impression d'espace. Si l'on accepte cette explication, les systèmes dépourvus de dispositifs leur permettant d'apprendre ces régularités entre action et sensation, ne peuvent pas se constituer de monde extérieur, et sont donc réduits à l'état de monade sans espace (et sans temps). Si l'on étudie rigoureusement le monde pour-soi sensori-moteur, il est probable que nous découvrirons un temps et un espace très différents de ceux auxquels nous sommes

habitués. Ce temps ressemblera probablement beaucoup au « présent remémoré » proposé par Edelman, et l'espace sera morcelé.

La constitution du monde pour soi spécifiquement humain, établissant un monde intersubjectif dans lequel prend place une conscience de soi et un temps autobiographique demande en effet à nouveau de faire appel à d'autres concepts, dont la formalisation rigoureuse semble encore plus lointaine.

2.4 Conscience de soi et empathie

Plusieurs auteurs s'accordent sur cette distinction entre une conscience « sensori-motrice », et une conscience de plus haut niveau, appelée d'ordre supérieur par Edelman (Edelman 1992), autobiographique par Damasio (Damasio 1995), ou encore « l'exister » par Todorov (Todorov 1995). Mais là s'arrête le consensus. En effet, on trouve peu d'accord dans la littérature sur les mécanismes explicatifs de ce niveau de conscience. Edelman considère que le langage est particulièrement important et permet de sortir du présent remémoré de la conscience primaire, sensori-motrice. Damasio récuse cette explication et pense que la différence ne constitue pas vraiment une rupture, mais vient d'un accroissement de la capacité de mémoire et de l'efficacité de son archivage. La controverse sur ce point n'en est qu'à ses débuts.

Nous avons proposé des pistes différentes pour expliquer ce changement de niveau. Ces réflexions sont fortement influencées par les travaux de J.P. Dupuy, qui lui-même revisite les thèses de A. Smith à la lumière de travaux récents sur la complexité. Il s'agit notamment d'amorcer la définition d'un cadre conceptuel plus précis, qui intègre l'un des fondements de la famille humaniste : le « je » et le « tu » naissent en même temps et se codéterminent. Ces idées sont en partie développées dans (Deffuant 1998) et surtout (Deffuant 2004). J'en résume ici les principes, sans en développer les conséquences sur les structures sociales (se reporter à (Deffuant 2004) sur ce point).

2.4.1 Composition d'empathie fusionnelle et distanciée

Nous proposons de poursuivre la voie ouverte par A. Smith et réempruntée par Dupuy (Dupuy 1992; Dupuy 1999), qui considère l'empathie (ou la sympathie dans le langage d'A. Smith) comme l'opérateur central des interactions humaines. Nous proposons de prolonger cette réflexion en distinguant deux opérateurs d'empathie⁵ :

- L'empathie fusionnelle est l'identification (partielle) de soi à l'autre, c'est à dire l'adoption de ce qu'on imagine être son point de vue et ses émotions. Cela correspond à ce que M. Scheler nomme « fusion affective ». Mme de Sévigné écrivant à sa fille « j'ai mal à votre poitrine », constitue un joli exemple d'empathie fusionnelle. A. Smith remarque que la découverte d'un livre par une autre personne peut nous donner un plaisir renouvelé à entendre sa lecture, alors que le lire seul nous aurait ennuyé. Ceci correspond à nouveau pour nous à de l'empathie fusionnelle. L'existence des neurones « miroir », indique une probable base biologique et inconsciente à l'empathie fusionnelle. Nous représenterons l'empathie fusionnelle par un rectangle en pointillé autour de la perception action ressentie chez l'autre (voir figure 2.1 sur la gauche). La valeur associée par B (du

⁵ Ces deux catégories recouvrent sans doute partiellement celles qui sont visées dans le débat entre « théorie de la simulation » et « théorie de la théorie ». Cependant, il se peut que les choses ne soient pas aussi claires. L'empathie distanciée pourrait être issue d'une simulation, qui prend ensuite ses distances. Inversement l'empathie fusionnelle pourrait être issue d'une reconstitution théorique, qui est ensuite totalement reprise à son propre compte. Donc il ne s'agit pas pour nous de prendre parti dans le débat qui oppose théorie de la simulation et théorie de la théorie.

point de vue de A), est reprise à son compte par A, et donc partagée par les deux systèmes (selon A).

- L'empathie distanciée est l'imagination du point de vue de l'autre, mais sans adopter les émotions correspondantes. Ainsi, le parent qui cherche à rassurer son enfant effrayé par l'obscurité, s'imagine-t-il les sentiments de l'enfant, mais sans les reprendre à son compte. De même, les concurrents, ou les ennemis imaginent-ils constamment leurs intentions ou sentiments mutuels, sans se les approprier. L'historien du totalitarisme pourra avoir une connaissance fine de la psychologie de Hitler ou Staline, imaginer leurs espoirs, leurs obsessions, leurs drames, sans ressentir lui-même directement ces émotions. La pitié peut s'analyser comme une empathie distanciée : la conscience du malheur de l'autre engendre une valeur négative, mais aussi un refus de s'identifier à sa situation. On peut supposer que l'empathie distanciée vient d'une analyse conceptuelle de la situation de l'autre, associée à un blocage plus ou moins volontaire de l'empathie fusionnelle, à la suite d'un refus d'échanger les rôles. Nous représenterons l'empathie distanciée par rectangle en traits pleins autour de la perception action attribuée à l'autre (voir figure 2.1 sur la droite), en outre, contrairement à l'empathie fusionnelle, les valeurs associées ne sont plus partagées : A associe une valeur à sa perception globale de l'état de B qui comprend un cycle perception action et une valeur associée. Ces valeurs peuvent donc être différentes.

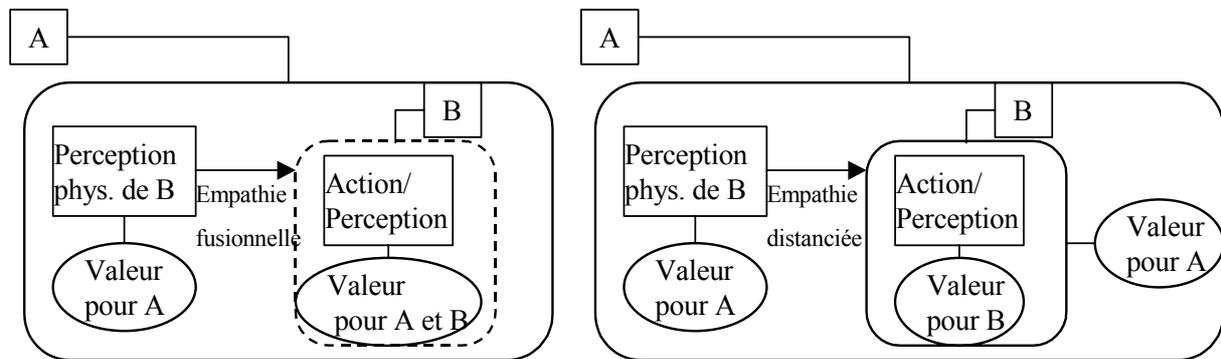


Figure 2.1 : Empathie fusionnelle et distanciée. A gauche l'empathie fusionnelle de A avec B le mène à partager ce qu'il imagine être les perceptions et valeurs de B. A droite l'empathie distanciée de A pour B distingue les perceptions et valeurs de B des siennes. En outre, A peut associer une valeur à ce qu'il imagine être le couple perception /valeur de B.

Il est probable que l'empathie fusionnelle s'établit facilement lorsque la fusion est bénéfique, c'est à dire lorsque la situation de l'autre semble égale ou préférable. Au contraire, l'empathie distanciée suppose un refus d'entrer en l'autre, de se mettre véritablement à sa place, même si sa situation est un sujet de préoccupation ou de satisfaction. Le refus de la fusion serait dû au désagrément ou au déclassement qu'elle impliquerait.

Nous supposons que l'humain se caractérise par l'utilisation de ces deux opérateurs, et de leurs combinaisons. En réalité, la distinction tranchée que nous proposons ici est sans doute artificielle, car il existe probablement une continuité entre empathies fusionnelle et distanciée. Cependant, cette simplification nous semble utile. En effet, en suivant l'approche de Dupuy, nous examinons différents modes de composition de ces opérateurs. Nous insistons plus particulièrement sur une figure composant empathie fusionnelle et distanciée, et permettant d'établir une empathie fusionnelle avec l'image de soi perçue en l'autre. Cette figure nous semble caractéristique d'une forme de conscience de soi. Nous appelons cette configuration : conscience de soi « par le haut ». Le prototype de cette configuration est la relation entre enfant et parent du point de vue de l'enfant. L'enfant a une empathie fusionnelle avec son parent qui a une empathie distanciée pour lui, ce qui donne à l'enfant la possibilité de s'identifier à l'image de lui-même ainsi perçue.

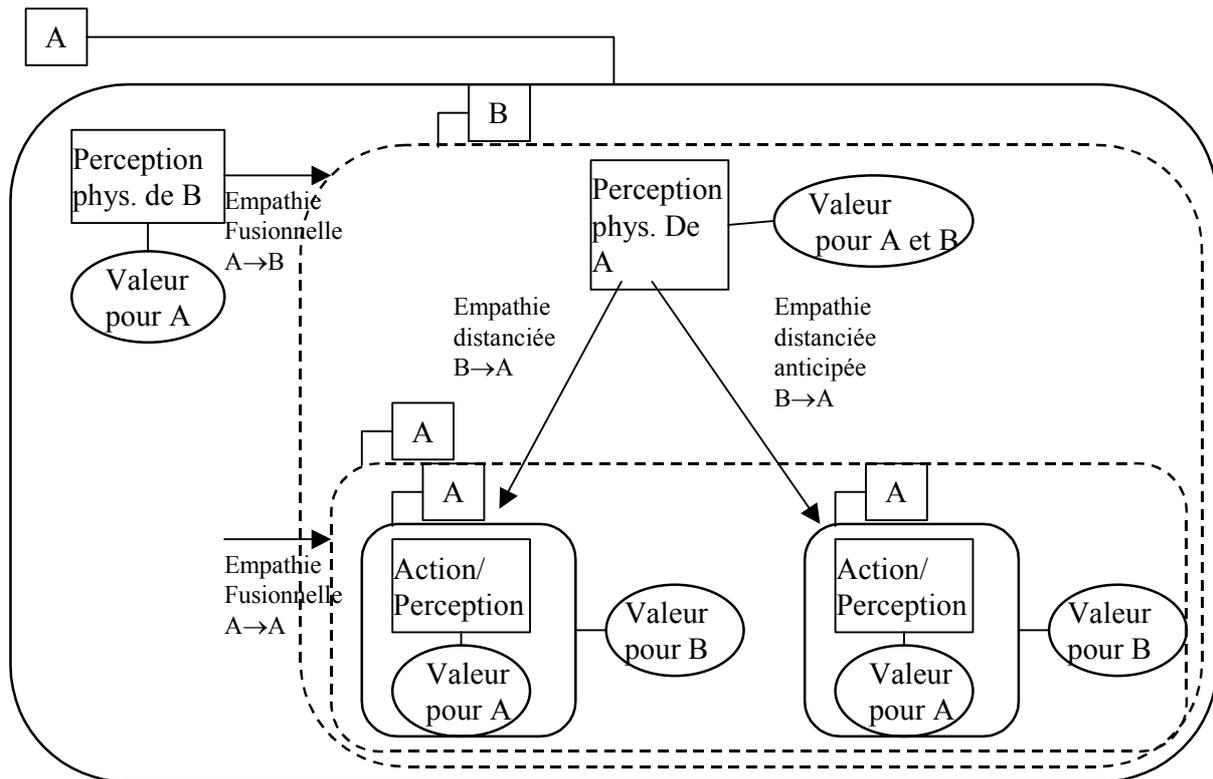


Figure 2.2 : Schéma de la conscience de soi « par le haut » comme composition d'empathie fusionnelle, puis distanciée, puis encore fusionnelle avec la vision de soi-même déployée dans le temps.

Il nous semble qu'une configuration de ce type est nécessaire pour sortir d'un monde pour soi sensori-moteur, limité par le « présent remémoré ». Dans (Deffuant 2004) j'utilise l'expérience de la croyance déçue, réinterprétée par Gopnik (Gopnik 1993), pour illustrer ce point. L'idée centrale est que la conscience d'un temps autobiographique nécessite un point de vue extérieur à l'anticipation sensori-motrice présente, qui permette d'organiser plusieurs épisodes dans un ensemble cohérent. Ce point de vue extérieur est donné par l'empathie distanciée.

La configuration composant empathie distanciée et fusionnelle dans l'ordre inverse permet aussi un déploiement de soi dans le temps, selon ce raisonnement. Cependant, cette image de soi est d'abord mise à distance par A lui-même, ce qui rend probablement un peu plus difficile son réinvestissement fusionnel que dans le cas précédent. Nous appellerons cette configuration : conscience de soi « par le bas ». Le prototype de cette configuration est aussi la relation entre parent et enfant, mais du point de vue du parent. La perception distanciée d'une empathie fusionnelle confère une forme de conscience de soi, d'existence (qui semble prémunir du suicide par exemple).

La conscience de soi, ou l'existence sociale, résulte dans notre cadre de configurations complexes, composant trois niveaux d'empathie : composition d'empathie fusionnelle et distanciée (dans un ordre ou dans l'autre), et ensuite investissement fusionnel de l'intériorité déployée dans le temps ainsi constituée.

La composition d'empathies de même type présente une grande richesse, mais ne permet pas, selon nous, d'établir une véritable conscience de soi. La composition d'empathies distanciées a été formalisée par la théorie des jeux (en général plusieurs anticipations sont considérées, comme par exemple dans le cas du dilemme du prisonnier). Cette configuration empathique est celle qui est la plus libre, car tous les niveaux sont issus d'une imagination distanciée. L'autre est donc manipulable à merci. Par ailleurs, un réinvestissement fusionnel de l'image de soi ainsi obtenue semble particulièrement difficile, car il contredit la double distance

introduite. Cette configuration permet aussi un déploiement du temps, dans la même logique que précédemment, puisqu'elle permet la vision de l'intériorité à plusieurs instants. Cependant la source de ce temps est une activité interne, activable selon les désirs de A, ce qui donne une illusion de maîtrise du temps. Ce processus serait à rapprocher d'une tendance à rejouer les scènes passées, en arrangeant ses propres attitudes et celles des autres à son avantage. Cette configuration peut probablement expliquer la constitution d'un temps complexe, et notamment des projections dans le passé, et donc une distinction peu claire entre passé et avenir⁶.

2.4.2 Autres virtuels

Notre cadre retrouve l'une des caractéristiques principales de l'humanisme selon Todorov : l'autre apparaît comme la source profonde de soi. Plus précisément, le déploiement temporel des séquences de la conscience primaire requiert un point d'appui extérieur à cette conscience, par nature prisonnière de chacune des séquences. L'autre joue le rôle de ce point d'appui extérieur.

Mais la présence effective d'autrui n'est pas indispensable. En effet, l'autre peut être virtuel, imaginaire. Cette capacité à simuler la présence d'autres virtuels est attestée dès l'âge de 18 mois. En fait, il nous semble pertinent de distinguer deux sortes d'autres virtuels : l'autre virtuel miroir et l'autre virtuel commun.

L'autre virtuel miroir joue le rôle de point d'appui extérieur imaginaire, potentiellement constamment présent. Comme le propose A. Smith avec l'homme intérieur (« the man within »), il s'agirait d'un amalgame d'autres réels, souvent composite. L'empathie plutôt fusionnelle ou plutôt distanciée avec l'autre virtuel miroir définit des formes différentes de conscience de soi. Nous examinons un peu plus en détails dans (Deffuant 2004) les conséquences de l'instabilité naturelle de l'autre virtuel miroir, soumis à la pression constante des autres réels, et de l'instabilité probable des formes d'empathie qui leur sont attribuées. Nous réfléchissons aussi au lien probable avec la création de personnages mythologiques et divins comme autres virtuels miroir plus stable et commun à l'ensemble d'une communauté.

L'autre virtuel miroir doit selon nous être distingué de l'autre virtuel commun. Ce dernier ne constitue pas la conscience de soi, mais une forme d'objectivité ou d'intersubjectivité (il n'y a rien de vraiment original, on trouve des idées similaires chez Husserl et Kant notamment). Le monde objectif correspondrait non seulement à nos propres perceptions et anticipations, mais aussi au fait que tout un chacun anticiperait ou percevrait la même chose⁷. Ainsi, la perception de l'espace est objectivée par la conviction que les autres en ont une similaire. De même, la perception du sens d'un mot s'accompagne de la conviction que l'ensemble des locuteurs en a une perception similaire. Dans les deux cas, il s'agit d'une empathie fusionnelle un ou des autres virtuels, que tous partagent. C'est pourquoi nous l'appelons autre virtuel commun, par opposition à l'autre virtuel miroir, qui a plutôt tendance à être particulier à chacun. En outre la configuration empathique engendrée par l'autre virtuel commun est plus simple, car elle peut se limiter à une empathie fusionnelle. Dans (Deffuant 2004), nous suggérons que l'activité scientifique établit des autres virtuels communs particuliers, rationnels et non empathiques, qui se traduisent par des paradigmes partagés au sein d'une communauté.

⁶ Cette discussion sur le temps pourrait utilement être rapprochée de la distinction proposée par J.P. Dupuy entre temps du projet et temps de l'histoire.

⁷ A ce niveau, il ne s'agit pas du « common knowledge » (CK), tel que le présente Dupuy, car l'empathie pour les autres n'est opérée qu'une fois, alors que ces opérations sont répétées à l'infini dans le CK. Nous soutenons donc que cette récursion à l'infini n'est pas nécessaire à la constitution d'un monde commun.

Dans cette perspective, le monde individuel et le monde social sont intimement imbriqués. En effet, même la conscience de soi passe par l'autre. Cette définition de la conscience de soi permet d'ailleurs de construire une vision particulière des institutions sociales, que nous évoquons très partiellement dans (Deffuant 2004) et qui serait hors de propos ici. En effet, nous avons déjà fait une excursion bien téméraire en dehors des limites scientifiques habituelles.

2.4.3 Hors des limites scientifiques ?

Force est de constater que l'empathie n'est présente dans aucune des conceptions scientifiques les plus structurantes de l'homme moderne. Elle a évidemment été éliminée de toutes les sciences fondamentales (physique, chimie, biologie..). Mais elle est également majoritairement contournée dans les sciences de l'homme, qui subissent l'hégémonie des sciences de la nature. L'idéologie récente considère l'homme comme une machine à « traiter de l'information », ce qui exclut l'empathie ou la tient dans le rôle marginal d'un module parmi d'autres. Pour toutes les anthropologies et sociologies qui sont fondamentalement influencées par cette vision, et c'est la majorité, l'empathie est donc hors du champ.

Selon nous cette tendance à éradiquer l'empathie fusionnelle a une cause très profonde : sa contradiction apparente avec la logique mathématique. Par logique mathématique, nous référons à ce que Castoriadis appelle la « logique ensembliste identitaire », qui structure tout discours scientifique. L'une des caractéristiques fondamentales de cette logique est que deux choses sont soit égales, soit distinctes. L'empathie fusionnelle se trouve donc en porte à faux dans un tel formalisme, car son principe inclut justement à la fois une séparation et une fusion.

L'anecdote suivante nous semble en fournir une illustration particulièrement frappante (tirée de (Barrow 1992)). Lors d'une conférence, B. Russel utilisait la thèse bien connue selon laquelle si un système logique contient une proposition fautive, alors on peut l'utiliser pour prouver n'importe quelle proposition. Un auditeur l'interrompt et le met au défi de démontrer que celui qui l'interpelle et le pape ne font qu'un à partir de l'affirmation $2+2=5$. B. Russel répond du tac au tac : « si $2+2 = 5$ alors $4 = 5$, si j'ôte 3 alors $1 = 2$, donc le pape et vous êtes deux personnes mais pouvez aussi bien n'en faire qu'une ». La possibilité de fusion (être à la fois un et deux) est donc directement associée à une aberration mettant en péril l'ensemble de l'édifice logique.

Notre figure de la conscience de soi qui met en oeuvre une telle fusion à deux niveaux différents, ne peut donc être qu'une hérésie totale dans une vision ensembliste identitaire⁸. L'autre virtuel miroir, qu'il soit public ou privé, ne peut donc avoir droit de cité explicite dans les conceptions scientifiques modernes, pas plus que l'autre virtuel « commun », car cela impliquerait de reconnaître l'empathie fusionnelle comme centrale. Selon nous ces évictions ne sont qu'idéologiques, et ne changent évidemment rien à la réalité d'un monde intersubjectif porté par une figure virtuelle partagée, ni à la nécessité d'une forme de fusion et de distanciation pour que la conscience de soi voie le jour.

Nous pensons que le caractère non scientifique du cadre que nous proposons n'est pas une fatalité. En améliorer la rigueur formelle, ainsi que le lien à des données précises observables nous semble une tâche de longue haleine, mais réalisable. Une étape majeure de cette démarche sera donc de dépasser l'incompatibilité prétendue entre empathie fusionnelle et principes scientifiques, qui n'est selon nous qu'apparente. Peut-être faudra-t-il pour cela

⁸ Notons que l'empathie distanciée, qui conserve plus clairement la séparation des monades, a trouvé un cadre scientifique dans la théorie des jeux, bien que ce développement soit relativement récent

repousser certaines limites habituelles de la rationalité. Ce défi nous paraît des plus importants et des plus stimulants.

2.5 Conclusion

Ce chapitre parcourt de grands espaces conceptuels à bien trop vive allure, et assène de manière péremptoire quelques idées qui paraîtront probablement inhabituelles, voire inacceptables à beaucoup. Il peut en résulter une certaine confusion dont je suis très conscient. Il faudrait en effet leur consacrer un ouvrage entier pour les argumenter et les articuler correctement. Le lecteur aura sans doute deviné que j'ai l'espoir de parvenir un jour à achever une telle tâche.

Cependant, dans ce document à vocation plus académique, mon objectif était de brosser en quelques pages un large paysage, dans lequel s'insèrent les travaux plus techniques qui vont suivre. Il me semble important en effet d'évaluer ces travaux plus précis dans cette perspective large, afin d'éviter des interprétations erronées.

Je tire une leçon de prudence et de modestie de cet exercice. A ma connaissance, aucun système artificiel n'est véritablement parvenu à s'extraire d'un monde pour soi monadique, c'est-à-dire celui des organismes privés de cortex, des végétaux, sans temps et sans espace. Les modèles de perception en effet ne se fondent en général pas sur l'apprentissage de boucles sensori-motrices qui pourraient laisser espérer l'accession à un embryon de conscience primaire. L'expression « intelligence artificielle » apparaît au mieux comme prétentieuse, et les simulations sociales d'agents informatiques actuelles ne peuvent capturer que des aspects grossiers et relativement superficiels des dynamiques sociales les plus profondes.

Les travaux plus techniques que nous présentons dans les chapitres prochains se nourrissent des convictions exprimées dans ce chapitre. Mais ils ne font pas exception et en restent aux toutes premières étapes de notre programme. Au moins sont-ils conçus avec une pleine conscience de leurs limites, et des étapes qu'il reste à franchir. Nous espérons que parfois, ils ouvrent quelques pistes qui permettront un jour d'y parvenir. Les chapitres 3 et 4 portent sur l'utilisation de la géométrie et des statistiques pour détecter certains aspects saillants d'ensemble de données dans de espaces euclidiens. Les chapitres 6 et 7, consacrés à des modèles sociaux, utilisent fortement la distinction entre faits et valeurs, et se focalisent plus particulièrement sur l'étude de dynamiques d'évolution de valeurs. En annexe, nous présentons un logiciel d'aide à la construction de valeurs au sein d'un groupe.

3 Des hyperplans séparateurs dans des espaces déployés : les Support Vector Machines

Nous considérons le problème typique de la théorie de l'apprentissage statistique : des données sont engendrées par une distribution de probabilité stable dans le temps (par exemple un phénomène physique), et se répartissent en deux classes. On désire utiliser au mieux un échantillon fini de ces données classées, pour construire une loi générale permettant de classer des points nouveaux tirés selon la même distribution.

Les « support vector machines » (*svm*) sont fondées sur l'objet géométrique le plus simple permettant de tracer une frontière dans l'espace : l'hyperplan. Il s'agit donc d'une saillance élémentaire. L'interprétation géométrique est explicite, car dans le cas où les données sont linéairement séparables, il s'agit de définir l'hyperplan qui sépare les points selon leur classe, en prenant les plus grandes marges possibles. La position du séparateur est donc obtenue à l'aide d'une optimisation quadratique.

Plusieurs recherches récentes ont relancé très fortement l'intérêt pour cette technique depuis 1995 (Vapnik 1998; Cristianini et Shawe-Taylor 2000). Tout d'abord, l'approche peut être étendue à des « espaces déployés », définis de manière très générale par un produit scalaire dans l'espace initial (qui définit une distance). Ainsi, les hyperplans dans l'espace déployé permettent-ils de définir des séparations non-linéaires de formes très diverses dans l'espace initial. L'optimisation dans un espace déployé requiert de faire une optimisation quadratique dont la matrice est une transformation linéaire simple de la matrice des produits scalaires des données, et permet de traiter à très peu de frais le cas d'échantillons non linéairement séparables (dans l'espace déployé). Cette dernière opération introduit un paramètre qui s'interprète comme un paramètre de « régularisation ».

L'argument déterminant en faveur des *svm* est qu'elles permettent de définir des surfaces séparatrices dans des espaces de dimension très élevée (voire infinie), en utilisant un nombre de degrés de liberté en général faible (la normale à l'hyperplan résultat est définie comme une combinaison linéaire de certains points de l'échantillon projetés dans l'espace déployé). Cette particularité leur donne une très grande capacité à rendre compte de caractéristiques robustes de la séparation, bien que parfois très irrégulières dans l'espace initial. Ainsi, les performances des *svm* sont-elles très souvent meilleures que celles d'autres techniques qui résolvent moins bien le dilemme de l'apprentissage.

Outre ces qualités, l'interprétation géométrique explicite de cette technique la rend évidemment très séduisante dans la perspective d'allier statistique et géométrie. Mais curieusement, je l'ai abordée initialement à partir d'une autre perspective géométrique : celle de la recherche d'algorithmes rapides de projection d'un point sur un polyèdre convexe quelconque dans l'espace. Ce problème m'était posé par des manipulations de polyèdres convexes évoqués dans le chapitre suivant. En fait, il est apparu que ce problème est très proche de celui de l'optimisation d'une *svm*. Après plusieurs années de travail récurrent sur ce problème, j'ai proposé très récemment un algorithme qui a la particularité d'être direct. Cette proposition est particulièrement intéressante lorsque le nombre final de « support vectors » est faible. Cet algorithme m'a permis également de proposer une technique automatique de pondération des support vectors, qui rend le résultat moins sensible aux points aberrants

(*outliers*). Enfin, j'ai pu montrer que l'approche était généralisable à un plus large ensemble de problèmes d'optimisation quadratique convexe. Pour simplifier la présentation, je commence par ce résultat général, que j'applique ensuite aux *svm*. Tous ces résultats sont récents et en cours de soumission.

3.1 La convergence de la stratégie de décomposition (*chunking*)

La stratégie de décomposition (*chunking*) est souvent utilisée dans les algorithmes d'optimisation des *svm*, surtout lorsque les échantillons sont grands. En effet, la forme quadratique à optimiser est définie par la matrice des produits scalaires des points d'apprentissage, donc de taille le carré du cardinal de l'échantillon. Lorsqu'on considère des échantillons de plusieurs dizaines de milliers d'éléments, cela peut devenir très difficile.

L'idée de la décomposition est de réaliser plusieurs optimisations sur des sous parties de l'échantillon, à la place d'une seule optimisation sur l'ensemble de l'échantillon. On considère donc un sous-ensemble courant de l'échantillon, sur lequel on réalise l'optimisation, ce qui permet d'identifier des éléments qu'on peut éliminer du sous-ensemble courant, et d'autres qu'on peut ajouter. La méthode a d'abord été proposée par Vapnik, et a fait l'objet de plusieurs améliorations et généralisations (Mangasarian et Musicant 1998; Smola et al. 1998; Kaufmann 1999). Cependant, la démonstration de sa convergence n'a été faite que récemment, dans le cas particulier des *svm* (Hush et Covel 2003).

Nous proposons une démonstration de convergence plus générale concernant un grand ensemble de problèmes d'optimisation, et qui nous permet ensuite de proposer un algorithme d'optimisation direct particulier (il en existe d'autres, notamment à partir des méthodes de « pivot complémentaire » (Murty 1997)).

3.1.1 Notations et hypothèses

Nous considérons un problème classique de minimisation quadratique convexe dans un espace euclidien \mathbb{R}^n soumis à l contraintes linéaires :

$$\begin{aligned} & \underset{\mathbf{x} \in \mathbb{R}^n}{\text{Minimiser}} \quad \mathbf{x}^T Q \mathbf{x} + R \mathbf{x} \\ & \text{Sous contrainte : } \mathbf{a}_i \cdot \mathbf{x} + b_i \geq 0, \quad i \in I = \{1, \dots, l\}, \end{aligned} \quad (1)$$

Où Q est une matrice semi-définie positive, R un vecteur de \mathbb{R}^n , et \cdot est le produit scalaire standard.

Nous considérons un sous ensemble J de I l'ensemble des contraintes, nous appelons P_J le polyèdre convexe défini par les contraintes de J :

$$P_J = \{ \mathbf{x} \in \mathbb{R}^n, \forall j \in J, \mathbf{a}_j \cdot \mathbf{x} + b_j \geq 0 \}. \quad (2)$$

Nous considérons que pour tout sous-ensemble J de I , le minimum de la fonction quadratique dans P_J existe et est unique. Soit \mathbf{x}_J^* ce point :

$$\mathbf{x}_J^* = \arg \min_{\mathbf{x} \in P_J} \mathbf{x}^T Q \mathbf{x} + R \mathbf{x} \quad (3)$$

Le point \mathbf{x}_J^* peut être caractérisé à l'aide de la théorie lagrangienne. Le lagrangien du problème est :

$$L_J = \frac{1}{2} \mathbf{x}^t Q \mathbf{x} + R \mathbf{x} - \sum_{j \in J} \alpha_j (\mathbf{a}_j \cdot \mathbf{x} + b_j). \quad (4)$$

Le point \mathbf{x} solution du problème, est donné par les équations suivantes, obtenue notamment en égalant les dérivées du lagrangien à 0 :

$$Q \mathbf{x} + R = \sum_{j \in J} \alpha_j \mathbf{a}_j \quad \text{avec: } \forall j \in J, \alpha_j \geq 0. \quad (5)$$

$$\forall j \in J, \alpha_j (\mathbf{a}_j \cdot \mathbf{x} + b_j) = 0. \quad (6)$$

$$\forall j \in J, \mathbf{a}_j \cdot \mathbf{x} + b_j \geq 0. \quad (7)$$

Géométriquement, ces équations expriment que \mathbf{x}_j^* se trouve sur une facette du polyèdre. Nous notons J^* l'ensemble des contraintes définissant cette facette. Ce sont les contraintes dont les multiplicateurs de Lagrange sont non nuls:

$$J^* = \{j \in J, \alpha_j > 0\}. \quad (8)$$

Dans la suite, pour simplifier l'écriture, nous utilisons la notation :

$$\varphi(\mathbf{x}) = \frac{1}{2} \mathbf{x}^t Q \mathbf{x} + R \mathbf{x}. \quad (9)$$

3.1.2 La procédure de décomposition

Pour réaliser la décomposition, nous supposons dans un premier temps que nous disposons d'une procédure d'optimisation. La procédure de décomposition considère un ensemble courant de contraintes J , initialisé arbitrairement. A chaque itération, on réalise une optimisation de l'ensemble courant, on élimine de l'ensemble courant les contraintes dont les multiplicateurs de Lagrange sont nuls, puis on lui ajoute un nombre arbitraire de contraintes non satisfaites au point rendu par l'optimisation. La procédure s'arrête lorsque le résultat de l'optimisation sur l'ensemble courant satisfait toutes les contraintes. Plus précisément, la procédure est la suivante :

```

chunking( $Q, R, I$ )
 $J_0 := \text{arbitrarySubset}(I)$ 
( $J^*_0, \mathbf{x}^*_0$ ):= minimum( $Q, R, J_0$ )
while ( $dJ_k := \text{chooseNonSatisfiedConstraints}(I, \mathbf{x}^*_k)$  not nil )
    ( $J^*_{k+1}, \mathbf{x}^*_{k+1}$ ):= minimum( $Q, R, J^*_k \cup dJ_k$ )
     $k := k+1$ 
end while
return ( $J^*_k, \mathbf{x}^*_k$ )

```

Les fonctions appelées dans la procédure sont :

- *chooseNonSatisfiedConstraints*(I, \mathbf{x}^*_k): choisit un nombre un nombre de contraintes non satisfaites, c'est à dire telles que : $\mathbf{a}_j \cdot \mathbf{x}_k + b_j < 0$.

- $\text{minimum}(Q, R, J)$: calcule le minimum de la fonction quadratique définie par Q, R et l'ensemble de contraintes J , retourne ce point \mathbf{x}^* et l'ensemble de contraintes correspondant J^* .

On peut aisément vérifier que si la procédure de décomposition converge, alors elle donne le résultat attendu. En effet, on peut voir que le résultat satisfait nécessairement les conditions imposées par théorie Lagrangienne. Le théorème 1 exprime la convergence de la procédure.

Theorème 1 : *Quelle que soit la procédure utilisée pour choisir les contraintes non satisfaites, si tout les ensembles J de I définissent des polyèdres P_J tels que le minimum de la fonction quadratique considérée existe et est unique, alors la procédure de décomposition converge.*

Démonstration du théorème 1 :

Nous considérons les valeurs J_k^* et \mathbf{x}_k^* calculées à chaque itération k de l'algorithme, et nous allons montrer que pour un couple (m, n) tel que $m > n$, nous avons nécessairement : $J_m^* \neq J_n^*$. Nous allons d'abord montrer que : $\varphi(\mathbf{x}_m^*) > \varphi(\mathbf{x}_n^*)$.

Pour tout k , $J_k^* \subset J_{k+1}$, donc $P_{J_k^*} \subset P_{J_{k+1}}$. Donc, puisque \mathbf{x}_k^* est le minimum de φ dans $P_{J_k^*}$, et \mathbf{x}_{k+1}^* le minimum de φ dans $P_{J_{k+1}^*}$, nous avons : $\varphi(\mathbf{x}_{k+1}^*) \geq \varphi(\mathbf{x}_k^*)$.

Supposons $\varphi(\mathbf{x}_{k+1}^*) = \varphi(\mathbf{x}_k^*)$. Puisque $P_{J_k^*} \subset P_{J_{k+1}}$, nous avons : $\mathbf{x}_{k+1}^* \in P_{J_k^*}$. De plus, puisque le minimum est unique $\mathbf{x}_{k+1}^* = \mathbf{x}_k^*$. C'est impossible puisque la procédure suppose que J_{k+1} est obtenu en ajoutant des contraintes non satisfaites par \mathbf{x}_k^* . Donc $\varphi(\mathbf{x}_{k+1}^*) > \varphi(\mathbf{x}_k^*)$, pour tout k , ce qui implique : $\varphi(\mathbf{x}_m^*) > \varphi(\mathbf{x}_n^*)$.

Or, $\varphi(\mathbf{x}_m^*) > \varphi(\mathbf{x}_n^*)$ implique $J_m^* \neq J_n^*$, puisque le minimum est unique. Le nombre de sous-ensembles J_k de I étant fini, l'algorithme s'arrête nécessairement. \square

3.2 La minimisation récursive de facettes

3.2.1 Calcul du minimum dans l'espace d'une facette

Nous considérons maintenant les facettes du polyèdre P_I . Soit J des sous-ensembles de I . Nous appelons espace d'une facette F_J le sous-espace de \mathbb{R}^n défini par l'intersection des hyperplans définis par les contraintes correspondantes dans J :

$$F_J = \left\{ \mathbf{x} \in \mathbb{R}^n \text{ tel que } \forall i \in J, \mathbf{a}_i \cdot \mathbf{x} + b_i = 0 \right\} \quad (10)$$

Nous recherchons le minimum dans le polyèdre P_I en faisant des minimisations dans les espaces de facette. Pour être sûr que la solution existe toujours, nous ajoutons artificiellement des variables de détente $(\xi_i)_{1 \leq i \leq l}$, et un paramètre C , et réécrivons le problème de la manière suivante :

$$\begin{aligned} & \underset{(\mathbf{x} \in \mathbb{R}^n, \xi_i \in \mathbb{R}^+)}{\text{Minimiser}} \quad \mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{R} \mathbf{x} + \sum_{i=1}^l \xi_i^2 \\ & \text{sous contrainte : } \mathbf{a}_i \cdot \mathbf{x} + b_i + C \|\mathbf{a}_i\| \xi_i \geq 0, \quad i \in I = \{1, \dots, l\}. \end{aligned} \quad (11)$$

Le choix de C proche de 0 (par exemple 0.000001) garantit de résoudre une approximation très proche du problème initial. Pour simplifier la présentation, nous conservons la notation de (1), puisque la réécriture du problème n'en change pas la forme.

L'optimum de φ dans l'espace d'une facette peut être obtenu directement, par la méthode du gradient conjugué, en utilisant le gradient de φ projeté sur l'espace de facette. Pour faire cette projection, nous construisons une base orthogonale (par la méthode de Schmidt) $U=(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k)$ à partir de la base $A=(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k)$ (la transformation du problème assure que A est toujours une base). Le gradient conjugué est initialisé par la projection d'un point quelconque de l'espace sur l'espace de facette.

Lorsque la procédure de gradient conjugué a convergé, le gradient projeté est nul, donc le gradient est une combinaison linéaire des vecteurs $(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k)$. Les coefficients de cette combinaison linéaire sont les multiplicateurs de Lagrange pour une minimisation avec des contraintes d'égalité. Ces multiplicateurs ne sont donc pas nécessairement positifs.

Le minimum de φ dans l'espace de facette F_J est noté $\hat{\mathbf{x}}_J$.

3.2.2 L'algorithme de minimisation récursive de facettes

La procédure est récursive. Elle est appliquée à un couple d'ensembles de contraintes J et K , tels que $K \subset J$. Elle utilise le calcul du minimum dans l'espace de facette pour calculer les multiplicateurs de Lagrange et éliminer ceux qui sont négatifs. L'ensemble K retient les contraintes qu'il est interdit d'éliminer à l'étape considérée pour empêcher la procédure de boucler.

```

recursiveMinimization( $J, K$ )
 $J_0 := J, K_0 := K,$ 
 $(\hat{\mathbf{x}}_0, \hat{J}_0) = \text{eliminateNegativeMultipliers}(J_0, K_0)$ 
while  $(i_n := \text{leastSatisfiedConstraint}(\hat{\mathbf{x}}_n, J))$  not nil
     $J_{n+1} := \hat{J}_n \cup \{i_n\}, K_{n+1} := K \cup \{i_n\}$ 
     $(\hat{\mathbf{x}}_{n+1}, \hat{J}_{n+1}) := \text{recursiveProjection}(J_{n+1}, K_{n+1})$ 
     $n := n+1$ 
end while
return  $(\hat{\mathbf{x}}_n, \hat{J}_n)$ 

```

Les deux procédures appelées par la minimisation récursive :

- La procédure *eliminateNegativeMultipliers*(J, K) calcule le minimum dans l'espace de facette J , élimine le multiplicateur de Lagrange le plus négatif, n'appartenant pas à K , jusqu'à ce que J ne contienne que des multiplicateurs strictement positifs.
- La procédure *leastSatisfiedConstraint*(\mathbf{x}, J) sélectionne la contrainte de J qui est la moins satisfaite en \mathbf{x} . Si toutes les contraintes sont satisfaites, la procédure retourne *nil*.

Nous allons montrer que la procédure renvoie le minimum de φ dans l'intersection du polyèdre P_J et de la facette F_K . Notons ce minimum \mathbf{x}_{J-K}^* , et $J^* \sim K$ la facette de P_J à laquelle ce point appartient.

Theorème 2 : Soit J un sous-ensemble de I et K un sous-ensemble de J , la minimisation récursive appliquée à (J, K) donne \mathbf{x}_{J-K}^* , le minimum de φ dans l'intersection du polyèdre J et du sous-espace de facette K et $J^* \sim K$, la facette à laquelle ce point appartient, en un nombre fini d'itérations.

Démonstration du théorème 2 :

Appelons profondeur de $recursiveMinimization(J, K)$ le nombre maximum d'appels imbriqués à elle-même au cours de son exécution.

Tout d'abord, notons que la profondeur est majorée supérieurement par la taille de $J-K$, car dès qu'une contrainte est choisie, elle est ajoutée à K , ce qui empêche qu'elle soit choisie à nouveau.

Il faut donc montrer que ces séquences sont en nombre fini, et que la procédure donne le bon résultat. Il s'agit d'une démonstration par récurrence sur la profondeur de $recursiveMinimization(J, K)$.

Notons $P(m)$ le fait que, si $recursiveMinimization(J, K)$ est inférieur m , alors la procédure retourne $J^* \sim K$.

$P(0)$ est vrai, parce si la profondeur est 0 $recursiveMinimization(j, K) = J_0$. Donc toutes les contraintes de J sont satisfaites pour $\hat{\mathbf{x}}_0$, et tous les multiplicateurs dont les indices sont dans J_0-K sont strictement positifs (garanti par la fonction $eliminateNegativeMultipliers(\hat{\mathbf{x}}_0, J)$). Le point satisfait donc les conditions de la théorie Lagrangienne, donc $\hat{J}_0 = J^* \sim K$, et $\hat{\mathbf{x}}_0 = \mathbf{x}_{J-K}^*$.

Supposons maintenant $P(m)$ vrai, et que la profondeur de $recursiveMinimization(J, R)$ est au plus $m+1$. Les appels à l'intérieur de la procédure sont donc au plus de profondeur m , et nous pouvons leur appliquer l'hypothèse de récurrence. Pour chacun de ces appels (supposé à l'étape n), nous avons donc :

$$recursiveMinimization(J_{n+1}, K_{n+1}) = (\mathbf{x}_{J_{n+1} \sim K_{n+1}}^*, J_{n+1}^* \sim K_{n+1}). \quad (12)$$

En utilisant le même raisonnement que lors de la démonstration du théorème 1, nous obtenons:

$$i_n \in J_{n+1}^* \sim K_{n+1}. \quad (13)$$

Le multiplicateur de Lagrange correspondant à cette contrainte dans $J_{n+1}^* \sim K_{n+1}$ est donc strictement positif. Donc, puisque $K_{n+1} = K \cup \{i_n\}$:

$$J_{n+1}^* \sim K_{n+1} = J_{n+1}^* \sim K. \quad (14)$$

Donc :

$$recursiveMinimization(J_{n+1}, K_{n+1}) = (\mathbf{x}_{J_{n+1} \sim K}^*, J_{n+1}^* \sim K) \quad (15)$$

Nous sommes donc dans les conditions du théorème 1, à l'intérieur de l'espace de facette F_K : La procédure d'ajout de contrainte est limitée à celle qui satisfait le moins le minimum courant. Le théorème 1 assure que la procédure se termine en un temps fini. \square

3.2.3 Décomposition un par un

La procédure décrite précédemment pourrait être lancée avec $J=I$ et $K = \{\}$. Mais cela obligerait à faire beaucoup de calculs pour éliminer les multiplicateurs de Lagrange négatifs, en partant d'une facette définie par l'ensemble de toutes les contraintes. Au contraire, pour limiter ces calculs, il est préférable de limiter le plus possible le nombre de contraintes à

chaque décomposition. Cela donne donc une procédure dans laquelle chaque décomposition contient uniquement une contrainte.

```

oneByOneChunking( $I$ )
( $\hat{\mathbf{x}}_0, \hat{J}_0$ ) := initialize( $I$ )
while ( $i_n :=$  leastSatisfiedConstraint( $\hat{\mathbf{x}}_n, I$ ) not nil )
    ( $\hat{\mathbf{x}}_{n+1}, \hat{J}_{n+1}$ ) := recursiveProjection( $\hat{J}_n + i_n, i_n$ )
     $n := n+1$ 
end while
return ( $\hat{\mathbf{x}}_n, \hat{J}_n$ )

```

La fonction *initialize* consiste, lorsque la forme quadratique est définie positive, à faire une minimisation sans contrainte de la forme quadratique. Dans ce cas, \hat{J}_0 est l'ensemble vide. Mais dans certains cas particuliers, notamment lorsque la forme quadratique est semi définie, il est nécessaire de faire une initialisation particulière. C'est par exemple le cas lorsque nous utilisons cette méthode pour calculer des *svm*. Nous savons que cette procédure converge, comme toutes les procédures de chunking, dans les conditions indiquées.

3.2.4 Quelques tests

R. Lalain a implémenté l'algorithme de minimisation récursive au cours d'un stage de l'ISIMA. Il l'a testé sur le problème de la projection d'un point sur un polyèdre convexe. Le tableau 3.1 présente le nombre de projections nécessaires pour obtenir le résultat (moyennes sur 10 répétitions), en utilisant des contraintes générées aléatoirement (tout en s'assurant que leur intersection n'est pas vide).

Nombre de contraintes		Dimension de l'espace			
		500	1000	5000	10000
10	nombre de projections :	46	67	95	111
	tps de calcul (en s) :	<1	<1	<1	<1
50	nombre de projections :	112	224	490	611
	tps de calcul (en s) :	<1	3	31	75
100	nombre de projections :	239	363	682	887
	tps de calcul (en s) :	14	33	200	465

Tableau 3.1 : Résultats des tests de performances (moyenne sur 10 répétitions)

Il apparaît clairement que la méthode peut être intéressante pour des problèmes comportant de nombreuses contraintes de petite dimension. Par exemple R. Lalain fait des tests supplémentaires sur 500 000 contraintes en dimension 10, pour lesquels l'algorithme donne des résultats en quelques secondes.

Des tests supplémentaires seraient évidemment nécessaires pour mieux apprécier l'intérêt de cette méthode, notamment pour des formes quadratiques dont la matrice est plus compliquée.

3.3 Application aux *svm*

La procédure de minimisation récursive peut être appliquée à la minimisation des *svm*, bien que la matrice de la forme quadratique ne soit pas définie. En effet le noyau de l'application est de rang 1, et la configuration du problème assure que le minimum de la forme quadratique considérée dans l'espace de la facette est unique pour toute facette définie par au moins un hyperplan.

Par ailleurs, les *svm* introduisent naturellement des variables « de détente » qui permettent de traiter les échantillons non linéairement séparables. L'idée est de minimiser la distance entre ces points mal classés et le séparateur. La distance *L1* est souvent utilisée pour les *svm*, car elle demande en général un nombre de vecteurs support moins élevé. Dans le formalisme classique des *svm*, le passage au dual pour résoudre la minimisation aboutit à des problèmes très similaires pour les distances *L1* ou *L2*, au point que les mêmes algorithmes de résolution peuvent être utilisés. Ce n'est pas le cas avec notre approche. En effet, la distance *L1* introduit des contraintes supplémentaires, exprimant que les variables de détente doivent rester positives. Cependant, cette complexité supplémentaire est compensée par le fait que le calcul du minimum dans une facette est plus simple que dans le cas *L2*.

3.3.1 Notations, présentation du principe géométrique dans le cas linéairement séparable

Nous considérons un échantillon $S = ((\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_l, y_l))$ de l vecteurs de \mathbb{R}^d , chacun étant associé à une classe ($y_i = 1$ or $y_i = -1$). I est l'ensemble des indices : $I = \{1, \dots, l\}$.

Le principe des *svm* est de rechercher l'hyperplan qui sépare les points avec la plus grande marge. Un hyperplan définit un demi espace par une inégalité du type :

$$\mathbf{w} \cdot \mathbf{x} + b \geq 0. \quad (16)$$

La marge est la distance de l'hyperplan au point \mathbf{x} qui lui est le plus proche, son expression est :

$$\frac{|\mathbf{w} \cdot \mathbf{x} + b|}{\|\mathbf{w}\|}. \quad (17)$$

Il est possible de fixer arbitrairement, pour le point le plus proche :

$$|\mathbf{w} \cdot \mathbf{x} + b| = 1, \quad (18)$$

Car nous avons un degré de liberté disponible dans l'équation de l'hyperplan. Dès lors, maximiser la distance du point le plus proche à l'hyperplan revient à minimiser la norme de \mathbf{w} . Le problème s'exprime donc de la manière suivante :

$$\underset{(w, b)}{\text{Minimiser}} \|\mathbf{w}\|^2 \quad (19)$$

Sous contraintes : $y_i(w \cdot x_i + b) \geq 1, i \in I.$

Mais toute la puissance des *svm* s'exprime lorsqu'on projète le problème dans un espace de traits, grâce à la définition d'un noyau.

3.3.2 Projection dans un espace de traits et ajout de variables « de détente »

Nous supposons que le problème est projeté dans un espace déployé E de dimension D (éventuellement infinie), par une fonction Ψ qui est définie (parfois de manière uniquement implicite) par un noyau K . C'est-à-dire qu'on a :

$$\psi(\mathbf{x}_1) \cdot \psi(\mathbf{x}_2) = K(\mathbf{x}_1, \mathbf{x}_2) \quad (20)$$

Où ' \cdot ' est le produit scalaire dans E . De plus, on ajoute des variables « de détente », notées ξ_i permettant de traiter des ensembles non linéairement séparables, et un paramètre C dit de régularisation. Il faut minimiser la valeur de ces variables, et deux possibilités ont été utilisées pour calculer cette distance : la distance $L1$ ou la distance $L2$.

L'espace de recherche est noté $\Omega = E \times \mathbb{R}^{l+1}$. Le problème d'optimisation d'une *svm* pour la distance $L2$ est le suivant :

$$\begin{aligned} & \underset{(\mathbf{v}, \xi_i, b)}{\text{Minimiser}} \quad \|\mathbf{v}\|^2 + \sum_{i=1}^l \xi_i^2 \\ & \text{Sous contraintes : } y_i(\mathbf{v} \cdot \Psi(\mathbf{x}_i) + b) + C\xi_i \geq 1, \quad i \in I. \end{aligned} \quad (21)$$

Ou bien, pour la distance $L1$:

$$\begin{aligned} & \underset{(\mathbf{v}, \xi_i, b)}{\text{Minimiser}} \quad \|\mathbf{v}\|^2 + \sum_{i=1}^l \xi_i \\ & \text{Sous contraintes : } y_i(\mathbf{v} \cdot \Psi(\mathbf{x}_i) + b) + C\xi_i \geq 1, \quad \xi_i \geq 0, \quad i \in I. \end{aligned} \quad (22)$$

Dans les deux cas, il s'agit de problèmes de minimisation quadratique convexe, du type de ceux évoqués en première partie de ce chapitre. La différence importante est l'introduction d'un espace « déployé ».

Cependant, le cas de la distance $L2$ semble a priori plus simple car il y a moins de contraintes, et que la matrice de la forme quadratique est très proche de la matrice identité. Nous l'examinons donc en premier.

3.3.3 Application de l'algorithme de minimisation récursive pour la norme $L2$

Montrons tout d'abord que nous nous trouvons dans les conditions d'application de la méthode de minimisation récursive.

La matrice de la forme quadratique est diagonale, avec des 1 partout sauf sur la composante b . La forme quadratique est donc semi-définie positive. Il nous faut donc en outre vérifier que la minimisation sur un espace de facette quelconque existe et est unique. On voit que cette condition est toujours vérifiée car les hyperplans de contraintes ne peuvent jamais être parallèles à l'axe b (il y a toujours au moins une composante en ξ_i non nulle).

Il s'avère que ce passage à un espace déployé (même s'il est de dimension infinie), n'empêche pas d'utiliser la méthode, car elle peut être adaptée de manière à n'utiliser que le noyau K appliqué à des éléments de l'espace initial. En effet, le calcul de la minimisation dans l'espace d'une facette est particulièrement direct dans ce cas, car il est obtenu à partir d'un seul pas du gradient conjugué (puisque le noyau de la matrice est de dimension 1).

Pour simplifier les notations, nous posons :

$$\mathbf{u}_i = (y_i \psi(x_i), y_i, 0, \dots, 0, C, 0, \dots, 0). \quad (23)$$

Où le C correspond à la composante sur l'axe des ξ_i , donc après $i-1$ zéros. Et nous définissons l'espace d'une facette définie par les contraintes de l'ensemble J inclus dans I par :

$$F_J = \{\boldsymbol{\omega} \in \Omega \text{ tels que } \forall j \in J, \mathbf{u}_j \cdot \boldsymbol{\omega} = 1\}. \quad (24)$$

Théorème 3 : Soit J un sous-ensemble de I , et soient :

\mathbf{b} le vecteur directeur de l'axe des b , et $\mathbf{0}$ l'origine de l'espace Ω ,

$\boldsymbol{\beta}$ la projection orthogonale du vecteur \mathbf{b} sur l'espace F_J ,

$\boldsymbol{\omega}_0$ la projection orthogonale du point $\mathbf{0}$ sur F_J .

Nous avons :

$$\hat{\boldsymbol{\omega}}_J = \boldsymbol{\omega}_0 + \frac{\boldsymbol{\omega}_0 \cdot \mathbf{b}}{1 - \boldsymbol{\beta} \cdot \mathbf{b}} \boldsymbol{\beta}. \quad (25)$$

Démonstration :

Calculons le gradient en $\hat{\boldsymbol{\omega}}_J$, il est égal à ce point auquel on enlève sa composante sur l'axe b :

$$\nabla(\hat{\boldsymbol{\omega}}_J) = \hat{\boldsymbol{\omega}}_J - (\hat{\boldsymbol{\omega}}_J \cdot \mathbf{b}) \mathbf{b}. \quad (26)$$

Ce gradient projeté sur la facette est donné par :

$$p(\nabla(\hat{\boldsymbol{\omega}}_J)) = p(\hat{\boldsymbol{\omega}}_J) - (\hat{\boldsymbol{\omega}}_J \cdot \mathbf{b}) \boldsymbol{\beta}. \quad (27)$$

$$p(\nabla(\hat{\boldsymbol{\omega}}_J)) = p(\boldsymbol{\omega}_0) + \frac{\boldsymbol{\omega}_0 \cdot \mathbf{b}}{1 - \boldsymbol{\beta} \cdot \mathbf{b}} \boldsymbol{\beta} - \left(\boldsymbol{\omega}_0 \cdot \mathbf{b} + \frac{\boldsymbol{\omega}_0 \cdot \mathbf{b}}{1 - \boldsymbol{\beta} \cdot \mathbf{b}} \boldsymbol{\beta} \cdot \mathbf{b} \right) \boldsymbol{\beta} = 0. \quad (28)$$

Donc le gradient projeté s'annule en $\hat{\boldsymbol{\omega}}_J$, et puisque ce point appartient à l'espace facette, il s'agit bien du minimum. \square

Le calcul de $\hat{\boldsymbol{\omega}}_J$ peut être exprimé uniquement à l'aide des vecteurs \mathbf{x}_i et des expressions $K(\mathbf{x}_p, \mathbf{x}_q)$ avec p et q dans J . Ce résultat fait l'objet du théorème 4.

Remarquons que nous avons (δ_{pq} étant le symbole de Kronecker):

$$\mathbf{u}_p \cdot \mathbf{u}_q = y_p \cdot y_q K(\mathbf{x}_p, \mathbf{x}_q) + y_p \cdot y_q + \delta_{pq} C^2. \quad (29)$$

Théorème 4 : Le calcul de $\hat{\boldsymbol{\omega}}_J$ peut se mettre sous la forme :

$$\hat{\boldsymbol{\omega}}_J = \sum_{j \in J} \alpha_j \mathbf{u}_j + \hat{b}_J \mathbf{b}. \quad (30)$$

Où α_i et \hat{b}_j sont des fonctions de $\mathbf{u}_p \cdot \mathbf{u}_q$ avec p et q dans J .

Démonstration:

L'orthogonalisation de Schmidt de la base $U=(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k)$ donne une base orthogonale $R=(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_k)$ dont les éléments sont des combinaisons linéaires des \mathbf{u}_i , avec des coefficients fonction de $\mathbf{u}_p \cdot \mathbf{u}_q$.

$$\mathbf{r}_1 = \mathbf{u}_1 \text{ et } \mathbf{r}_i = \sum_{j=1}^{i-1} \lambda_{ij} \mathbf{r}_j + \lambda_{ii} \mathbf{u}_i \quad \text{pour } i = 2, \dots, k. \quad (31)$$

$$\lambda_{i1} = 1, \lambda_{ii} = -\frac{\mathbf{u}_1 \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_i} \text{ et } \lambda_{ij} = -\lambda_{ii} \frac{\mathbf{u}_i \cdot \mathbf{r}_j}{\mathbf{r}_j \cdot \mathbf{r}_j} \text{ for } j = 2, \dots, i-1. \quad (32)$$

Cette base permet de calculer $\boldsymbol{\beta}$ et $\boldsymbol{\omega}_0$ comme combinaisons linéaires de \mathbf{u}_i et \mathbf{b} car, (en supposant que $R=(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_k)$ est normée):

$$\boldsymbol{\beta} = \mathbf{b} - \sum_{i=1}^k (\mathbf{r}_i \cdot \mathbf{b}) \mathbf{r}_i. \quad (33)$$

$$\boldsymbol{\omega}_0 = \sum_{i=1}^k \mu_i \mathbf{r}_i \text{ avec } \mu_1 = \frac{1}{\mathbf{r}_1 \cdot \mathbf{u}_1} \text{ et } \mu_i = \frac{1 - \sum_{j=1}^{i-1} \mu_j (\mathbf{r}_j \cdot \mathbf{u}_i)}{\mathbf{r}_i \cdot \mathbf{u}_i}. \quad (34)$$

En remplaçant $\boldsymbol{\beta}$ et $\boldsymbol{\omega}_0$ par leur valeur dans l'expression du théorème 3, on obtient le résultat recherché. \square

L'algorithme de minimisation récursive peut donc être appliqué directement au problème d'optimiser une *svm* en norme $L2$. Il fournit une solution analytique à ce problème, sans passer par le dual.

3.3.4 Tests de la méthode pour la norme $L2$

Dans un premier temps, nous nous limitons à des tests dans lesquels l'espace déployé est le même que l'espace initial. Nous considérons le nombre de tests de facettes nécessaire à l'obtention du résultat.

3.3.4.1 Un ensemble linéairement séparable en dimension 2

La figure 3.1 représente la progression de l'algorithme dans le cas d'un échantillon linéairement séparable de dimension 2 comprenant 50 points. La *svm* est représentée par ses marges, l'hyperplan séparateur se trouve donc en leur milieu.

Pour converger, l'algorithme utilise 8 minimisations de facette. On peut remarquer que dans ce cas, il y a au plus un point dont le multiplicateur de Lagrange est négatif, et que son élimination conduit directement à un ensemble dont tous les multiplicateurs sont strictement positifs. Dans cet exemple, la profondeur des récursions est toujours 0. Ce ne sera plus le cas lorsque la dimension de l'espace initial sera plus grande.

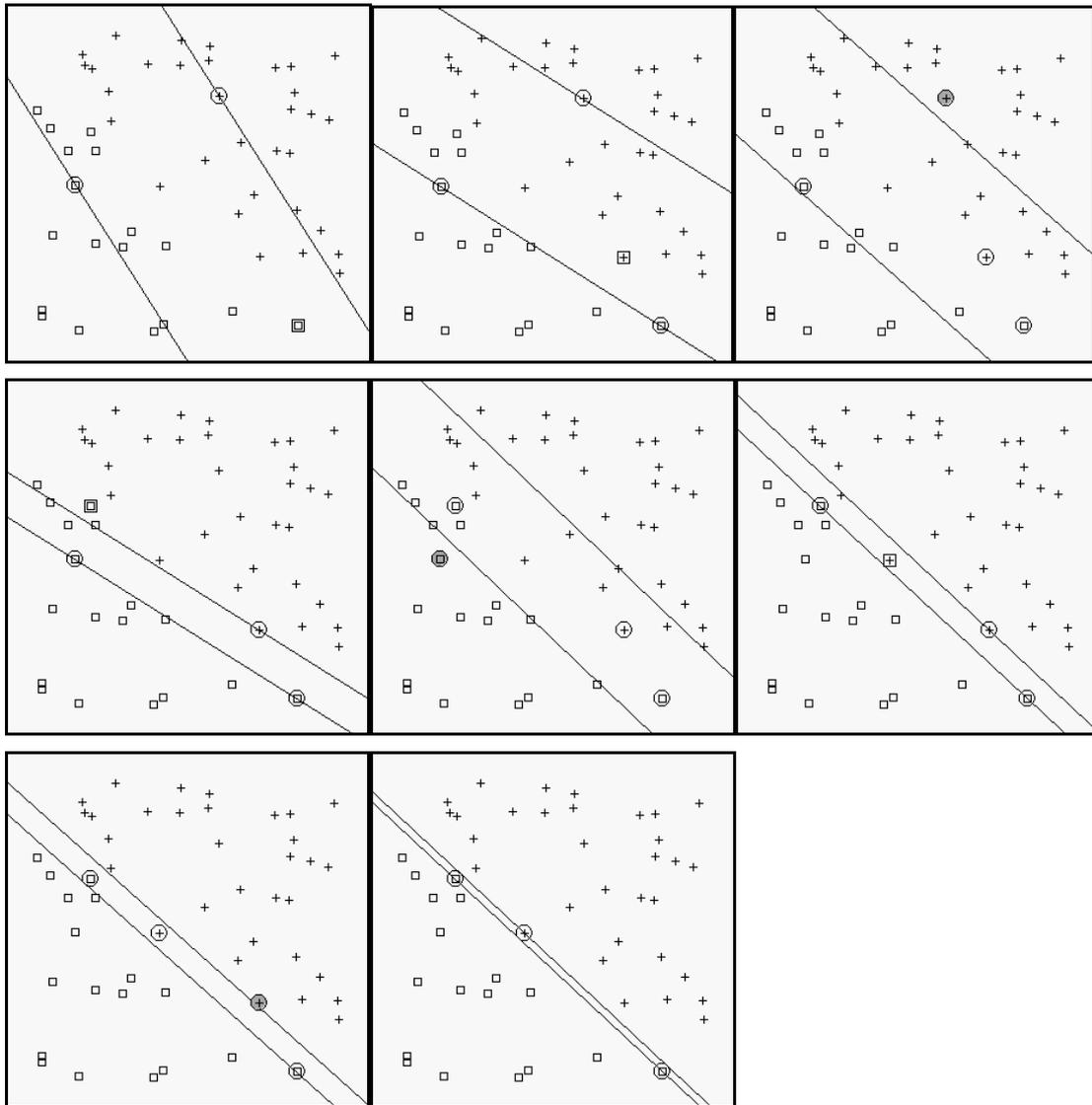


Figure 3.1 : L'algorithme de minimisation récursive appliqué à l'optimisation d'une *svm* en norme L2. Les points définissant la facette en cours de test sont entouré d'un cercle. Un nouveau point sélectionné, qui va s'ajouter à ceux qui définissent la facette, est représenté entouré d'un carré. Le point qui va être éliminé de ceux qui définissent la facette est représenté par un disque grisé.

3.3.4.2 Echantillons linéairement séparables à d'autres dimensions

Nous considérons maintenant des échantillons linéairement séparables de différentes tailles (de 500 à 5000) dans des espaces de différentes dimensions (de 10 à 100). Les échantillons sont tirés selon une distribution uniforme dans $[0,5]^d$, d étant la dimension de l'espace, et leur classe est calculée à l'aide d'une séparation linéaire définie au hasard. La figure 3.2 présente les résultats moyennés sur 20 échantillons de chaque taille et chaque dimension. Nous avons choisi C le paramètre de régularisation égal à $1.0e10^{-2}$ fois la dimension de l'espace.

Ces résultats appellent plusieurs commentaires :

- Le nombre de minimisations dans une facette croît moins vite que linéairement avec la taille de l'échantillon. Cela est dû au fait que ce nombre croît avec le nombre de vecteurs support nécessaires à la solution finale.
- Le nombre de minimisations est au plus 3,5 fois le nombre de décompositions, indiquant qu'en moyenne, pour qu'un nouveau vecteur support soit inclus, moins de 3,5 récursions sont nécessaires,

- Le nombre maximum de minimisations de facettes (relativement au nombre de décompositions) apparaît lorsque le nombre de vecteurs support est proche de la dimension de Vapnik Chervonenkis (VC) de l'espace. Au contraire, lorsque le nombre de vecteurs supports est soit très inférieur, soit très supérieur à la dimension de VC, le nombre de décompositions et le nombre de minimisations sont très proches.

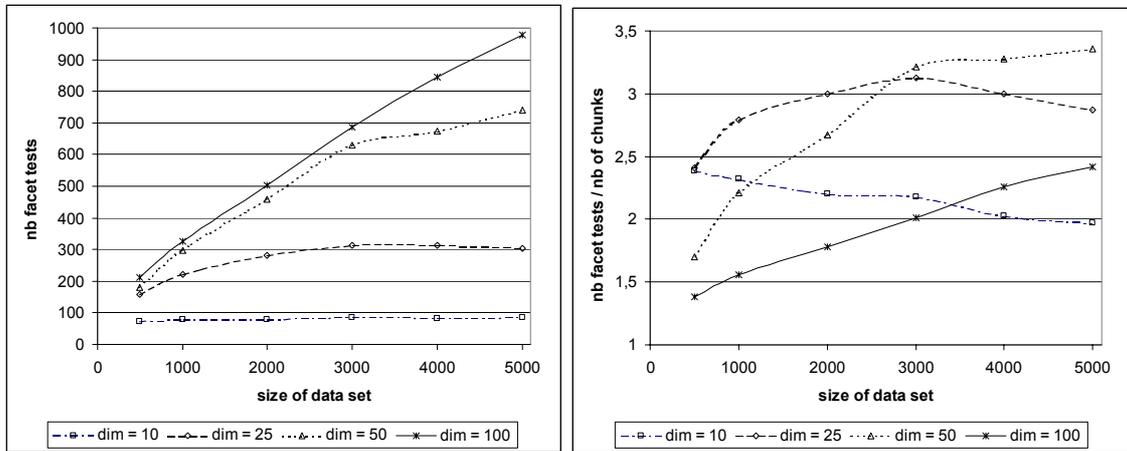


Figure 3.2 : Echantillons linéairement séparables de différentes tailles et différentes dimensions. A gauche, le nombre de minimisations de facettes. A droite, le rapport du nombre de minimisations au nombre de chunks. Les points sont des résultats moyens sur 20 répliques.

3.3.4.3 Séparation linéaire bruitée

Afin d'obtenir des échantillons non linéairement séparables, en partant des échantillons précédents, nous avons modifié la classe de certains points choisis au hasard, avec une probabilité gaussienne autour de la séparation. En tout, nous avons modifié la classe de 5% de l'échantillon. Comme nous l'attendions, le nombre de vecteurs support augmente très fortement, plus de 350 vecteurs support pour un échantillon de taille 2000 en dimension 50. Ceci rend le nombre de minimisations très proche du nombre de décompositions, mais allonge considérablement le temps de calcul, qui est en degré 3 du nombre de vecteurs supports.

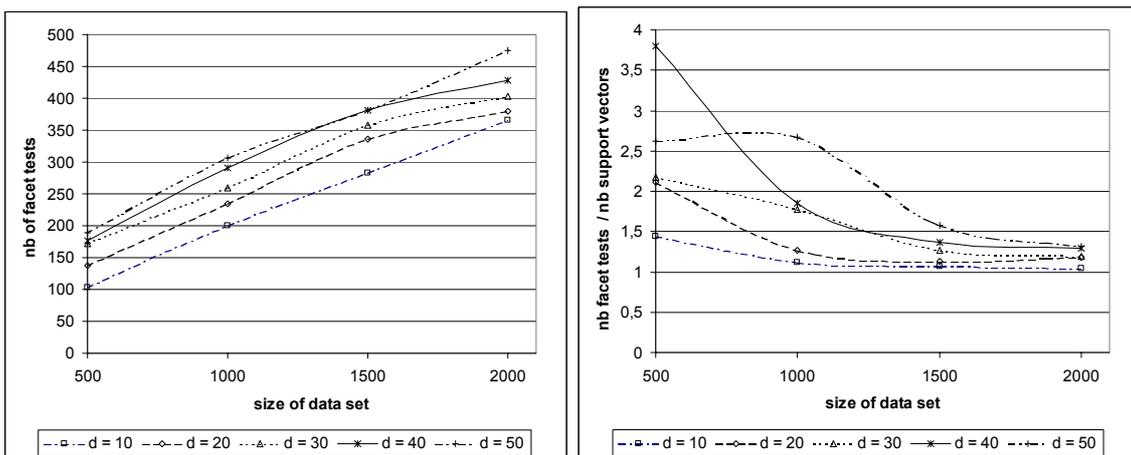


Figure 3.3 : Echantillons non linéairement séparables (5% de bruit gaussien autour de la séparation). A gauche, le nombre de minimisations en fonction de la dimension de l'espace et de la taille de l'échantillon. A droite, le rapport du nombre de minimisations au nombre de chunks. Chaque point des courbes est la moyenne de 20 répliques.

3.3.4.4 Conclusion de ces premiers tests

La méthode paraît intéressante lorsque le nombre de vecteurs support n'est pas trop grand. En effet, lorsque ce nombre augmente elle devient très peu performante. Ce défaut est très rapidement perceptible avec des *svm* en norme L_2 , car ils sont réputés pour établir un grand nombre de vecteurs supports dès que les données sont non linéairement séparables.

3.3.5 Application de la méthode de minimisation récursive pour la norme L_1

Le problème de l'optimisation d'une *svm* en norme 1 ne permet pas d'appliquer directement la méthode de minimisation récursive. En effet, pour certaines configurations de facettes, le problème n'a pas de solution. Nous devons donc introduire de nouvelles variables ξ'_i , et un nouveau paramètre C' , afin de construire le problème approché, selon la méthode indiquée au paragraphe 3.2.1. En fait, il n'est pas nécessaire d'ajouter ces variables pour tous les indices de I . En effet, dans la définition d'une facette, il est opportun de distinguer des indices pour lesquels à la fois la contrainte de position ($y_i(\mathbf{v} \cdot \Psi(\mathbf{x}_i) + b) + C\xi_i = 1$) et la contrainte $\xi_i = 0$ doivent être respectées (nous appellerons conventionnellement cet ensemble d'indice J), et les indices où seule la contrainte de position doit être respectée (nous appellerons cet ensemble d'indice K). Les variables ξ'_i ne sont nécessaires que pour pallier l'annulation de ξ_i , et donc seulement pour l'ensemble J . Le problème de minimisation dans une facette s'écrit donc de la manière suivante :

$$\begin{aligned} & \underset{(\mathbf{v}, \xi, \xi', b)}{\text{Minimiser}} \quad \|\mathbf{v}\|^2 + \sum_{j \in J} \xi_j'^2 + \sum_{k \in K} \xi_k \\ & \text{Sous contraintes : } y_j(\mathbf{v} \cdot \Psi(\mathbf{x}_j) + b) + C' \xi'_j = 1, \quad \xi_j = 0, \quad j \in J, \\ & y_k(\mathbf{v} \cdot \Psi(\mathbf{x}_k) + b) + C \xi_k = 1, \quad k \in K. \end{aligned} \quad (35)$$

Le Lagrangien de ce problème s'écrit :

$$\begin{aligned} L = & \frac{1}{2} \left(\|\mathbf{v}\|^2 + \sum_{j \in J} \xi_j'^2 \right) + \sum_{k \in K} \xi_k - \sum_{j \in J} \alpha_j (y_j(\mathbf{v} \cdot \Psi(\mathbf{x}_j) + b) + C' \xi'_j - 1) \\ & - \sum_{k \in K} \alpha_k (y_k(\mathbf{v} \cdot \Psi(\mathbf{x}_k) + b) + C \xi_k - 1) - \sum_{j \in J} \beta_j \cdot \xi_j \end{aligned} \quad (36)$$

En égalant ses dérivées à 0, on obtient :

$$\mathbf{v} = \sum_{j \in J} \alpha_j y_j \Psi(\mathbf{x}_j) + \sum_{k \in K} \alpha_k y_k \Psi(\mathbf{x}_k), \quad (37)$$

$$C \alpha_k = 1, \quad k \in K, \quad (38)$$

$$\sum_{j \in J} \alpha_j y_j + \sum_{k \in K} \alpha_k y_k = 0, \quad (39)$$

D'où, on obtient :

$$\mathbf{v} = \sum_{j \in J} \alpha_j y_j \Psi(\mathbf{x}_j) + \frac{1}{C} \cdot \sum_{k \in K} y_k \Psi(\mathbf{x}_k), \quad (40)$$

Dans cette expression, on retrouve la distinction entre les vecteurs supports des *svm L1* qui sont à la limite des marges définies par le séparateur (ceux qui sont dans J), et ceux qui sont à l'intérieur des marges (appartenant à K), et dont le coefficient est toujours $1/C$.

Dès que l'ensemble J est non vide, le problème de minimisation de facette ainsi posé a toujours une solution unique. Dans l'algorithme, nous forçons donc toujours à avoir au moins un élément dans J . Ce n'est pas restrictif, car le calcul donne finalement une valeur égale ou très proche de $1/C$ pour le multiplicateur de Lagrange, dans le cas où cet élément appartient en fait à K .

Nous pouvons donc appliquer la méthode de minimisation récursive. En outre, on peut montrer que la solution du problème peut se calculer directement, avec une méthode très proche de celle utilisée dans le cas *L2*. C'est ce qui est exprimé par le théorème 5.

Pour simplifier les notations, nous posons :

$$\mathbf{u}'_j = (y_j \psi(\mathbf{x}_j), y_j, 0, \dots, 0, C', 0, \dots, 0), \text{ pour } j \in J. \quad (41)$$

Où le C' correspond à la composante sur l'axe des ξ'_i , donc après $i-1$ zéros. Et nous définissons l'espace d'une facette définie par les contraintes de l'ensemble J inclus dans I par :

$$F_J = \{\boldsymbol{\omega} \in \Omega \text{ tels que } \forall j \in J, \mathbf{u}'_j \cdot \boldsymbol{\omega} = 1\}. \quad (42)$$

Nous définissons en outre :

$$\mathbf{u}_k = (y_k \psi(\mathbf{x}_k), y_k, 0, \dots, 0, C, 0, \dots, 0), \text{ pour } k \in K \quad (43)$$

Où le C correspond à la composante sur l'axe des ξ_k , donc après $l+k-1$ zéros.

La facette $F_{J,K}$ est définie par :

$$F_{J,K} = \{\boldsymbol{\omega} \in \Omega \text{ tels que } \forall j \in J, \mathbf{u}'_j \cdot \boldsymbol{\omega} = 1 \text{ et } \zeta'_j = 0 \text{ et } \forall k \in K, \mathbf{u}_k \cdot \boldsymbol{\omega} = 1\}. \quad (44)$$

Théorème 5 : Soient J et K deux sous-ensembles de I , F_J et $F_{J,K}$ des espaces de facettes définis comme indiqué plus haut. Soient :

\mathbf{b} le vecteur directeur de l'axe des b ,

$$\mathbf{o}_K = \frac{1}{C} \cdot \sum_{k \in K} \mathbf{u}_k,$$

$\boldsymbol{\beta}$ la projection orthogonale du vecteur \mathbf{b} sur l'espace F_J ,

$\boldsymbol{\omega}_{J,K}$ la projection orthogonale du point \mathbf{o}_K sur $F_{J,K}$.

Le point $\hat{\boldsymbol{\omega}}_{J,K}$ défini par l'équation suivante :

$$\hat{\omega}_{J,K} = \omega_{J,K} + \frac{\omega_{J,K} \cdot \mathbf{b}}{1 - \beta \cdot \mathbf{b}} \beta, \quad (45)$$

est le point où le minimum de la forme quadratique est atteint.

Démonstration du théorème 5 :

On procède de la même manière que pour le théorème 3. Le gradient en $\hat{\omega}_{J,K}$, est égal à ce point auquel on enlève sa composante sur l'axe b , et ses composantes sur les axes des ξ_k , (on appelle δ_k le vecteur directeur de l'axe correspondant), et auquel on ajoute la somme des δ_k , correspondant à la dérivée du terme linéaire de la forme quadratique :

$$\nabla(\hat{\omega}_{J,K}) = \hat{\omega}_{J,K} - (\hat{\omega}_{J,K} \cdot \mathbf{b})\mathbf{b} - \sum_{k \in K} (\hat{\omega}_{J,K} \cdot \delta_k)\delta_k + \sum_{k \in K} \delta_k. \quad (46)$$

On peut montrer que les termes en δ_k s'annulent dans cette somme. Après projection sur la facette, il ne reste que le terme en β .

$$p(\nabla(\hat{\omega}_{J,K})) = -(\hat{\omega}_{J,K} \cdot \mathbf{b})\beta, \quad (47)$$

dont le développement, similaire à celui du théorème 3, est nul. Donc le gradient projeté s'annule en $\hat{\omega}_{J,K}$, et puisque ce point appartient à l'espace de la facette, il s'agit bien du minimum. \square

Le calcul de l'optimum dans une facette est donc finalement moins lourd dans le cas des *svm* $L1$, puisque le calcul des multiplicateurs de Lagrange, exigeant le calcul d'une base orthonormale, n'est nécessaire que pour les éléments de J . Pour les éléments de K , ce calcul est immédiat. Comme les vecteurs support de J sont ceux qui sont sur les limites des marges, ils sont souvent moins nombreux que ceux de K .

Bien que la recherche des contraintes $\xi_j = 0$ demande de faire de plus nombreuses minimisations de facettes que pour la norme $L2$, en général le temps de calcul est plus court pour les minimisations en norme $L1$.

3.3.6 Tests de la méthode pour la norme L1

3.3.6.1 Echantillon linéairement séparable en dimension 2

La figure 3.4 montre un exemple d'optimisation d'une *svm* en norme $L1$ pour un échantillon linéairement séparable de 50 points. Le résultat a nécessité de tester 10 contraintes pour 21 minimisations de facettes.

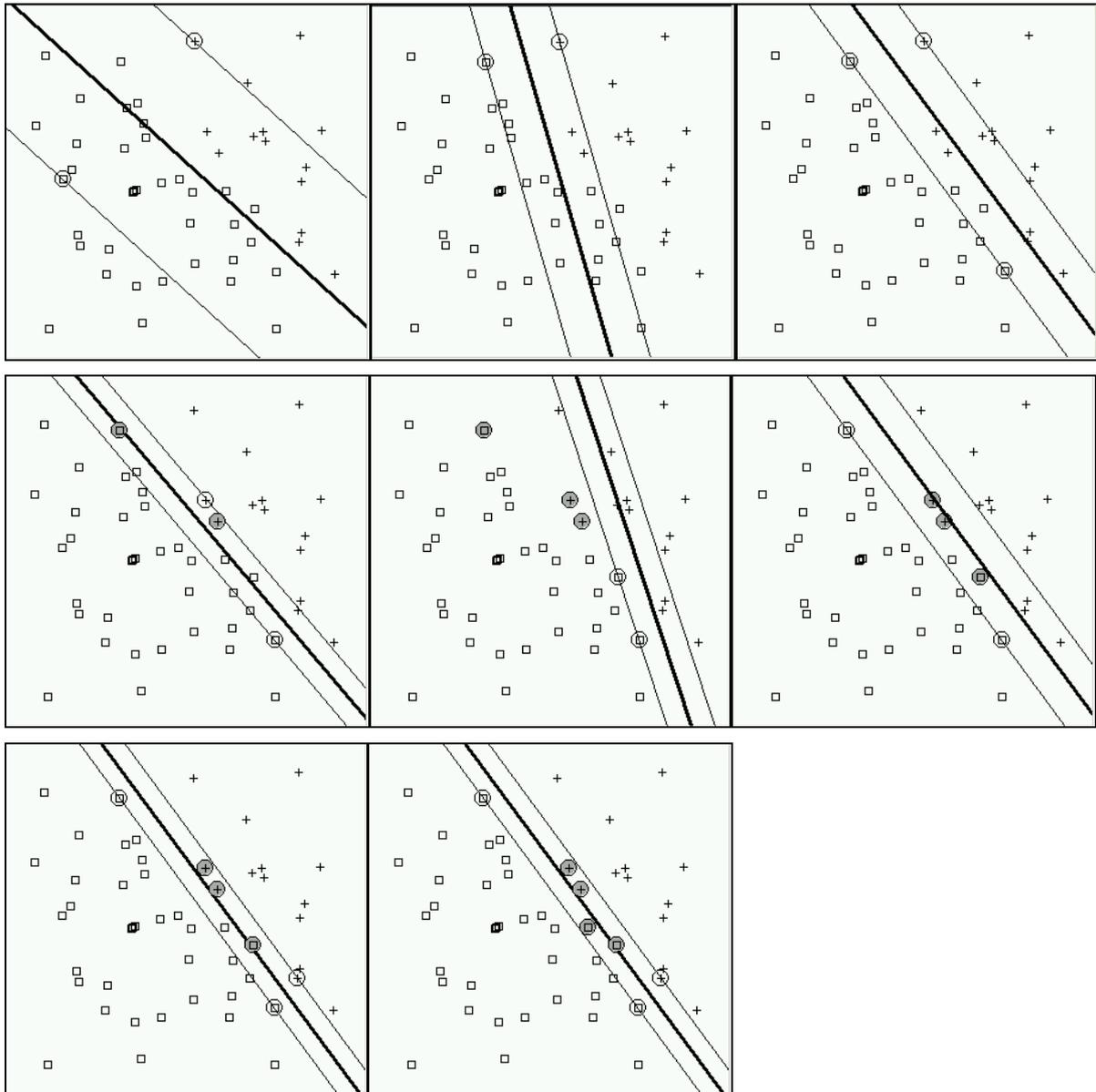


Figure 3.4: Exécution de l'algorithme de minimisation récursive pour l'optimisation d'une *svm* en norme L_1 . Les vecteurs supports sont représentés entourés d'un cercle, et grisés s'ils ont $1/C$ comme multiplicateur de Lagrange. L'échantillon est linéairement séparable de taille 50.

3.3.6.2 Bruit uniforme sur des échantillons linéairement séparables

Le tableau 3.2 ci-dessous donne quelques résultats de tests de la méthode sur des échantillons initialement linéairement séparables, dont on modifie aléatoirement les classes de certains points (un pourcentage donné, ici 10%). Les résultats sont des moyennes sur 20 répétitions.

dimension	10	20	30	40	50
% erreur test	4.25	6.52	7.64	9.09	10.14
Nbre sup. vec.	213	214	210	211	212
Nb fac.minim.	1658	1835	1832	1842	1869

Tableau 3.2 : *svm* en norme $L1$ pour un échantillon linéairement séparable bruité à 10% de taille 500, dans des espaces de différentes dimensions (de 10 à 50). Les trois lignes du tableau donnent de haut en bas : le pourcentage d'erreur en test (sur 10 000 points tiré selon la distribution sans le bruit), le nombre de vecteurs support, le nombre de minimisations de facettes.

On vérifie que le nombre de vecteurs supports est bien moindre que pour la norme $L2$. Autre point remarquable, le nombre de minimisations de facettes n'est pas beaucoup influencé par la dimension de l'espace.

3.4 Diminuer la sensibilité des *svm* aux points aberrants

3.4.1 La sensibilité des *svm* aux points aberrants

L'un des défauts des *svm* reste leur sensibilité aux points aberrants, surtout pour les *svm* en distance $L2$. En effet cette méthode accorde un poids maximal aux points qui sont les plus mal classés (c'est-à-dire le plus loin à l'intérieur des marges définies par la machine), et tend à élargir les marges, ce qui peut amener à définir un nombre excessif de vecteurs support. La figure 3.5 illustre ce comportement.

Ce problème est moindre pour les *svm* en norme $L1$, mais il demeure. En effet, le poids accordé aux points qui sont en dehors des marges est égal à celui donné aux points qui sont à l'intérieur des marges. C'est justement l'idée de donner moins de poids aux vecteurs support qui sont en dehors des marges, que nous explorons maintenant.

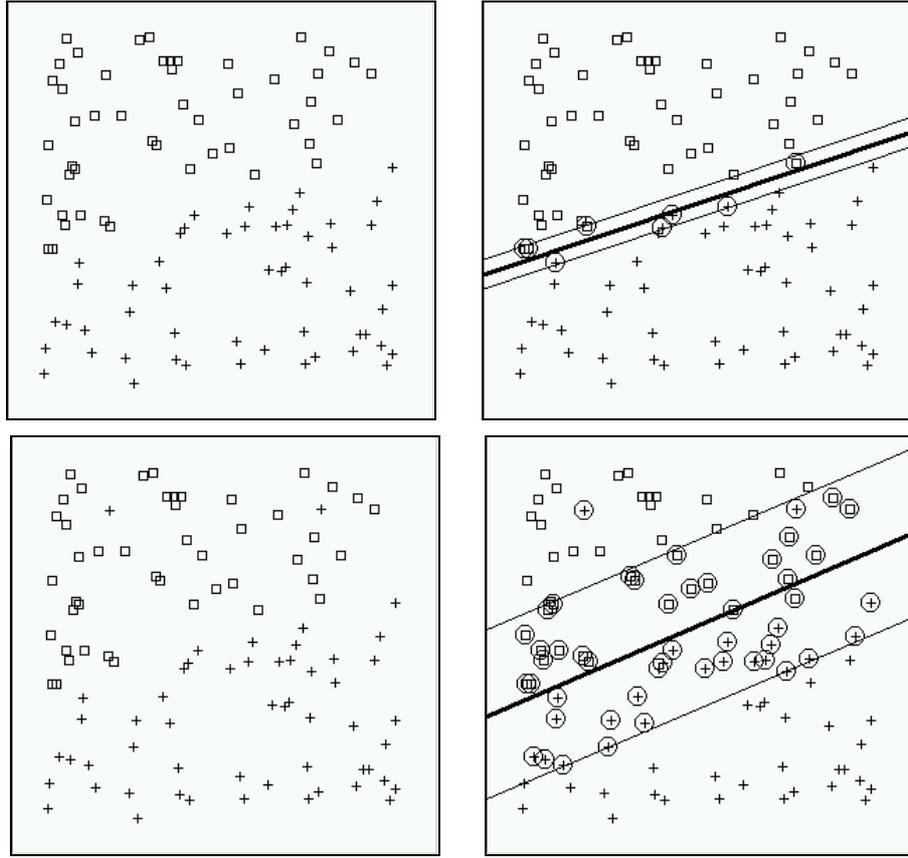


Figure 3.5 : une *svm* en norme 2 est fortement perturbée par l'ajout de 2 points bruités. En haut, l'échantillon est linéairement séparable, et la *svm* comporte 8 vecteurs supports. En bas, avec les mêmes paramètres et un échantillon dans lequel les classes de deux points ont été modifiées, on obtient une *svm* qui comprend 50 vecteurs supports, et dont la direction est sensiblement modifiée.

3.4.2 Donner moins de poids aux vecteurs support situés en dehors des marges

Afin de réduire ce poids des vecteurs support situés au delà des marges, nous introduisons des contraintes supplémentaires dans le problème. La difficulté est que nous devons maintenant minimiser une forme quadratique non plus sur un polyèdre convexe, mais sur l'union de deux polyèdres. Le problème n'est donc plus convexe. Cependant, l'approche de la minimisation récursive offre des possibilités pour résoudre ce type de problèmes.

Nous ajoutons en effet de nouvelles contraintes avec un nouveau paramètre de régularisation C' , et dont nous cherchons à limiter l'effet aux vecteurs support qui sont au delà de la marge. Il nous faut donc modifier également l'ordonnée à l'origine des hyperplans qui définissent ces contraintes, de manière à ce qu'elles intersectent les anciennes contraintes au niveau des marges (c'est à dire lorsque $\omega \cdot \mathbf{u}_i = -1$). Le nouvel ensemble de contraintes est donc du type :

$$(y_i(\mathbf{v} \cdot \Psi(\mathbf{x}_i) + b) + C\xi_i \geq 1) \vee \left(y_i(\mathbf{v} \cdot \Psi(\mathbf{x}_i) + b) + C'\xi_i \geq \frac{2C'}{C} - 1 \right), i \in I, \quad (48)$$

auquel s'ajoutent les contraintes $\xi_i \geq 0$ dans le cas des *svm* $L1$. La figure 3.6 représente la projection de ces contraintes selon des axes ξ_i et $-y_i(\mathbf{v} \cdot \Psi(\mathbf{x}_i) + b)$:

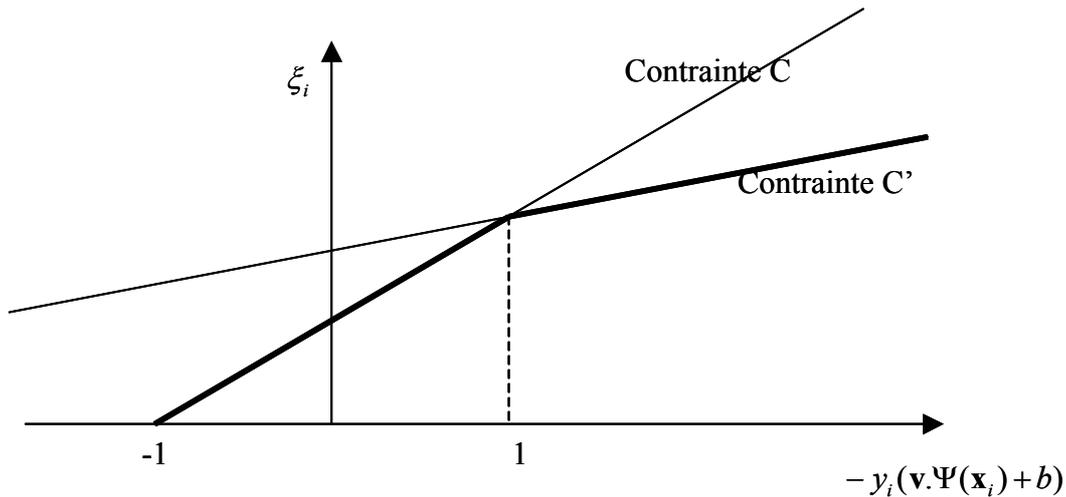


Figure 3.6 : La contrainte supplémentaire dont le coefficient de régularisation est C' , est active pour les points qui sont hors des marges. A l'intérieur des marges, c'est la contrainte de coefficient de régularisation C qui est active. Le problème n'est donc plus convexe.

La non convexité de l'ensemble des contraintes implique des difficultés supplémentaires. Nous essayons de les résoudre en utilisant les caractéristiques particulières du processus d'optimisation par décomposition un par un.

3.4.3 Modification de l'algorithme d'optimisation

L'observation du fonctionnement de l'algorithme de décomposition un par un suggère en effet une voie possible pour résoudre le problème. Lorsque l'échantillon est non linéairement séparable, les premiers pas de l'algorithme donnent lieu à des oscillations très brutales. Ce comportement s'explique par l'ajout de la contrainte la moins bien satisfaite à chaque pas, ce qui amène la machine à se concentrer sur les points aberrants, qui sont en général parmi les premiers à être intégrés. Au bout d'un moment, l'hyperplan se stabilise dans une position assez proche de sa position finale, et la marge décroît progressivement (cf. figure 3.7). C'est uniquement lorsque la stabilisation commence qu'il est possible d'identifier véritablement les vecteurs supports qui sont hors des marges. Auparavant cela peut mener à de graves erreurs.

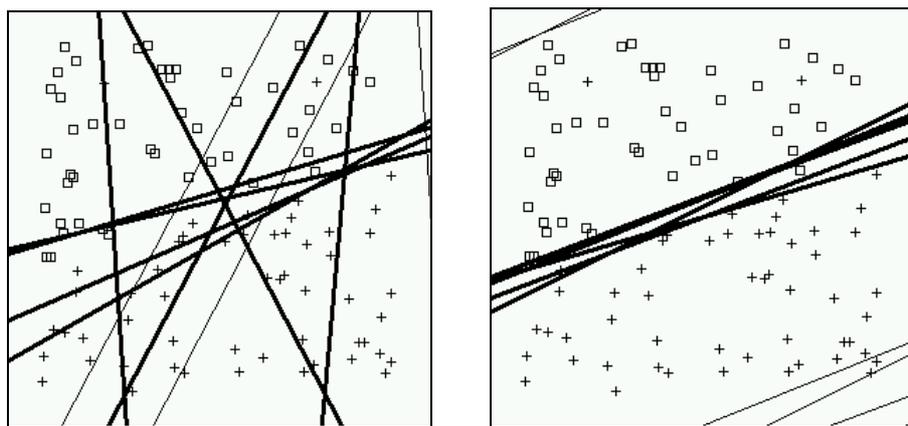


Figure 3.7: Evolution du *chunking* un par un. A gauche, la trace de l'hyperplan engendre au cours des 10 premiers pas de l'algorithme. Sur la droite, la trace de l'hyperplan du 15^{ème} au 20^{ème} pas. La séparation se stabilise, et les marges diminuent. Au cours des premiers pas (sauf pour le premier) ces marges sont très larges.

Le principe est donc de commencer l'optimisation uniquement avec les contraintes C , et de remplacer les contraintes C en C' uniquement lorsque la stabilisation est suffisante pour être sûr, ou presque sûr qu'on ne reviendra pas en arrière. Le critère heuristique que nous adoptons

est de comparer si le remplacement de la contrainte C en une contrainte C' représente un gain supérieur à celui obtenu par le dernier recrutement d'une contrainte C . Mais on ne recrute jamais que des contraintes C .

Ainsi l'algorithme devient :

```

doubleConstraintOneByOneChunking( $I$ )
( $\hat{\mathbf{x}}_0, \hat{J}_0$ ) := initialize( $I$ )
while (( $i_n, d_n$ ) := leastSatisfiedCConstraint( $\hat{\mathbf{x}}_n, I$ ) not nil )
    ( $\hat{\mathbf{x}}_{n+1}, \hat{J}_{n+1}$ ) := recursiveProjection( $\hat{J}_n + i_n, i_n$ )
    changeConstraints ( $\hat{J}_{n+1}, d_n$ )
     $n := n+1$ 
end while
return ( $\hat{\mathbf{x}}_n, \hat{J}_n$ )

```

Les différences essentielles avec l'algorithme de décomposition un par un standard sont :

- La fonction *leastSatisfiedCConstraint()* renvoie non seulement l'indice de la contrainte, mais aussi la distance correspondante (d_n).
- On a ajouté la fonction *changeConstraints()*, qui va changer une contrainte C en une contrainte C' ou inversement, selon que ce changement fait gagner plus que d_n ou pas.

Notons que cette méthode ne requiert pas d'utiliser la minimisation récursive. Tout autre méthode d'optimisation pour chaque *chunk* pourrait convenir. Cependant, cette approche convient particulièrement bien à la minimisation récursive pour les *svm L1*, car le calcul des multiplicateurs de Lagrange après le passage d'une contrainte de C à C' ne nécessite pas de changer la base orthonormale. Pour les *svm L2*, la situation est moins favorable puisque le passage d'une contrainte de C à C' oblige à recalculer la base orthogonale (au moins une partie), ce qui peut être très coûteux.

Pour l'instant, la méthode ne garantit pas que le passage d'un type de contrainte à l'autre soit irréversible. En effet, il peut se produire des allers et retours entre une contrainte C et son alter ego C' . Nous n'avons pas non plus démontré la convergence de cet algorithme, bien que nous n'ayons jamais rencontré de bouclage au cours des plusieurs centaines de tests que nous avons effectués.

3.4.4 Tests de la méthode

Nous reprenons tout d'abord notre exemple initial en dimension 2, puis nous présentons une batterie plus large de tests dans des dimensions différentes.

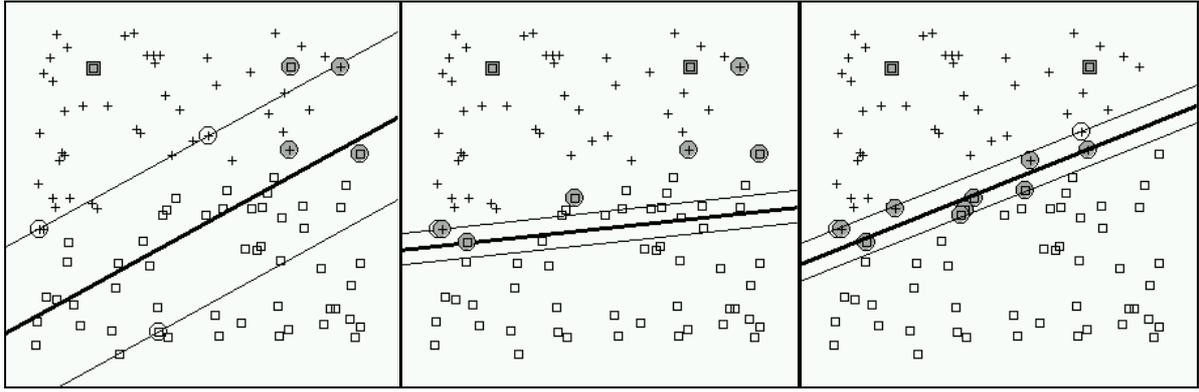


Figure 3.8: Décomposition un par un avec double contrainte pour une *svm* L1. L'échantillon est le même que celui des figures précédentes. Les deux premières figures représentent l'état de la machine lorsque des contraintes sont changées de C en C' (ici $C = 4$ et $C' = 1$), le point correspondant à la contrainte qui change est représenté entouré d'un carré grisé. La figure de droite représente l'état final. Les autres vecteurs support sont entourés d'un cercle, vide lorsque les vecteurs support sont sur le bord de la marge, grisé lorsqu'ils sont à l'intérieur. La machine finale comprend 13 vecteurs support (2 sur les bords, 9 C , 2 C').

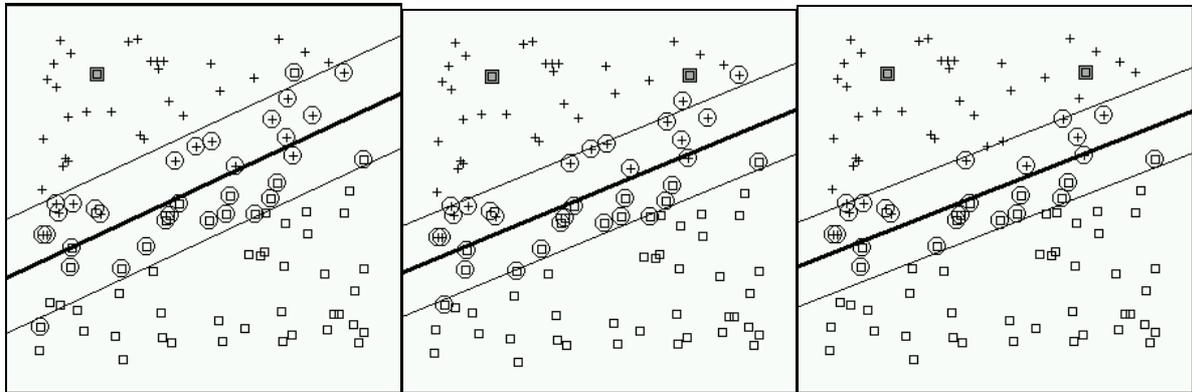


Figure 3.9: Chunking un par un avec double contrainte pour une *svm* L2. L'échantillon est le même que celui des figures précédentes. Les deux premières figures représentent l'état de la machine lorsque des contraintes sont changées de C en C' (ici $C = 10$ et $C' = 1$), le point correspondant à la contrainte qui change est représenté entouré d'un carré grisé. La figure de droite représente l'état final. Les autres vecteurs support sont entourés d'un cercle. La machine finale comprend 27 vecteurs support (25 C et 2 C').

La figure 3.10 indique une comparaison des performances en généralisation sur des apprentissages faits à partir d'échantillon de 500 points bruités uniformément, avec différents taux de bruit (5%, 10%, 15%), et de dimensions différentes (10, 20, 30, 40, 50). Il s'agit de *svm* L2. On constate que la méthode en double contrainte apporte une amélioration sensible.

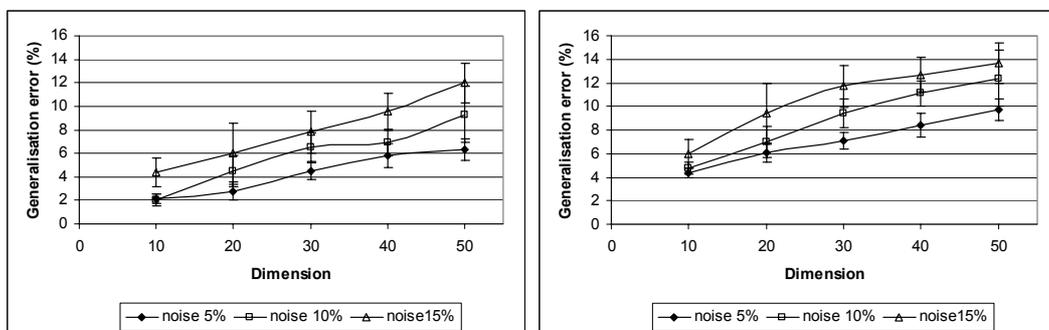


Figure 3.10: A gauche : apprentissage avec double contrainte, à droite apprentissage standard (*svm* L2). L'échantillon comprend 500 points et les résultats sont moyennés sur 20 répliques (l'écart type est figuré par les barres verticales autour des points).

L'utilisation de la méthode de double contrainte permet d'obtenir des améliorations des performances en généralisation pour un bruit uniforme.

Il faut cependant signaler deux bémols importants :

- Lorsque le bruit sur les données est gaussien, l'amélioration des performances devient non significative, en grande dimension, on peut même constater une dégradation des performances. L'analyse montre que cette dégradation est due à une forme de sur-apprentissage : l'erreur sur les données d'apprentissage est en général beaucoup plus faible avec l'algorithme à double contrainte. Seulement, lorsque le bruit est gaussien, les exemples bruités sont souvent près de la frontière, et la machine tend à les prendre excessivement en compte.
- Le temps de calcul est souvent sensiblement augmenté par le passage en double contrainte, même pour les *svm* $L1$, car pour ces derniers, le nombre de tests de facette peut augmenter considérablement dans certains cas. Des analyses plus poussées sont nécessaires pour en identifier les causes.

3.5 Conclusion de ce chapitre

Ce travail sur une approche géométrique de l'optimisation quadratique et son application aux *svm* est encore en devenir. Il faudrait poursuivre les tests pour mieux en identifier l'intérêt pratique, au delà de l'originalité de proposer une nouvelle méthode directe, qui ne passe pas par le dual. Il apparaît quand même dès maintenant que l'algorithme est particulièrement intéressant lorsque le problème comporte de nombreuses contraintes (plusieurs milliers) en faible dimension (jusqu'à 15 environ).

La méthode de double contrainte, en limitant le nombre de vecteurs support peut améliorer les performances de l'apprentissage dans certains cas. Cette méthode peut sans doute être encore améliorée, et son étude théorique reste à faire.

L'intérêt de contrôler ce nombre de vecteurs support peut être encore plus net dans un contexte d'apprentissage « avec oracle », par exemple pour déterminer les frontières de certaines propriétés d'un système dynamique. Dans ce cas, ce serait plutôt la capacité des *svm* à représenter des formes complexes à partir d'un nombre limité de points qui serait le plus important. Il nous semble donc important de poursuivre ces efforts. Nous reviendrons sur ce point dans les perspectives.

4 Les réseaux géométriques dynamiques

4.1 Vers des organismes mathématiques ?

Ce chapitre aborde la question suivante : Peut-on construire des systèmes dynamiques qui, à partir d'éléments simples, constituent des formes et rendent compte de propriétés statistiques intéressantes de distributions de points ? Dans la perspective du mouvement de recherche intitulé « vie artificielle », l'idée est de construire des « organismes mathématiques » qui se développent dans des espaces multidimensionnels sur des distributions de points. Mais le développement de ces organismes doit permettre de détecter des caractéristiques saillantes et robustes des échantillons considérés. Trois thèses ont été consacrées à ces recherches (T. Fuhs (Fuhs 1998), E. Monneret (Monneret 1995) et la mienne), qui tentent de tisser des liens entre différents domaines (statistiques, vie artificielle, optimisation), tout en gardant un intérêt particulier pour la forme, la géométrie.

Parmi ces influences, mentionnons celle du concept « d'autopoïèse », proposé dans les années 70 par H. Maturana et F. Varela (Varela 1989). L'ambition de ces deux chercheurs était d'identifier les dynamiques fondamentales de la vie, et de les transcrire dans un modèle abstrait. Grossièrement, un système autopoïétique possède deux propriétés fondamentales : il se définit par un réseau de dynamiques de génération de ses composants, et ces dynamiques créent une unité topologique qui contraint de manière essentielle les dynamiques de génération de ses composants. La définition rappelle certaines visions de la complexité, qui se focalisent sur la constitution spontanée de structures exerçant une action en retour sur les dynamiques élémentaires. Notons en outre que la définition de l'autopoïèse mentionne explicitement l'idée d'une forme (unité topologique), comme manifestation de ce niveau d'organisation supérieur. Cette définition a provoqué énormément de débats, dont certains ont animé la communauté Vie Artificielle, que nous ne pouvons reprendre ici. L'une des difficultés nous semble être liée au caractère trop général de la définition, et il nous semblait intéressant de chercher à la préciser, quitte à en affaiblir un peu la portée.

La définition des réseaux géométriques dynamiques (RGD) cherche à répondre à cette préoccupation (Deffuant 1993; Deffuant et Fuhs 1993; Deffuant et Monneret 1993; Deffuant et al. 1995). L'idée centrale est que la structure de niveau supérieur se décrit comme une forme géométrique issue de l'assemblage de formes élémentaires, qui sont soumises à des dynamiques de génération, d'élimination et éventuellement de liaison. Ainsi, il nous apparaissait que cette idée de niveaux pouvait prendre un contenu mathématique plus précis, tout en gardant une grande richesse de comportements dynamiques.

Comme je l'ai montré dans ma thèse (Deffuant 1992), cette définition des RGD fournit aussi un point de vue différent sur beaucoup de modèles qui se situent à la croisée des mathématiques et de l'informatique. Par exemple, certains modèles de fractals, comme la courbe de Koch, entrent dans la définition des réseaux géométriques dynamiques, car ils sont fondés sur des dynamiques de transformation simples d'objets géométriques. Ce n'est pas le cas d'autres fractals, comme les ensembles de Julia, qui reposent sur le comportement d'un système dynamique particulier dont on fait varier les paramètres. Beaucoup de modèles connexionnistes peuvent aussi avoir une interprétation géométrique, comme notamment les

réseaux de Kohonen (Kohonen 1982). Nous verrons à la fin de ce chapitre que les *svm* peuvent également être considérées comme des réseaux géométriques dynamiques.

Les réseaux géométriques dynamiques (RGD) sont aussi à rapprocher des modèles individus-centrés. En effet, comme ces derniers, ils se définissent à partir de la dynamique individuelle de composants, qui évoluent en fonction de leur environnement (comprenant notamment d'autres composants). On s'intéresse particulièrement à la manière dont ces dynamiques individuelles se combinent pour former des comportements collectifs cohérents. Les RGD se caractérisent par des composants géométriques simples, qui évoluent dans un espace euclidien. L'évolution peut consister en déplacement ou déformation des composants géométriques, mais aussi des dynamiques de création, de destruction ou de liaison de ces composants. On espère ainsi pouvoir caractériser mathématiquement les formes globales obtenues par leurs interactions.

On le voit, la famille des RGD est très large. Trop peut-être. Identifier cette famille se justifiera si ce point de vue permet de développer des méthodes assez générales pour étudier et éventuellement contrôler le fonctionnement de ces modèles. Dans ce chapitre, nous présentons deux nouveaux modèles, les membranes de perceptrons et les fibres adaptatives, directement suggérés par ce cadre. Nous utilisons ces modèles pour définir certaines saillances de distributions de points dans l'espace, et nous insistons sur les avantages que procure la perspective RGD pour gérer le « dilemme de l'apprentissage ». Ce chapitre s'inspire des thèses de E. Monneret, T. Fuhs et du papier écrit en collaboration avec F. Varela (Deffuant 1995) sur les réseaux semi-algébriques. La méthode pour définir les ensembles d'apprentissage des membranes de perceptrons afin de minimiser l'erreur de classement est cependant nouvelle.

4.2 Le dilemme de l'apprentissage

Pour mener à bien efficacement cette tâche de définition de saillances, les modèles doivent gérer correctement le « dilemme de l'apprentissage ». Dans la perspective des RGD, ce dilemme peut être exprimé de la manière suivante :

- Il est important de permettre la création de nombreux composants qui puissent s'adapter aux caractéristiques particulières de la distribution de points, sinon, ils risquent d'en donner une image trop caricaturale,
- Il faut en même temps que la forme globale qu'ils définissent soit la plus simple possible, de manière à rendre compte uniquement des régularités significatives de la distribution.

Ce dilemme peut s'exprimer plus rigoureusement sous la forme du dilemme « biais – variance ». L'erreur moyenne faite par une procédure d'apprentissage sur un nouveau point \mathbf{x} (tiré selon la même distribution que celle dont est issu l'échantillon) se décompose en la somme de (voir (Fuhs 1998) pour des précisions et la bibliographie):

- la différence entre la moyenne de la réponse en \mathbf{x} , calculée en faisant varier tous les échantillons de taille donnée l , tirés selon la distribution considérée, et la moyenne de la distribution en ce point (le biais).
- La moyenne des différences entre la réponse obtenue en \mathbf{x} et le tirage correspondant selon la distribution (la variance), en faisant varier tous les échantillons de taille donnée l , tirés selon la distribution considérée.

Cette expression mathématiquement plus précise rejoint la description précédente du dilemme. En effet, ajouter des degrés de liberté au modèle accroît ses chances d'être proche en moyenne de la réponse moyenne de la distribution cible, donc de diminuer le biais. Cependant, cet accroissement des degrés de liberté du modèle a de bonnes chances d'en accroître la variabilité selon les échantillons, et donc la variance, ce qui diminuera ses chances d'être performant.

La gestion de ce dilemme se traduit dans les RGD par la convergence vers un équilibre entre des dynamiques de création de composants (diminution du biais) et d'autres dynamiques qui tendent à les éliminer ou à les simplifier (diminution de la variance).

4.3 Les fibres adaptatives : des composants de dimension 1

Les fibres adaptatives, développées par E. Monneret, ont pour objectif de rendre compte de la saillance de distributions de points autour d'une courbe à une dimension. On suppose donc que l'on dispose d'un échantillon de points $S = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_l)$ de l vecteurs de \mathbb{R}^d , disposés autour d'une courbe quelconque de dimension 1.

Les composants sont des segments de dimension 1, chacun défini par un couple de points dans l'espace. On définit donc la fibre à un instant t par :

$$F(t) = \{(\mathbf{a}_1, \mathbf{b}_1), (\mathbf{a}_2, \mathbf{b}_2), \dots, (\mathbf{a}_n, \mathbf{b}_n)\}. \quad (49)$$

Les segments se déforment sous l'influence de la distribution des points, par l'intermédiaire d'une descente de gradient stochastique, qui a pour effet de minimiser la distance entre le segment et les points qui sont dans son entourage. En outre des dynamiques de création, destruction et liaison des segments sont définis de manière à obtenir de bons compromis du dilemme de l'apprentissage.

4.3.1 Dynamiques de déformation des segments

Cette dynamique a lieu sur un temps court, et s'inspire des algorithmes de gradient stochastiques souvent utilisés dans les réseaux connexionnistes. Les points de l'échantillon sont parcourus dans un ordre aléatoire, et modifient les paramètres du segment le plus proche. Ainsi, la procédure est très proche de celle utilisée en « quantification vectorielle ». La différence est que la quantification vectorielle utilise des points au lieu de segments. La modification du segment est faite de manière à minimiser une fonction E de la distance entre le point et le segment.

$$E(\mathbf{x}, (\mathbf{a}_i, \mathbf{b}_i)) = \exp(-\lambda d(\mathbf{x}, (\mathbf{a}_i, \mathbf{b}_i))). \quad (50)$$

Où $d(\mathbf{x}, (\mathbf{a}_i, \mathbf{b}_i))$ est la distance du point \mathbf{x} au segment, et λ un paramètre de l'algorithme. La déformation provoquée par le point \mathbf{x} est donnée par :

$$\begin{aligned} \mathbf{a}_i &:= \mathbf{a}_i - \mu \frac{\partial E(\mathbf{x}, (\mathbf{a}_i, \mathbf{b}_i))}{\partial \mathbf{a}_i}, \\ \mathbf{b}_i &:= \mathbf{b}_i - \mu \frac{\partial E(\mathbf{x}, (\mathbf{a}_i, \mathbf{b}_i))}{\partial \mathbf{b}_i}. \end{aligned} \quad (51)$$

Où μ est un paramètre de l'algorithme.

On définit un cycle par la répétition de l (taille de l'échantillon) tirages aléatoires des points, et les modifications correspondantes.

4.3.2 Les dynamiques de renouvellement des composants

Les procédures de création et de duplication de segments sont appliquées après chaque cycle de modification des segments, avec une probabilité décroissante au cours du temps. Les procédures d'élimination et de liaison sont systématiquement appliquées après chaque cycle. L'idée métaphorique est que les segments ont besoin des points de la distribution pour subsister. Ils sont en compétition pour recouvrir le maximum de points. Ceux qui ont un nombre de points insuffisant (inférieur à un seuil m , paramètre de l'algorithme) dans leur voisinage sont éliminés. Nous décrivons maintenant ces procédures plus en détails.

- La création des composants doit permettre de rendre compte de parties de la distribution qui ne sont pas recouvertes par des segments. Pour cela, on teste si un segment créé à partir d'un point et de son plus proche voisin aurait un nombre de points supérieur à un seuil (paramètre de l'algorithme). Si c'est le cas, on crée effectivement le segment.
- De plus, les segments sont dupliqués, avec la même probabilité décroissante, c'est à dire qu'un segment en donne deux nouveaux, en se cassant en son centre.
- Inversement, la procédure d'élimination élimine un segment qui n'a pas suffisamment de points dans son voisinage. Allié à la duplication, l'élimination empêche la croissance intempestive de segments dont certaines parties ne seraient par « nourries » par des points de la distribution.
- La procédure de liaison fusionne les extrémités des segments, lorsqu'elles sont à une distance inférieure à deux fois la distance moyenne d'un point à son plus proche voisin dans la distribution. Cette procédure a pour effet de « régulariser » la forme globale obtenue, et constitue une différence importante vis à vis de la quantification vectorielle (où ce type de procédure n'est évidemment pas possible puisque les objets manipulés sont des points). De plus, cette régularisation a des effets très importants sur la qualité du compromis entre biais et variance.

L'algorithme global d'apprentissage des fibres adaptatives est donc le suivant :

ApprentissageFibres ($S = (x_1, x_2, \dots, x_l)$)

$F :=$ Initialisation(S)

$T := T_0$

Tant que $T > 0$

Répéter l (taille de l'échantillon) fois :

choisir x_i au hasard

ajuster le segment correspondant

avec la probabilité T :

générer de nouveaux segments

dupliquer des segments

éliminer les segments

lier les segments

$T := T - \alpha/n$ (α est un paramètre et n le nombre de segments)

Fin tant que

Comme on le constate sur la figure 4.1, les fibres adaptatives s'ajustent rapidement à des distributions de points autour de courbes de dimension 1.

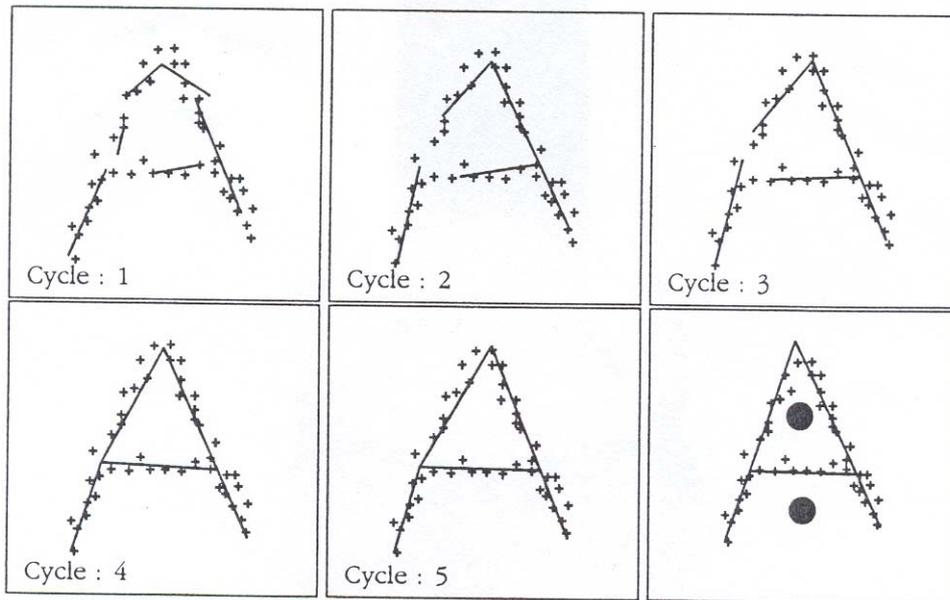


Figure 4.1 : Les fibres adaptatives. Au premier cycle, les segments sont créés, et ensuite se modifient et la figure se simplifie grâce aux dynamiques de liaisons. Les deux points noirs de la dernière figure repèrent certains ensembles de singularités.

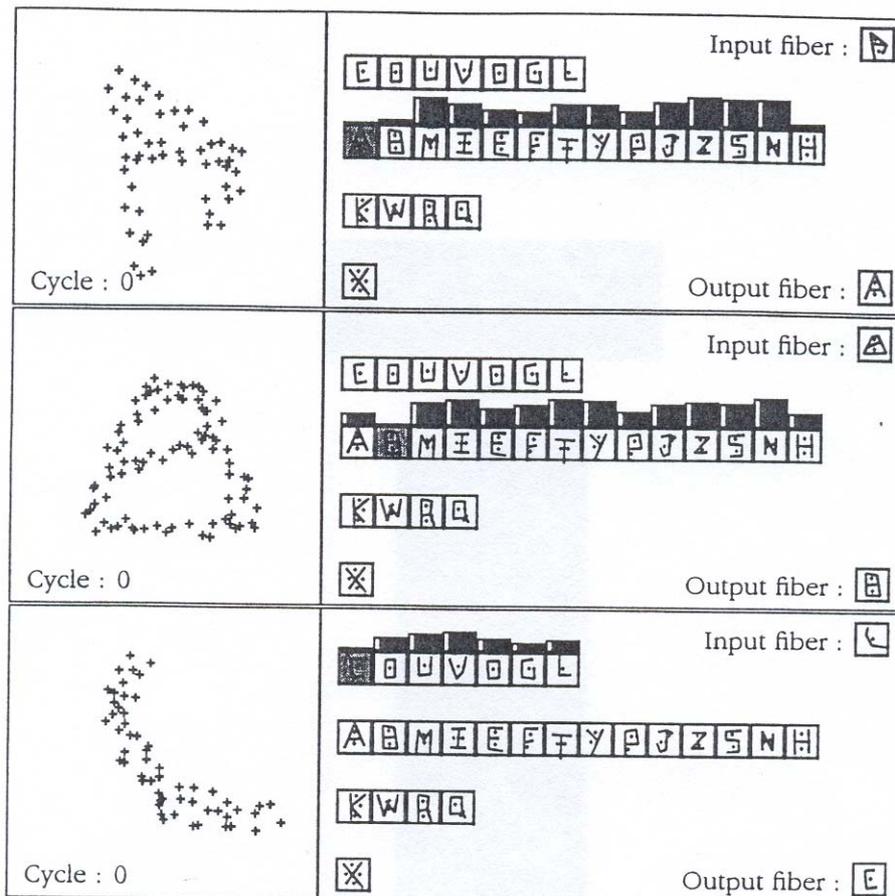


Figure 4.2 : Mémorisation des formes par groupes présentant des caractéristiques similaires. Calcul de reconnaissance et activation des formes selon leur ressemblance à l'entrée.

4.3.3 Mémorisation des formes, et reconnaissance selon leurs singularités

Une fois qu'une fibre adaptative est stabilisée sur un échantillon, il est possible d'en extraire certaines caractéristiques géométriques, notamment ses singularités : les extrémités appartenant à 1 seul segment (singularité 1), à 2, 3 etc...

Etienne Monneret a développé une méthode inspirée elle-même des RGD pour faire cette catégorisation. En effet, il génère de nouveaux composants géométriques qui interagissent maintenant non plus avec les points de la distribution, mais avec les segments de la fibre après convergence. Ces composants géométriques sont des points qui sont attirés par les singularités (les extrémités des segments). Ils sont éliminés lorsqu'ils traversent un segment et ils fusionnent lorsqu'ils sont proches les uns des autres. A convergence, cet ensemble de points rend compte de grandes caractéristiques morphologiques de la distribution de points.

Ces caractéristiques globales sont utilisées pour établir des regroupements entre les formes, afin ensuite d'en accélérer la reconnaissance. La reconnaissance en tant que telle dépend à la fois de ces points, et des singularités auxquelles ils se rattachent.

4.4 Les membranes de perceptrons

J'ai commencé à travailler sur les membranes de perceptrons durant ma thèse, et c'est à partir de leur exemple que nous avons collectivement (avec E. Monneret et T. Fuhs principalement) imaginé leur généralisation aux RGD. Les membranes de perceptrons visent à résoudre un problème de discrimination. L'échantillon dont on dispose est donc du type : $S = ((\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_l, y_l))$ de l vecteurs de \mathbb{R}^d , chacun étant associé à une classe. Nous supposons dans un premier temps que la classe ne peut prendre que deux valeurs ($y_i = 1$ ou $y_i = -1$).

Le terme « perceptron » fait référence à différents travaux qui remontent au tout début du connexionnisme. Le perceptron de Rosenblatt, développé dans les années 50, comprenait une rétine et une cellule de décision qui prenait en entrée les regroupements de détecteurs de la rétine. Minsky et Papert ont fortement critiqué ce dispositif en 1969, en soulignant la limitation à des problèmes linéaires de l'algorithme d'apprentissage de la cellule de décision. Dans les années 80, l'algorithme de rétropropagation du gradient permet d'ajuster plusieurs couches d'un réseau, et donc des fonctions non linéaires. La rétropropagation du gradient résout en effet le problème de la répercussion d'une erreur sur les différents poids du réseau (credit assignment problem).

Les membranes de perceptrons offrent une solution géométrique au problème de l'attribution des responsabilités d'une erreur à différents composants. T. Fuhs et moi avons montré que cette attribution géométrique de l'erreur permettait une optimisation globale de l'erreur considérée. En outre, ces modèles permettent naturellement une évolution de leur structure, afin de l'ajuster au mieux aux caractéristiques de l'échantillon et espérer ainsi gérer correctement le dilemme biais – variance.

Je présente tout d'abord l'objet géométrique et ses dynamiques d'ajustement, puis les dynamiques de renouvellement des composants. Il s'agit d'une variante un peu différente de celle qui est présentée dans (Deffuant 1995), qui se rapproche des listes de décision, et permet de simplifier considérablement la structure de la membrane dans certains cas. La parenté avec les arbres de décision apparaît aussi plus clairement dans cette variante.

4.4.1 Définition des membranes de perceptrons

Une membrane de perceptrons est définie par une liste ordonnée de polyèdres convexes auxquels sont associés une classe. Une classe dite « externe » est en outre spécifiée. Chaque polyèdre est défini comme l'intersection de demi-espaces, eux-même définis par une séparation linéaire (un hyperplan). Les séparations linéaires peuvent être partagées par plusieurs polyèdres, constituant des liaisons entre eux. Ainsi, on définit la membrane par l'ensemble des hyperplans, puis les polyèdres convexes par des ensembles d'indices des hyperplans qui les constituent (un signe (+1 ou -1) indiquant le sens dans lequel la séparation est prise est aussi nécessaire, mais nous l'omettons ici pour simplifier la présentation). Précisons les notations :

$$H = \{(s_i = (\mathbf{a}_i, b_i)), i \in \{1, \dots, n\}\}. \quad (52)$$

Est l'ensemble des hyperplans, qui sont définis par un vecteur normé et un scalaire. Pour un séparateur $s_i = (\mathbf{a}_i, b_i)$ et un point \mathbf{x} de l'espace, nous notons :

$$s_i(\mathbf{x}) = \mathbf{a}_i \cdot \mathbf{x} + b_i. \quad (53)$$

A partir de ces séparateurs, on définit un ensemble de polyèdres :

$$P = \{(P_k, c_k), P_k \subset \{1, \dots, n\}, c_k \in \{1, -1\}, k \in \{1, \dots, m\}\}. \quad (54)$$

Chaque polyèdre est associé à une classe c_k . Un point \mathbf{x} appartient à un polyèdre défini par l'ensemble d'indices P_k si et seulement si :

$$s_i(\mathbf{x}) \geq 0, i \in P_k. \quad (55)$$

Enfin, on définit la classe « externe » : $c_e \in \{1, -1\}$. Une membrane M est définie par le triplet suivant :

$$M = \{H, P, c_e\}. \quad (56)$$

L'algorithme par lequel la membrane associe une classe à un point est le suivant :

Classe(M, \mathbf{x})

Pour k de 1 à m faire :

Si P_k contient \mathbf{x} , alors retourner c_k

Retourner c_e .

Cet algorithme est similaire à celui des listes de décision. On renvoie la classe du premier polyèdre (dans l'ordre de la liste P) qui contient le point \mathbf{x} . Si aucun des polyèdres ne contient le point, on renvoie la « classe externe » c_e . On peut le voir comme un arbre de décision qui n'aurait qu'une seule branche et dont les nœuds seraient des polyèdres au lieu de séparations linéaires.

4.4.2 Dynamiques d'ajustement des hyperplans

Le principe de l'attribution du crédit géométrique est de définir des ensembles d'apprentissage pour chacun des séparateurs. Nous avons proposé deux manières de définir ces ensembles d'apprentissage, qui dépendent du critère global à optimiser : cela peut être

l'erreur de classement ou une erreur fondée sur la distance des points à la surface de décision et leur classe.

Remarquons que ce problème d'ajustement d'une surface linéaire par morceaux est difficile (il est NP complet), et fortement non-linéaire. En effet, la relation entre un point et un séparateur connaît des changements brusques lorsque ce point bouge dans l'espace. Bennett a proposé de le résoudre (dans le cas des arbres de décision) en utilisant des algorithmes de type tabou, par exemple.

La solution que nous proposons est différente. Elle consiste à décomposer un problème non-linéaire en une itération de sous-problèmes linéaires. Le problème étant de définir ces sous problèmes afin qu'ils garantissent de résoudre le problème global.

4.4.2.1 Décomposition de l'erreur

Nous supposons que nous définissons un ensemble d'apprentissage L_i associé à chaque séparation (\mathbf{a}_i, b_i) , et nous examinons la manière dont change l'erreur si la séparation est modifiée. Nous nous focalisons donc sur une séparation particulière s et son ensemble d'apprentissage L , nous supposons qu'elle est modifiée suite à une optimisation et devient s' , son ensemble d'apprentissage devenant L' . Le problème est en effet d'éviter que l'optimisation de la séparation sur l'ensemble d'apprentissage provoque des erreurs supplémentaires en dehors de l'ensemble d'apprentissage. L'évolution de l'erreur globale est donnée par la somme des évolutions sur l'intersection des deux ensembles, le complémentaire de leur réunion dans l'espace entier Ω , et le complémentaire de leur intersection dans leur réunion.

$$\Delta E_M(\Omega) = \Delta E_M(L \cap L') + \Delta E_M((L \cup L') - (L \cap L')) + \Delta E_M(\Omega - (L \cup L')). \quad (57)$$

Dans cette équation, l'indice M indique que la fonction d'erreur est calculée en utilisant les réponses fournies par la membrane.

Pour permettre une équivalence entre l'optimisation locale et l'optimisation globale, une première condition est que l'erreur définie par la membrane totale et par la séparation soient les mêmes sur les ensembles d'apprentissage :

$$E_M(L) = E_s(L). \quad (58)$$

En effet, sinon, l'optimisation de la séparation sur l'ensemble d'apprentissage n'a aucune garantie d'apporter une optimisation globale.

Cette condition implique :

$$\Delta E_M(L \cap L') = \Delta E_s(L \cap L'). \quad (59)$$

D'autre part, il faut essayer de limiter la portée des deux derniers termes de la décomposition, si possible les annuler. En effet, s'ils sont nuls, on est sûr que l'optimisation de la séparation entraîne l'optimisation globale de l'erreur, car alors :

$$\Delta E_M(\Omega) = \Delta E_s(L \cap L'). \quad (60)$$

La méthode de définition d'ensembles d'apprentissage dans le cas de l'erreur de classement satisfait cette condition. Dans le cas d'une erreur qui fait intervenir la distance entre l'hyperplan et les points d'apprentissage, nous proposons une définition de l'ensemble d'apprentissage qui garantit que les deux derniers termes restent petits devant le premier, dans certaines conditions.

4.4.2 Définition de l'ensemble d'apprentissage pour l'erreur de classement

Nous supposons ici qu'il est possible de minimiser l'erreur de classement faite par un séparateur seul. Cette hypothèse de travail n'est satisfaite de manière exacte que si l'on se limite a priori à des séparations parallèles aux axes. En effet, dans ce cas, il est possible de parcourir exhaustivement toutes les possibilités de séparations et de choisir la meilleure. Si l'on considère un séparateur linéaire quelconque, la recherche exhaustive devient impossible. En général, le problème n'est résolu que de manière approchée, en utilisant des minimisations de fonctions des erreurs qui sont différentiables. Les méthodes étudiées au chapitre précédent pour optimiser les *svm* en sont un exemple, et notre proposition pour traiter les points aberrants permet de se rapprocher encore de l'optimisation de l'erreur de classement. Nous examinons donc dans ce paragraphe une situation idéale un peu théorique, qui réclamera quelques précautions pour se décliner en pratique.

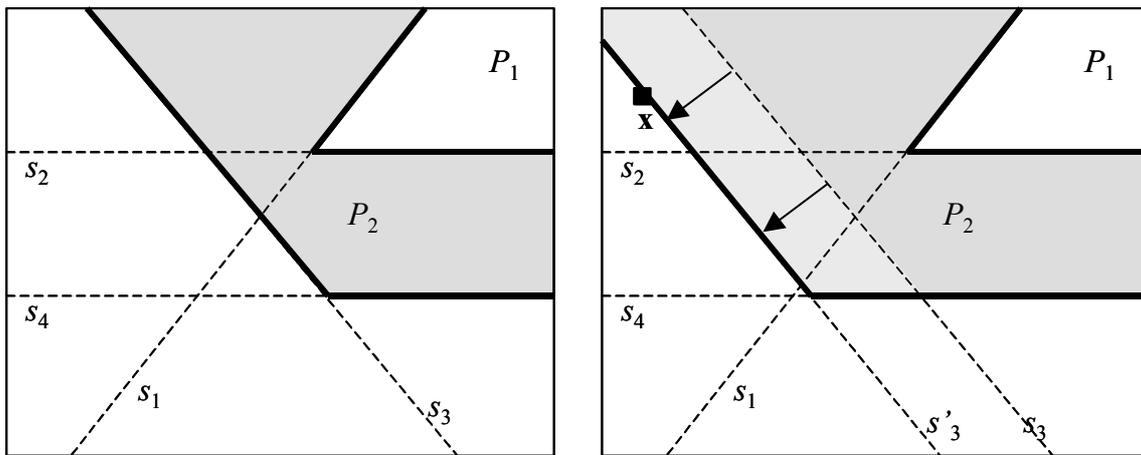


Figure 4.3 : La liste de polyèdres de la membrane est $\{(P_1, 1), (P_2, -1)\}$, $c_e=1$. Le point x appartient à l'ensemble d'apprentissage du séparateur s_3 , car lorsqu'on le translate jusqu'au point x , ce dernier appartient à la nouvelle frontière ainsi définie.

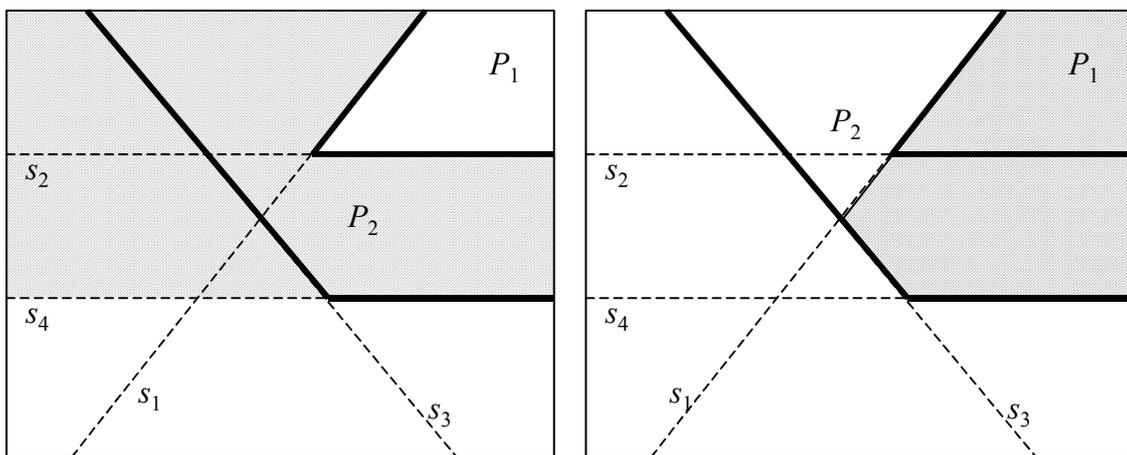


Figure 4.4 : Exemples d'ensembles d'apprentissage pour l'erreur de classement. A gauche celui du séparateur s_3 , à droite, celui du séparateur s_2 . Les ensembles d'apprentissage peuvent se recouvrir, et leur réunion ne donne pas l'ensemble de l'espace.

Nous adoptons une définition de l'ensemble d'apprentissage qui est formellement un peu différente de celle adoptée dans (Fuhs 1998) pour les arbres de décision. Mais il ne s'agit que d'une simplification de la procédure de calcul de l'ensemble, qui donne le même résultat. Une manière de présenter cette définition est de considérer que, pour déterminer si un point x

appartient au domaine d'apprentissage d'un séparateur s , nous translatons le séparateur jusqu'à ce qu'il atteigne \mathbf{x} , si \mathbf{x} est alors sur la frontière de la nouvelle membrane ainsi définie, alors le point est dans l'ensemble d'apprentissage, sinon il n'en fait pas partie.

Nous devons tester la réponse de la membrane en \mathbf{x} sans calculer $s(\mathbf{x})$, mais en supposant que $s(\mathbf{x}) > 0$ puis ensuite en supposant que $s(\mathbf{x}) < 0$. Nous noterons :

- $M(\mathbf{x}, s_+)$ la réponse de M en \mathbf{x} , en supposant que $s(\mathbf{x}) > 0$
- $M(\mathbf{x}, s_-)$ la réponse de M en \mathbf{x} , en supposant que $s(\mathbf{x}) < 0$

Définition : On appelle ensemble d'apprentissage de la séparation s d'une membrane M pour l'erreur de classement, et on note $L(s)$, l'ensemble des points \mathbf{x} de l'échantillon tels que :

$$M(\mathbf{x}, s_+) \neq M(\mathbf{x}, s_-). \quad (61)$$

L'ensemble d'apprentissage ainsi défini a la propriété de ne pas dépendre de la position de s . En effet, sa définition ne dépend pas de la position de s , mais uniquement des autres séparateurs. Nous avons donc :

$$\Delta E_s(L \cap L') = \Delta E_s(L). \quad (62)$$

Bien sûr, s contribue à définir l'ensemble d'apprentissage d'autres séparateurs, et donc lorsqu'il varie, ces ensembles d'apprentissage varient.

D'autre part, en dehors de l'ensemble de l'apprentissage, la réponse de la membrane ne varie pas lorsque s varie. En effet, pour ces points, par définition la valeur de la sortie en s est indifférente.

Nous avons donc :

$$\Delta E_M(\Omega) = \Delta E_s(L \cap L') = \Delta E_s(L). \quad (63)$$

Ce qui garantit que la diminution de l'erreur locale induit directement la même diminution de l'erreur globale (voir Fuhs 98 pour une démonstration complète du théorème).

Un cycle d'optimisation de la membrane s'écrit donc :

CycleAjustement(M)

Pour tous les séparateurs s_i faire :

$$s_i := \text{optim}(s_i, L(s_i))$$

Fin

Où $\text{optim}(s_i, L(s_i))$ est l'algorithme d'optimisation de l'erreur pour un séparateur sur l'ensemble d'apprentissage. (Fuhs 1998) montre que l'application répétée de ces cycles aboutit nécessairement à une stabilisation (l'argument est simplement que l'erreur de classement globale ne peut que baisser si elle ne fait que baisser localement), et donne une majoration du nombre de cycles nécessaires à la stabilisation (l'erreur maximale).

Une telle définition de l'ensemble d'apprentissage offre donc un moyen intéressant pour optimiser globalement une surface linéaire par morceaux. Cependant, il faut bien signaler que la procédure ne garantit pas de parvenir au minimum global. En effet, cette procédure peut être bloquée dans des minima locaux car elle est « aveugle » à certaines parties de l'espace, dont elle ne tient pas compte (les angles morts).

Il est possible d'éviter ces « angles morts » en définissant les ensembles d'apprentissage d'une manière légèrement différente. Cependant, cette définition a des propriétés un peu moins agréables pour la minimisation.

4.4.2.3 Ensemble d'apprentissage pour une erreur fonction de la distance à la membrane. Approximation par la distance de dilatation

Comme nous l'avons souligné au début du paragraphe, la plupart des algorithmes qui optimisent une séparation linéaire se fondent sur une fonction de la distance des points à l'hyperplan (comme les *svm* par exemple). Dans cette perspective, le calcul de l'ensemble d'apprentissage d'un séparateur pourrait se fonder sur le calcul de la distance des points à la surface de décision : la présence dans l'ensemble d'apprentissage serait liée à la présence du séparateur dans le calcul de la distance. Une telle approche est théoriquement possible, mais présente des difficultés pratiques. En effet, le calcul de la distance à la surface de décision implique de calculer la projection d'un point sur de nombreux polyèdres convexes, qui peuvent être longs à déterminer. Nous proposons une approximation de cette distance, dont le calcul est beaucoup plus facile : la distance de dilatation.

Le principe de ce calcul s'apparente à celui qui est utilisé pour déterminer les ensembles d'apprentissage pour l'erreur de classement. Ainsi, pour calculer cette approximation de la distance d'un point \mathbf{x} à la membrane, on considère un par un les séparateurs s et on calcule la distance algébrique $s(\mathbf{x})$, on considère ensuite une membrane fictive dont tous les séparateurs sont translétés de cette valeur. Si le point \mathbf{x} est sur la frontière définie par cette membrane fictive, alors il fait partie de l'ensemble d'apprentissage de s et sa distance de dilatation à la membrane est $s(\mathbf{x})$.

Décrivons plus précisément la membrane dilatée en fonction de s et de \mathbf{x} . Elle est obtenue en tenant compte de la classe de sortie de la membrane au point \mathbf{x} ($M(\mathbf{x})$) :

dilatation(M, s, \mathbf{x})

$M' :=$ copie de M

$c := M(\mathbf{x})$

Pour tous les séparateurs s_i faire :

 Si s_i appartient à un polyèdre P_k tel que $c_k = c$, alors :

$$s'_i = (\mathbf{a}_i, b_i - s(\mathbf{x}))$$

 remplacer s_i par s'_i dans M'

 sinon

$$s'_i = (\mathbf{a}_i, b_i + s(\mathbf{x}))$$

 remplacer s_i par s'_i dans M'

fin pour tous

retourner M'

Pour chaque séparateur on regarde si le point \mathbf{x} est à la frontière de la membrane dilatée. Ce test est semblable à celui qui était pour les erreurs de classement : on force la valeur de $s(\mathbf{x})$ à une valeur positive, puis à une valeur négative, simulant ainsi un passage de la frontière en \mathbf{x} . Plus précisément, l'algorithme est le suivant (illustré par les figures 4.5 et 4.6).

distanceDilatation(M, \mathbf{x})

pour tout séparateur s faire :

$M' := \text{dilatation}(M, s, \mathbf{x})$

Si $M'(\mathbf{x}, s_+) \neq M'(\mathbf{x}, s_-)$ alors retourner $(s, s(\mathbf{x}))$.

fin

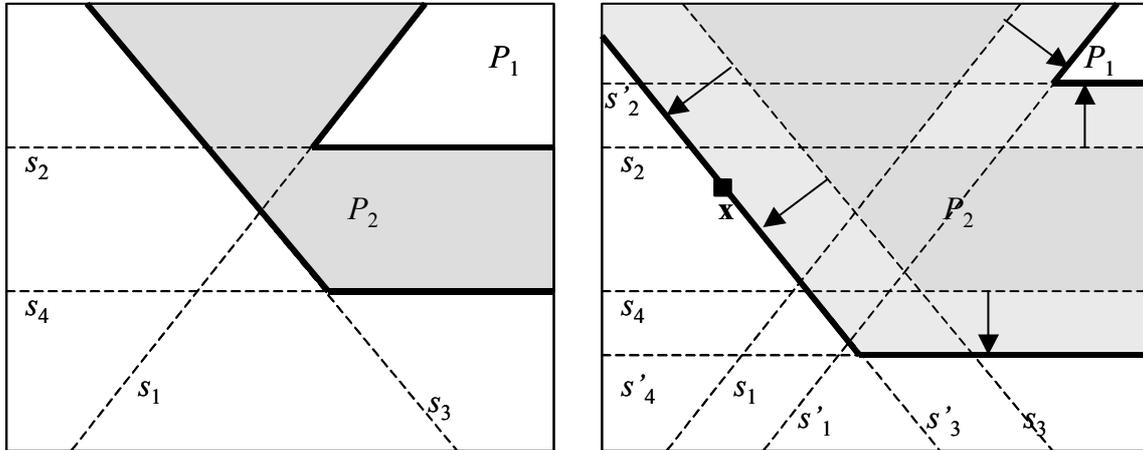


Figure 4.5 : Dilatation de la membrane, déterminée par \mathbf{x} et s_3 . Le convexe P_2 , d'une classe différente de la réponse en \mathbf{x} se dilate alors que le convexe P_1 de même classe que la réponse, se contracte.

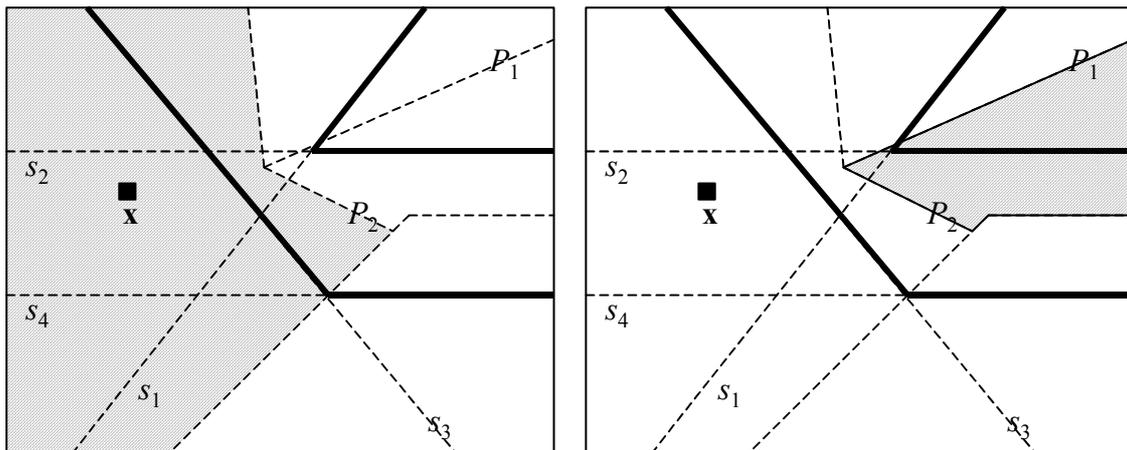


Figure 4.6 : Ensembles d'apprentissages définis par la distance de dilatation. A gauche celui du séparateur s_3 , à droite, celui du séparateur s_2 . Les ensembles d'apprentissage de tous les séparateurs forment une partition de l'espace.

Les ensembles d'apprentissage ainsi définis forment une partition de l'espace, ce qui évite les « angles morts » induits par la méthode précédente. Cependant, si le séparateur est modifié, son ensemble d'apprentissage l'est aussi, puisque cet ensemble est limité par les hyperplans médians qu'il forme avec les autres hyperplans. L'équivalence entre l'évolution locale des erreurs et son évolution globale n'est donc pas obtenue avec cette définition. Cependant, on peut montrer qu'en général les différences entre l'évolution de l'erreur globale et celle de l'erreur locale sont faibles, à condition de procéder par petites modifications des séparateurs.

L'argument central à la démonstration de cette propriété (voir (Fuhs 1998) pour les détails de la démonstration du théorème) est que les différences entre distance globale et distance locale se trouvent dans les ensembles $(L \cup L') - (L \cap L')$, dont le volume est du premier ordre par

rapport au volume de $(L \cap L')$ pour de petits déplacements du séparateur (voir illustration figure 4.7).

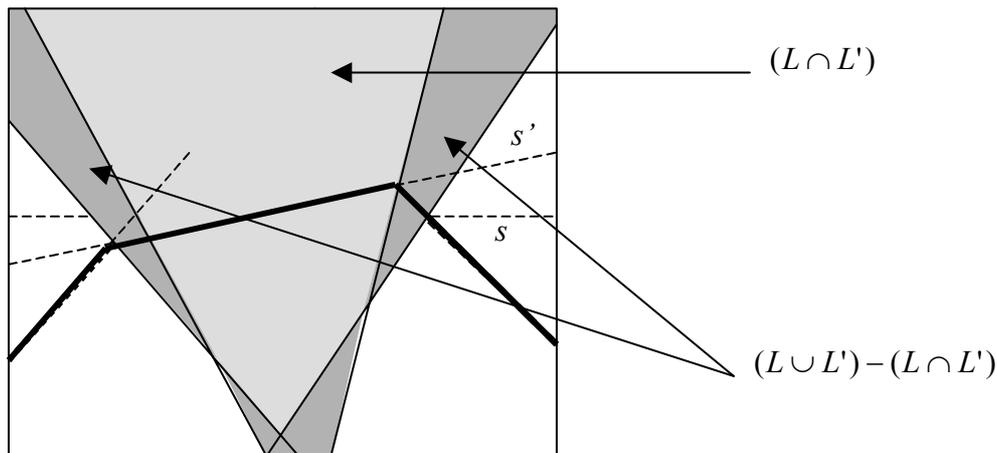


Figure 4.7 : Le séparateur s est transformé en s' , et nous figurons la modification de l'ensemble d'apprentissage qui en résulte. La distance à la membrane est différente de la distance aux séparateurs dans cette zone, mais cette différence est du même ordre que les modifications des séparateurs.

L'utilisation de cette fonction d'erreur implique donc que les méthodes d'optimisation locales sont progressives, comme des descentes de gradient par exemple. Pour définir le nouvel algorithme d'optimisation globale, nous supposons donc qu'on peut définir un cycle d'optimisation locale. Un cycle d'ajustement s'écrit donc :

CycleAjustement(M)

Pour tous les séparateurs s_i faire :

$$s_i := \text{cycleOptim}(s_i, L(s_i))$$

Fin

Où $\text{cycleOptim}(s_i, L(s_i))$ est un cycle d'optimisation locale du séparateur sur l'ensemble d'apprentissage du séparateur.

Si l'on s'intéresse à cette optimisation pour elle-même, la condition d'arrêt n'est plus si immédiate que dans le cas de l'erreur de classement. Le théorème garantit cependant une stabilisation progressive. Il est donc possible de définir un seuil de modification des séparateurs en dessous duquel on considère que la convergence est effective. En pratique, la méthode fonctionne très correctement avec des méthodes de descente de gradient sur une erreur logistique par exemple.

4.4.3 Renouvellement des composants

Les dynamiques de renouvellement des composants des membranes de perceptrons suivent les mêmes principes que ceux exposés dans le paragraphe consacré aux fibres adaptatives : des composants sont générés afin d'accroître la plasticité du modèle, ces dynamiques sont en concurrence avec des dynamiques simplificatrices, qui éliminent les composants paraissant redondants ou inutiles.

Le principe de l'élimination repose sur un seuil minimum du nombre de points de chaque côté de la séparation dans l'ensemble d'apprentissage d'un séparateur. En deçà de ce minimum, le rendement du séparateur est jugé trop faible. Notons que cette évaluation du rôle des séparateurs dans l'ensemble du modèle est possible grâce à la décomposition en ensembles

d'apprentissage. En effet, une telle évaluation est beaucoup moins directe dans d'autres modèles comme les réseaux connexionnistes ou les arbres de décision.

Signalons une différence importante entre ces dynamiques de renouvellement et celles qui animent les fibres adaptatives : l'élimination d'un séparateur peut provoquer des modifications importantes de la fonction de décision globale, même si le nombre de points qu'il discrimine est très faible. En effet, cette élimination peut modifier sensiblement la géométrie d'un ou plusieurs polyèdres, de manière plus ou moins indésirable. Nous avons développé des procédures particulières pour limiter de tels effets, et qui conduisent à lier les polyèdres entre eux en partageant des séparateurs. De même que pour les fibres adaptatives, ces liaisons tendent à simplifier, à régulariser la surface, et donc augmentent probablement ses performances.

Les procédures de renouvellement des composants (pour les détails, voir (Deffuant 1995) sont donc :

- Génération de nouveaux polyèdres convexes. On considère un point de l'échantillon dont la majorité des plus proches voisins sont de la même classe. On construit le polyèdre en ajoutant des hyperplans qui isolent le point considéré des points de classe différente. Fuhs 98 envisage d'utiliser des algorithmes inspirés du développement d'arbres de décision pour créer des polyèdres. Cette dernière approche peut avoir l'avantage de créer moins de séparations, mais il faudrait la tester plus avant.
- Duplication de séparateurs. Un nouveau séparateur peut être créé et ajouté à la membrane, de façon à accentuer la convexité du polyèdre (ajout à la liste du polyèdre), ou au contraire créer une concavité (ajout d'un nouveau polyèdre qui est égal à celui dont il est issu, mis à part pour un séparateur qui est légèrement différent).
- Liaison entre polyèdres convexes en partageant un séparateur. La procédure de création de liaisons inspecte tous les couples de polyèdres dont l'intersection est non vide, et teste tous les séparateurs non partagés afin de déterminer si leur partage améliore la surface. Le critère d'amélioration est que la partie de membrane amputée à la suite du partage contient un taux d'erreurs supérieur à un seuil, ou un nombre de points inférieur au seuil d'élimination.
- Elimination de séparateurs. On élimine les séparateurs dont l'un des côtés de l'ensemble d'apprentissage contient un nombre de points inférieur au seuil d'élimination. Une procédure de recherche des liaisons est lancée après chaque élimination, afin de limiter les perturbations engendrées par une élimination.

De même que pour les fibres, on diminue progressivement la génération de polyèdres et la duplication de séparateurs, de manière à garantir la convergence du processus. La procédure est donc la suivante :

ApprentissageMembrane ($S = ((x_1, y_1), (x_2, y_2), \dots, (x_l, y_l))$)

$M :=$ Initialisation(S)

$T := T_0$

Tant que $T > 0$

CycleAjustement(M)

avec la probabilité T :

générer de nouveaux polyèdres

dupliquer des séparateurs

éliminer les séparateurs

lier les polyèdres

$$T := T - \alpha/n \text{ (}\alpha \text{ est un paramètre et } n \text{ le nombre de séparateurs)}$$

Fin tant que

La figure 4.8 donne un exemple de développement de membranes. D'autres exemples sont disponibles dans (Deffuant 1995).

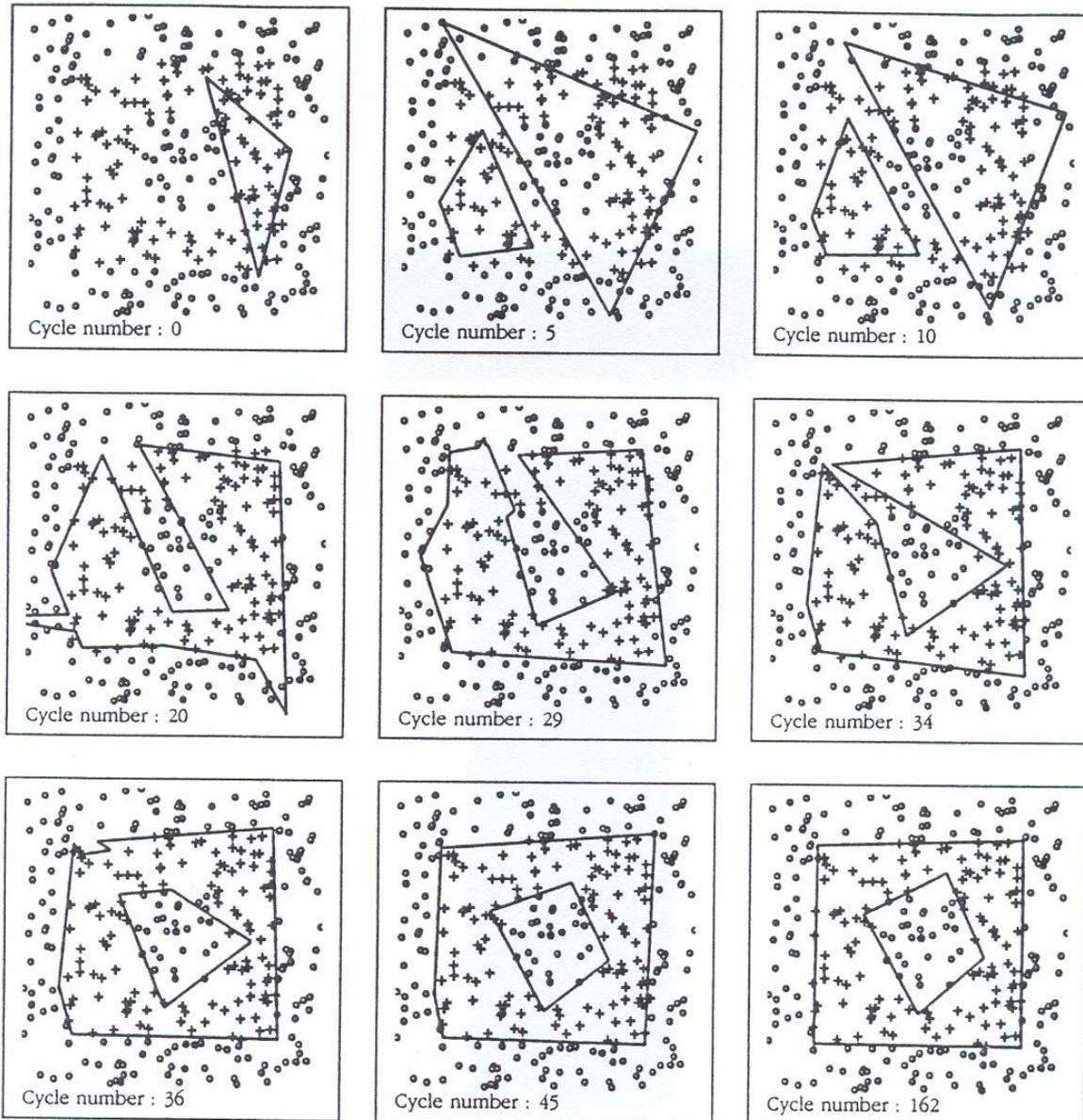


Figure 4.8 : développement d'une membrane de perceptrons sur un échantillon de dimension 2 ayant la forme d'un carré percé. Dans ce cas, on utilise une erreur de dilatation et un ajustement à l'aide d'un algorithme de gradient. Comme pour les fibres, l'apprentissage comprend deux phases : tout d'abord de nombreux polyèdres et séparateurs sont créés, puis dans la seconde phase, la forme se régularise et se simplifie.

4.5 Conclusion de ce chapitre

La partie la plus aboutie de ce travail est sans doute l'optimisation des surfaces linéaires par morceaux pour la discrimination. En effet, elle bénéficie de plusieurs résultats théoriques (sur

l'erreur de classement et sur l'erreur de dilatation) qui donnent une solution originale à un problème difficile. Cette réussite illustre l'intérêt d'une approche géométrique de ces problèmes. Sans doute il faudrait faire beaucoup de tests supplémentaires pour mieux évaluer son intérêt pratique dans des problèmes concrets.

Une autre idée structurante de ce travail est le renouvellement des composants, en particulier la compétition entre génération et élimination pour parvenir à un équilibre, permettant de trouver de bons compromis entre biais et variance. Malheureusement il nous faut convenir que cette idée n'a pas été non plus poussée jusqu'au terme de ses possibilités. Avec T. Fuhs, nous avons tenté de la formaliser comme une coopération entre induction et abduction. En effet, la génération et l'élimination de composants permet de définir un espace de recherche, dans lequel l'ajustement des composants trace un chemin et affine une solution. Cette formulation me semble riche conceptuellement, mais peut être un peu trop générale pour servir de guide rigoureux au développement de nouveaux modèles et de nouvelles techniques. Il est cependant sans doute trop tôt pour tirer toutes les leçons de cette tentative de s'inspirer du vivant pour développer des techniques mathématiques efficaces.

Troisième idée structurante : utiliser la géométrie pour définir rigoureusement les caractéristiques du niveau global (collectif) émergeant de la coopération des comportements individuels. Ici aussi, la démonstration n'est qu'esquissée (avec le deuxième niveau des fibres). Il aurait été parfaitement possible de développer l'équivalent sur les membranes. Les singularités ou les caractéristiques topologiques à considérer auraient simplement été un peu différentes. Mais il y a là un domaine de recherche qui englobe les statistiques en quelque sorte. L'enjeu est de construire des outils qui classifient les distributions statistiques auxquels ils sont confrontés, et sont éventuellement capables d'établir des rapprochements, des analogies. Il s'agit donc d'établir des analogies entre échantillons. Ici encore, il est probablement trop tôt pour conclure sur la pertinence de cette direction de recherche.

Notons enfin que les algorithmes d'optimisation des *svm* du chapitre précédent les rapprochent des réseaux géométriques dynamiques. En effet, la surface de décision est définie par un ensemble de points de l'échantillon (les vecteurs support courants) qui grossit et diminue en fonction des caractéristiques courantes de la surface. On retrouve donc des dynamiques similaires à celles des réseaux géométriques dynamiques, dans lesquels un objet géométrique global (la surface de décision), est construit à partir de composants géométriques simples qui s'assemblent et s'ajustent entre eux. Il n'est pas sûr que la construction de membranes à partir de séparateurs de type *svm* soit très efficace, même si c'est une idée qui vient naturellement (soulignons qu'elle est rendue possible par les algorithmes de minimisation géométriques, car sinon les algorithmes des *svm* seraient inadaptés par exemple dans un arbre de décision). En effet, le passage non linéaire dans des espaces déployés, notamment grâce à des noyaux gaussiens, joue un rôle similaire à celui des membranes, avec probablement plus d'efficacité.

5 Modèles géométriques dynamiques pour la diffusion d'opinions : comparer le modèle individus-centré à son approximation agrégée

Les modèles individus-centrés sont l'objet d'un intérêt croissant en sciences sociales, en particulier afin d'identifier des mécanismes qui engendrent des comportements globaux structurés à partir de dynamiques individuelles. En effet, les organisations humaines sont à la fois engendrées par des comportements individuels, mais aussi structurent fortement ces comportements. Des modèles de populations d'individus, qui engendrent collectivement des structures pourraient aider à comprendre cette caractéristique du fonctionnement social. Ce chapitre est consacré à deux modèles très simples, qui se situent dans cette perspective.

Il s'agit de modèles d'évolution d'opinions sous l'influence des interactions sociales. L'opinion considérée est simplement l'effet d'un processus d'interaction, sans aucune référence à des faits extérieurs. Il s'agit donc d'une dynamique sur des valeurs, a priori déconnectées de faits. De manière sous-jacente, nous retrouvons donc la distinction forte entre faits et valeurs, soulignée au chapitre 2.

Techniquement, ces modèles se rapprochent des réseaux géométriques dynamiques présentés au chapitre précédent. En effet, leurs éléments sont des segments géométriques, de même que dans les fibres adaptatives. Cependant ces segments n'interagissent plus avec une distribution de points, mais se modifient entre eux. Ils sont interprétés comme des opinions incertaines d'individus qui se rencontrent et s'influencent.

La simplicité de ces modèles est un de leurs avantages car elle permet de comprendre leur fonctionnement assez précisément. En effet, des études systématiques de l'espace de paramètres sont possibles, ce qui permet de « cartographier » leur comportement dynamique. De plus, il est possible de développer des modèles approchés de l'évolution de certaines variables agrégées, qui apportent un éclairage complémentaire.

Cette simplicité est aussi une faiblesse, qui a été soulignée par des collègues dans un débat qui a suivi l'une des publications (Deffuant et al. 2003). En effet, il est évidemment extrêmement réducteur de rendre compte de l'influence sociale par de simples dynamiques sur des segments. Nous souscrivons totalement à cette critique, et même souvent au-delà de nos contradicteurs (voir chapitre 2). Mais il nous semble raisonnable de postuler que ces dynamiques peuvent représenter une partie du processus, influence quasi-inconsciente, et qui devient très proche d'un mécanisme d'habituation au sens neuronal. Le chapitre suivant propose un modèle plus complet, qui utilise ces modèles d'influence, couplés à d'autres mécanismes, prenant notamment en compte un état de faits extérieurs. Nous verrons alors qu'il était important de mener les études sur les modèles simples pour pouvoir interpréter le comportement du modèle plus complet.

Les deux modèles présentés dans ce chapitre ont été élaborés durant le projet IMAGES. La paternité du premier, appelé modèle de « confiance limitée » (*Bounded confidence*, BC), fait l'objet de débats. Il semble que les premiers à l'avoir proposé et étudié sont Hegselman et Krause, et un peu plus tard, indépendamment, il est issu d'une collaboration entre E. Chattoe,

G. Weisbuch et moi dans le cadre du projet IMAGES. J'ai proposé le second modèle, appelé modèle de l'accord relatif, comme une amélioration du précédent, permettant d'éliminer certaines caractéristiques contre intuitives du modèle BC. Nous avons particulièrement utilisé ce modèle pour rendre compte de l'influence d'opinions minoritaires extrêmes (Deffuant et al. 2002). Ces modèles présentent deux particularités notables : ils sont fondés sur des opinions continues, alors que les modèles de la littérature sont le plus souvent discrets, et ils prennent en compte une incertitude sur cette opinion, qui joue un rôle dans la dynamique d'influence.

Nous présentons d'abord chacun des modèles et son étude en simulation. Puis nous proposons un modèle d'évolution de distribution de probabilité qui vise à approcher le comportement du modèle individus-centré, et nous tentons d'en tirer des enseignements. Nous reviendrons en conclusion sur l'intérêt de comparer ces études en simulation sur des modèles individus-centrés à des modèles agrégés d'évolutions de distributions de probabilité.

5.1 Le modèle de confiance limitée (BC)

5.1.1 Principe du modèle et équations d'influence

Dans le modèle de « confiance limitée » (BC) (Krause 2000; Deffuant et al. 2001; Hegselmann et Krause 2002), les agents ont des opinions continues et les interactions sont non linéaires : deux agents modifient alors leur point de vue uniquement si la distance entre leurs opinions est au-dessous d'un seuil donné. Ce seuil peut être interprété comme une incertitude, ou une largeur de vue autour de l'opinion, puisque les agents ne tiennent pas compte des opinions situées en dehors de leur intervalle d'incertitude. Le modèle BC peut être vu comme une version non-linéaire de modèles de dynamiques d'opinions continues existants (Chatterjee et Seneta 1977; Friedkin et Johnsen 1990).

Nous considérons une population d'individus définis par une opinion x tirée selon une distribution uniforme entre -1 et 1, et une incertitude constante u . L'opinion d'un individu est donc définie par un segment dont les bornes sont $x+u$ et $x-u$. Nous tirons des couples d'individus au hasard et les faisons interagir.

CycleEvolutionBC $((x_1, x_2, \dots, x_n), u)$

Répéter n fois

$(x, x') :=$ couple tiré au hasard dans la population

si $|x - x'| < u$

$x := x + \mu(x' - x)$

$x' := x' + \mu(x - x')$

fin si

fin

Le paramètre μ règle l'intensité des interactions.

Dans ce modèle très simple, les segments ne se déforment pas, ils se déplacent seulement. Il s'agit donc d'un RGD très simple.

5.1.2 Résultat théorique pour u petit en connexion complète

Ce résultat est dû à J.P. Nadal. On suppose que la distribution des opinions $\rho(x)$ est régulière et le reste lors des interactions, et que u est suffisamment petit pour autoriser un

développement limité. Un tel développement limité permet d'approcher l'évolution de la densité de probabilité $\delta\rho(x)$:

$$\delta\rho(x) = \frac{u^3}{2} \mu(\mu - 1) \frac{\partial(\rho^2)}{\partial x^2}. \quad (64)$$

Si l'on applique ce résultat à une distribution gaussienne, la distribution s'accroît à l'intérieur de l'intervalle formé par la moyenne plus ou moins l'écart type, et diminue à l'extérieur de l'intervalle.

Plus généralement, le comportement qualitatif de la dynamique est de renforcer les maxima locaux de la distribution, et au contraire de diminuer les zones où la distribution est faible. L'ensemble doit donc converger vers un ensemble fini de groupes localisés sur les maxima locaux initiaux. La détection de ces maxima dépend de la valeur de u , qui détermine une sorte de résolution (ou d'échelle) à laquelle les maxima locaux sont détectés. A ce stade d'analyse, il apparaît que ce modèle fait un peu l'inverse d'une moyenne glissante.

Cependant, en le testant en simulation sur des distributions uniformes, des phénomènes un peu différents apparaissent, notamment la propagation d'un effet de bord, que nous avons pu mettre en évidence, avec T. Faure, en étudiant un modèle d'évolution de distributions de probabilité.

5.1.3 Résultats en simulation pour une distribution uniforme des opinions

Les tests en simulations montrent effectivement que les opinions convergent vers plusieurs groupes de même opinion dont le nombre varie en fonction de u . Le nombre de pics final varie en effet expérimentalement comme :

$$p_{\max} \approx \frac{1}{u}. \quad (65)$$

De plus la configuration finale est en général symétrique (voir figure 5.1).

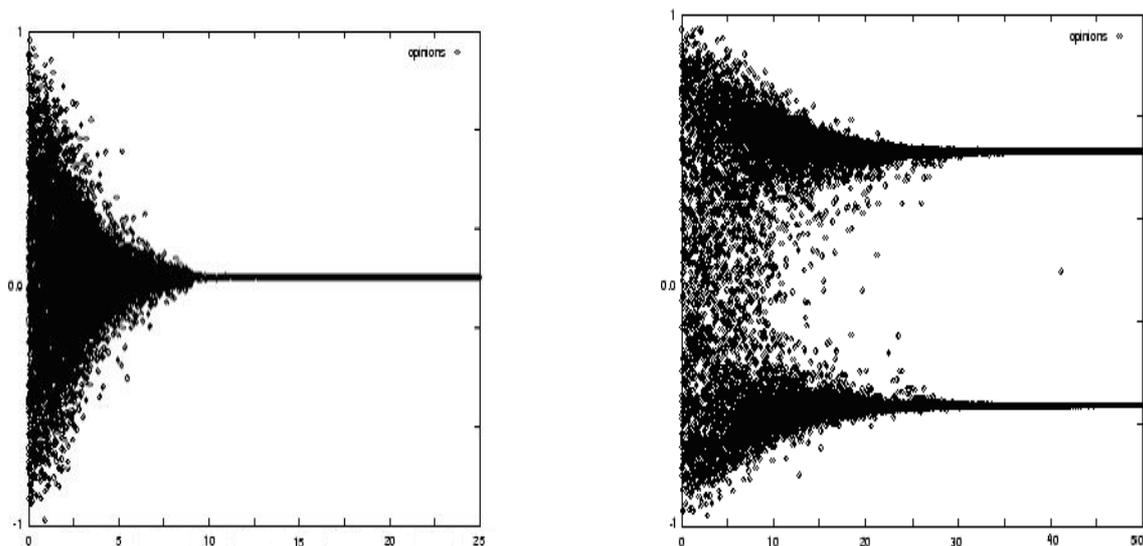


Figure 5.1: Evolution des opinions. Intensité des interaction : $\mu = 0.5$ taille de la population : $N = 400$. A gauche l'incertitude $u = 1$, à droite $u = 0.5$. L'axe horizontal représente le nombre de cycles d'interactions, l'axe vertical les opinions des individus.

Dans (Weisbuch et al. 2002), nous étudions en simulation l'évolution de populations dans le cas où deux incertitudes sont présentes, ce qui donne lieu à des résultats robustes fondés sur la relation empirique du nombre de pics en fonction de l'incertitude. Nous avons par ailleurs cherché à mieux établir cette relation, en étudiant un modèle approché faisant évoluer une distribution de probabilité d'opinions.

5.1.4 Un modèle d'évolution de distributions de probabilité

T. Faure et moi avons développé une approximation de l'évolution de la distribution de probabilité des opinions, au delà des petits u . Pour cela, nous avons utilisé une démarche inspirée de la physique, que certains appellent « sociophysique » (Weidlich 2000), voir aussi (Nadal et al. 1996).

Le principe est de définir un ensemble d'état et des probabilités de présence de la population dans ces états, puis de décrire l'évolution de la distribution de probabilité par une « équation maîtresse » qui définit les flux de probabilité entre les états. Cette approche requiert donc de discrétiser notre ensemble d'opinions possibles, et d'écrire les équations de flux ces valeurs discrètes d'opinions. Nous découpons donc l'intervalle $[-1,1]$ en cases de taille ε . Soit i l'indice d'une case d'opinion et $\rho(i)$ la probabilité de présence des individus dans cette case. L'équation des flux s'écrit :

$$\frac{d\rho(i)}{dt} = (\text{flux entrant dans } i) - (\text{flux sortant de } i). \quad (66)$$

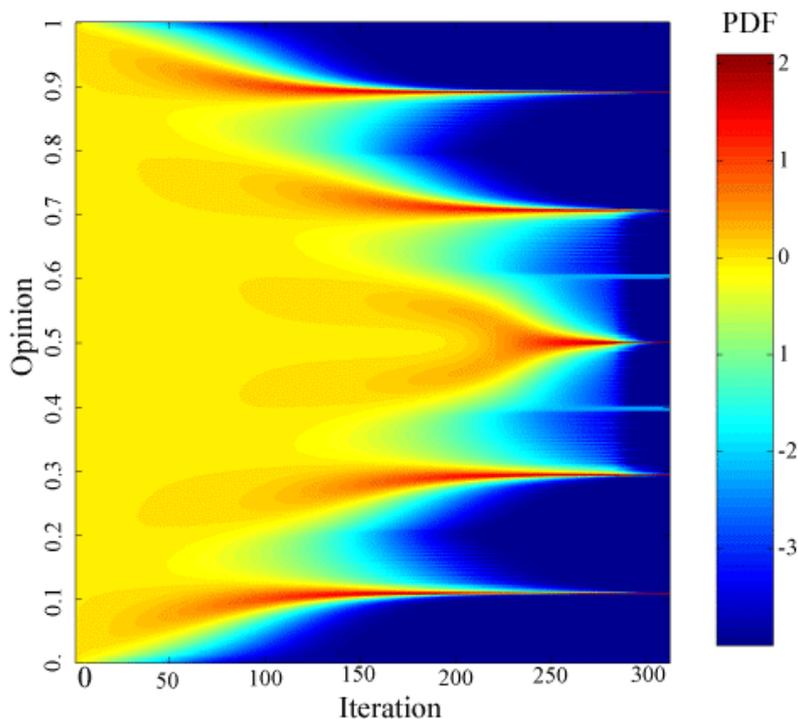


Figure 5.2 : Approximation du modèle par l'évolution d'une distribution de probabilité (1000 pas de discrétisation sur l'axe des opinions). Les pics de densité sont rouges, les densités nulles sont bleues. On peut remarquer que la perturbation se propage à partir des bords de la distribution.

Le flux entrant dans la case i s'écrit en sommant sur les couples de cases qui sont à une distance d inférieure à u l'une de l'autre, et dont l'une des deux est à une distance de l'ordre de μd de la case i (à ε près). Le flux sortant s'écrit en sommant sur toutes les cases à une distance d de i inférieure à u , et telles que $\mu d > \varepsilon$. On obtient ainsi une approximation de

l'évolution de la probabilité de présence dans une case. Ensuite, nous simulons l'évolution de la distribution de probabilité par une approximation au premier ordre :

$$\rho(i, t+1) = \rho(i, t) + \mu' \frac{d\rho(i, t)}{dt}. \quad (67)$$

Où μ' est un paramètre de cette nouvelle simulation.

La figure 5.2 donne un exemple d'évolution d'une distribution de probabilité parfaitement uniforme initialement. Il apparaît que les regroupements sont initiés par le changement fort de densité aux extrémités de l'intervalle des opinions. Ensuite, la perturbation se propage comme une onde. Dans cet exemple, on voit que deux pics centraux issus de cette propagation sont trop proches l'un de l'autre pour se maintenir et ils finissent par fusionner.

Nous pouvons vérifier si cette approximation du modèle donne la même relation entre le nombre de pics d'opinions finaux et l'incertitude que celle que nous avons établie empiriquement par simulation du modèle individu-centré. En fait, les simulations du modèle d'évolution de distribution montrent que les situations où le nombre de pics est impair sont plus stables que celles où il est pair. Nous avons proposé une explication de cette différence dans (Faure et Deffuant 2003), qui est liée à la présence d'un pic central.

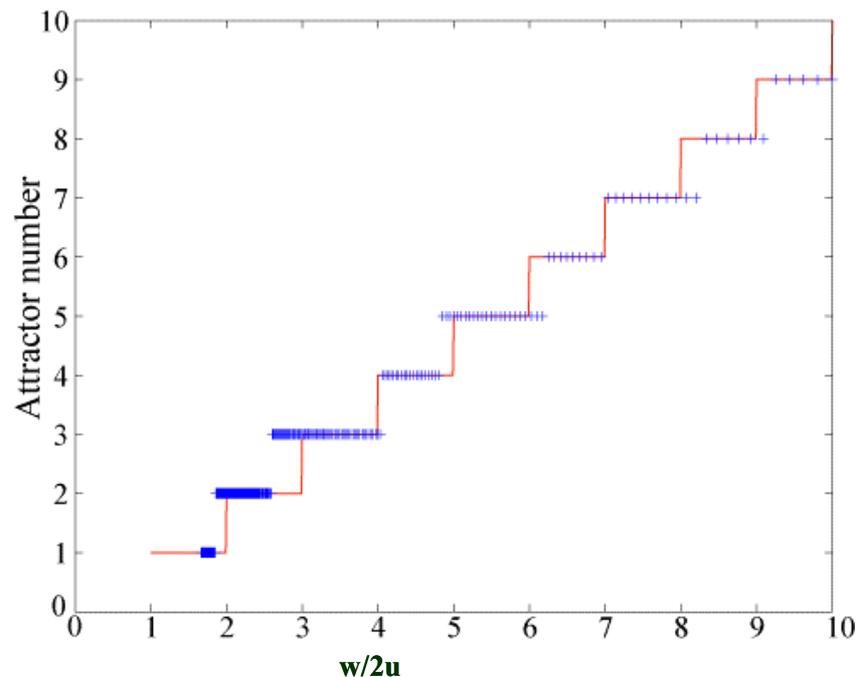


Figure 5.3 : Nombre d'attracteurs en fonction de $1/u$. La relation empiriquement établie par les simulations n'est pas tout à fait vérifiée par le modèle d'évolution de distribution. Les paliers correspondant à un nombre d'attracteurs impair sont plus stables que ceux dont le nombre est pair.

Cette étude confirme l'intérêt de comparer des simulations d'un modèle individu-centré à une approximation sous forme d'évolution de distribution de probabilité. Cette approximation gomme de nombreuses fluctuations aléatoires, puisque toutes les mises à jour sont faites en même temps, ce qui permet de faire apparaître des phénomènes qui sont difficiles à identifier en observant le modèle individu-centré.

5.2 Le modèle d'accord relatif (AR) et la propagation d'opinions extrêmes

Le modèle d'accord relatif vise à répondre à deux critiques qu'on peut adresser au modèle BC :

- l'influence a un profil qui contredit l'intuition. En effet, cette influence augmente linéairement avec la distance entre les opinions, tant que cette distance est inférieure à u , et brusquement elle tombe à 0. Cette discontinuité semble peu vraisemblable.
- On s'attendrait à ce que les personnes plus sûres d'elles soient plus influentes, en général. Ce n'est pas le cas dans le modèle BC simple.

Le modèle AR repose sur l'hypothèse que l'influence d'un agent sur un autre est proportionnelle au recouvrement (accord) relatif entre leurs segments d'opinion. Au cours des interactions, les agents influencent aussi bien les incertitudes que les opinions des individus avec lesquels ils sont en interaction. Ceci permet de résoudre les deux problèmes mentionnés précédemment :

- l'influence change comme l'inverse de l'incertitude de l'agent influent, ce qui rend les extrémistes (moins incertains selon nos hypothèses) plus influents,
- l'influence varie continûment en fonction de la distance entre les opinions.

Le comportement du modèle AR est étudié ici sur des distributions d'opinions initiales uniformes dans lesquelles nous introduisons un ensemble d'extrémistes : i.e. la proportion d'agents ayant leurs opinions en marge de la distribution auxquels on associe une incertitude inférieure à celle de la majorité. Dans un premier temps, nous considérons le cas où les agents peuvent interagir librement les uns avec les autres (connexion complète), puis nous étudions l'influence de structures de réseaux sociaux de type small-world sur les résultats obtenus.

Nous proposons un modèle de dynamique d'opinions dans lequel, en fonction des paramètres, une population peut évoluer vers des opinions extrêmes ou, au contraire, résister à leur influence. Nous nous intéressons particulièrement aux conditions qui mènent à l'une ou l'autre de ces situations.

5.2.1 Définition du modèle

Nous supposons donc maintenant que chaque individu i a une opinion x_i et une incertitude u_i . Nous appellerons le segment $s_i = [x_i - u_i, x_i + u_i]$ le segment d'opinion de l'agent i . L'ensemble des opinions de la population est initialisé à partir d'une distribution uniforme sur $[-1 ; 1]$.

Nous supposons que des interactions ont lieu aléatoirement par paires d'agents (i, j) : le changement d'opinion x_j de l'agent j sous l'influence de l'agent i est proportionnel à la superposition entre leurs deux segments d'opinions (l'accord), divisée par l'incertitude de l'agent i . Quand les agents ont des incertitudes différentes, l'influence réelle est alors dissymétrique, du fait de la division par l'incertitude de l'agent influençant (Figure 5.4).

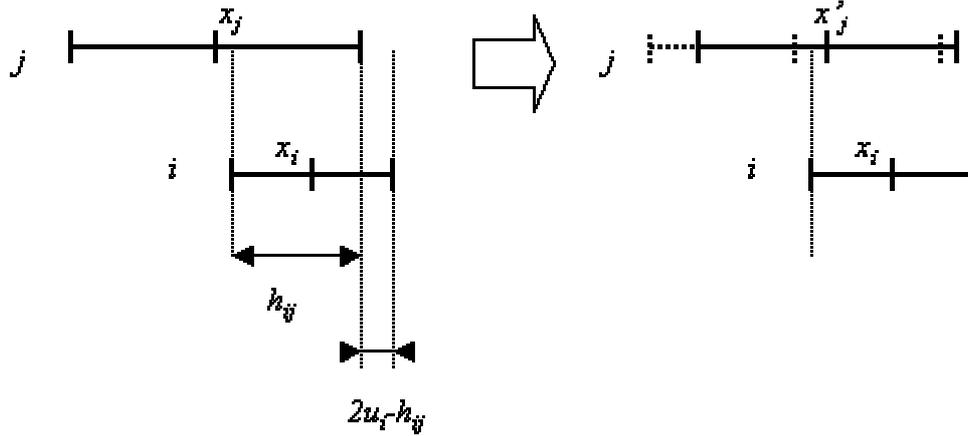


Figure 5.4 : L'agent i (d'opinion x_i et d'incertitude u_i) influence l'agent j (d'opinion x_j et d'incertitude u_j). h_{ij} est le recouvrement de i et j et $2u_i - h_{ij}$ est la partie de j non-recouverte avec i . A droite, les lignes en pointillés représentent la position du segment avant interaction, et les lignes continues après interaction.

Plus précisément, considérons les segments d'opinions $s_i = [x_i - u_i, x_i + u_i]$ et $s_j = [x_j - u_j, x_j + u_j]$, nous définissons l'accord de l'agent i avec j (il n'est pas symétrique) comme la différence entre la partie recouverte s_i et s_j , et la partie non-recouverte de s_i (Fig.1).

Le recouvrement h_{ij} est :

$$h_{ij} = \min(x_i + u_i, x_j - u_j) - \max(x_i - u_i, x_j + u_j). \quad (68)$$

Et la largeur de la partie non-recouverte est donc: $2u_i - h_{ij}$

L'accord de i avec j est donc la différence entre la partie recouverte et celle qui ne l'est pas, soit: $h_{ij} - (2u_i - h_{ij}) = 2 \cdot (h_{ij} - u_i)$. L'accord relatif correspond alors à l'accord, divisé par la longueur du segment s_i :

$$\frac{2 \cdot (h_{ij} - u_i)}{2 \cdot u_i} = \frac{h_{ij}}{u_i} - 1. \quad (69)$$

Si $h_{ij} > u_i$, on applique alors les modifications suivantes:

$$x_j := x_j + \mu \cdot \left(\frac{h_{ij}}{u_i} - 1 \right) \cdot (x_i - x_j), \quad (70)$$

$$u_j := u_j + \mu \cdot \left(\frac{h_{ij}}{u_i} - 1 \right) \cdot (u_i - u_j). \quad (71)$$

où μ est un paramètre constant dont l'amplitude contrôle la vitesse de la dynamique.

Si $h_{ij} \leq u_i$, il n'y a pas d'influence de i sur j .

Les principales caractéristiques de ce modèle sont les suivantes :

- Au cours des interactions, les opinions et les incertitudes associées peuvent être modifiées. Les segments modifient donc à la fois leur taille et leur position, ce qui en fait de ce modèle un RGD un peu plus authentique que le modèle BC.

- L'influence n'est pas symétrique lorsque les deux agents ont des incertitudes différentes, les agents plus "sûrs d'eux" (faible incertitude) sont plus influents.
- Les modifications de x_j et u_j varient de manière continue quand x_j , u_j , x_i et u_i varient de manière continue, ce qui n'est pas le cas pour le modèle BC..

5.2.2 Incertitude identique pour toute la population

Le comportement général du modèle est similaire à celui du modèle BC lorsque l'incertitude est identique pour toute la population : les maxima locaux de densité d'opinions tendent s'amplifier en absorbant les opinions situées à une distance u . Le modèle converge vers la formation d'un ensemble de clusters d'opinions, plus ou moins également distribués (en fonction de la distribution initiale des maxima locaux) (Fig. 5.5).

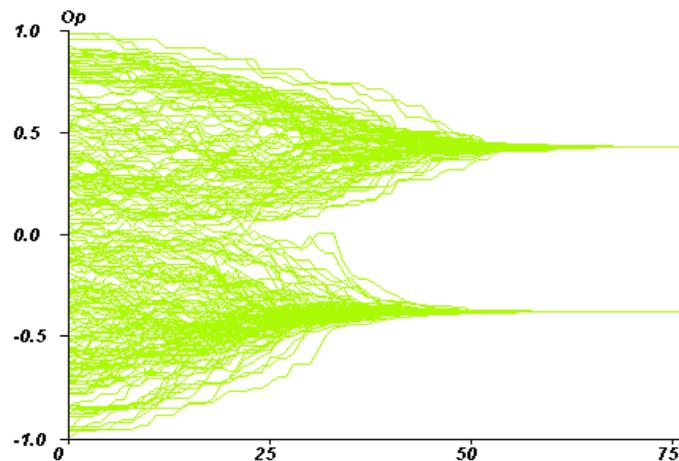


Figure 5.5: Simulation avec une incertitude constante pour toute la population ($u=0.4$).

Dans le cas d'une incertitude initiale u constante pour l'ensemble de la population, et pour une distribution initiale uniforme de largeur w , le nombre moyen de clusters est alors proche de $w/2u$ (Deffuant et al. 2002).

5.2.3 Introduction d'extrémistes dans le modèle d'accord relatif : cas de convergence

5.2.3.1 Introduction des extrémistes

Nous cherchons maintenant à modéliser la propagation d'opinions extrêmes dans une population. Il s'agit par exemple de modéliser l'influence des agriculteurs « bio » sur la population agricole, ce qui a initialement motivé ce travail (Deffuant 2001), mais le modèle peut être appliqué dans d'autres contextes.

Les tentatives qui ont jusque là été entreprises pour modéliser ce phénomène, considèrent généralement des opinions binaires et concernent souvent la capacité d'une minorité à se maintenir ou à maintenir sa puissance (Galam et al. 1997; Latané et Nowak 1997; Galam 2000). Dans ces modèles, l'hypothèse d'opinions binaires ne permet pas de distinguer les opinions modérées des opinions extrêmes. C'est la raison pour laquelle nous considérons ici une dynamique d'opinions continues. De plus, il nous faut tenir compte du contexte général (politique, socio-économique) qui concourt à une montée de l'extrémisme. Ce dernier intervient dans notre modèle au moyen de l'incertitude associée aux opinions. Un contexte de crise favorable à un essor de l'extrémisme, correspondra alors à des incertitudes très élevées.

Nous introduisons les extrémistes dans notre population en supposant que ces derniers, situés aux extrémités de la distribution d'opinion, sont plus sûrs d'eux (leur incertitude est plus

faible). Soit u_e l'incertitude des extrémistes, et U , l'incertitude des modérés, on suppose donc que $u_e < U$.

Soit p_e la proportion d'extrémistes dans la population, p_+ et p_- sont les proportions d'extrémistes respectivement situés aux extrémités positives et négatives de la distribution d'opinions. La différence relative entre les proportions d'extrémistes positifs et négatifs, notée δ , est alors:

$$\delta = \frac{|p_+ - p_-|}{p_+ + p_-}. \quad (72)$$

En pratique, nous déterminons les opinions initiales de notre population en suivant une loi uniforme, puis nous associons aux Np_+ opinions les plus positives et aux Np_- les plus négatives, l'incertitude u_e , les autres ayant l'incertitude U .

5.2.3.2 Les attracteurs: bipolarisation ou polarisation unique

Ce modèle exhibe trois régimes dynamiques différents. Ainsi, nous observons des zones de paramètres pour lesquelles les extrémistes ont une faible influence sur le reste de la population, et d'autres pour lesquelles au contraire, la majorité de la population devient extrémiste, en convergeant vers les deux extrêmes, ou vers un seul.

L'ensemble des figures qui suit exhibe ces différents régimes dynamiques. Pour chacune d'elle, l'axe des abscisses représente le temps (en nombre d'itérations), l'axe des ordonnées représente les opinions, et les couleurs l'incertitude. Chaque trajectoire permet ainsi de suivre l'évolution en opinion et en incertitude d'un agent. Les paramètres communs à l'ensemble de ces figures sont : $\mu = 0.5$, $\delta = 0$, $u_e = 0.1$ et $N = 200$. Le paramètre d'incertitude U des modérés de la population augmente de la figure 5.6 à la figure 5.8.

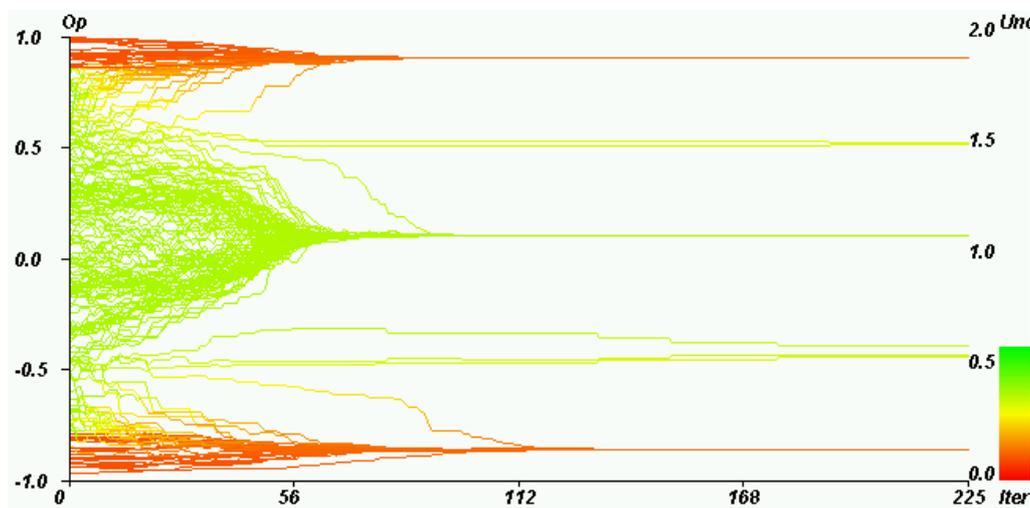


Figure 5.6: Exemple de convergence centrale. $p_e = 0.2$, $U = 0.4$. La majorité (96%) des agents modérés n'est pas attirée par les extrêmes.

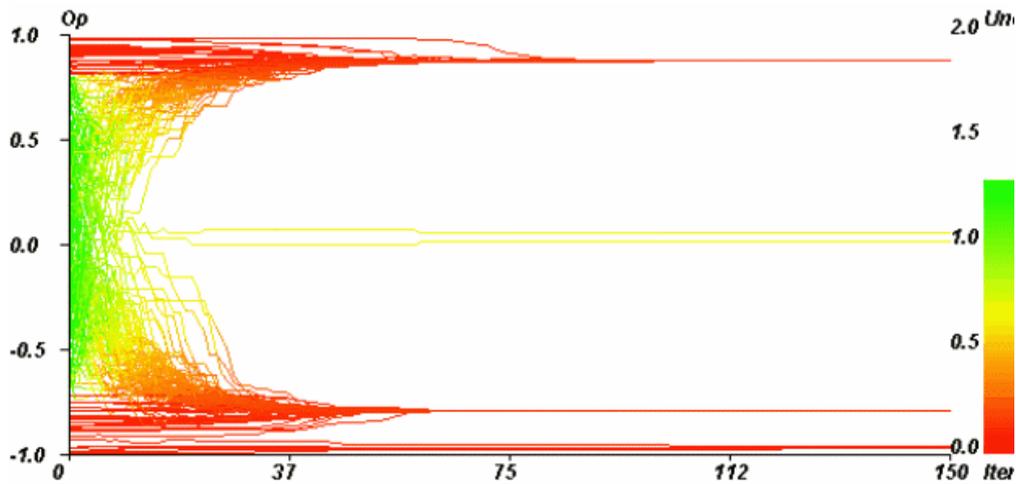


Figure 5.7: Exemple de convergence vers les deux extrêmes. $p_e = 0.25$, $U = 1.2$. Les agents modérés se séparent et deviennent extrémistes (43% du côté positif, 56% du côté négatif).

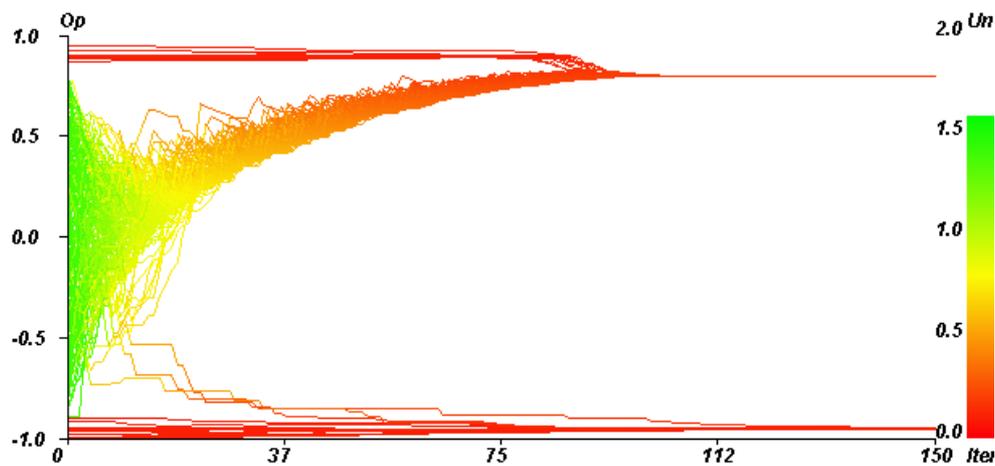


Figure 5.8: Exemple de convergence vers un seul extrême. $p_e = 0.1$, $U = 1.4$. La majorité (98.33%) des agents initialement modérés est attirée par l'extrême positif.

D'autres simulations pour le même jeu de paramètres que celui de la figure 5.8, conduisent cependant à une convergence centrale. Le type de convergence peut donc dépendre de l'aléa du tirage selon la distribution et de l'ordre aléatoire des interactions.

5.2.3.3 Formes typiques de répartition des types d'attracteurs dans l'espace des paramètres

Pour exprimer les résultats de l'exploration plus aisément, nous utilisons un indicateur de type de convergence, noté y :

$$y = p'_+{}^2 + p'_-{}^2. \quad (73)$$

Où p'_+ et p'_- sont les proportions d'agents initialement modérés qui deviennent extrémistes (respectivement positifs et négatifs). Ainsi :

- Si aucun des agents modérés ne devient extrémiste (cas de convergence centrale), alors p'_+ et p'_- sont nuls et $y = 0$.
- Si les agents modérés convergent équitablement vers les deux extrêmes, nous avons alors $p'_+ = 0.5$ et $p'_- = 0.5$, d'où $y = 0.5$.

- Si les agents modérés convergent vers un seul extrême, par exemple le positif, nous avons $p'_+ = 1$ et $p'_- = 0$, d'où $y = 1$.

Les valeurs intermédiaires de y correspondant à des situations intermédiaires, la valeur de cet indicateur nous permet aisément de déterminer le type de convergence.

Le comportement du modèle dans l'espace des paramètres peut être présenté lisiblement sous la forme des variations de moyennes de y en fonction de U et p_e . Nous observons alors une forme typique de y , pour $\delta = 0$ (extrémistes initialement équilibrés), et une forme différente pour $\delta > 0$ (déséquilibre initial entre les extrémistes).

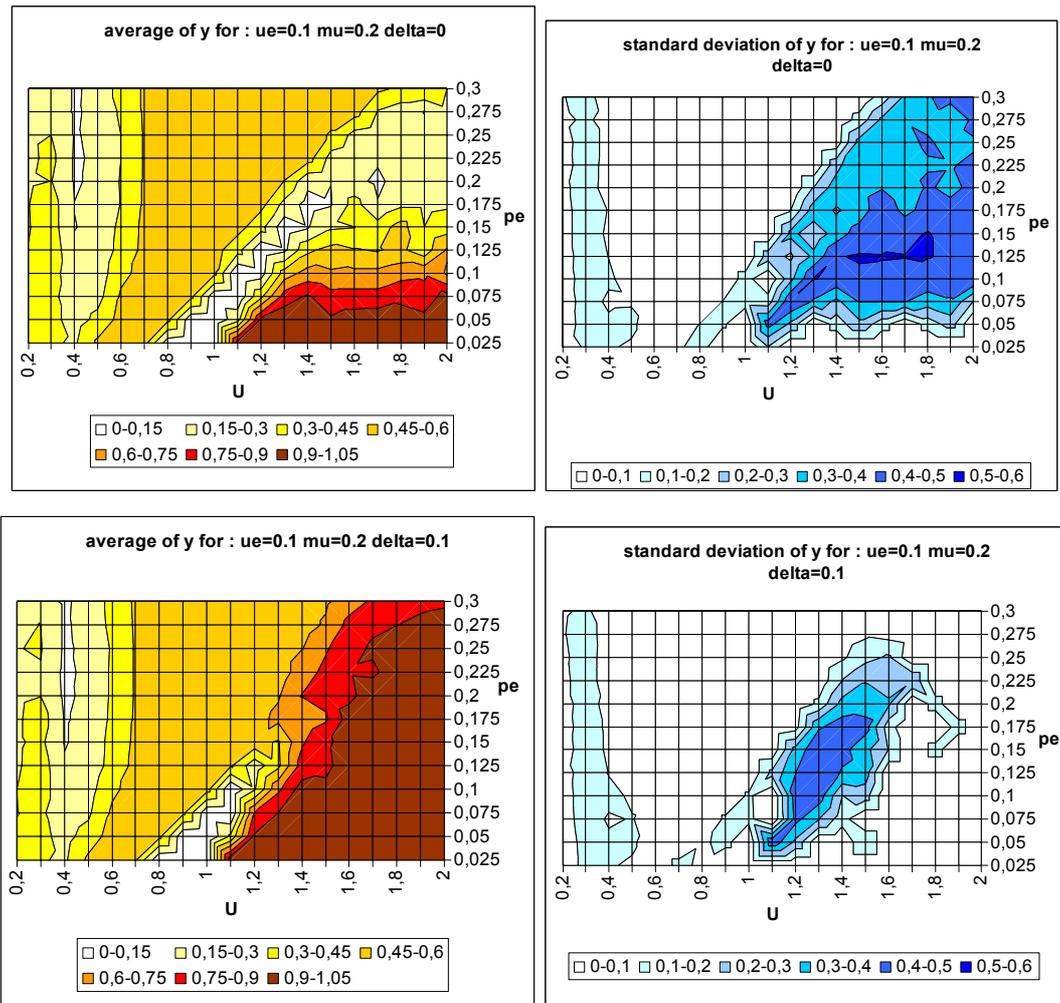


Figure 5.9: Formes typiques de la moyenne de y et de son écart-type pour 50 simulations (à chaque point de la carte) en fonction de l'incertitude des agents modérés (U) et de la proportion d'extrémistes (p_e) pour $\delta = 0$ (en haut) et $\delta = 0.1$ (en bas). Avec $u_e = 0.1$, $\mu = 0.2$, $N = 1000$.

On peut alors identifier quatre régions déterminées par les régimes dynamiques. Les deux zones blanches et jaunes correspondent ainsi à des convergences centrales, la zone orange à des convergences vers les deux extrêmes et la zone marron vers un seul extrême.

Les diagrammes (Fig. 5.9) montrent cependant de larges zones dans lesquelles les valeurs de y sont intermédiaires. Les régimes dynamiques purs ($y = 0$; 0.5 ou 1) sont ainsi séparés par des zones où les valeurs intermédiaires de y associées à fort écart-type peuvent être dues:

- Soit à une distribution bimodale d'attracteurs purs, causée par la stochasticité de la distribution initiale et de la sélection des interactions;

- Soit à une distribution unimodale d'attracteurs plus complexes ayant des clusters d'agents de tailles différentes.

L'histogramme des valeurs de y prises sur la ligne horizontale $p_e = 0.125$ supprime toute ambiguïté sur la nature des attracteurs dans ces zones intermédiaires (Fig. 5.10). Ainsi, nous observons que les zones intermédiaires pour U faible correspondent à une distribution unimodale de y entre 0 (convergence centrale) et 0.5 (convergence vers deux extrêmes), alors que pour U élevé, elles correspondent à une distribution bimodale de convergences centrales (0) et vers un seul extrême (1).

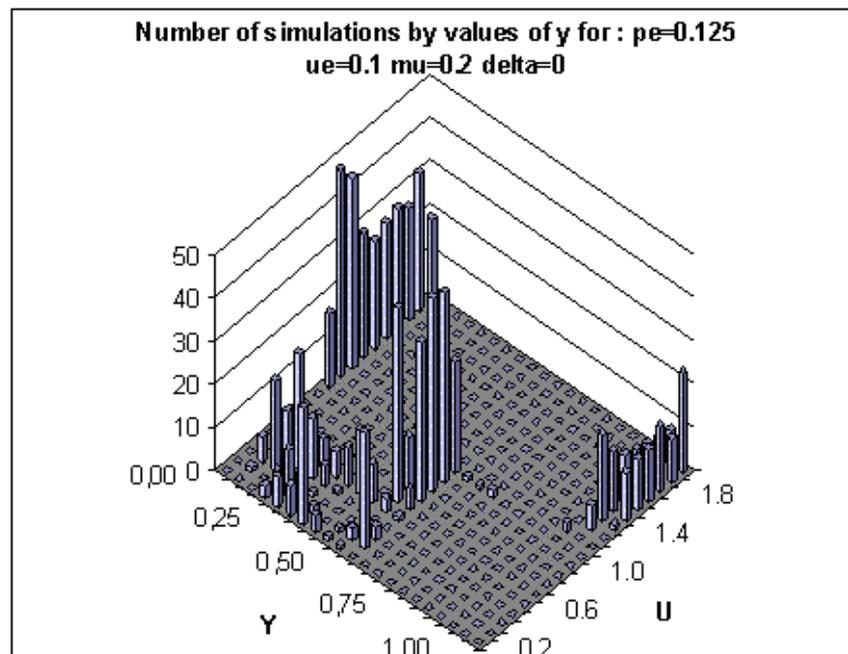


Figure 5.10: Histogramme des valeurs de y sur la ligne $p_e = 0.125$ de la figure 7 avec $\delta = 0$.

Nous avons réalisé une exploration systématique de l'espace des paramètres pour identifier l'influence des différents paramètres sur les différents types de convergence. Cette exploration (Deffuant et al., 2002) a conduit aux résultats suivants :

- Quand l'intensité des interactions (μ) augmente, la zone de convergence vers deux extrêmes s'étend et la zone de convergence vers un seul extrême rétrécit. De plus, dans le cas d'un nombre équilibré d'extrémistes des deux côtés ($\delta = 0$), la zone de convergence centrale située entre ces deux dernières rétrécit.
- Quand le rapport initial entre les deux extrêmes (δ) augmente, les zones de convergence vers deux extrêmes rétrécit au profit de la zone de convergence vers un seul extrême.
- Il n'y a pas d'influence significative de l'incertitude des extrémistes (u_e) sur les patterns observés tant que la relation $u_e < U$ est conservée.

5.2.4 Analyse du processus de convergence avec extrémistes

5.2.4.1 Méthode d'analyse

Nous proposons maintenant une analyse plus précise du processus de convergence sur des cas représentatifs de chacun des régimes dynamiques. Pour ce faire, nous introduisons un indicateur représentant la contribution des extrémistes à l'influence moyenne de la population sur l'opinion de chaque agent. Ainsi, pour chaque agent j à chaque itération, nous calculons E_j , qui est le rapport entre l'influence due aux extrémistes et l'influence globale moyenne :

$$E_j = \frac{\sum_{i=\text{extremist}} \delta x_{ij}}{\sum_i \delta x_{ij}}. \quad (74)$$

Où δx_{ij} est la modification que subirait l'opinion de l'agent j sous l'influence de l'agent i . Dans la suite, nous représentons E_j par une couleur et nous bornons artificiellement ses valeurs négatives par -1. Ainsi, lorsque δx_{ij} est proche de 1, l'influence globale moyenne est en majeure partie due aux extrémistes, et lorsque δx_{ij} est proche de -1, les extrémistes ont une influence qui est opposée à l'influence globale moyenne.

5.2.4.2 Convergence centrale pour une incertitude initiale faible des modérés ($U=0.4$)

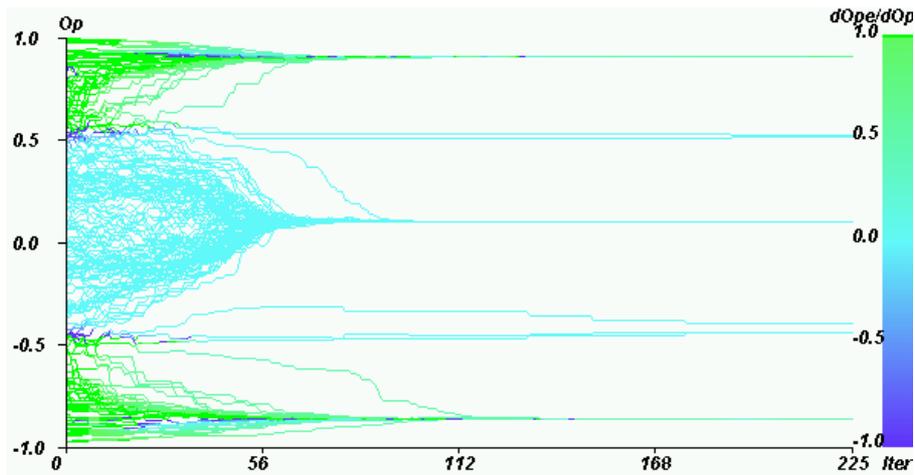


Figure 5.11: Poids des extrémistes dans l'influence moyenne que l'opinion pour $U = 0.4$, $p_e = 0.2$, $\delta = 0$, $N = 200$.

La figure 5.11 exhibe deux grands types de trajectoires: les vertes, proches des extrêmes, sont principalement dues à l'influence des extrémistes, et les turquoise qui ne subissent pas leur influence. A la limite de la zone d'influence des extrémistes, on voit de plus quelques trajectoires bleues qui indiquent que l'influence moyenne globale et celle des extrémistes sont opposées, et peuvent se neutraliser. Ceci est la raison pour laquelle quelques agents continuent horizontalement à partir de cette zone.

Cette situation correspond au cas intuitif où les extrémistes attirent uniquement les parties de la population qui sont proches de leurs opinions.

5.2.4.3 Convergence vers deux extrêmes

La convergence vers deux extrêmes peut intervenir pour différentes valeurs d'incertitude des agents modérés. La figure 5.12 illustre un premier exemple pour une incertitude relativement faible ($U = 0.8$), et la figure 5.13, pour une incertitude plus élevée ($U = 1.6$).

On s'attendrait à ce que la convergence vers deux extrêmes corresponde à une extension de l'attraction des extrémistes observée en figure 5.11. L'analyse de la figure 5.12 montre que ce n'est pas exactement le cas. Les agents qui sont principalement attirés par les extrêmes (trajectoires vertes) ont initialement des opinions autour de 0. Au contraire, les agents qui sont proches des extrêmes ont tendance à être davantage attirés par le centre (trajectoires bleu foncé).

La raison en est qu'avec une incertitude initiale de 0.8, les agents qui sont proches des extrêmes, ont une large partie de leur segment qui se trouve en dehors de la largeur de la

distribution initiale. Ils subissent donc une influence plus forte de la part de la population centrale, influence qui n'est pas compensée par la présence des extrémistes. Au contraire, les agents qui ont une opinion plus proche du centre ont des extrémistes situés à une extrémité de leur segment et des agents modérés partout ailleurs, ce qui constitue un déséquilibre en faveur des extrémistes.

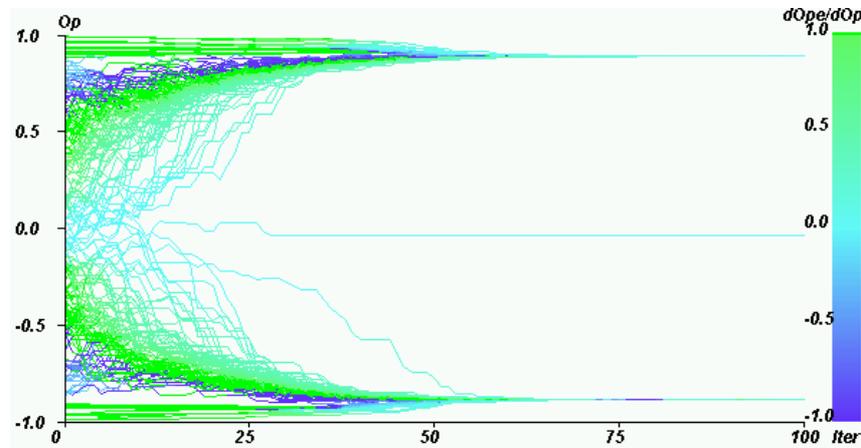


Figure 5.12: Influence moyenne dans un cas de convergence vers deux extrêmes. $p_e = 0.15$, $U = 0.8$, $\mu = 0.5$, $\delta = 0$, $N = 200$.

La figure 5.13 montre une convergence vers deux extrêmes pour une valeur plus élevée de l'incertitude ($U=1.6$) et pour une plus large proportion d'extrémistes. Le processus est alors bien différent:

- Au départ, les agents modérés ont tendance à se regrouper au centre, et l'influence des extrémistes est faible (bleu turquoise).
- Dans un second temps, il y a une forte compétition entre l'influence des extrémistes et celle des modérés (parties bleu foncé). La raison en est que l'incertitude des agents diminue globalement pour atteindre des valeurs pour lesquelles les agents ont un seul extrême à l'extrémité de leur segment d'opinion, ce qui augmente l'influence des extrémistes.

Finalement, les extrémistes gagnent cette compétition et divisent la population en deux parts égales. Si la proportion d'extrémistes avait été plus faible, le regroupement au centre aurait pu l'emporter comme on le verra dans le paragraphe suivant.

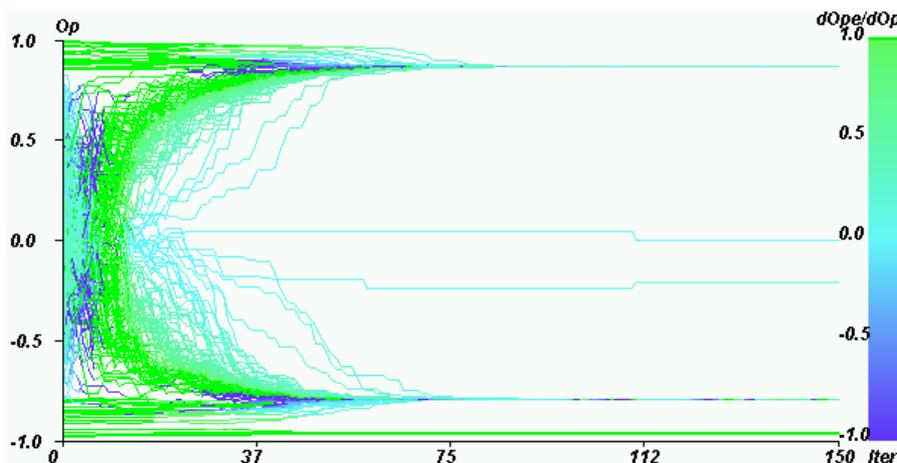


Figure 5.13: Poids des extrémistes sur l'influence moyenne pour $U=1.6$, $p_e=0.2$, $\mu=0.5$, $\delta=0$, $N=200$.

5.2.4.4 Convergence vers un seul extrême

La figure 5.14 montre un exemple de convergence vers un seul extrême. Dans ce cas, la proportion d'extrémistes est faible et leur rôle est très limité au départ (une majorité de trajectoires turquoises), à l'exception du centre où des parties vertes apparaissent.

Une compétition entre regroupement et dispersion apparaît alors (parties bleu foncé), qui est remportée par le processus de regroupement au centre. Au cours de cette phase un nombre limité d'agents modérés rejoint les extrêmes.

On remarque que le processus n'est pas parfaitement symétrique, du fait de la stochasticité introduite par la distribution initiale et par la sélection aléatoire des interactions. En particulier, à la fin de la phase de regroupement central, la partie inférieure est turquoise, indiquant que les extrémistes négatifs ne sont plus en contact avec le centre. Seuls les extrémistes positifs ont alors une influence sur le regroupement central et l'attirent à eux progressivement.

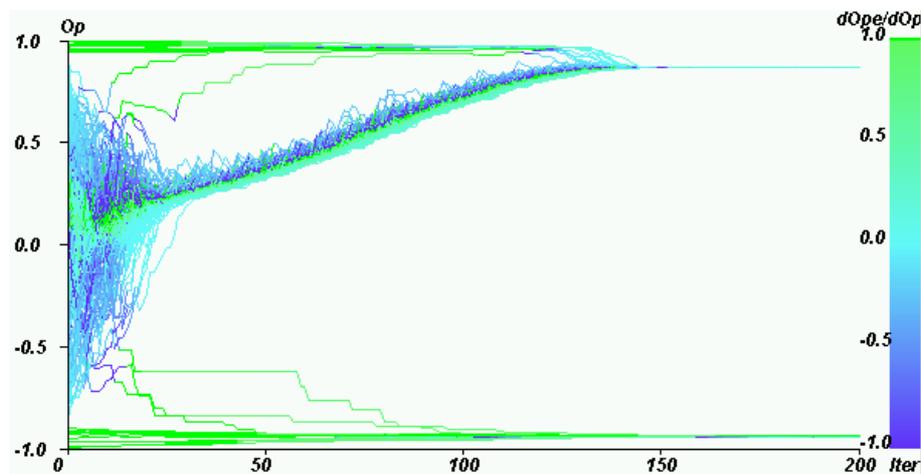


Figure 5.14: Poids des extrémistes dans l'influence moyenne pour $U = 1.4$, $p_e = 0.1$, $\mu = 0.5$, $\delta = 0$.

Au cours de la dérive progressive vers l'extrême positif, on peut distinguer trois grands types de trajectoires: vertes en bas, bleu foncé en haut et au centre, ainsi qu'un ensemble de trajectoires bleu turquoise qui se dirigent vers la partie haute (Fig.5.14). L'interprétation en est la suivante :

- Dans la moitié inférieure du regroupement, les individus ont tendance à monter, sous l'influence des extrémistes positifs et sous la tendance au regroupement. L'influence de la majorité est alors faible (en moyenne) comparée à celle des extrémistes.
- Dans la moitié supérieure du regroupement (bleu foncé), les individus sont soumis à des influences contradictoires : ils ont tendance à la fois à descendre vers le centre de gravité du regroupement et à monter sous l'influence des extrémistes.
- Les individus qui sont attirés par les extrémistes s'écartent du regroupement, et ont alors tendance à retourner en son centre, car l'influence moyenne de ce regroupement devient alors prépondérante. Malgré cela, en se rapprochant du centre, ils attirent aussi vers le haut quelques agents. Ceci conduit à une dérive progressive vers l'extrême positif (Fig. 5.15).

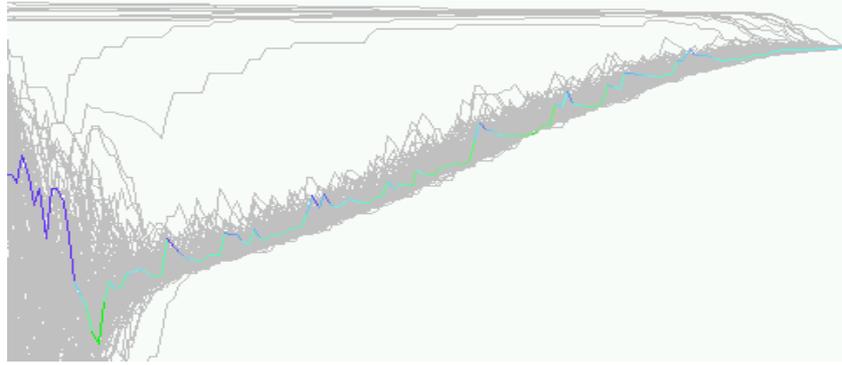


Figure 5.15: Zoom sur une trajectoire individuelle au cours de la simulation de la figure 12.

5.2.4.5 Convergence centrale à la frontière entre convergence vers deux extrêmes et convergence vers un seul extrême

La figure 5.16 montre un exemple de convergence centrale obtenue pour les mêmes paramètres que ceux de la figure 12 ($U = 1.4, p_e = 10\%$).

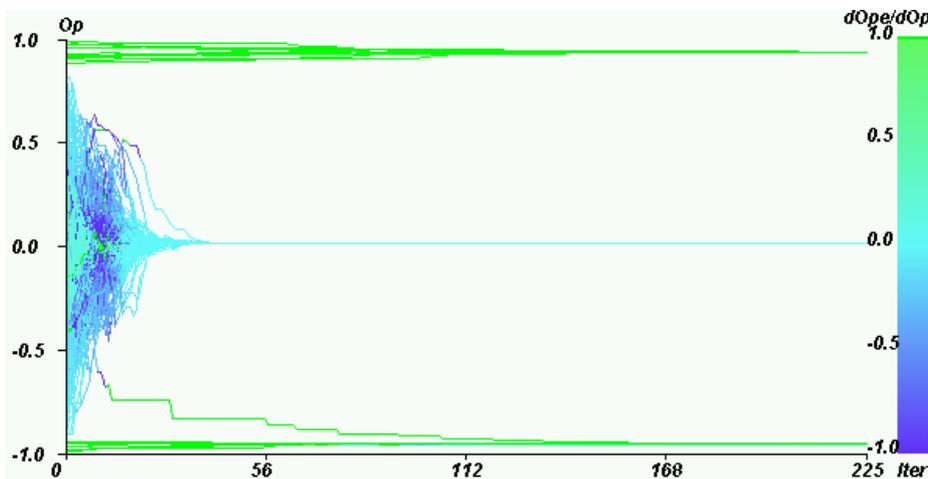


Figure 5.16: Poids des extrémistes dans l'influence moyenne pour $U=1.4, p_e = 0.1, \mu = 0.5, \delta = 0$.

Le début du processus est identique à celui de la figure 5.15 : les extrémistes ont une faible influence au départ, puis, lorsque l'incertitude diminue, l'attraction des extrêmes entre en compétition avec celle du centre (parties bleu foncé). Finalement, ici, c'est le processus de convergence centrale qui l'emporte. Dans ce cas, le regroupement central est très proche de l'opinion 0 et le contact avec les extrêmes est perdu du fait de la diminution des incertitudes (les trajectoires deviennent turquoise).

Les fluctuations des positions autour de l'opinion 0 expliquent que pour des valeurs de paramètres identiques, la population converge soit vers un extrême, soit au centre.

5.2.5 Modèle approché d'évolution de distribution de probabilités

De même que pour le modèle BC, T. Faure a développé une approximation du modèle RA sous forme d'évolution de distributions de probabilités. Le principe du modèle est le même que celui utilisé pour approcher le modèle BC. La différence essentielle tient à ce que la distribution est décrite en deux dimensions : opinion et incertitude. Les équations de flux sont donc écrites entre les éléments d'une grille correspondant à la discrétisation de ces deux axes.

Ce modèle approché permet de retrouver les différents types de convergence, pour des zones de paramètres similaires à celles qui ont été identifiées en étudiant le modèle individus-centré. Une différence doit être cependant soulignée : la convergence vers un seul extrême n'est

possible que lorsqu'une asymétrie est introduite dans la distribution initiale. Cependant, même si cette asymétrie est très légère, la convergence vers le simple extrême a lieu.

Toutefois, il semble que quelques différences de comportement apparaissent entre les deux modèles, que nous n'avons pas complètement terminé d'analyser.

5.2.6 Influence de la topologie du réseau

Nous cherchons maintenant à déterminer les conditions d'existence de la convergence simple lorsque l'on introduit un réseau social de type small-world. Frédéric Amblard a utilisé le modèle β (Watts 1999). Le principe est le suivant : à partir d'une structure régulière (sur un cercle) de connectivité k , on considère chaque lien, et avec la probabilité β , on casse ce lien en une extrémité et on le reconnecte au hasard.

Nous avons donc sélectionné une valeur des paramètres qui menant à une convergence vers un seul extrême dans le cas complètement connecté : une forte incertitude de la population $U=1.8$, et une faible proportion d'extrémistes $p_e=0.05$. Nous avons alors effectué des simulations pour des réseaux obtenus pour β allant de 0 à 1, (donc de réseaux réguliers à des réseaux aléatoires) et k allant de 2 à 256 en suivant les puissances de 2. Au dessus d'une connectivité de 256 (en moyenne chaque individu est connecté à 25% de la population), le comportement du modèle est le même que dans le cas complètement connecté, quelle que soit la valeur de β .

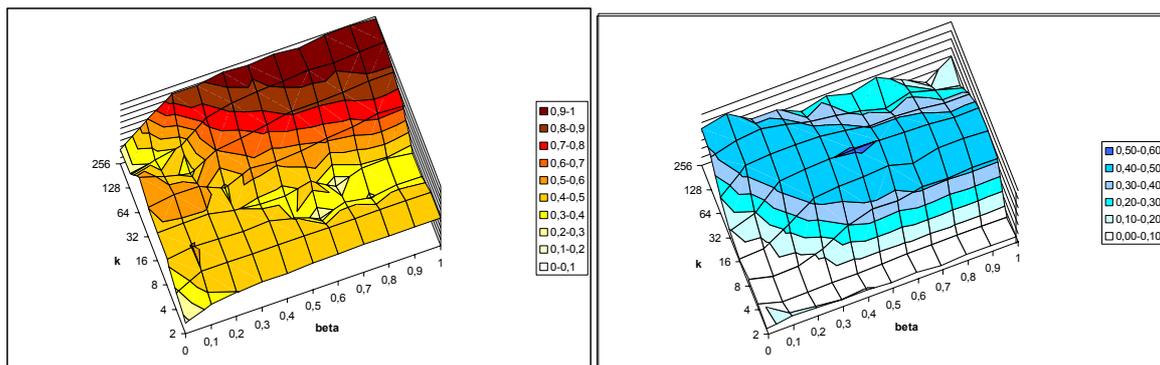


Fig. 5.17. Moyenne (à gauche) et écart-type (droite) de y pour différents réseaux de type small-worlds (β et k variant) pour $U=1.8$, $u_e=0.1$, $N=1000$, et $p_e= 0.05$ (convergence vers un seul extrême dans le cas complètement connecté).

Lorsque k augmente, nous observons une transition entre des convergences vers deux extrêmes et des convergences vers un seul extrême. Dans la zone de transition, l'écart type important de y correspond à un mélange entre convergences centrales et convergences vers un seul extrême. Ainsi, pour de faibles connectivités, chaque extrémiste influence son voisinage qui devient rapidement extrémiste. Nous obtenons alors plusieurs groupes chacun étant en général contrôlé par un extrémiste, conduisant à une convergence vers deux extrêmes. Quand la connectivité atteint un seuil critique, la population tend d'abord vers une opinion moyenne, ce qui donne la possibilité d'un déplacement global vers un seul extrême. Les cas de convergence centrale apparaissent quand la majorité au centre perd contact avec les deux extrêmes, situation favorisée par de faibles connectivités. La convergence vers un seul extrême apparaît pour des connectivités plus fortes. De plus, la transition apparaît pour des connectivités plus élevées quand β diminue i.e. quand les réseaux sont plus réguliers. Notre hypothèse est que la régularité du réseau renforce l'effet local qui favorise une propagation locale forte de l'influence de chaque extrémiste initial, conduisant à une convergence vers deux extrêmes. Cette transition a été également observée pour d'autres couples (U, p_e) , les

types de convergence pour des connectivités élevées correspondant alors à ceux du cas complètement connecté.

Nous avons donc exhibé ici un seuil critique de connectivité qui permet l'apparition d'une convergence vers un seul extrême pour des topologies de type small-worlds. Ce seuil augmente lorsque la régularité du réseau augmente. Ce résultat peut être expliqué par la nécessité d'une première phase de regroupement global pour que la convergence vers un seul extrême apparaisse. Une faible connectivité et une forte régularité du réseau favorisent une propagation locale rapide de chacun des extrêmes, ce qui empêche le regroupement central global de la population.

5.3 Conclusion de ce chapitre

Ce chapitre explore le comportement de deux modèles d'influence sociale fondés sur des opinions continues et des incertitudes. Ils se rapprochent des réseaux géométriques dynamiques car ils se fondent sur des segments et les dynamiques en utilisent la géométrie. Ils s'en distinguent par l'absence de renouvellement des composants. Mais ajouter de telles dynamiques semblerait parfaitement légitime pour rendre compte des arrivées et départs d'individus dans la population. L'inspiration venant des RGD pourrait donc être poursuivie plus avant.

Pour mieux comprendre le comportement dynamique de ces modèles, nous en avons développé des approximations de type « socio-physique », dans lesquelles nous définissons une équation maîtresse régissant l'évolution d'une distribution de probabilité de présence dans certains états. Ce modèle approché est aussi plus général, car il rend compte d'un comportement moyen, faisant abstraction de nombreux aspects aléatoires du modèle individu-centré. Il permet ainsi d'analyser plus finement certains mécanismes, comme la création d'un pic de densité à partir des bords de la distribution uniforme lorsque tous les individus ont la même incertitude. Ces résultats seront utilisés dans le modèle plus complet développé au chapitre prochain.

Cette méthode était praticable pour les deux modèles considérés, lorsque l'on suppose les individus tous connectés. En effet, les modèles ne font pas intervenir beaucoup de dimensions (une ou deux), ce qui permet de conserver une taille raisonnable à l'espace d'état. Construire de tels modèles dans des cas plus difficiles, notamment lorsque la population n'est pas totalement connectée nous semble être un chantier important.

M. Edwards, doctorante au LISC encadrée par F. Goreaud, a commencé à s'attaquer à ce chantier (Edwards et al. 2003), pour un modèle de diffusion de l'innovation de type automate cellulaire, et elle est parvenue à traiter de manière satisfaisante le cas de réseaux aléatoires. Mais l'extension de l'approche à des réseaux d'un autre type (small-world ou invariants d'échelle) reste ouverte.

En collaboration avec G. Weisbuch et moi-même, une autre doctorante au LISC, Sophie Martin, a travaillé au développement des équations maîtresses pour rendre compte du comportement d'une version binaire du modèle BC. Ce modèle considère des opinions sous forme de chaînes binaires, et les individus s'influencent si leurs opinions ont une distance de Hamming inférieure au seuil. L'influence consiste à adopter la valeur de son interlocuteur sur l'un des bits différents. Pour l'instant, les solutions développées deviennent vite impraticables lorsque la dimension des chaînes binaires augmente. En effet, l'espace d'état à considérer dans l'équation maîtresse augmente de manière exponentielle.

Le développement de modèles approchés des modèles individus-centrés simples nous semble une méthode de travail particulièrement opportune pour mieux comprendre le comportement dynamique de ces modèles. Mais il faut pouvoir décliner les techniques appropriées pour y parvenir selon les caractéristiques du modèle individus-centré, et ce n'est pas toujours aisé.

6 Projet Européen IMAGES : un modèle individus-centré de propagation de l'innovation

L'objectif du projet Européen IMAGES était de développer un modèle « multi-agents » de la diffusion des mesures agri-environnementales dans les populations d'agriculteurs, afin d'en mieux comprendre les mécanismes. Nous avons choisi de rapprocher cette diffusion de la diffusion de l'innovation proposée par (Valente 1995): « *la diffusion d'innovations est la diffusion au sein de la société d'idées, d'opinions nouvelles ou de produits nouveaux. Ainsi, la diffusion est un processus de communication par lequel les adoptants persuadent ceux qui n'ont pas encore adopté d'adopter* ». Le modèle de diffusion de l'innovation développé au cours du projet s'appuie sur le modèle d'influence sociale fondé sur « l'accord relatif » présenté au chapitre précédent. Le modèle comprend d'autres aspects, qui ont été imposés par les études sur le terrain, notamment : une fonction de décision à trois états, un module de « calcul rationnel » qui utilise de l'information. Ce chapitre reprend principalement un article en cours de soumission (Deffuant et al. 2003).

Le modèle prend en compte deux dimensions influant sur le processus de diffusion d'une innovation : les aspects socio-culturels et l'anticipation rationnelle et personnelle d'un individu. Par exemple, dans l'adoption de mesures agri-environnementales par des agriculteurs, l'intérêt pour l'environnement et la perception de son état sont de nature socio-culturelle ; l'anticipation du travail supplémentaire, du fait que la subvention compensera les coûts engendrés par l'adoption, est plus personnelle et appréciée de façon plus rationnelle. De même, l'adoption du téléphone portable peut s'apprécier d'un point de vue socio-culturel au travers du statut social que sa possession induit, autant que d'un point de vue plus personnel et financier ou même du risque éventuel pour la santé.

Ce découpage peut être interprété comme une forme grossière de distinction entre faits et valeurs. Ici, les aspects techniques et économiques sont considérés comme des états de faits, sur lesquels les individus pratiquent des tests et communiquent des résultats. L'appréciation socio-culturelle relève de la sphère des valeurs, et son évolution est supposée indépendante des faits qu'elle qualifie. Cependant, cette évolution pourra influencer radicalement la diffusion de la connaissance factuelle. Ce jeu d'influence entre les deux sphères est au cœur du modèle.

La diffusion de l'innovation a suscité un vaste champ de recherche en sociologie. Les modèles les plus répandus sont des modèles à seuils (Granovetter 1978). Ils supposent qu'un individu s'informe en observant ses pairs adoptants, et réduit ainsi son incertitude sur les conséquences de l'adoption (Rogers 1983). Ainsi, une augmentation du nombre d'adoptants au sein du réseau des pairs est une incitation à l'adoption. La capacité de résistance à cette incitation est le seuil d'adoption. Pour un individu, il se définit comme la proportion d'adoptants dans son groupe qui provoque son adoption. Cette théorie a fait l'objet d'applications dans de nombreux domaines : agriculture, politique, langage, ... (voir (Rogers 1983) pour une synthèse).

Or, ces modèles à seuils des sociologues sont un cas particulier d'automates cellulaires que les informaticiens connaissent et étudient depuis longtemps. Les chercheurs en simulation sociale se sont souvent tournés vers ces modèles simples afin de reproduire les dynamiques

sociales, et notamment la diffusion de l'innovation. Chaque automate représente un individu dans son contexte social ; les liens entre les cellules représentent les différents types de relations entre les individus. Le comportement est soit l'adoption, soit la non-adoption de l'innovation. (Blume 1993; Ellison 1993; Blume 1995) interprètent le seuil comme un coût interne propre à chaque individu. La définition de l'utilité à adopter d'un agent est alors fonction d'un profit intrinsèque et de la proportion de ses voisins ayant adopté. Plusieurs variantes de ce modèle ont été proposées (Goldenberg et al. 1999). L'étude de ce formalisme a donné lieu à d'intéressants résultats théoriques (Weisbuch et Boudjema 1999; Young 1999).

Cependant, les enquêtes menées dans le projet IMAGES font apparaître les limites d'un modèle de diffusion de l'innovation basé sur un simple effet de contagion. Cette constatation nous a amenés à incorporer des éléments supplémentaires à notre modèle. Notre approche de modélisation apparaît comme un hybride entre l'informatique, l'approche « agent cognitif » (Conte et Castelfranchi 1995; Ferber 1995; Müller 1998), et les modèles d'automates cellulaires inspirés de la physique (Weisbuch 1991; Young 1998). Notre modèle est inspiré de versions proposées par (Chattoe et Gilbert 1998) et peut être comparé aux approches multi-agents dans la mesure où l'individu est doté d'un comportement élaboré.

Nous présentons tout d'abord le modèle général, puis nous proposons une étude de ses comportements possibles, à partir d'un plan d'expérience en simulation. Nous interprétons ces résultats de simulation à l'aide de comportements typiques des dynamiques d'influence sociale et de propagation d'information⁹. Nous retrouvons ainsi des cas typiques d'innovations.

6.1 Le modèle individu-centré de diffusion de l'innovation

6.1.1 Hypothèses du modèle

6.1.1.1 Limites des modèles à seuil

Notre modèle est construit sur la base de modèles à seuil de la diffusion de l'innovation (Granovetter 1978; Valente 1996). Au sein de ces derniers, les individus sont supposés liés entre eux par un réseau social et leur décision d'adoption est fonction :

- d'une pression sociale, liée au nombre de proches ayant adoptés l'innovation,
- d'un profit intrinsèque lié à l'adoption de l'innovation et variable selon les individus.

L'innovation, adoptée dans un premier temps par les individus pour lesquels le profit intrinsèque est suffisamment important pour compenser à la pression sociale contre l'adoption, se propage par effet de contagion dans un second temps. Dans ce modèle l'état de l'individu est binaire (adoption/non-adoption). Notre étude (Deffuant 2001) sur la diffusion des mesures agri-environnementales parmi les agriculteurs révèle des points importants que les modèles à seuil ignorent :

- le rôle de l'incertitude est important dans la décision et dans les discussions,
- l'attitude envers l'information change selon l'intérêt a priori,
- l'importance du volume d'information disponible permet aux agriculteurs d'évaluer plus ou moins bien leur intérêt personnel à adopter,

⁹ Nous utilisons ici le terme information dans le sens de connaissance factuelle.

- les institutions communiquent sous différentes formes pour fournir de l'information (réunions, visites, médias) ; le choix de la forme de communication peut avoir un impact important sur le processus de décision,
- le processus de décision implique plus de deux étapes (oui/non) ; il semble en particulier exister une période intermédiaire de réflexion et d'hésitation.

Il s'agit ici d'explicitier plus avant le fait que l'adoption est basée sur une compensation entre la valeur sociale ou culturelle de l'innovation et son avantage potentiel d'un point de vue strictement individuel.

6.1.1.2 Distinction entre opinion socio-culturelle et individuelle-rationnelle

La valeur sociale ou culturelle est posée ici comme une opinion non raisonnée. Elle est plutôt le fait d'une perception du voisinage social et évolue au travers des interactions entre individus qui s'influencent. Cette hypothèse est très différente de celle posée dans les modèles à seuil qui postulent que la valeur sociale est directement déterminée par le nombre d'adoptants. Cette nouvelle hypothèse permet la représentation de situations courantes, telles qu'un individu très convaincu ayant une évaluation sociale différente de celle de ses voisins, ou un individu ayant une opinion sociale très favorable bien que nul n'ait encore adopté l'innovation. Il est difficile de représenter ces situations dans le modèle classique à seuil.

Le versant individuel de la décision porte sur les conséquences probables de l'adoption, d'un point de vue strictement individuel (c'est-à-dire en termes d'avantages économiques ou de désagréments opposés au confort). Il correspond, par exemple, à l'évaluation des services rendus par un téléphone mobile au regard de son prix ou à l'efficacité d'une méthode contraceptive comparée à ses risques pour la santé. Nous voyons dans ces exemples que cette évaluation exige un minimum d'informations sur l'innovation (son efficacité, son prix, les conditions de son utilisation, ...) et nécessite que l'individu les rapporte à sa situation personnelle (son budget, ses besoins, ...). Cette évaluation suppose par conséquent que l'individu soit suffisamment intéressé par l'innovation pour prêter attention aux informations sur les conséquences probables de l'adoption. Nous supposons par ailleurs que cette opinion n'est pas influencée par les autres. Avant son calcul, l'individu n'a aucune évaluation (c'est un changement comparé à (Deffuant et al. 2002)). Après son calcul, il ne peut pas être influencé par les autres sur cet aspect strictement personnel. Cette opinion est donc supposée plus privée, moins objet de communication et d'influence.

Cette séparation stricte entre une valeur socio-culturelle et une évaluation strictement personnelle des conséquences de l'adoption est une simplification forte. Nous ne prétendons pas qu'une distinction aussi claire corresponde à la réalité. Nous pensons cependant que la stricte séparation de ces deux composantes est une simplification utile à la compréhension des processus complexes réellement mis en œuvre lors d'une décision.

6.1.1.3 L'incertitude, la conviction et la largeur d'esprit représentées par une unique variable

Il est clair que l'incertitude, la conviction et la largeur d'esprit jouent un rôle important dans la diffusion d'une innovation. La forte incertitude d'un individu sur le coût d'un téléphone mobile ou sur sa capacité de fonctionnement dans sa ville l'incitera probablement à ne pas en acquérir. De même, une femme fortement convaincue que la contraception est moralement interdite décidera probablement de ne pas y avoir recours. La largeur d'esprit définit la capacité d'un individu à prendre en compte l'avis d'autrui et son niveau d'intérêt pour ce qui est nouveau, ce qui peut évidemment être déterminant dans le processus d'adoption.

Ces notions sont par ailleurs connexes : un individu très certain de ses opinions est en général convaincu et peu ouvert à celles d'autrui ; réciproquement, un individu très ouvert est probablement incertain quant à ses propres opinions et relativement peu convaincu. Naturellement ces règles simplificatrices admettent probablement des exceptions.

Nous supposons, dans un but de simplicité, d'agrèger ces notions au sein d'une variable unique nommée "incertitude" par commodité. Cette variable est associée aux opinions relatives à l'innovation (sociales ou personnelles) et influe sur la définition du niveau d'intérêt de l'individu, sa décision et l'amplitude de sa capacité à être influencé par autrui (pour l'opinion sociale seulement). La dynamique de l'opinion sociale est donc particulièrement sensible à l'incertitude (c'est le modèle d'influence d'accord relatif, décrit dans (Deffuant et al. 2002) et dans le chapitre précédent. Le processus de décision est également sensible à l'incertitude des opinions sociales et personnelles.

6.1.1.4 Trois niveaux d'intérêt régissant l'attitude face à l'information et à la décision

Nous définissons trois niveaux d'intérêt à partir de l'agrégation des opinions sociales et personnelles et les incertitudes correspondantes: «non intéressé», «incertain», «intéressé». A chaque niveau d'intérêt correspond une attitude particulière, notamment vis à vis des informations.

Le niveau d'intérêt régit la propagation de l'information. Nous supposons qu'un individu ne prend en compte l'information que s'il est «incertain» ou «intéressé». A l'état «non intéressé», il ignore toute information concernant l'adoption. Nous simulons ainsi une sorte de filtre perceptif.

La possession d'informations sur l'innovation permet une évaluation factuelle de l'intérêt personnel d'adopter, qui peut modifier l'appréciation initiale.

Enfin, nous supposons qu'un individu n'adopte que lorsqu'il est resté intéressé durant un certain laps de temps après une telle évaluation personnelle.

6.1.1.5 Le modèle en deux temps de propagation des discussions initiées par les institutions

La propagation des discussions est inspirée par le modèle en deux temps de diffusion décrit par (Valente 1995). L'hypothèse est que les institutions diffusent des messages aux individus. Ces messages déclenchent des échanges d'opinions et d'information dans la population. Une fonction simple d'amortissement diminue progressivement la propagation des discussions dans le réseau.

6.1.2 L'initialisation du modèle et le lien aux données

6.1.2.1 Initialisation de l'opinion sociale

L'état initial des opinions sociales reflète le degré d'acceptabilité culturelle de l'innovation au sein de la population. Nous supposons en effet qu'il existe une opinion a priori sur la valeur sociale de l'innovation. Pour la représenter simplement, nous supposons qu'elle suit une distribution normale dont la moyenne reflète l'acceptabilité culturelle moyenne et l'écart-type, l'hétérogénéité de cette acceptabilité.

De même, nous faisons des hypothèses simplificatrices pour initialiser l'incertitude des individus autour de leur degré d'acceptabilité. Nous considérons deux cas :

- tous les individus ont la même incertitude U ;
- une majorité d'individus a la même incertitude U tandis qu'une minorité est dotée d'une incertitude plus faible. Cette minorité, davantage convaincue de la valeur sociale de

l'innovation, représente des extrémistes dont la présence peut, dans certains cas, avoir un impact fort sur la dynamique du modèle. Il s'agit du modèle avec extrémistes présenté au chapitre précédent.

6.1.2.2 Initialisation de l'évaluation (personnelle) économique

Seule l'évaluation personnelle, utilisant des informations factuelles permet à un individu « incertain » ou « intéressé » d'attribuer une valeur à son opinion économique. Cette opinion n'a donc aucune valeur initiale.

De façon à explorer la dynamique du modèle, nous supposons que le résultat de l'évaluation personnelle suit une distribution normale dont la moyenne reflète l'intérêt économique moyen de l'innovation et l'écart-type les variations de cet intérêt en fonction de la diversité des situations personnelles. Ainsi, un écart-type faible indique que l'innovation présente un intérêt à peu près similaire pour tous les individus de la population. A l'inverse, un écart-type fort traduit l'importance des différences de situations personnelles.

6.1.2.3 Les réseaux sociaux des individus

Les réseaux sociaux ont fait l'objet de nombreuses recherches ((Wasserman et Faust 1994) pour une revue) dont certaines sont plus spécifiquement liées à la diffusion de l'innovation. Ces dernières ont essentiellement pour but de tenter de corréler la probabilité d'adoption d'une innovation aux caractéristiques d'un réseau. Notre modèle permet de tester l'impact sur l'adoption de différents types de réseaux sociaux. L'importance de cet impact est mis en regard de l'influence d'autres paramètres.

Nous nous restreignons ici aux réseaux issus de (Deffuant 2001) et inspirés du concept «Small world » (Watts et Strogatz 1998). Les individus sont situés dans un espace social. La probabilité d'une relation entre deux individus dans cet espace est d'autant plus grande que la distance qui les sépare est faible. Un individu peut toutefois, dans une bien moindre mesure, être lié à un individu lointain.

6.1.2.4 Entrées de la dynamique : les scénarios institutionnels

Dans le modèle, les institutions promeuvent ou dénigrent l'innovation. Le modèle postule que son utilisateur définit au préalable toutes les actions qui vont être mises en œuvre par une institution. Le scénario institutionnel est donc l'ensemble des actions que l'institution mène pour diffuser l'innovation selon un calendrier donné. L'institution délivre, selon une fréquence définie, un message sur l'innovation, au travers d'interactions régulières avec certains individus de la population, ou de moyens de communication de masse (réunion, publication, radio). Les récepteurs de la diffusion représentent un pourcentage de la population, dont les individus sont choisis au hasard. Les individus qui reçoivent ces messages en parlent avec leurs voisins. Ces derniers vont à leur tour échanger avec leurs proches ; la discussion se propage.

Dans les tests que nous présentons plus bas, l'institution diffuse uniquement l'information nécessaire à l'évaluation personnelle sans délivrer d'opinions culturelle ou sociale. Le réseau de l'institution est la population entière.

6.1.3 La dynamique du modèle

6.1.3.1 Modèle de propagation des discussions

Le principe est que la réception d'un message de l'institution entraîne une cascade de discussions au sein du réseau des individus. Les destinataires directs du message de l'institution propagent plus que leurs interlocuteurs, ces derniers plus que les leurs, etc. (voir

figure 6.1) . Le paramètre γ régit la diminution du nombre de discussions de la manière suivante :

- Un individu contacté directement par l'institution déclenche un nombre de conversations avec ses voisins égal à : $(1 - \gamma) * \text{la taille de son réseau}$;
- Chaque voisin contacté déclenche à son tour un nombre de conversation égal à : $(1 - 2\gamma) * \text{la taille de leur réseau}$;
- Chaque voisin contacté par les précédents déclenche un nombre de conversations égal à : $(1 - 3\gamma) * \text{la taille de leur réseau}$;
- et ainsi de suite.

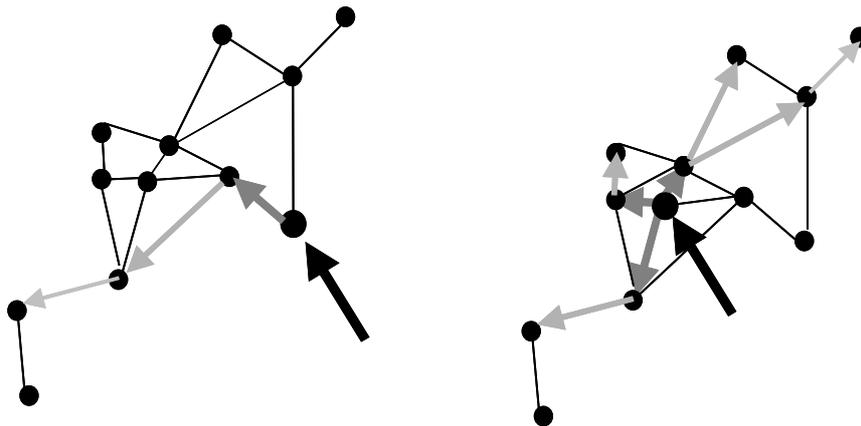


Figure 6.1: Deux exemples de propagation des discussions dans le même réseau avec $\gamma = 0,2$. Les nœuds du graphe représentent les individus et les liens correspondent à leurs relations. La flèche noire indique l'individu qui a reçu le message de l'institution ; les interactions conséquentes sont figurées par les flèches grises dont l'intensité décroissante de gris illustre le phénomène d'extinction de la propagation des discussions

Les discussions à propos de l'innovation sont considérées comme des cas particuliers d'interactions usuelles entre les individus, dont les fréquences sont données.

Lors des rencontres entre les individus, les opinions sociales sont modifiées selon le modèle d'accord relatif, présenté au chapitre précédent.

6.1.3.2 Niveaux d'intérêt des individus

Pour définir le niveau d'intérêt d'un individu pour l'innovation, nous comparons le minimum et le maximum agrégés des opinions sociales et personnelles à un seuil τ . Si l'opinion personnelle n'est pas calculée, l'opinion agrégée se réduit à l'opinion sociale. Si l'opinion personnelle est calculée, le maximum agrégé est la somme (la valeur moyenne plus l'incertitude) des valeurs les plus élevées des deux opinions. Le minimum agrégé est l'équivalent pour les valeurs les plus basses (moyenne sans l'incertitude).

Les règles définissant le niveau d'intérêt sont les suivantes (voir figure 6.2) :

- si le minimum agrégé est supérieur à τ alors l'individu est « intéressé »,
- si le maximum agrégé est inférieur à τ alors l'individu est « non intéressé »,
- si τ est entre le minimum agrégé et le maximum agrégé, alors l'individu est « incertain ».

La valeur du niveau d'intérêt est essentielle dans le processus de décision (voir figure 6.3). On suppose que les individus ont une aversion au risque : l'état intéressé est atteint lorsque l'évaluation pessimiste dépasse le seuil.

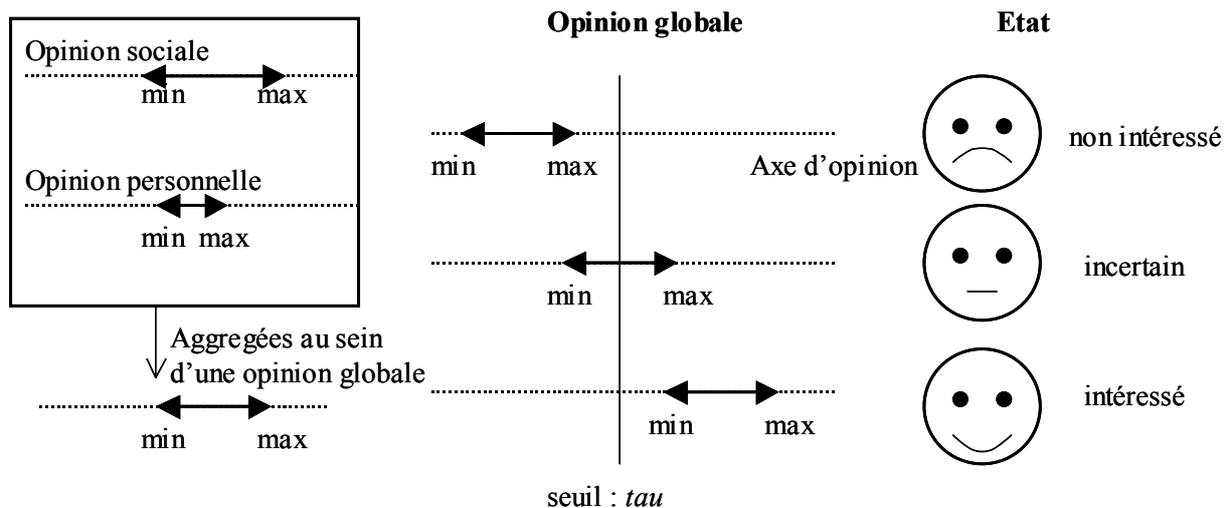


Figure 6.2 : Définition du niveau d'intérêt. Les opinions et les incertitudes sociales et personnelles sont agrégées dans une opinion et une incertitude globale, comparées à un seuil τ . Cette comparaison définit le niveau d'intérêt.

6.1.3.3 La mise à jour de l'information et l'évaluation personnelle des individus

Les individus ne recueillent l'information que s'ils sont « intéressés » ou « incertains » et avec une probabilité ω . Le paramètre ω représente la difficulté de transmission de l'information relative à l'innovation (sa complexité technique par exemple).

Dès qu'un individu obtient cette information, il évalue l'impact personnel de l'adoption de l'innovation.

6.1.3.4 Le processus de décision

Le processus de décision est représenté par le diagramme d'état de la figure 6.3. Il prend en compte le niveau d'intérêt et le fait que l'individu possède ou non l'information. Si l'individu n'a pas d'information, il n'a, selon notre hypothèse, aucune opinion. Dans cet état, deux événements peuvent se dérouler :

- l'individu est « non intéressé », alors son état de décision est « *non-concerné* ». L'individu ne prête ici aucune attention à l'information qu'on lui propose ;
- l'individu est « incertain » ou « intéressé », son état de décision est alors « *information requise* », l'individu reçoit l'information. On suppose donc ici que l'information est immédiatement disponible. Cependant, l'individu a une probabilité ω de la comprendre et de l'utiliser pour calculer son évaluation factuelle, personnelle.

Dès qu'un individu a réalisé une évaluation personnelle, son état de décision change, et selon son état d'intérêt, deux possibilités apparaissent encore :

- L'individu est « non intéressé » ou « incertain » et son état de décision est « *pas d'adoption* », l'évaluation personnelle est suffisamment négative pour entraîner une réduction de l'intérêt global. Cet état n'est pas définitif, des messages porteurs d'opinions sociales très favorables peuvent changer le niveau d'intérêt en « intéressé ».
- L'individu est « intéressé », son état de décision devient alors « *pré-adoption* ». L'individu est prêt à adopter mais réfléchit quelque temps avant de prendre une décision définitive. Durant la période de réflexion, l'influence d'autrui peut modifier son opinion et ramener

son niveau d'intérêt à «non intéressé» ou «incertain» et, donc son état de décision à « pas d'adoption ».

Lorsque l'individu demeure intéressé durant toute la période de réflexion (nombre de pas de temps donné), il adopte et son état de décision devient « adoption »

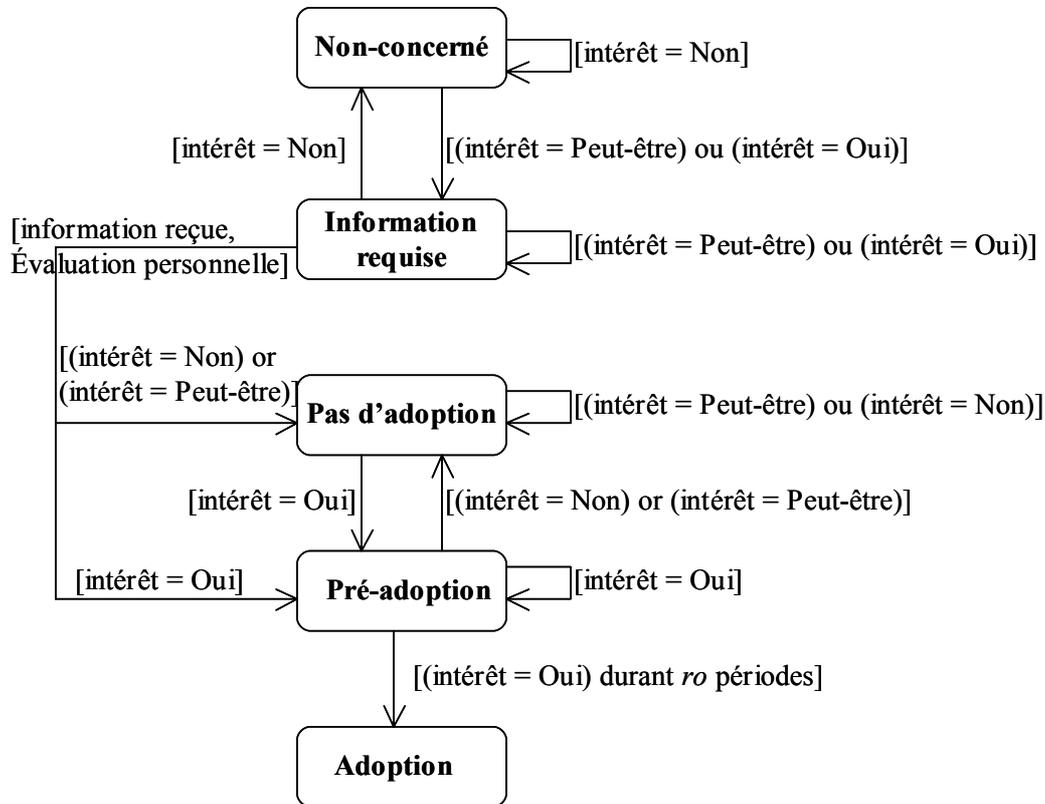


Figure 6.3 : Diagramme d'état du processus de décision. Les places représentent les états et les commentaires près des flèches les conditions pour les transitions. L'état d'intérêt régit l'attitude envers l'information ainsi que la décision finale (la décision intervenant alors que l'individu est doté d'une évaluation personnelle). Le niveau d'intérêt est ainsi résumé : non-intéressé = non ; incertain = peut-être ; intéressé = oui).

Illustrons processus par quelques exemples, représentés en figure 6.4. Nous considérons trois individus qui reçoivent les mêmes messages (message 1 et message 2).

L'individu 1 a un segment d'opinion sociale initiale totalement négatif (le segment est défini par l'opinion plus ou moins l'incertitude). Il n'a pas d'opinion personnelle donc son intérêt initial est «non intéressé» puisque son segment global d'opinion se réduit à son segment d'opinion sociale. Son état de décision est donc « non-concerné ». Il n'accorde par conséquent aucune attention au message 1 (message d'information de l'institution) dont la réception ne change pas son niveau d'intérêt ou de décision. La réception du message 2, d'opinion sociale en provenance d'un autre individu, ne change rien non plus car le segment d'opinion du message n'a aucune intersection avec son propre segment social d'opinion.

L'individu 2 a initialement un segment d'opinion sociale qui contient le seuil de décision. Son niveau d'intérêt initial est «incertain» et son état de décision est « information requise ». Nous avons supposé que l'information donnée par le message 1 est réellement transmise à l'individu 2 (avec la probabilité ω) qui effectue donc son évaluation personnelle. Son évaluation personnelle inclut, elle-aussi, le seuil de décision, son niveau d'intérêt demeure donc

«incertain», et son état de décision devient « pas d'adoption ». La réception du message 2 améliore sa perception sociale de l'innovation, mais ne change pas son niveau d'intérêt.

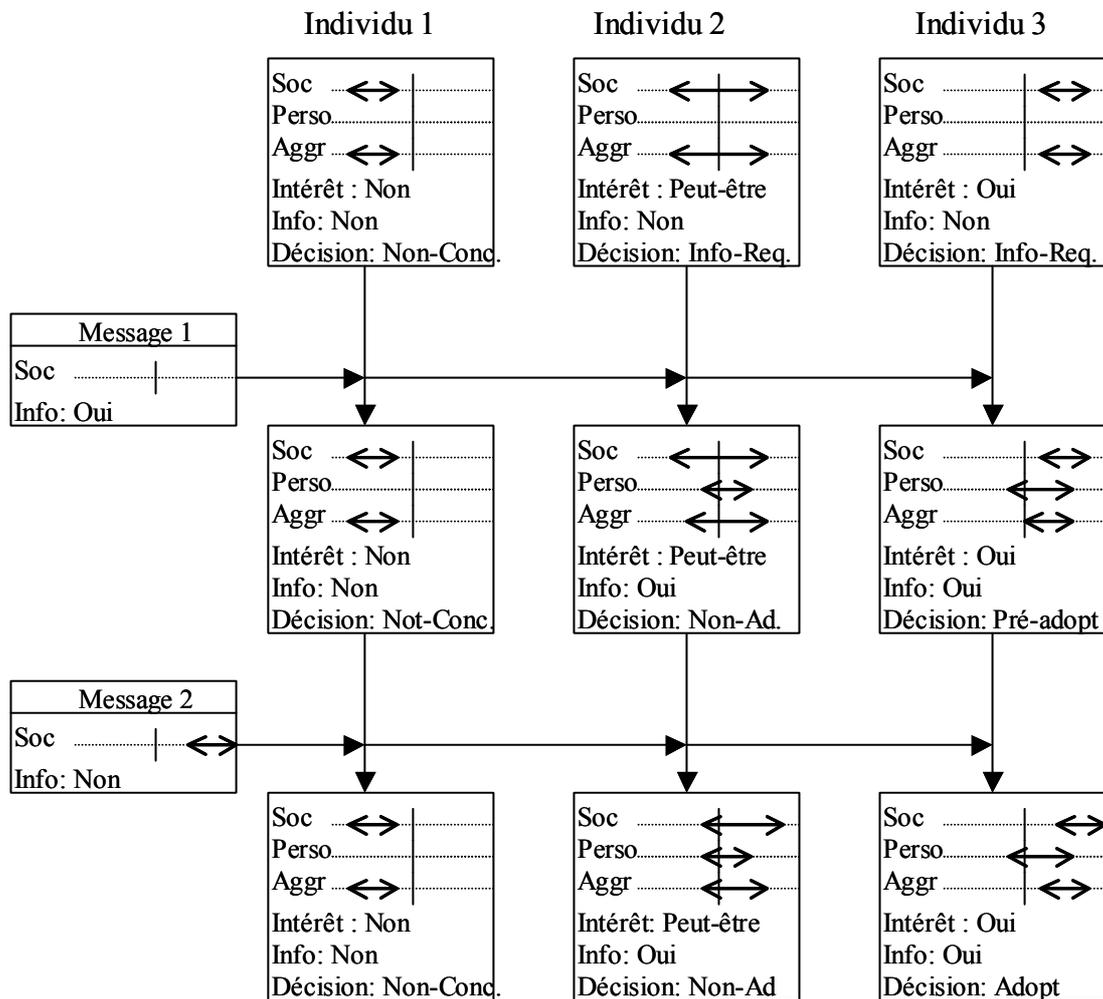


Figure 6.4 : Exemples de différentes évolutions. Les individus reçoivent les mêmes messages et réagissent différemment selon leur niveau d'intérêt. Celui-ci est ainsi résumé : non-intéressé = non ; incertain = peut-être ; intéressé = oui). Voir le texte pour plus de commentaire

L'individu 3 a initialement un segment d'opinion sociale totalement positif. Son niveau d'intérêt initial est «intéressé» et son état de décision est « information requise ». Nous supposons que le message 1 transmet l'information permettant à l'individu 3 de réaliser son évaluation personnelle. Celle-ci est légèrement négative mais ne modifie pas son intérêt ; l'état de décision devient alors « pré-adoption ». Le message 2 délivre une opinion sociale très positive et renforce l'avis social positif de l'individu. Nous supposons que l'individu 3 reste « intéressé » pendant τ périodes après lesquelles son état de décision devient « adoption ».

6.2 Analyse du comportement global du modèle

Pour rendre compte du comportement global du modèle, nous étudions la proportion d'adoptants, la proportion d'individus détenteurs de l'information ou la distribution d'opinion sociale, après un temps de simulation correspondant à une période raisonnable d'exposition à l'innovation. Nous utilisons un plan d'expérience dans lequel varient la plupart des

paramètres du modèle, et nous observons l'impact de ces modifications sur les variables globales de résultat en fin de simulation.

Pour faciliter l'analyse de la proportion d'adoptants, nous distinguons deux processus : l'évolution de l'opinion sociale et la propagation de l'information (et par conséquent de l'évaluation personnelle). Nous vérifions par ailleurs ponctuellement nos interprétations par de nouvelles expériences non incluses dans le plan d'expérience initial.

6.2.1 Evolutions typiques des opinions et de la propagation de l'information

6.2.1.1 Typologie de l'évolution de l'opinion sociale

Nous utilisons des comportements typiques de l'évolution de l'opinion sociale qui sont proches de ceux proposés dans le chapitre précédent. La principale différence réside dans l'introduction d'un comportement typique supplémentaire au sein duquel les opinions demeurent stables, et qui est provoqué par des réseaux sociaux très peu denses, et dans lequel l'opinion et l'incertitude demeurent quasiment inchangées durant tout le processus. Ce cas est appelé : "*stabilité hétérogène*", car elle est due à une incapacité des individus à s'influencer mutuellement.

Les deux autres comportements typiques sont :

- La "*convergence centrale*" dans laquelle les opinions tendent toutes vers la même valeur.
- La "*convergence extrême positive*", dans laquelle les opinions tendent toutes vers l'extrême positif.

Il n'y a pas lieu de considérer la convergence vers deux extrêmes puisque dans ce plan d'expérience, les extrémistes sont présents uniquement du côté positif (quand il y en a).

Le tableau suivant récapitule les choix des paramètres qui favorisent les différents comportements typiques.

Type d'évolution sociale	Facteurs favorisant ce type d'évolution
Stabilité	- Pas ou très peu d'interactions (faible fréquence des messages institutionnels, γ proche de 1, faible connectivité) - Incertitude faible comparée à l'écart type de la distribution
Convergence centrale	- Forte incertitude comparée à l'écart type, aucun extrémiste, nombreuses interactions (forte fréquence des messages institutionnels, γ faible, forte connectivité)
Convergence extrême positif	- Présence d'extrémistes, forte incertitude comparée à l'écart type de la distribution et nombreuses interactions (forte fréquence des messages, γ faible, forte connectivité)

Tableau 6.1 : Types de convergences des opinions sociales et valeurs des paramètres qui les favorisent.

6.2.1.2 L'information et la propagation de l'évaluation personnelle

La propagation de l'information dépend initialement des états d'intérêt, qui dépendent eux-mêmes des opinions sociales (l'opinion personnelle n'étant pas encore calculée). Pour simplifier l'analyse, nous considérons deux cas extrêmes :

- *l'information ne se propage pas* : une minorité d'individus détient l'information. Dans ce cas, la plupart des individus n'a pas d'opinion économique. On rencontre cette situation typiquement lorsque l'opinion sociale initiale est négative pour une majorité d'individus qui demeurent alors dans un état de décision « non-concerné » ;

- *l'information s'est propagée : toute la population en dispose.* Ce cas se produit lorsque l'opinion sociale initiale sous-tend l'état « information requise » pour une part importante de la population. Si l'institution fournit l'information, elle déclenche alors l'évaluation personnelle.

Les situations intermédiaires sont naturellement possibles. Les conditions initiales favorisant l'une ou l'autre des extrémités sont (tableau 6.2) :

Situation typique	Facteurs favorisant la situation typique
Propagation faible (seule une minorité dispose de l'information)	- Distribution initiale de l'opinion sociale négative - Pas ou peu d'interactions (faible fréquence des messages institutionnels, γ proche de 1, faible connectivité), - Innovation techniquement difficile (ω proche de 0)
Propagation forte (la population entière dispose de l'information)	- Distribution initiale de l'opinion sociale positive - Nombreuses interactions (haute fréquence des messages institutionnels, γ proche de 0, forte connectivité) - Innovation techniquement facile (ω proche de 1)

Tableau 6.2 : Types de propagation de l'information et valeurs des paramètres qui les favorisent.

6.2.1.3 Plan d'expérience

Afin d'analyser la variété des comportements du modèle, nous avons élaboré un plan d'expérience qui définit les configurations de paramètres à tester.

Les dimensions principales de ce plan d'expériences sont :

- les états initiaux de l'opinion sociale et de la distribution de l'évaluation personnelle ;
- les paramètres affectant l'évolution de l'opinion sociale et la propagation de l'information (le réseau d'échange entre les individus, la fréquence de diffusion de l'information par l'institution).

Les principaux paramètres de la dynamique interne n'étant pas l'objet de cette étude sont fixés de la façon suivante :

ω (probabilité de transmission de l'information nécessaire au calcul de l'évaluation personnelle)	0.5
γ propagation d'une discussion au sein du réseau	0.3
μ intensité de l'influence sociale	1
ρ temps de réflexion (nécessaire à la décision définitive d'adoption)	15 itérations

Tableau 6.3 : valeurs fixes des paramètres de la dynamique

Nous fixons par ailleurs l'écart-type de la distribution de l'évaluation personnelle, son incertitude, de même que l'incertitude des extrémistes :

Ecart-type de la distribution de l'évaluation personnelle	0.1
Incertitude sur l'évaluation personnelle	0.01
Incertitude des extrémistes sociaux	0.01

Tableau 6.4 : valeurs fixes des paramètres d'initialisation des opinions et de l'évaluation factuelle.

6.2.1.4 Valeurs prises par les paramètres testés

Les valeurs considérées sont résumées dans le tableau 6.5.

Moyenne de l'opinion sociale	-0.2; -0.15; 0.15; 0.2
Ecart-type de la distribution initiale de l'opinion sociale	0.1; 0.3
Incertitude des non-extrémistes sur leur opinion sociale	0.05; 0.3
Pourcentage d'extrémistes "pro" ayant une incertitude de 0.01	0%; 15%
Nombre de voisins d'un individu au sein du réseau social	1; 4
Probabilité par itération d'émission d'un message institutionnel	0.02; 0.38

Tableau 6.5 : différentes valeurs prises par les paramètres testés

Une simulation dure 500 itérations pour une population de 1000 individus. 15 répliques sont exécutées pour chaque jeu de paramètres.

6.2.2 Résultats et analyse

La figure 6.5 présente les résultats agrégés sur la population entière en fin de simulation. Pour les analyser nous distinguons plusieurs cas en fonction des distributions initiales des opinions sociales d'une part et des évaluations personnelles d'autre part.

6.2.2.1 Les moyennes des distributions de l'opinion sociale initiale et de l'évaluation personnelle sont toutes deux négatives

La proportion d'adoptants est ici la plupart du temps très basse (entre 0 et 2 %) et ce n'est guère étonnant. Le modèle peut cependant engendrer un taux non négligeable d'adoption (autour de 10 %) lorsque l'écart-type de la distribution initiale des opinions sociales est grand ; cet effet est renforcé en présence d'extrémistes. En effet, dans ce cas, certains individus sont initialisés avec une opinion sociale très élevée. Il est possible de faire croître l'effet « présence d'extrémiste » en portant la valeur de l'incertitude de l'opinion sociale U à 0,7. Dans ce dernier cas, on atteint plus de 50 % d'adoptants.

Cette dynamique simule des innovations qui sont *a priori* très mal perçues et qui finalement peuvent avoir une large diffusion grâce à l'influence d'une minorité très convaincue. Ainsi, certaines sectes ou mouvements religieux ne sont initialement soutenus que par une très faible minorité. Y adhérer peut représenter pour les premiers convertis un risque personnel fort (évaluation personnelle négative) mais leur conviction en la valeur sociale de l'innovation leur permet de l'assumer.

6.2.2.2 Les moyennes des distributions de l'opinion sociale initiale et de l'évaluation personnelle sont toutes deux positives

La proportion d'adoptants est ici importante (plus de 70 %) lorsque l'incertitude sur l'opinion sociale est faible ($U = 0,05$). Cette faible incertitude associée à une valeur positive de l'opinion permet aux individus d'être majoritairement « intéressés » dès l'initialisation. Il s'agit donc d'innovations dont la valeur sociale est positive et présentant un intérêt factuel pour les individus. Le téléphone mobile en est un exemple.

Lorsque l'incertitude est plus forte, les individus sont plutôt « incertains » au départ et tendent à le rester, car l'incertitude forte favorise une convergence centrale des opinions sociales. La proportion d'adoptants est alors autour de 55 %. Les innovations considérées ici sont socialement et globalement identifiées comme bonnes, mais avec une incertitude élevée (part d'inconnu importante pour des enjeux sociaux vitaux). L'introduction des organismes génétiquement modifiés en agriculture pourrait en être un exemple : bien qu'elle puisse apporter des améliorations en terme de rendement, de résistance des plantes au manque d'eau et aux maladies, ses conséquences potentiellement néfastes sur l'écosystème global et la santé humaine sont peu connues et inspirent une crainte qui peut en bloquer la diffusion.

U	Extrém.	Sigma	Réseau - Fréquence	Type évolution opinions	Moyenne d'adoptants (%)								Moyenne de possesseurs de l'information (%)							
					0	5	47	97	0	14	22	96	7	7	99	99	16	16	97	98
0,05	0	0,1	faible	stabilité	0	5	47	97	0	14	22	96	7	7	99	99	16	16	97	98
0,05	0	0,1	fort	centrale	0	5	50	98	0	14	17	97	7	8	99	99	17	17	98	98
0,05	0	0,3	faible	stabilité	10	28	49	76	9	35	36	72	31	31	80	80	36	36	75	74
0,05	0	0,3	fort	stabilité	7	27	48	78	8	36	35	75	31	31	80	80	39	37	75	76
0,05	0,15	0,1	faible	stabilité	0	2	49	97	0	9	24	96	3	3	99	99	9	9	97	97
0,05	0,15	0,1	fort	centrale + extrême	0	3	50	98	0	9	22	97	3	3	99	99	9	9	98	98
0,05	0,15	0,3	faible	stabilité	11	27	48	76	11	35	37	73	31	30	79	80	37	36	75	75
0,05	0,15	0,3	fort	stabilité	9	28	49	78	9	37	36	74	32	32	80	81	37	38	75	75
0,3	0	0,1	faible	centrale	0	0	2	57	0	1	0	56	92	92	100	100	97	97	100	100
0,3	0	0,1	fort	centrale	0	0	0	69	0	0	0	70	100	100	100	100	100	100	100	100
0,3	0	0,3	faible	stabilité	2	11	20	54	2	17	13	54	66	67	95	95	72	72	94	94
0,3	0	0,3	fort	centrale	1	6	9	74	1	9	6	73	92	92	97	97	92	92	96	96
0,3	0,15	0,1	faible	centrale + extrême	0	15	39	69	0	31	28	69	72	72	100	100	87	87	100	100
0,3	0,15	0,1	fort	extrême	0	42	94	99	0	84	71	99	62	63	100	100	90	89	100	100
0,3	0,15	0,3	faible	stabilité	12	19	24	53	12	22	20	52	65	66	95	95	72	73	94	94
0,3	0,15	0,3	fort	centrale + extrême	14	41	43	70	13	44	38	68	87	84	97	97	93	94	96	96
Opinion personnelle					-0,2	0,15	-0,2	0,15	-0,2	0,2	-0,2	0,2	-0,2	0,15	-0,2	0,15	-0,2	0,2	-0,2	0,2
Opinion sociale					-0,2		0,2		-0,15		0,15		-0,2		0,2		-0,15		0,15	

Figure 6.5 : Résultats des simulations réalisées dans le cadre du plan d'expérience. "U" est l'incertitude initiale sur l'opinion sociale ; "extrém." correspond à la proportion d'extrémistes « pro » dans la population ; "sigma" est l'écart type de la distribution de l'opinion sociale initiale ; l'"opinion personnelle" est la moyenne de la distribution normale de l'évaluation personnelle ; l'"opinion sociale" est la moyenne de la distribution normale initiale de l'opinion sociale.

6.2.2.3 Les situations intermédiaires ne sont pas symétriques

Nous considérons à présent les situations intermédiaires qui opposent une distribution des opinions sociales positive à une distribution des opinions personnelles négative, ou l'inverse. Ces configurations de paramètres nous permettent de souligner l'asymétrie des rôles des opinions sociales et personnelles. Cette asymétrie est due à l'hypothèse d'existence *a priori* d'une opinion sociale qui régit le niveau d'intérêt initial.

Considérons les résultats correspondant à :

- cas 1 : la moyenne de la distribution de l'opinion sociale (ms) est égale à 0,2 et la moyenne de la distribution de l'opinion personnelle (mp) est égale à -0,15 ;
- cas 2 : $ms = -0,15$ et $mp = 0,2$.

La somme des moyennes est identique dans les cas 1 et 2. Cependant, l'adoption dans le cas 1 ($ms = 0,2$) est significativement plus élevée que dans le cas 2 (quand $ms = -0,15$). Ce phénomène est particulièrement accentué lorsque l'incertitude sur l'opinion sociale est faible ($U = 0,05$), puisqu'on observe que l'adoption est de l'ordre de 50 % dans le cas 1 alors qu'elle se situe entre 9 % et 37 % dans le cas 2. Il s'explique par le fait que dans le cas 2, lorsque l'écart-type de l'opinion sociale est faible, l'information se propage mal car la majorité est «non intéressé». Dans le cas 2 et pour une incertitude faible ($U = 0,05$), la part d'individus ayant l'information est très proche de celle des individus ayant adoptés : cette proximité nous indique que la propagation de l'information est bien ici le facteur limitant l'adoption. Dans le cas 1, au contraire, l'information est possédée par la presque totalité de la population (au moins 80 %). La différence de propagation de l'information est donc le facteur explicatif de la différence entre les deux cas.

Le modèle rend compte ici de la difficulté de diffuser des innovations culturellement mal acceptées bien qu'elles apportent surcroît de bien-être individuel. Par exemple le préservatif est culturellement mal accepté culturellement dans certaines sociétés alors qu'une partie importante de la population est séropositive ou atteinte d'autres maladies sexuellement transmissibles. A l'inverse, certains comportements dangereux pour l'individu diffusent rapidement dans certaines communautés parce qu'ils bénéficient d'une image sociale positive : les rites d'initiation en sont un exemple.

Dans le modèle, la prépondérance de l'opinion sociale a cependant des limites. Lorsque l'évaluation personnelle devient trop négative au regard du gain social, la diffusion est très limitée : lorsque $ms = 0,15$ et $mp = -0,2$, la diffusion culmine autour 35% (sauf en présence d'extrémistes « pro » où elle peut croître jusqu'à 70%).

6.3 Discussion - conclusion

Nous avons proposé une extension du modèle à seuil de diffusion de l'innovation, qui conserve deux hypothèses essentielles des modèles à seuil :

- une distinction entre influence sociale et profit personnel de l'adoption,
- la décision d'adoption est basée sur une synthèse de ces deux aspects permettant de juger de l'intérêt global d'une l'innovation.

Nous avons présenté plusieurs hypothèses supplémentaires qui ont été suggérées par notre étude sur la diffusion des mesures agri-environnementales. Nous considérons en particulier une opinion sociale continue dotée d'une incertitude (qui peut également être interprétée comme la conviction ou la largeur d'esprit) et une dynamique des influences sociales basées

sur cette incertitude. Cette dynamique offre la possibilité de simuler la diffusion d'opinions extrêmes au sein d'une population. Nous avons introduit une variable « information relative à l'innovation » nécessaire au calcul du profit intrinsèque de l'adoption pour un individu. Nous avons par ailleurs supposé que l'attitude envers cette information dépend de l'opinion sociale.

Les hypothèses de ce modèle sont critiquables : le choix de la dynamique sociale, de la dynamique de la discussion, de la séparation forte entre aspects sociaux et aspects strictement personnels, du type de réseaux sociaux, ...

D'un côté, certaines des hypothèses reposent sur des bases empiriques ou théoriques relativement faibles et peuvent être vus par des sociologues ou des psychologues comme des simplifications outrancières ou erronées. Nos propositions sont donc des approximations qui mériteraient d'être raffinées, voire modifiées si elles s'avéraient fausses.

D'un autre côté, le modèle peut être critiqué du fait de sa complexité : il peut en effet sembler trop compliqué pour être efficacement lié aux données quantitatives. Nos tentatives d'initialisation de la distribution initiale des opinions à partir de données empiriques (Deffuant 2001) n'ont pas été entièrement satisfaisantes. Il nous a été par ailleurs difficile d'attribuer des valeurs concrètes à certains paramètres de la dynamique. Il est plus facile, à cet égard, d'utiliser le modèle à seuil.

Néanmoins, la richesse des différents comportements dynamiques du modèle et les possibilités d'interprétation qu'ils offrent nous semblent des arguments en faveur de notre modèle. Ces possibilités d'interprétations indiquent que sa complexité demeure humainement abordable et qu'il est susceptible de fournir de nouvelles explications des phénomènes sociaux. N'est-ce pas là ce que nous pouvons attendre d'un modèle de comportements sociaux?

Rappelons les principaux résultats issus de l'analyse des simulations menées dans le cadre du plan d'expérience :

- La présence d'extrémistes « pro » peut engendrer une diffusion significative des innovations considérées au départ comme négatives d'un point de vue social ou personnel.
- Une forte incertitude de la population sur la valeur sociale de l'innovation peut en limiter la fortement la diffusion en laissant la population longtemps dans l'expectative.
- Une mauvaise opinion sociale a priori peut bloquer la diffusion d'une innovation individuellement bénéfique, en bloquant la diffusion de l'information factuelle à son propos.

Ces résultats illustrent la richesse du modèle et sa capacité à simuler différents types d'innovations et différents contextes de diffusion. Retenons notamment que l'influence de l'opinion sociale sur la diffusion d'information permet de rendre compte des difficultés de la diffusion d'innovations culturellement négatives bien qu'elles soient sources de bien-être personnel. Le modèle apporte une typologie nouvelle de la diffusion de l'innovation. Cette typologie peut aider à caler le modèle sur des données réelles, même si la détermination des valeurs de certains paramètres doit faire l'objet d'une exploration.

La structure de ce modèle est également adaptable à l'étude de phénomènes de diffusion plus spécifiques tels que ceux de comportements de vote ou d'achat d'un produit particulier. Le rôle des opinions sociales diffusées par les institutions dans les médias (cas non étudié ici) de même que la structure du réseau social seraient alors d'intéressants sujets de recherche.

7 Conclusion, perspectives

Les premiers chapitres de ce document sont consacrés à des modèles statistiques géométriques, alors que les derniers portent sur des modèles individus-centrés d'influence sociale et de diffusion de l'innovation. Même si l'idée de composants géométriques est utilisée dans les deux modèles, ils répondent à des préoccupations différentes, et posent des problèmes de méthode très différents. En apparence, les modèles sociaux se situent à un niveau beaucoup plus élevé dans la progression proposée au chapitre 2. En effet, il y est question de prise de décision, d'influence sociale, de communications, bref de caractéristiques spécifiquement humaines. Les modèles statistiques semblent comparativement très rudimentaires.

Certes, nous avons la conviction que notre modèle social parvient à rendre compte d'un aspect pertinent de certaines dynamiques de propagations de comportements. Mais mise à part la distinction entre faits et valeurs qui structure fortement la séparation entre opinion sociale (qui est de l'ordre des valeurs) et l'évaluation individuelle (qui est de l'ordre des faits), les mécanismes spécifiquement humains que nous évoquions au chapitre 2 sont négligés. En effet les dynamiques de modification des segments d'opinion sont similaires à celles des réseaux géométriques dynamiques. La dynamique des autres variables (information, intérêt) donne un comportement dynamique plus riche à l'ensemble, mais mathématiquement, la complexité de ces modèles sociaux n'est pas très différente de celle des membranes de perceptrons ou des fibres adaptatives. On peut même soutenir que les dynamiques de renouvellement des composants de ces derniers les rendent plus complexes.

En fait, un autre point les rapproche, de manière plus fondamentale : la recherche de régularités statistiquement robustes. Pour les modèles statistiques, il s'agit d'établir ces régularités à partir d'un échantillon issu d'une distribution, mais aussi, dans le cas de l'apprentissage à double niveau des fibres, de les détecter en examinant l'état du modèle. Nous avons également recherché des régularités statistiques sur les comportements typiques du modèle social. Pour cela, nous avons mobilisé différentes techniques comme les plans d'expériences, la construction de modèles agrégés approchés.

La recherche de régularités statistiquement robustes dans des dynamiques complexes nous paraît donc pouvoir établir un trait d'union entre les directions de recherche présentées. Je souhaite développer plus avant mon activité dans cette direction à l'avenir.

Mais avant de le préciser plus concrètement, je dois présenter le projet de l'équipe que j'anime. Ce projet d'équipe est encore en discussion et peut évoluer sur des orientations mineures. Mon projet plus personnel s'inscrit dans ce projet collectif.

7.1 Le projet d'équipe : les modèles individus-centrés comme source d'expérimentation en écologie et sciences sociales

Notre projet est lié à l'intérêt croissant que suscitent les modèles individus-centrés en écologie et en sciences sociales (Gilbert et Conte 1995; Gilbert 1998; Grimm 1999). Les dynamiques écologiques ou les dynamiques sociales sont en effet naturellement conçues comme l'effet

cumulé des interactions entre des individus, qui présentent une certaine variabilité. Prendre en compte cette variabilité, à la fois du positionnement des individus dans l'espace considéré, et de leurs dynamiques d'interactions, peut engendrer des dynamiques globales sensiblement différentes de celles de modèles qui ne considèrent que des individus moyens, tous identiques. Les modèles individus centrés, en représentant explicitement les individus dans le modèle, suscitent donc l'espoir d'apporter une meilleure compréhension de phénomènes complexes, notamment en écologie et en sciences sociales, grâce à une représentation plus fine des phénomènes et à l'utilisation d'outils informatiques adaptés à la construction de modèles multi-disciplinaires. Ces perspectives sont particulièrement intéressantes pour mieux définir des objectifs de développement durable et les politiques d'action correspondantes.

La réalisation concrète de cet espoir se heurte à une difficulté majeure : le modèle devient lui-même un objet complexe, au même titre que le phénomène qu'il représente. En effet, les travaux de ces dernières décennies sur les systèmes complexes montrent qu'un ensemble d'entités élémentaires en interaction, même très simples, peut engendrer une richesse a priori imprévisible de comportements collectifs globaux. Outre des comportements observés sur des systèmes dynamiques classiques (cycles, attracteurs de différentes formes, chaos), ces recherches ont identifié des comportements dynamiques plus typiques des systèmes complexes : états critiques auto-organisés (Bak 1997), structuration de formes loin de l'équilibre (Prigogine et Nicolis 1989), capacité d'adaptation ou résilience (Gell-Mann). Ces propriétés sont en général impossibles à déduire analytiquement à partir seulement des dynamiques individuelles et des interactions. Elles sont parfois délicates à établir directement par l'observation du fonctionnement du modèle. Les modèles individus-centrés peuvent donc être aussi opaques à la compréhension de l'utilisateur que les phénomènes qu'ils sont censés modéliser.

De plus, on peut difficilement attendre de grandes capacités de prédiction quantitatives de ces modèles. En effet, en écologie ou en sciences humaines, le modèle doit souvent prendre en compte un grand nombre d'incertitudes et d'imprécisions, car les données sur les phénomènes étudiés sont souvent imprécises et parcellaires. L'ensemble de ces imprécisions et incertitudes cumulées a donc peu de chances de fournir des prédictions quantitatives fines.

Progressivement cependant, l'intérêt des modèles individus-centrés est apparu dans une nouvelle perspective où l'objectif de la modélisation des systèmes complexes est d'identifier les grands types de comportements du système (qu'ils soient ceux des systèmes dynamiques classiques, ou plus caractéristiques des systèmes complexes), statistiquement robustes et les mécanismes responsables de leur apparition. Ainsi, par exemple, franchir des valeurs critiques de certaines variables peut mener la dynamique du modèle à changer radicalement, ce qui peut par exemple s'interpréter comme une situation de crise, plus ou moins irréversible, dans le système réel. C'est ce type de comportements typiques que nous avons mis en évidence dans notre étude du modèle d'influence sociale. Dans certains cas, notamment pour les modèles d'écosystèmes, les comportements statistiquement robustes pourront donner des prédictions quantitatives, avec une fiabilité évaluable.

Le modèle doit non seulement aider à prédire, mais aussi fournir les concepts permettant d'expliquer les comportements globaux. Pour des modèles qui ont vocation à être utilisés pratiquement, ces situations typiques et les concepts permettant de les expliquer doivent être élaborés en liaison avec les utilisateurs, de manière à prendre en compte leurs préoccupations et leurs objectifs.

L'intérêt des modèles individus-centrés est de permettre d'étudier l'émergence de ces comportements globaux dans une sorte de laboratoire virtuel. Généralement, cette étude est éclairée par le développement d'un modèle agrégé du comportement du modèle individus-

centré, qui permet plus facilement d'établir les mécanismes menant aux comportements globaux typiques (voir figure 7.1), comme nous l'avons fait pour le modèle d'influence sociale. Le comportement du modèle individu-centré devient donc un phénomène complexe à modéliser, à l'instar d'un phénomène naturel. La différence, très importante, est qu'une grande variété d'expérimentations est facilement accessible, à un coût très faible puisque les expérimentations sont des simulations du modèle individus-centré. Une telle richesse d'expérimentations est évidemment en général hors de portée sur le système réel. Le développement du modèle agrégé peut donc bénéficier d'un grand nombre de cycles alternant conception, étude théorique du modèle agrégé, et expérimentations, jusqu'à obtenir une représentation et une compréhension satisfaisante des comportements globaux.

Une fois que ce modèle du modèle agrégé présente un degré d'adéquation suffisant avec le modèle individu-centré, il peut être utilisé pour définir des expérimentations réelles, beaucoup plus ciblées, correspondant aux comportements typiques ou critiques attendus, ou encore être comparé à des observations existantes. La compréhension des liens entre les dynamiques individuelles et les comportements globaux apportée par le modèle agrégé permet en outre de faire évoluer de manière pertinente le modèle individus-centré au vu des résultats des expérimentations réelles.

Les directions de recherche de l'équipe s'articulent dans le cadre de cette caractérisation de comportements globaux dans une démarche expérimentale sur les modèles individus-centrés. Notre ambition est de progresser dans la résolution de plusieurs problèmes de méthode posés par cette démarche : méthodes pour définir des modèles agrégés pertinents selon le type de modèle individus-centré, pour définir des plans d'expériences efficaces permettant de confirmer ou réfuter ce modèle, pour intégrer les objectifs et les connaissances d'utilisateurs dans l'ensemble de la démarche.

Généralement nous développons les modèles à partir d'un cas appliqué, avec un objectif d'utilisation précis, en lien avec des équipes thématiques, et éventuellement en partenariat avec les futurs utilisateurs. Parfois, l'objectif du modèle est plus exploratoire. Le modèle individus-centré permet par exemple de relier des hypothèses de fonctionnement individuel à des « faits stylisés » de comportements dynamiques globaux. Plusieurs directions de recherche se dégagent dans ce cadre général.

- Concevoir des cadres informatiques et mathématiques pour les modèles individus-centrés. La conception d'un modèle individu pose des questions de recherche en informatique pour optimiser l'implémentation de certaines dynamiques, et l'évolutivité des applications, faciliter l'implication d'utilisateurs non spécialistes. Construire des distributions pertinentes à partir de diverses sources de données plus ou moins agrégées pose aussi des problèmes de mathématiques appliquées.
- Identifier des dynamiques globales typiques (comportements dynamiques classiques ou typiques de systèmes complexes), et les tests statistiques correspondants. Utiliser des connaissances des utilisateurs et les caractériser rigoureusement mathématiquement.
- Développer un logiciel générique d'exploration de modèles (SimExplorer). L'objectif de cet outil est de faciliter la définition de plans d'expérience sur les modèles, d'en améliorer la rigueur (Amblard et al. 2003; Deffuant et al. 2003). Ce logiciel doit notamment être couplé à la grappe de PC récemment acquise par le Cemagref au groupement de Clermont-Ferrand.

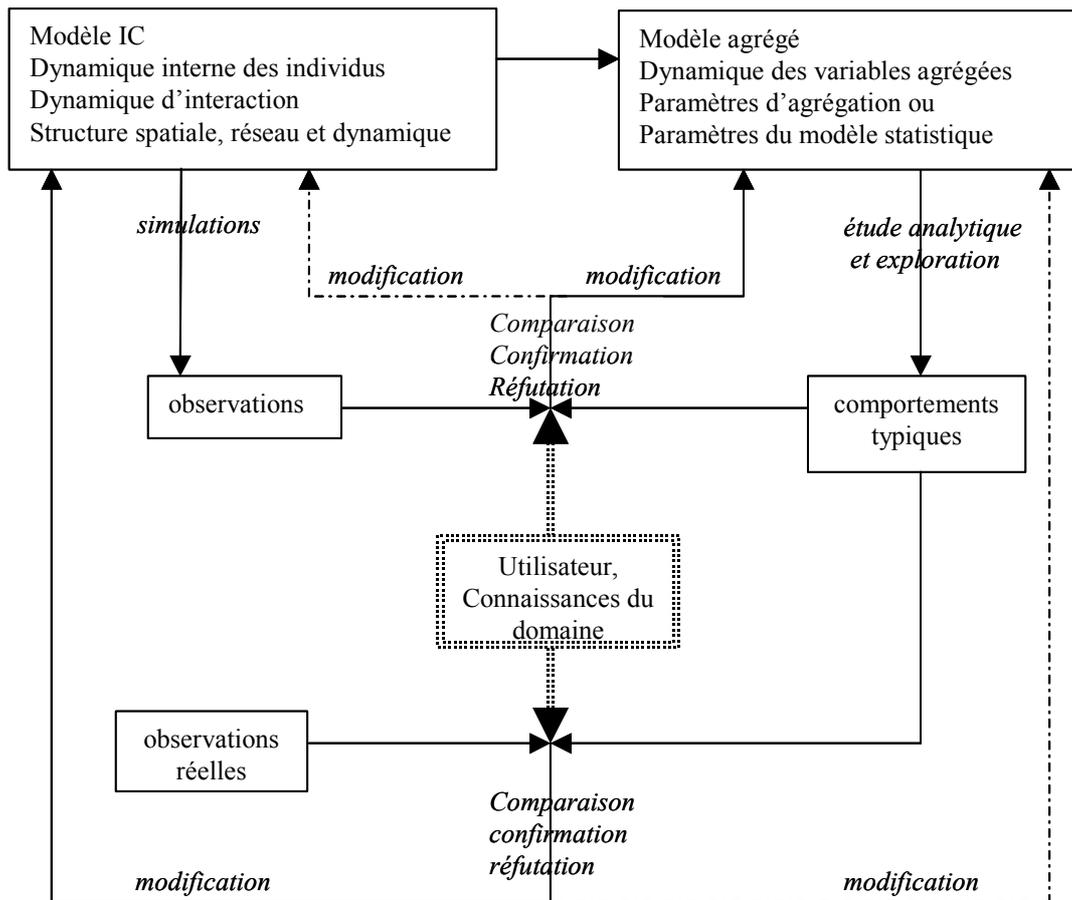


Figure 7.1 : Les modèles individus centrés (IC) sont utilisés comme source d'expérimentation permettant d'identifier des comportements globaux statistiquement robustes, et de les expliquer. Le modèle agrégé peut être statistique (le plus simple étant la moyenne et un intervalle de confiance d'une variable agrégée), ou exprimer directement les dynamiques de variables agrégées. La mise au point de ce modèle utilise des plans d'expériences sur le modèle IC. Lorsque les comportements statistiquement robustes sont identifiés de manière satisfaisante, ils peuvent être comparés à des observations sur le système réel. Cette comparaison peut suggérer des modifications du modèle individus-centré, grâce à la compréhension du lien entre dynamiques individuelles et comportements collectifs émergents apportée par le modèle agrégé.

- Développer des modèles des comportements globaux des modèles individus-centrés, afin de mieux comprendre et mieux prédire leur apparition. Nous visons plus particulièrement des méthodes issues de la physique, dans lesquelles une équation maîtresse définit l'évolution d'une distribution de probabilités sur les états du système considéré. Mais nous envisageons aussi différentes méthodes statistiques alliées à des plans d'expérience dynamiques (ce point est développé un peu plus dans mon projet personnel).
- Développer des méthodes permettant de transcrire les cartes de comportements typiques et leurs conditions d'apparition en des représentations utilisables par des scientifiques non spécialistes de modélisation ou des gestionnaires.

Toutes ces directions de recherche m'intéressent personnellement à des degrés divers, et j'espère collaborer avec les chercheurs de l'équipe au développement de plusieurs d'entre elles. Cependant, certaines sont plus proches de mes travaux antérieurs, et je compte m'y impliquer davantage.

7.2 Mon projet de recherche personnel :

7.2.1 Des outils statistiques pour définir les contours de comportements typiques

Mes recherches antérieures sur l'apprentissage statistique me conduisent en effet à considérer l'utilisation des méthodes développées dans ce domaine pour aider à définir, dans des espaces de paramètres, les contours de comportements typiques de modèles individus-centrés.

Ce contexte d'utilisation nouveau pour ces méthodes est riche de promesses. En effet, dans le cadre de la théorie de l'apprentissage statistique classique, on dispose d'un échantillon de taille donnée, issu d'une distribution de probabilités fixée. Dans ce nouveau contexte au contraire, il est possible d'acquérir de nouvelles données (ce sont des simulations), et de choisir les endroits où elles sont tirées. Ce passage a de profonds points communs avec le passage d'une science de l'observation et de la classification aux sciences expérimentales modernes. Le chercheur passe d'une position d'observateur passif à celle de juge qui pose des questions à la nature, comme dans un tribunal (pour reprendre une métaphore célèbre). On peut aussi faire un parallèle avec les expériences de Bach y Rita, et le changement provoqué par la possibilité de manipuler la caméra (voir chapitre 2). Ces parallèles laissent augurer de bonnes perspectives.

Cette nouvelle utilisation des outils de l'apprentissage pose le problème de définir un compromis entre exploration et exploitation. L'exploration consiste à faire des expériences dans des zones où l'on a peu de connaissances, alors que l'exploitation consiste au contraire à raffiner les conclusions obtenues à l'aide des expérimentations précédentes, et donc à se concentrer dans les zones déjà bien connues. Les recherches en programmation dynamique et apprentissage par renforcement, que j'ai côtoyées au cours de l'encadrement de trois thèses, posent un problème similaire: il s'agit de choisir entre des actions dont on ne connaît pas la récompense, et d'autres dont on sait qu'elle a une probabilité élevée d'être bonne. Il y aurait donc des sources d'inspirations utiles dans ces travaux.

Par ailleurs, les « support vector machines » (*svm*) que j'ai plus particulièrement étudiées (voir chapitre 3), semblent très intéressantes pour cette tâche de repérage progressif des frontières de comportements dynamiques. En effet, les *svm* définissent des marges à l'intérieur desquelles l'incertitude est plus grande, donc des zones dans lesquelles des expérimentations seraient pertinentes. On peut ainsi imaginer des procédures de raffinement progressif. Les marges définies à partir des expériences précédentes définiraient les zones où pratiquer préférentiellement les nouvelles. Cette opération peut être itérée en éliminant certains des points non nécessaires à la définition de la *svm* courante. On espère ainsi caractériser des frontières de plus en plus précisément, en utilisant un nombre raisonnable d'expériences.

De plus, les *svm* permettent de représenter des formes très diverses (non convexes, non connexes) avec peu de degrés de liberté. Cette propriété remarquable permet d'espérer caractériser des frontières dans des espaces de relativement grande dimension, à partir d'un faible nombre de points. Or beaucoup d'approches butent sur l'explosion exponentielle du nombre de points nécessaire pour caractériser des ensembles en grande dimension. Ces outils pourraient résoudre cette difficulté dans certains cas.

La résolution de ce problème peut avoir des implications pratiques importantes. En effet, les frontières identifiées peuvent représenter des limites de viabilité d'écosystèmes par exemple, et fournir des indications utiles pour leur gestion. Mais plus généralement, ces méthodes permettraient d'identifier les zones de comportements robustes dans des espaces de grande

dimension, et donneraient ainsi un accès à des résultats statistiquement fiables dans de tels espaces. Il peut s'agir typiquement de modèles présentant des couplages multiples (physique, biologique, social).

J'envisage le développement de cette direction de recherche à relativement court terme par le lancement d'au moins une thèse. Notons que ce travail entre en résonance avec plusieurs autres directions de travail envisagées par l'équipe, notamment le développement du logiciel SimExplorer, la comparaison des comportements typiques du modèle IC avec un modèle agrégé, l'étude de l'influence de structures spatiales ou de réseaux sur la dynamique, la collaboration avec des utilisateurs pour interpréter les résultats.

7.2.2 Contribuer à une compréhension « humaniste » de la complexité

Parallèlement à ces recherches relativement techniques, et probablement en dehors des heures de bureau, je souhaite poursuivre la réflexion que j'ai amorcée sur la complexité. J'envisage cette contribution à plusieurs niveaux, inspirés par les développements du chapitre 2 :

- Inclure des concepts et outils issus de la géométrie et des statistiques, et définir des boucles expérimentales dans une théorie plus large de la complexité que celle issue de l'universalité de la machine de Turing (complexité de Kolmogorov, Bennett). Dans un tel cadre, la complexité serait définie à partir d'un couple comprenant une machine « sensori-motrice » et un phénomène dynamique. La machine a la possibilité de faire des expérimentations et de construire des régularités statistiques sur les comportements typiques obtenus. Cette théorie renouvelée de la complexité devrait faire appel à des universaux de régularités définies à la fois statistiquement et géométriquement.
- Définir des méthodes permettant de raisonner scientifiquement sur le « monde pour soi » de systèmes dynamiques complexes. En particulier, déterminer à partir de quel type de dynamiques une machine se construit un monde extérieur et une forme de temps. Parvenir à des conclusions solides en ce domaine me semblerait un pas important dans la compréhension de la complexité.
- Formaliser des opérateurs d'empathie afin de pouvoir en faire des simulations informatiques dans des contextes simplifiés. Les dynamiques issues de tels modèles informatiques pourraient présenter des propriétés spécifiques des dynamiques sociales, que d'autres formes d'interactions (de type physique) ne pourraient pas engendrer.

Les deux derniers points s'inscrivent dans une démarche plus générale. Comme je l'ai suggéré au chapitre 2, il s'agit de contribuer au renforcement scientifique d'une pensée humaniste (au sens de Todorov) renouvelée. Les obstacles et les pièges ne manquent pas. Mais on ne se défait pas facilement du goût de l'aventure.

8 Références

- Amblard, F., D. Hill, S. Bernard, J. Truffot et G. Deffuant (2003). "MDA Compliant Design of SimExplorer A Software Tool to Handle Simulation Experimental Frameworks". *Summer Computer Simulation Conference*, Montréal, Canada.
- Arendt, H. (1961). *La condition de l'homme moderne*, Calman-Lévy.
- Austin (1962). *How to do things with words*. Oxford, Oxford University Press.
- Bach y Rita, P. (1994). "The brain beyond the synapse: A review." *NeuroReport*(5): 1553-1557.
- Bach y Rita, P. (1995). *Neurosynaptic Diffusion Neurotransmission and late Brain Reorganization*. New-York, Demos Pub.
- Bak, P. (1997). *How Nature Works: the science of self-organized criticality.*, OUP.
- Barrow, J. D. (1992). *Pourquoi le monde est-il mathématique ?*, O. Jacob.
- Blume, L. (1993). "The Statistical Mechanics of Strategic interaction." *Games and Economic Behavior* **4**: 387-424.
- Blume, L. (1995). "The Statistical Mechanics of Best-Response Strategy Revision." *Games and Economic Behavior* **11**: 111-145.
- Chatterjee, S. et E. Seneta (1977). "Toward consensus: some convergence theorems on repeated averaging." *Journal of Applied Probability* **14**: 89-97.
- Chattoe, E. et N. Gilbert (1998). Modelling the adoption of agri-environmental measures as an innovation diffusion process.
- Conte, R. et C. Castelfranchi (1995). *Cognitive and Social Action*. London, UCL Press.
- Cristianini, N. et J. Shawe-Taylor (2000). *Support Vector Machines and other kernel-based learning methods*, Cambridge University Press.
- Damasio, A. (1995). *l'Erreur de Descartes, la raison des émotions*, O. Jacob.
- Damasio, A. (1999). *Le sentiment même de soi*, O. Jacob.
- Deffuant, G. (1992). Réseaux connexionnistes auto-construits. *PARIS*. Paris, Université Paris VI: 134 p.
- Deffuant, G. (1993). "Réseaux morphodynamiques et modèles cognitifs". *colloque AFCET systémique et cognition*, Versailles.
- Deffuant, G. (1995). "An algorithm for building regularized piecewise linear discrimination surfaces : the perceptron membrane." *Neural Computation* **vol. 7**: 380-398.

- Deffuant, G. (1997). "L'autonomie est-elle création ou rupture de clôture ? tentative d'articulation entre les visions de l'autonomie de F. Varela et de C. Castoriadis." *Revue Internationale de Systémique* **11**(5): 515-534.
- Deffuant, G. (1998). "Les modèles cognitifs à l'épreuve des formes religieuses et sociales : propositions de directions de recherche centrées sur l'empathie." *Intellectica* **1-2**(26-27): 89-109.
- Deffuant, G. (2001). Final report of the IMAGES project: 193.
- Deffuant, G. (2004). "S'évader du présent remémoré et construire un temps social : le rôle central de l'empathie ?". *Journées de Rochebrune sur la complexité*, ENST.
- Deffuant, G., F. Amblard, R. Duboz et E. Ramat (2003). "Une démarche expérimentale pour la simulation individu-centrée". *Journées de Rochebrune sur la complexité*.
- Deffuant, G., F. Amblard, G. Weisbuch et T. Faure (2002). "How can extremism prevail? A study based on the relative agreement interaction model." *Journal of Artificial Societies and Social Simulations* **5**(4).
- Deffuant, G., F. Amblard, G. Weisbuch et T. Faure (2003). "Simple is beautiful...and necessary." *Journal of Artificial Societies and Social Simulations* **6**(1).
- Deffuant, G. et T. Fuhs (1993). "Un nouveau paradigme pour la vie artificielle : les réseaux morphodynamiques". *colloque international de systématique AFCET*, Prague.
- Deffuant, G., T. Fuhs, E. Monneret, P. Bourguine et F. Varela (1995). "Semi-algebraic networks : an attempt to design geometric autopoietic models." *Artificial Life* **vol. 2**(n°2): 157-177.
- Deffuant, G., S. Huet et F. Amblard (2003). "An agent-based model of innovation diffusion mixing social influence and rational anticipation." *American Journal of Sociology* (submitted).
- Deffuant, G., S. Huet, F. Amblard, J. P. Bousset, J. Henriot, G. Amon et G. Weisbuch (2002). "Agent based simulation of organic farming conversion in Allier département". *Complexity and Ecosystem Management*. M. A. Janssen, Edward Elgar: 158-187.
- Deffuant, G. et E. Monneret (1993). "Morphodynamic networks : the example of adaptative fibres". *ECAL 1993*, Brussels.
- Deffuant, G., D. Neau, F. Amblard et G. Weisbuch (2001). "Mixing beliefs among interacting agents." *Advances in Complex Systems* **3**: 87-98.
- Delahaye, J. P. (1999). *Information, complexité, hasard*, Hermès.
- Dumont, L. (1983). *Essais sur l'individualisme*, Le seuil.
- Dupuy, J. P. (1992). *Introduction aux sciences sociales*, Ellipses.
- Dupuy, J. P. (1999). *A l'origine des sciences cognitives*, Seuil.
- Edelman, G. (1992). *Biologie de la conscience*, O. Jacob.
- Edwards, M., S. Huet, F. Goreaud et G. Deffuant (2003). "Comparing individual-based model of innovation diffusion with its mean field aggregated approximation." *Journal of Artificial Societies and Social Simulations*.
- Ellison, G. (1993). "Learning, Social Interaction and Co-ordination." *Econometrica* **61**: 1047-1071.

- Faure, T. et G. Deffuant (2003). "A sociodynamic approach the BC influence model." *submitted to Physica A*.
- Ferber, J. (1995). *Les Systèmes Multi-Agents : Vers une intelligence collective*. Paris, InterEditions.
- Fodor, J. (1975). *The language of Thought*, Harvard University Press.
- Friedkin, N. E. et E. C. Johnsen (1990). "Social influence and opinions.". *Journal of Mathematical Sociology* **15**: 193-206.
- Fuhs, T. (1998). Les membranes linéaires par morceaux : une approche géométrique de la boucle abduction-induction dans les arbres et listes de décision . Caen.
- Galam, S. (2000). "Les réformes sont-elles impossibles ?" *Le Monde*: 18-19.
- Galam, S., B. Chopard, A. Masselot et M. Droz (1997). Competing Species Dynamics: Qualitative Advantage versus Geography: 7.
- Gardner, H. (1985). *Histoire de la révolution cognitive*, Payot.
- Gell-Mann, M. (1997). *Le quark et le jaguar*, Flammarion.
- Gilbert, N. (1998). "The simulation of social processes". *Systèmes Multi-Agents pour la Gestion de l'Environnement et des Territoires*, Clermont-Ferrand, France, Cemagref.
- Gilbert, N. et R. Conte, Eds. (1995). *Artificial Societies*, UCL Press.
- Goldenberg, J., B. Libai et E. Muller (1999). Modelling New Product Growth through Cellular Automata: is Consumers' Heterogeneity a Barrier to Aggregate Analysis.
- Gopnik, A. (1993). "How we know our mind : the illusion of first person knowledge of intentionality." *Behavioral and Brain Sciences*(16): 1-14.
- Granovetter, M. (1978). "Threshold models of collective behavior." *American Journal of Sociology* **83**: 1360-1380.
- Grimm, V. (1999). "Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future ?" *Ecological Modelling* **115**: 129-148.
- Hegselmann, R. et U. Krause (2002). "Opinion Dynamics and Bounded Confidence Models, Analysis and Simulation." *Journal of Artificial Societies and Social Simulation* **5**(3).
- Hush, D. et C. Scovel (2003). "Polynomial-Time Decomposition Algorithms for Support Vector Machines." *Machine Learning* **51**(1): 51-71.
- Kaufmann, L. (1999). "Solving the quadratic programming problem arising in support vector classification". *Advances in Kernel Methods - Support Vector Learning*. B. Schölkopf, C. J. C. Burges and A. J. Smola, MIT Press: 169-184.
- Kohonen, T. (1982). "Self organization of topologically correct feature maps." *Biological cybernetics* **43**: 59-69.
- Krause, U. (2000). "A discrete nonlinear and non-autonomous model of consensus formation". *Communications in Difference Equations*, Amsterdam, Gordon and Breach.
- Latané, B. et A. Nowak (1997). "The cause of polarisation and clustering in social groups." *Progress in Communication Sciences* **39**: 1-24.
- Lenay, C. (1997). "Invariants sensorimoteurs et perception". *Journées de Rochebrune sur la complexité*, ENST.

- Mangasarian, O. L. et D. R. Musicant (1998). Successive overrelaxation for support vector machines, University of Wisconsin in Madison.
- Monneret, E. (1995). Apprentissage multi-niveaux dans des réseaux connexionnistes auto-construits. Caen.
- Müller, J.-P. (1998). "Vers une méthodologie de conception de systèmes multi-agents de résolution de problèmes par émergence". *Journées Francophones de l'Intelligence Artificielle Distribuée et des Systèmes Multi-Agents*.
- Murty, K. (1997). *Linear complementarity, linear and non linear programming*.
- Nadal, J.-P., O. Chenevez, G. Weisbuch et A. P. Kirman (1996). A formal approach to market organization: choice functions, mean field approximation and maximum entropy principle.
- Petitot, J. (1992). *Physique du sens*, Editions du CNRS.
- Pinker, S. (2000). *Comment fonctionne l'esprit*, O. Jacob.
- Prigogine, I. et G. Nicolis (1989). *Exploring complexity*. San Francisco, W. H. Freeman & Co.
- Renaut, A. (1989). *L'ère de l'individu*, Gallimard.
- Rogers, E. M. (1983). *Diffusion of innovations*. New York, Free Press.
- Searle, J. (1992). *La redécouverte de l'esprit*, Gallimard.
- Segal, J. (2003). *Le Zéro et le Un, Histoire de la notion scientifique d'information*. Paris, Syllepse.
- Smola, A. J., B. Schölkopf et K. R. Müller (1998). "General cost functions for support vector regression". *9th Australian Conference on Neural Networks*, University of Queensland.
- Thom, R. (1988). *Esquisse d'une sémiophysique. Physique aristotélicienne et théorie des catastrophes.*, InterEditions.
- Todorov, T. (1995). *La vie commune. Essai d'anthropologie générale*, Seuil.
- Todorov, T. (1998). *Le jardin imparfait. La pensée humaniste en France.*, Grasset.
- Valente, T. W. (1995). *Network Models of the Diffusion of Innovations*. Cresskill, New Jersey, Hampton Press, Inc.
- Valente, T. W. (1996). "Social network thresholds in the diffusion of innovations." *Social Networks* **18**: 69-89.
- Vapnik, V. (1998). *Statistical Learning Theory*, Wiley.
- Varela, F. (1989). *Autonomie et connaissance*, Seuil.
- Wasserman, S. et K. Faust (1994). *Social Network Analysis : Methods and Applications*. New York and Cambridge, Cambridge University Press.
- Watts, D. J. et S. H. Strogatz (1998). "Collective dynamics of 'small-worlds' networks." *Nature* **393**.
- Weidlich, W. (2000). *SocioDynamics: A Systematic Approach to Mathematical Modelling*, Harwood.
- Weisbuch, G. (1991). *Complex systems dynamics*, Santa Fe Institute Publications.

- Weisbuch, G. et G. Boudjema (1999). "Dynamical aspects in the Adoption of Agri-environmental Measures." *Advances in Complex Systems* **2**: 11-36.
- Winograd, T. et F. Florès (1986). *Understanding Computers and Cognition: a new foundation for design*, Norwood.
- Young, P. (1998). *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*, Princeton University Press.
- Young, P. (1999). Diffusion in Social Networks.

9 Annexe : le logiciel SICOPTER d'aide à la formulation de problèmes

Cette annexe décrit une partie du travail de la thèse de Lambert Rousseau, réalisé dans le cadre du projet FIRMA. L. Rousseau, S. Bernard et moi avons conçu le logiciel, et S. Bernard en a développé l'essentiel. Il repose sur l'idée, défendue par L. Rousseau, d'utiliser la comparaison de points de vue comme moyen privilégié pour définir des problèmes collectifs. Ma participation concerne deux aspects particuliers :

- La définition des points de vue, dans laquelle j'ai beaucoup insisté pour distinguer le modèle de fonctionnement du modèle de préférence. Cette distinction permet ensuite de distinguer différents types de désaccords entre les participants, qui se résolvent différemment. Il s'agit en fait de la distinction entre faits et valeurs, que nous évoquons au chapitre 2. Dans le papier qui suit, il s'agit d'ailleurs uniquement d'organiser une discussion sur les valeurs, individuelles et collectives, d'un aménagement.
- La définition d'expression des préférences sous forme géométrique, qui permet ensuite la définition d'indicateurs globaux.

Par ailleurs, j'ai supervisé l'organisation des tests par les utilisateurs potentiels, qui ont été très riches d'enseignements et nous ont amenés plusieurs fois à faire évoluer l'application.

Confronting evaluations of scenarios via Internet

Lambert Rousseau, Stéphan Bernard, Guillaume Deffuant

Laboratoire d'Ingénierie des Systèmes Complexes

Cemagref - Aubière

Abstract : We propose a generic software tool supporting participatory decision processes based on simulation of scenarios. The main objective of the tool is to provide information about the state of the participatory process, by enhancing the current consensus and conflicts. The tool provides a generic interface, which is connected to a database of model results. The user can visualise the results corresponding to the scenarios for different attributes of objects located in the space, and a variety of indicators related to these variables. The user expresses evaluations of scenarios by specifying thresholds of acceptance of objects attribute values. The software provides then indicators of comparisons between the expressed evaluations, highlighting the differences of similarities between the expressed evaluation. A series of tests with a group of potential users is currently in progress.

Introduction

To increase and improve the participation of the variety of stakeholders in the decisions is an objective which is explicitly expressed by the European commission (Van Der Veen, A.,2000), and which is also a major aspect of the FIRMA project. Several directions are considered :

- animation techniques such as role games (Barreteau et al.,2001), (Hare et al., 2001),
- participative elaboration of agent based models (Bousquet et al., 2002) ,

using models as a basis of discussion among stakeholders

In all these directions modern software technologies and internet offer new potentialities. The development of Internet based role games, as well as the development a user friendly software environments for the rapid elaboration of agent models (Cormas, MIMOSE, SDML) are illustrations of this trend.

Our present work also follows this trend, in the context of a discussion over an already developed model. We suppose that the model is reasonably accepted and understood by the stakeholders, and that several possible scenarios of action for the future are identified. Our main hypothesis is that two nested dynamics must be considered in a group discussion trying to decide possible future actions (Rousseau 2001):

the private elaboration of a public position by each participant, which takes into account his own interests, and also the potential reaction of the other participants to his view. We consider that elaboration can be reactivated with the expression of new views by other participants, which may lead to a change in the public opinion.

the public elaboration of the group position, which is an interpretation of all the public opinions at a given moment. This interpretation may lead to the adoption of the decision, or to acknowledge a conflict situation, or even to agree on the exploration of a new option which appeared in the discussion.

These dynamics are linked by the exchange and the design of concrete representations of the problem called artefacts as described in the distributed cognition field [Grosjean, 2000 #206]. The whole process can be viewed as a problem formulation which progressively builds the decision problem, both at the collective and individual level [Landry, 1995 #27].

Our aim is to propose a software tool which helps the participants in both cycles. In particular, we aim at providing indicators and visual representations of the expressed opinions, which help to identify the areas of conflict or of agreement. These tools could be used by a single participant to help him to elaborate his public opinion, and by the whole group to get a better view of the state of the discussion.

Different uses can be envisaged for such a tool in the context of a role game or a real discussion about a decision. The possibility for distant interactions can modify significantly the process, as well as the role of a moderator / facilitator. We planned to test it firstly as a game role with a group of technicians (experts) of the Orb valley, using a hydrological model predicting the consequences of several scenarios of dike building for flood protection.

The paper is organised as follows : the first section provides the hypotheses and the conceptual framework, the second section describes the main indicators and visual representations of the expressed evaluations, the third section describes the main choices made for the first version of the software.

Hypotheses and conceptual framework

Private and public dynamics of discussion

We suppose that one can distinguish between a private and a public process in a negotiation or a discussion about possible options of actions and their consequences on the future :

the private process correspond to a reflection on the next public position to adopt. It takes into account the own personal interests of the participant, and his knowledge of the others views (to anticipate their reactions to a new position)

the public process is fed by the public expressions of the participants, and a more or less shared view over the state of this process, which may lead to different actions. For instance, all participants may agree that the discussion is in a dead end, or that a consensus is reached, or that a new option must be considered in the discussion, or other participants must be invited...

We note that both processes are deeply interconnected : the public process is constructed by the expressions of the private reflection, and in return the private reflection takes into account the public group state of the discussion.

In this framework, we consider that software tools can help the participants in both processes, especially by providing calculations and representations of the expressed opinions relatively to each other. Such tools may help to clarify the issues of a discussion and the formulation of the problem, and avoid wrong diagnosis about its state. We suppose that it can lead to reduce the effects of group-think, weak consensus and misunderstanding conflicts. Moreover, it could help the participants to be more efficient, by a better anticipation of the others reactions.

The approach is first designed to deal with problem formulation. There is evidence that it is a necessary step for negotiation or group problem solving for integrative management of natural resources because they are shared and used by a large variety of consumers. Then our tool might help to identify that other scenarios might be considered, or other participants, or other data...

We consider the development of such tools in a specific context : the availability of a model.

Distributed functioning through the internet

We suppose that the results of the model for the different scenarios are available in a data base, accessible through internet. The users progressively load their evaluations on the data base, as well as their comments, as shown on figure 1.

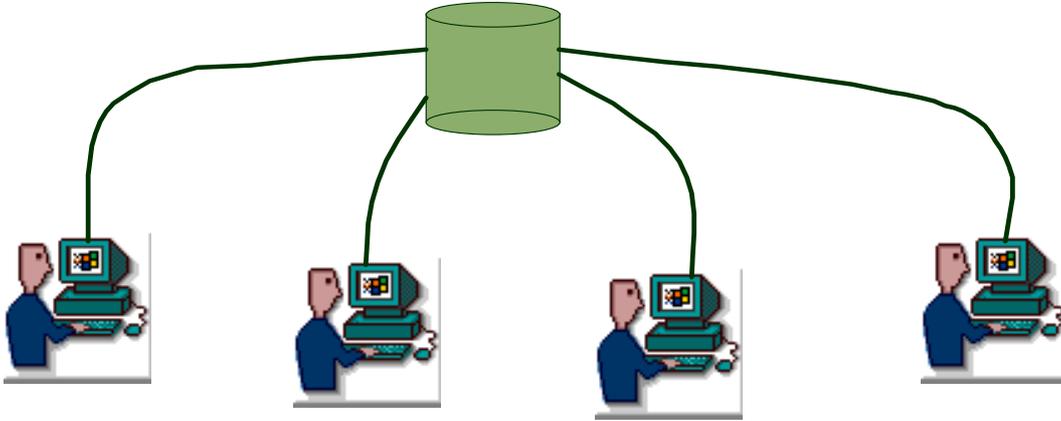


Figure 1 : the participants are related to a data base of model results, evaluations and comments through internet.

Model of territory

Our approach is relies on several important general assumptions about the model :

- the model of a territory is available,
- a set of possible scenarios S_1, S_2, \dots, S_n of action were identified, and can be simulated by the model,
- a set participants P_1, P_2, \dots, P_m accepted a discussion over the simulations of the scenarios by the model, and they have a reasonable confidence and knowledge of the model.

We shall not discuss here in depth the realism of these assumptions. However, let us note that hydrological or economic models are already used as a help to take decisions. One can expect this use to generalise with the development of computer power and modelling techniques.

Moreover, we make several more specific assumptions about the results of the model for each scenario :

- the model gives information on a set of objects or locations O_1, O_2, \dots, O_p of a territory (for instance : a dam, a location on a river or on its bank...), which can be represented on a map,
- a set of attributes A_1, A_2, \dots, A_q describes each object
- the model provides, for each scenario of action, for each attribute of each object, the corresponding temporal evolution : $S_i O_j A_k(t)$ (for instance : the water level during a particular flood at a given place for a scenario of dike building)

We suppose that the chosen attributes are relevant for the participants and that the values of these attributes over the time are an issue at stake for the discussion. In the following we will designate a set of temporal evolutions by *SOA*, meaning a choice of scenario(s), object(s) and attributes(s).

Participants specifying upper and / or lower levels of acceptance on attribute temporal evolution

We suppose that the participants are able to specify upper and /or lower levels of acceptance on some of the temporal evolution of object attributes. These upper and /or lower level may evolve with time, for instance, the constraints on water level can vary with the seasons.

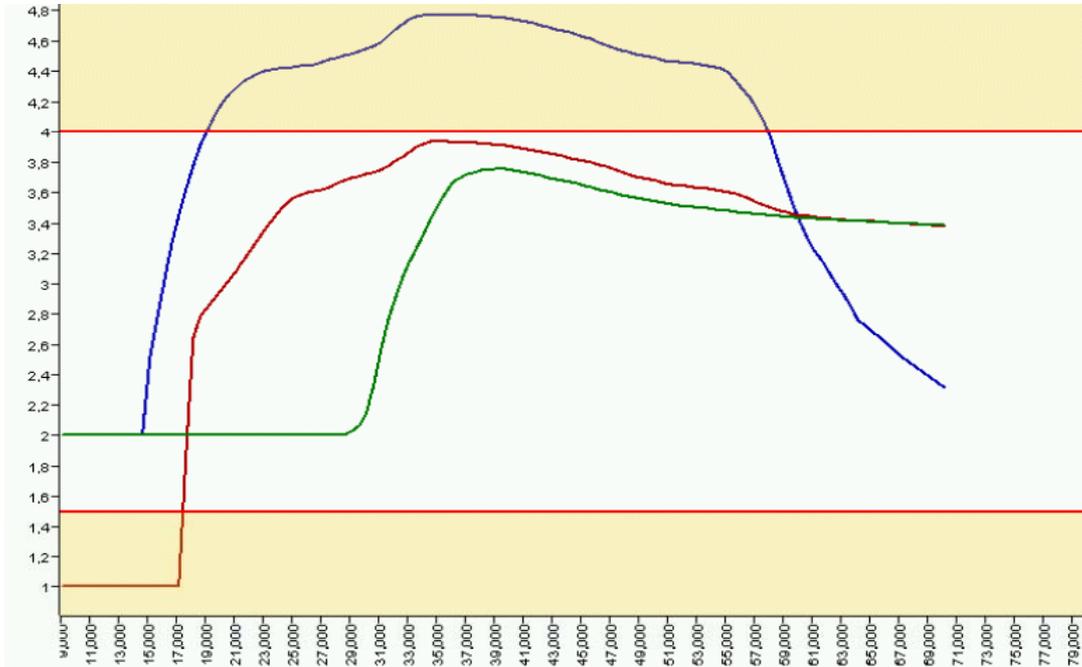


Figure 2: Example of evaluation. The curves represent the evolution of the water level (in meters) at particular places for the exceptional flood, the horizontal axis is the time in hours. The user defines the levels of the curves which he refuses (yellow zones).

We consider that the opinions of the participants are expressed by such zones of acceptance of the temporal values of attributes. The process of interactions with others leads the participants to consider the constraints of the others, and possibly to modify their view because of their knowledge of the others.

Formally, the evaluation of a participant P_i is noted $P_i S_i O_j A_k(t)$, and is a couple of piecewise constant functions (which might be void), representing the upper and the lower acceptance levels. The possibility for entering fuzzy levels of acceptance is currently under study. In the following, a set of evaluations by participants on particular a SOA will be denoted $PSOA$. This particular definition of an evaluation allows us to define a set of indicators and visual representations of the expressed opinions in systematically comparing the different evaluation on the same SOA .

Map indicators related to a sets of evaluations

The aim of the indicators and visual representations is to help the participants in the elaboration of their public evaluation or to understand the global state of the discussion. They give synthetic information about the state of the discussion, and can be considered at different levels of aggregation (participant, scenario, object, attribute, attribute of a given object). The principle of calculation is to define an operation at a time step, and then to aggregate over all the concerned time steps for the chosen aggregation level. We shall note $PSOA$ the support of

aggregation of the indicator, defining the participants, the scenarios, the objects and attributes on which the aggregation is performed.

One can see that the aggregation level has to be well decided to be relevant. Methodology to guide such aggregation can be found in the multi-criteria decision making literature (such as (Bouyssou,2000)).

This indicator proposes a measure of how much the participant(s) focused on a given aggregation support *PSOA*. At a given time step, the focalisation ϕ is defined by :

- $\phi(P_i S_i O_j A_k(t))=1$ if $P_i S_i O_j A_k(t) \neq \{ \}$ (there is an upper or a lower level of acceptance at this time step)
- $\phi(P_i S_i O_j A_k(t))=0$ if $P_i S_i O_j A_k(t) = \{ \}$ (there is an upper or a lower level of acceptance at this time step)

The aggregated value of focalisation on support *PSOA* is given by :

$$\phi(PSOA) = \frac{1}{K} \sum_{P_i S_i O_j A_k \in PSOA, t \in T} \phi(P_i S_i O_j A_k(t))$$

Where K is the number of considered time steps in *PSOA*, T is the set of time steps for the considered attribute time evolution.

Acceptance (refusal) of the scenarios

This indicator measures whether the temporal evolution corresponding to the chosen scenarios are compatible with the levels of acceptance expressed by the participants. In a first version we only consider a qualitative (bi-valued) evaluation. It can be 1 (for acceptance) and 0 (for reject). A continuous evaluation (valued in $[0,1]$) is under study. The acceptance of the scenarios, noted α is given by :

- $\alpha(P_i S_i O_j A_k(t))=1$ if $S_i O_j A_k(t) \in P_i S_i O_j A_k(t)$ (the value of the attribute is compatible with the acceptance constraints expressed by the participants at this time step)
- $\alpha(P_i S_i O_j A_k(t))=0$ if $S_i O_j A_k(t) \notin P_i S_i O_j A_k(t)$ (the value of the attribute is *not* compatible with the acceptance constraints expressed by the participants at this time step)

The aggregated value of acceptance of the scenarios is given by :

$$\alpha(PSOA) = \frac{1}{K} \sum_{P_i S_i O_j A_k \in PSOA, t \in T} \alpha(P_i S_i O_j A_k(t))$$

Where K is the number of considered time steps in *PSOA*, T is the set of time steps for the considered attribute time evolution.

The refusal indicator is calculated by the summation of $1-\alpha$ instead of α in the definition of the function at the time step level.

Conflict (agreement) between participants on the scenarios

This indicator supposes that at least two participants are chosen in *PSOA*. The agreement indicator expresses that the participants agree to accept or refuse the scenario at the same time. It is based on the Gini criterion used in decision trees (Breiman et al.). A similar

function based on information could be considered. We note κ the conflict indicator, for a set P of participants, it is calculated at a given $S_i O_j A_k(t)$ as follows :

Let $N = \{P_i \in P \mid \alpha(P_i S_i O_j A_k(t)) = 0\}$ be the set of participants of P who do not accept the value $S_i O_j A_k(t)$, and $|N|$ the cardinal of N , then,

$$\kappa(PS_i O_j A_k(t)) = 4 \cdot \frac{|N| \cdot (|P| - |N|)}{|P|^2}$$

One can verify that this function is maximum when half of the considered participants accept the scenario, and half of them refuse, and its value is 1. When all participants accept or refuse the scenario, it is 0.

The aggregated value of conflict of the scenarios is given by :

$$\kappa(PSOA) = \frac{1}{K} \sum_{\substack{S_i O_j A_k \in SOA \\ t \in T}} \kappa(PS_i O_j A_k(t))$$

Where K is the number of considered time steps in SOA , T is the set of time steps for the considered attribute time evolution.

The indicator of agreement between participants on the scenarios is given by : $1 - \kappa$.

Distance indicators

The definition of a distance between two evaluations at a given time step can be the basis of several indicators : the average distance between all the selected participants, maximum or minimum distance or any indicator of dispersion.

These indicators can be calculated at any aggregation level as the acceptance (refusal) indicator.

Moreover, such a distance would open the possibility to perform a hierarchical classification of the participants, based on the proximity of their expressions. This would open perspectives for supporting the identification of potential coalitions.

The definition of this distance is currently under study (several possibilities are available). It is particularly interesting to generalise it to the case of fuzzy acceptance constraints.

Privileging a participant

In the previous indicators, we considered equally all the participants. Consider now that we are more interested by a particular participant P_0 , and his relation to a set of other participants P relatively to a given SOA .

The conflict indicator can be adapted : consider now the set $N = \{P \in P \mid \alpha(PS_i O_j A_k(t)) \neq \alpha(P_0 S_i O_j A_k(t))\}$. Using the same calculation on this new set N , we obtain a measure of the isolation of a participant.

The distance indicators can be adapted as well : one can compute for each time step the distance to P_0 , and compute average, minimum and maximum distances or any indicator of dispersion.

Representations on time graphics

The point is now to give the participants a higher level of details, relatively to a particular attribute. The main difference with the indicators is that the aggregation can only be made on *SOAs* on which the evaluation of all the participants is constant. Otherwise, there are ambiguities which prevent an easy graphical representation. The software will have to identify automatically these *SOAs*, and propose a the corresponding visualisation to the user.

We now give some examples of such graphical representations. For all of them, we suppose that we consider a single *SOA* on which the evaluations are stable for all the considered participants. We suppose that:

Only one attribute is selected, but it may concern different objects, and several scenarios.

The possible values v of the attribute vary in a discrete manner between in the set V (the latter hypothesis is only for the convenience of presentation, it is not formally required for the definition of the representations).

Density of acceptance (refusal)

At each time step and each value of the attribute, we represent by a different colour the proportion of participants who accept the value of the attribute.

This representation highlights the zones (in time and attribute values) where all participants accept the values. Futures developments will take into account several degrees of agreement dues to the continuity of the evaluation function.

Density of conflict (agreement)

At each time step and each value of the attribute, we represent by a colour the product of proportion of participants who accept the value of the attribute by the proportion of participants who refuse the value of the attribute.

This representation underlines the zones (subparts of attribute temporal evolution) where the conflicts are the most salient (from the evaluations).

The software

We implemented a first version of the software, and performed first tests. This section describes the main features of this version.

At the beginning of the session, the participants login by clicking on their name in a login window (figure 3).

Then, two main types of windows are considered :

a discussion archive window, which offers the access to all already expressed evaluations and the connected comments. In this first version, we neglected the distinction between private and public, because we want to make a first test focused on the definition of the indicators and the public process.

Evaluation and comparison visualisation or edition. In the evaluation edition window, the possibility to add upper and lower acceptance levels is available.

Login window

The roles of the possible participants appear in the login window at the beginning of the session. All the evaluations and comments are signed by the login name.

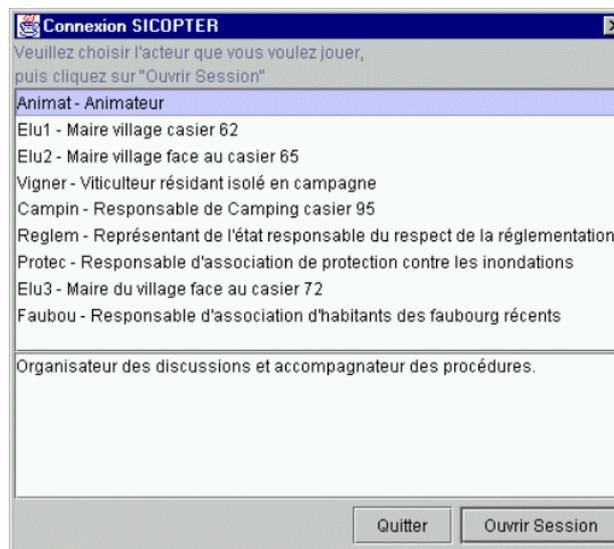


Figure 3 : login window. A list of participants appear on the top. When choosing one of them, a short description and possible comments appear below. When the role is chosen, the participant can open the session with the chosen role.

Opening the session opens the archive window, which is now being described.

Discussion archives

An example of this window devoted to the discussion archive is shown on figure 4.

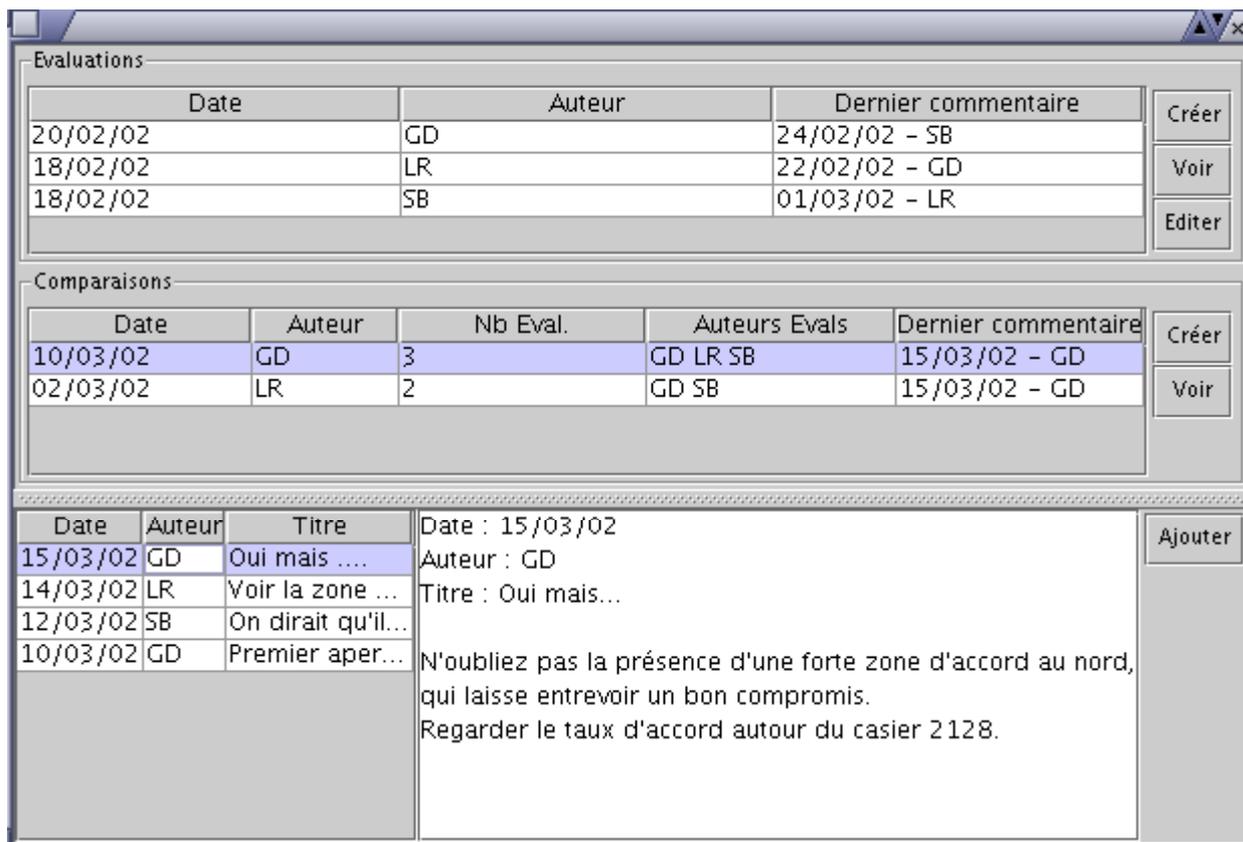


Figure 4 : Window of discussion archives

The window comprises three different parts :

On the top, appears a list of evaluations, designated by their author, the date of the evaluation and the date of the last comment about the evaluation, with the initials of the participant who made the comment. On the right, a series of buttons allow the user to create a new evaluation, to visualise or to edit a selected evaluation.

Just below, a similar list of comparisons, with the date of the comparison, the author, the number of concerned evaluations, the initials of their authors, and the date and author of the last comment about this comparison.

At the bottom left of the window, the list of comments about the selected evaluation or comparison appears. Each comment is designated by a date, an author and a title (optional). When selected, the comment appears on the right.

Note that the software invites the participants to communicate through comments on the different evaluations and comparisons. In a further version, we envisage the possibility to illustrate the comment by a particular view of the evaluation of comparison.

Visualisation or edition of evaluations

An example of the edition or visualisation window is shown on figure 5.

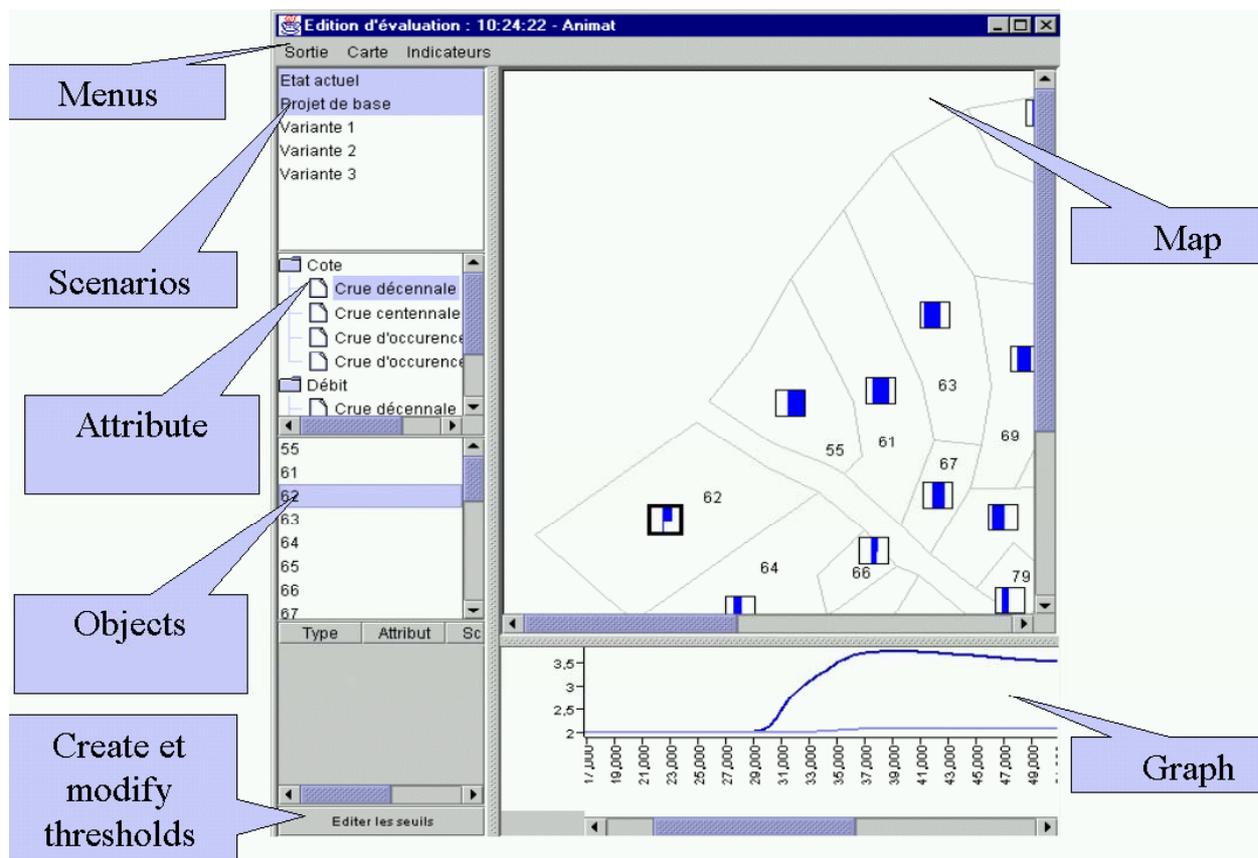


Figure 5 : evaluation edition window.

This window comprises two main graphic zones :

- the top one is the representation of the map of the territory, and the indicator values are represented as colours in icons.
- The bottom one is the representation of the threshold upper or lower levels for the considered selection of *SOA*.

On the left, the user can select the *SOA*. The objects can also be selected on the map.

Visualisation of data on the map

The tool allows the user to visualise the results of a model, and to compare these results for different scenarios. The choice of indicators is done with the menu. Several possibilities are provided :

- min/max values of the time series
- mean values of the time series
- maximum difference between scenarios

On each location corresponding to an object, in the example the objects are zones on which the model computed a water level, an icon provides the corresponding information for the chosen scenarios : the icons are horizontally separated, and each line corresponds to a scenario. This allows the user to compare rapidly the outputs of the chosen scenarios.

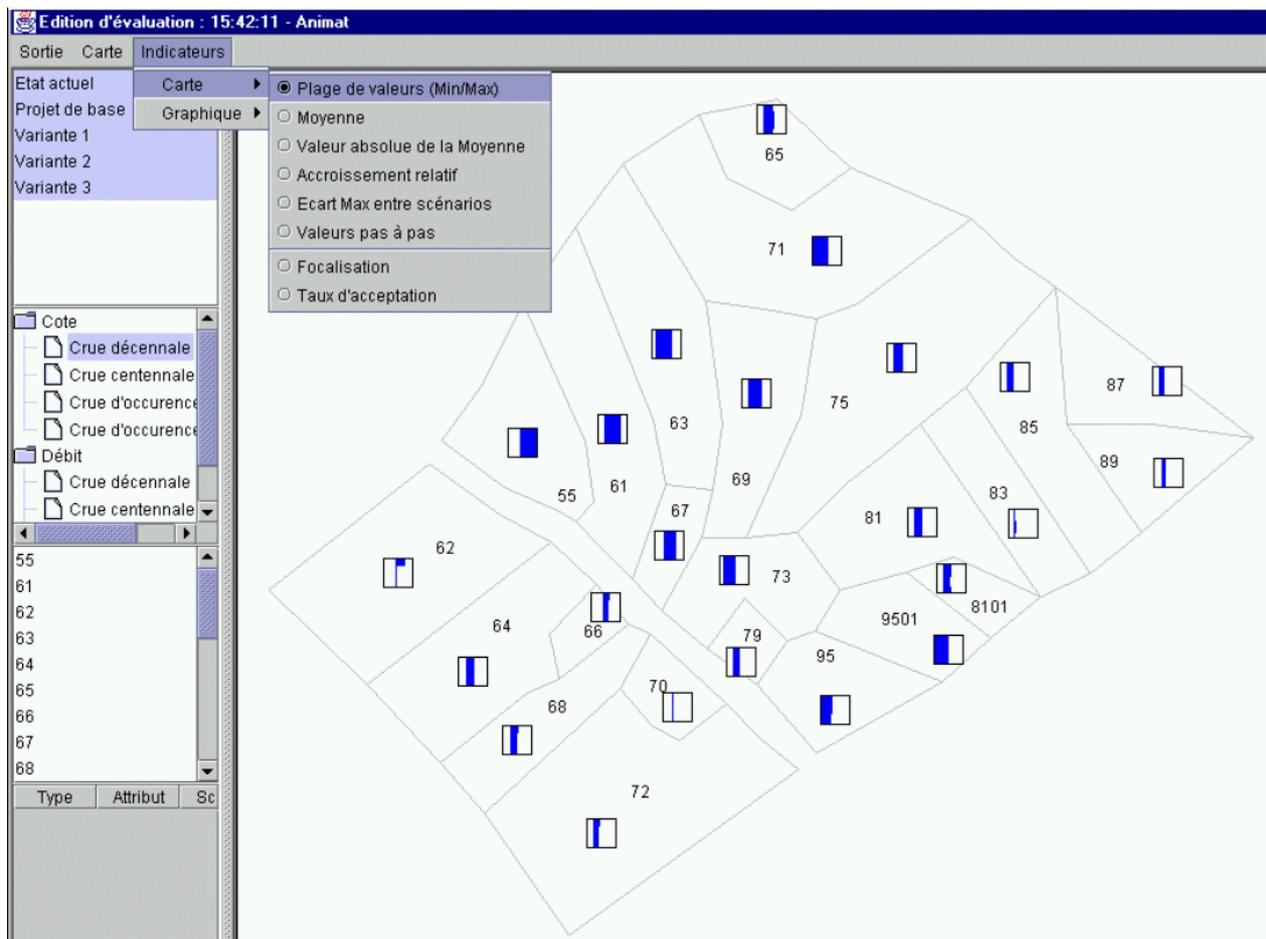


Figure 6: Example of visualisation of data on the map. The icons show the values of the chosen indicator (here min and max value) for the selected scenarios. On zone 62, we note the the first scenario (current situation) leads to a higher level of water than all the other scenarios.

Visualisation of the data on the graph

The graph below the map provides the time series of model results for the chosen objects and attributes. Colour codes allow the user to distinguish the objects and the scenarios :

- the colours attributed to the object in their order in the list window,
- the scenarios are lighter and lighter also in their order in the scenario window.

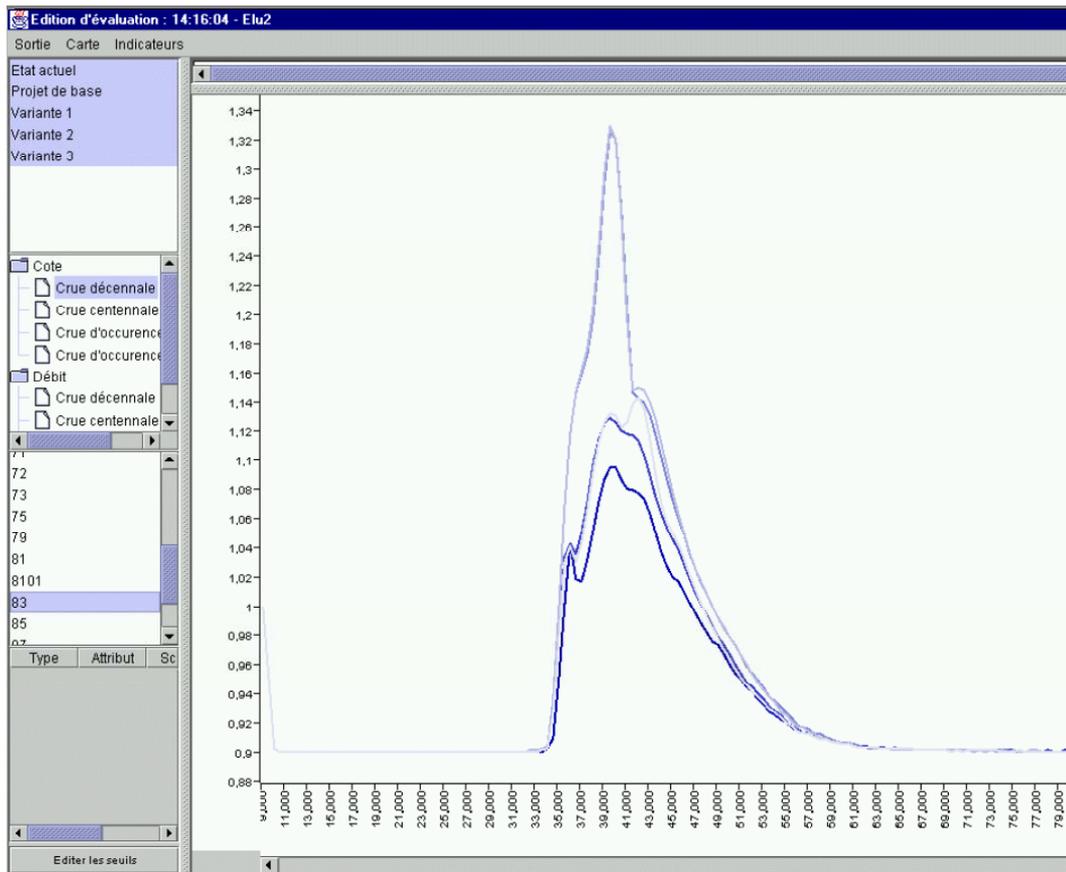


Figure 7 : example of time series visualisation. Here only one object is selected (zone 83) for the 5 scenarios, and the attribute water level for 10 year period flood. The darker time series correspond to the current state, and the lighter to the other scenarios.

Creation or modification of an evaluation

New threshold upper or lower levels of acceptance can be defined on the selected *SOA*, from the window of evaluation edition. This action opens a new window of the type presented on figure 8. The main noticeable point is that one must consider the other parts of the evaluation that could be affected by a change on the selected *SOA*. The *SOAs* on which an evaluation is already present are figured on the top left part. The user can visualise them and decide whether he wants to erase them or not.

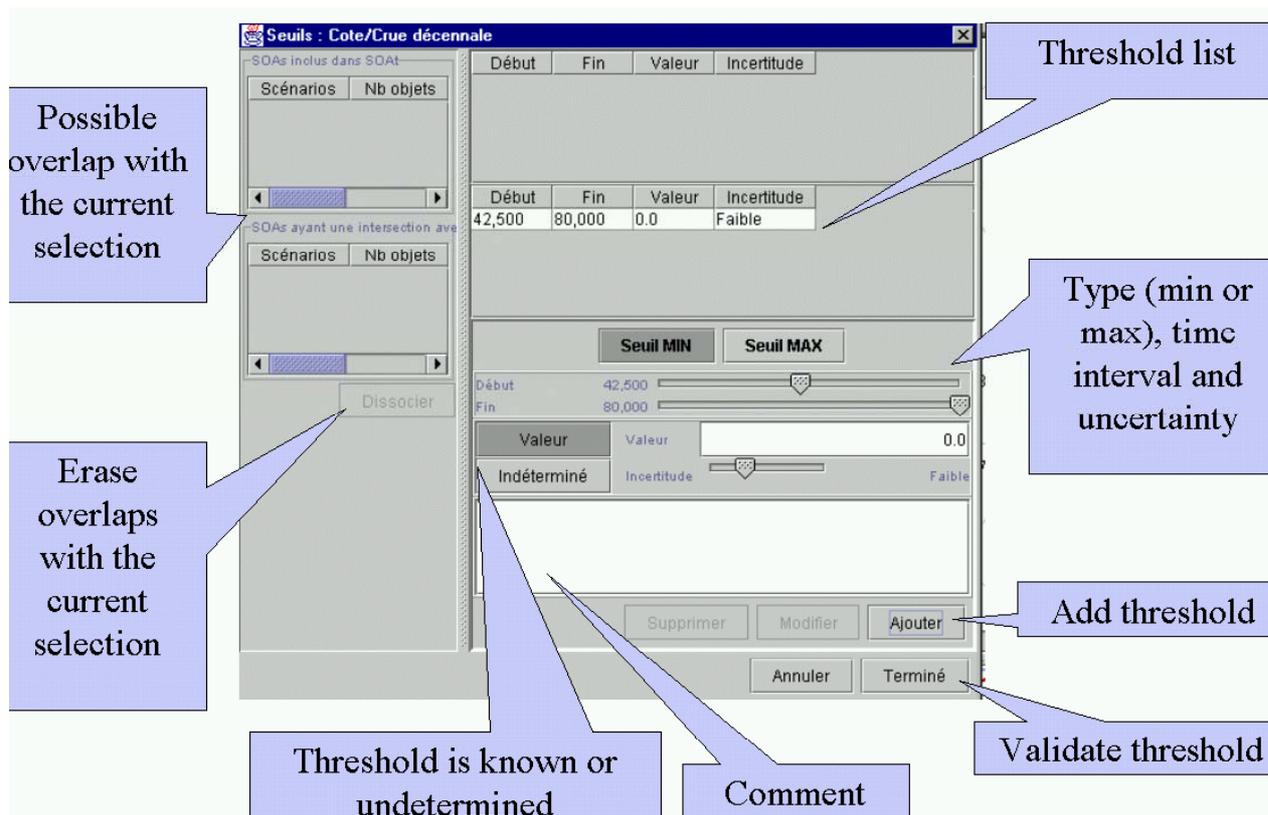


Figure 8 : Definition or modification of upper or lower levels of acceptance on an attribute.

Indicator of evaluation on the map

When evaluations have been expressed, it is possible to access some indicators about them on the map. In the current implementation, two are available :

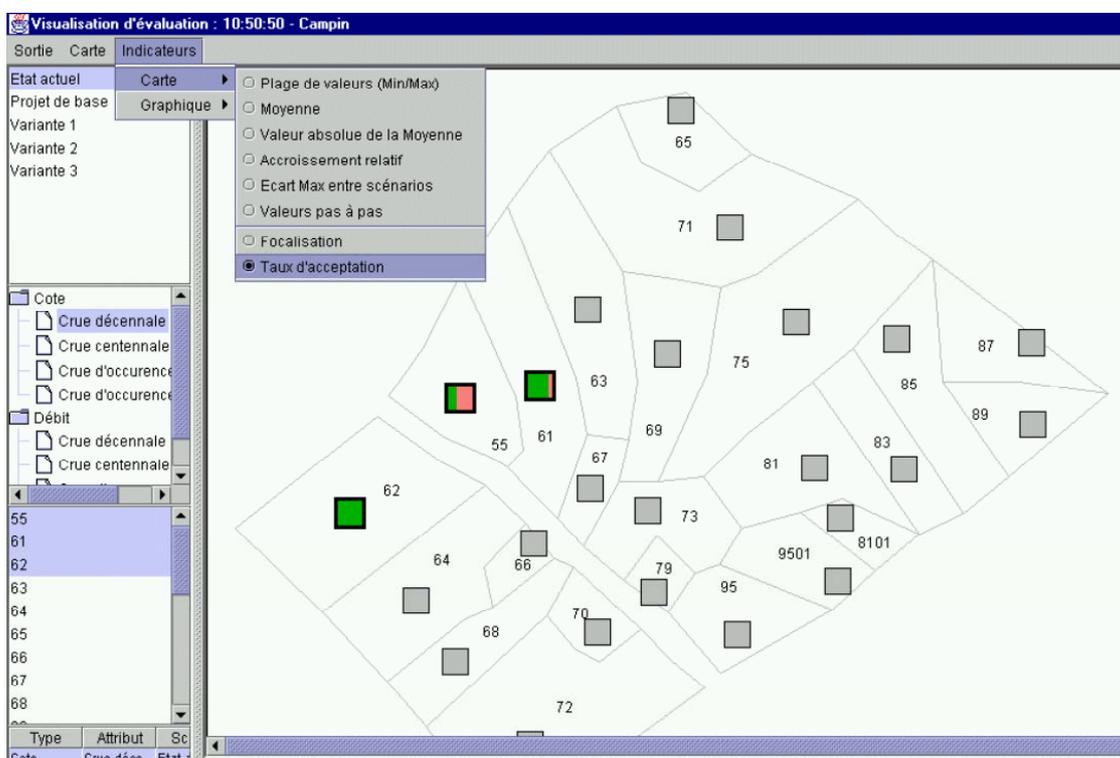


Figure 9 : indicator of evaluation on the map acceptance level.

- Focalisation which indicates the number of time steps on which the thresholds is put. This is visualised in the icons by a green horizontal bar.
- Acceptance level which indicates the number of time steps for which the scenario is in the acceptance zone defined by the user's thresholds. The proportion of accepted time steps is visualised as a green bar. The proportion of not accepted time steps is visualised in pink/red.

Visualisation of the evaluations on the graph

The current implementation offers two possibilities for visualising the evaluations on the graph :

- the thresholds on the time series : the refusal zones appear in yellow,
- the acceptance zones : the acceptance zones appear in blue.

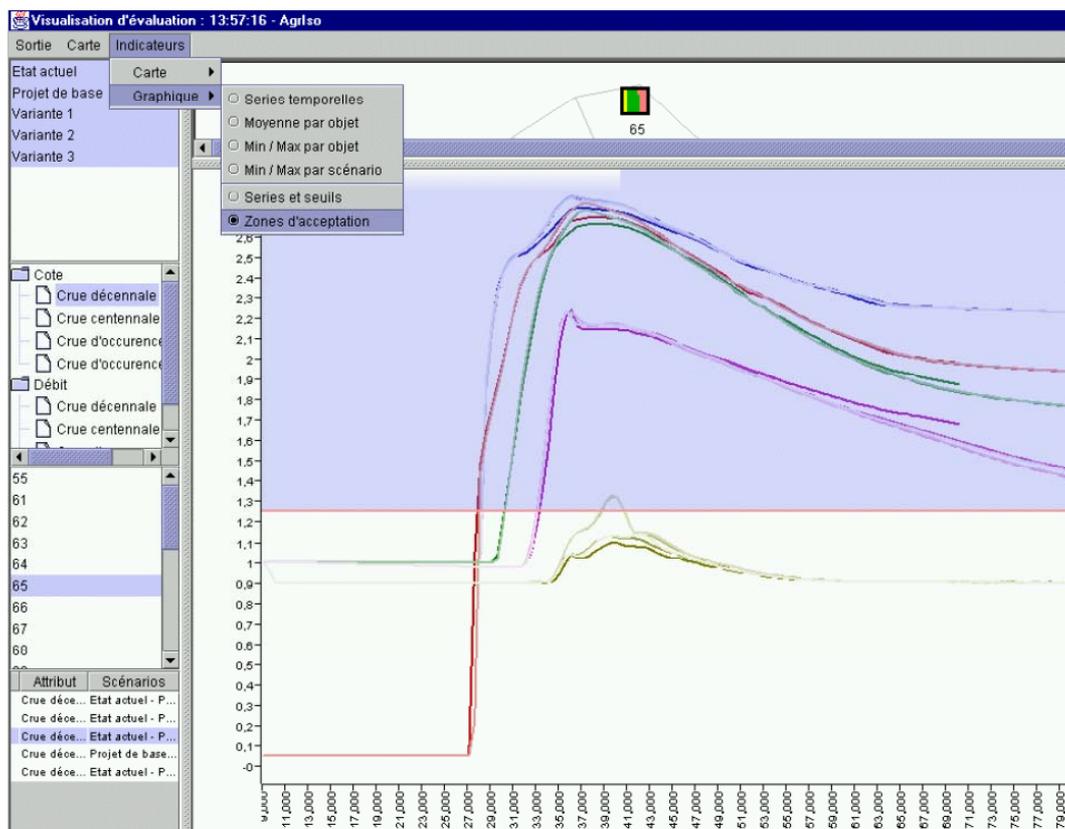


Figure 10 : acceptance zones for object 65. The corresponding icon appears on the map.

Visualisation of comparisons

A new comparison can be made from the archive window by choosing several evaluations and click the button “new comparison”. Then a comparison window appears.

An example of window for the visualisation of comparisons is shown on figure 11. This window is very similar to the window of evaluation visualisation, except that the available indicators are not exactly the same, and that a window on the top left reminds the list of evaluations on which the comparison is made. The user can visualise these evaluation directly by a double click. Note that the set of SOAs that can lead to different time graphical representations are computed from the evaluations as the selections on which all the evaluations are constant.

In the current implementation, the following indicators are available on the map :

- Focalisation, which indicates the proportion of times steps on which thresholds were put (for all the considered evaluations)
- Acceptance level : proportion of considered time steps for which the scenario is accepted,
- Agreement level : indicates whether the evaluations agree or not.
- On the graph, the following- visualisations are available (for homogeneous SOAs):
- Refusal zone : the zones excluded by the thresholds are visualised in yellow. The overlapping refusal zones are indicated by a more intense colour, showing that several people refuse in this zone,

Acceptance zones : the zones accepted by the thresholds are visualised in blue. The overlaps are also figured by a more intense colour.

Agreement level : the agreement indicator is computed at each time step, and represented by a colour (green indicates that people agree, red that they disagree).

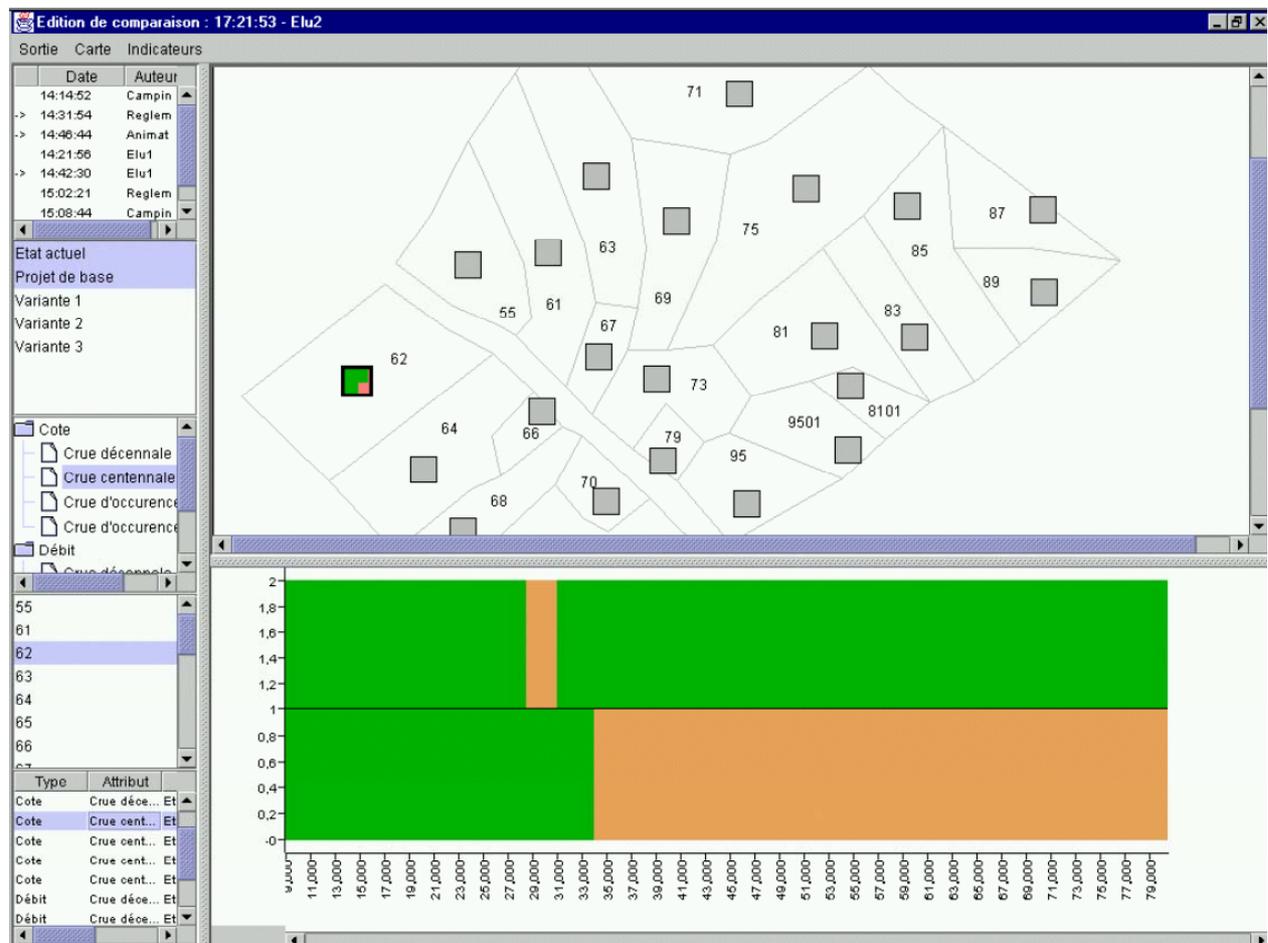


Figure 11 : window of comparison visualisation. Indicator of agreement level. On the graph window, one can see that for the second scenario, there is a lot of disagreement in the second part of the time series.

First tests

We performed two tests of the software using the results of a model developed to evaluate the consequences of dike positioning in the south of the Orb Valley. This model was developed by a specialised firm, and we were authorised by the group which financed its development to use its results in our research. However, for confidentiality reasons, they modified the map to make difficult to recognise (there are discussions currently going on about this project which are very sensitive and they don't want to take any risk).

The results of the model comprise the time series of water levels and flow between these zones, for different types of floods (10 or 100 year period), for different variants of a dyke project protecting one particular zone.

We defined a particular land use in the zone, comprising villages, camping zones, beaches, industries.

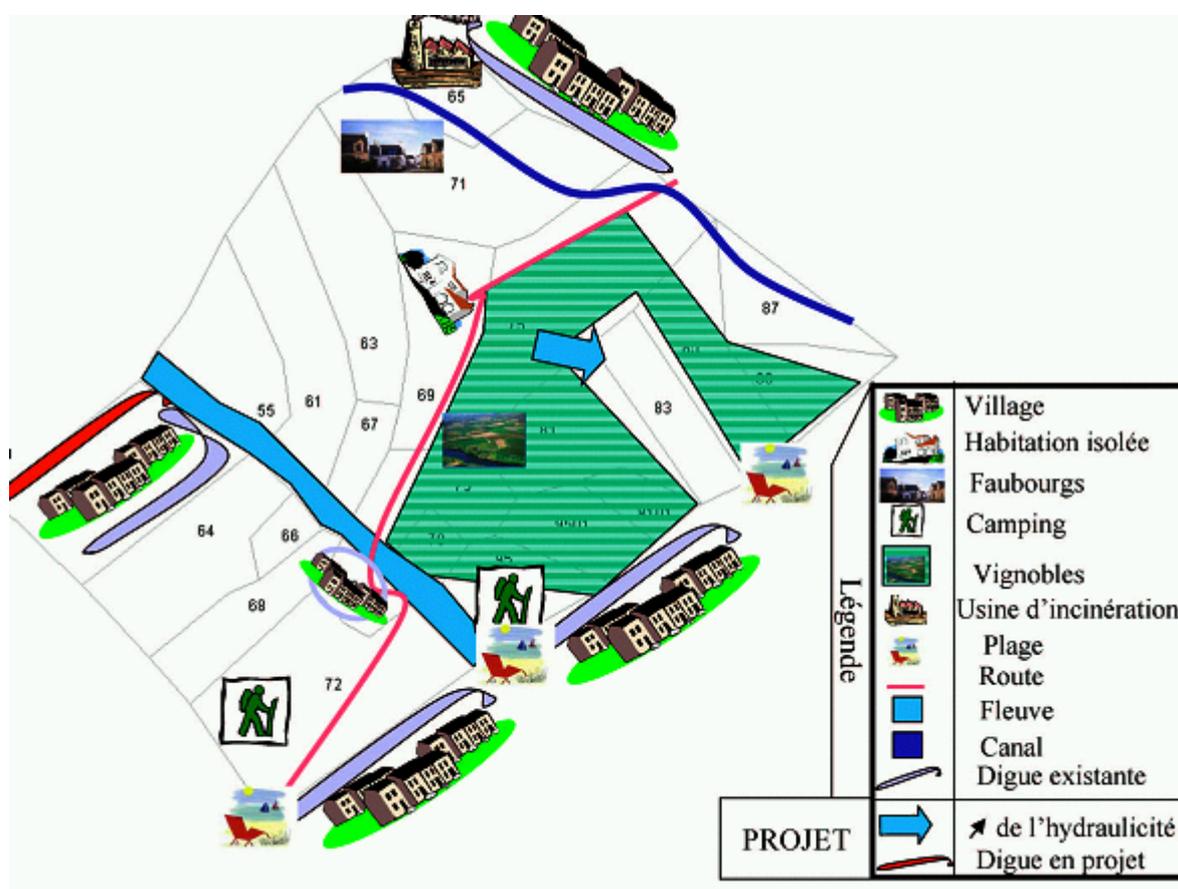


Figure 12 : land use hypothesis for the role playing test.

Moreover, we defined 8 potential roles including : the facilitator, mayors of villages, farmers, camping owner, representative of local flood protection association. The possibility to play these role was implemented in the software.

We made the first tests on a local network, with all the players in the same room.

The tests of the software were organised as role playing games, in which institutional stakeholders and some of researchers, put on roles of real actors of the basin, well known to be active in this issue of flood management in downstream part of the Orb valley. At the end of each test, a specific discussion with the players helped us to define the next test.

The first meeting allowed players to handle the tool and propose some modifications regarding the interface and some requirements of new specific indicators. This test showed us that the tool helps the players to quickly understand the results of the model and the evaluations¹⁰ of others.

The second meeting allowed to test the whole protocol of use of the tool itself. It did not appear to support the whole integration, but it facilitated the expression of evaluations clearly linked to the result of the model. The way to use the tool in the whole process of the dialogue facilitator, appears to be very important and to need some clarifications and refinements. However, the test of this tool led us already to learn on the way the facilitator uses it, since he had to uncover it in order to explain why he was not totally at ease with the introduction of the tool in the process. We then pointed out the importance of an animation scenario that define the way to use such a tool. It has to be seen as a part of the role playing scenario in the test.

The aim of the third meeting was to test a specific animation scenario. We alternated two kinds of phases. First a scenario was evaluated with the tool and second the facilitator used the viewpoints' visualisation to present a synthesis and opening further discussion on new scenarios to be tested. The results showed that the process works well and that the presentation of the viewpoints comparisons induces the actors to develop their evaluation in term of stakes. Then, with a limited set of tests, we showed that such a tool can be used both to progress in the choice of a globally preferred scenario and to induce a better mutual comprehension.

The two main results of these tests are :

- such a tool helps the players to understand the scenarios and each other's evaluation,
- the outputs are useful for the animation of the discussion.

Moreover, using these tool and methodology improves the participation of the actors in the definition of a common problem.

References

Asakawa, T. and N. Gilbert (2002), "Synthesizing Experiences: Lessons to be Learned from Internet-Mediated Simulation Games". FIRMA report.

Barreteau, O., Bousquet, F. (2001) "Des systèmes irrigués virtuels aux systèmes irrigués réels : retour par les jeux de rôles" in : Représentations spatiales et développement territorial, Lardon S., Maurel P., Piveteau V. (eds), p. 163-174, Hermès, Paris,

Bousquet, F., O. Barreteau, P. d'Aquino, M. Etienne, S. Boissau, S. Aubert, C. Le Page, D. Babin, J.C. Castella, (2002), "Multi-agent systems and role games: collective learning processes for ecosystem management". Edward Elgar Publishers

Bouyssou D., Marchant Th., Perny P., Pirlot M., Tsoukiàs A., Vincke Ph.(2000), "Evaluation and Decision Models: a critical perspective", Kluwer Academic, Dordrecht, 2000.

¹⁰ Evaluations themselves not the stakes they represent.