

Approches variationnelles et autres contributions en optimisation stochastique

Cyrille Strugarek

▶ **To cite this version:**

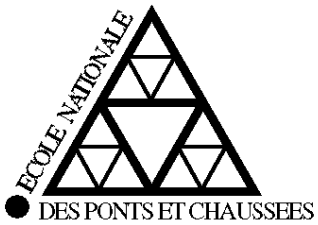
Cyrille Strugarek. Approches variationnelles et autres contributions en optimisation stochastique. Mathématiques [math]. Ecole des Ponts ParisTech, 2006. Français. NNT : . pastel-00001848

HAL Id: pastel-00001848
https://pastel.hal.science/pastel-00001848

Submitted on 17 Jul 2006

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

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



THÈSE

présentée pour l'obtention du titre de

Docteur de l'École Nationale des Ponts et Chaussées

Spécialité : Mathématiques Appliquées

par

Cyrille STRUGAREK

Sujet : *Approches variationnelles et autres contributions en optimisation stochastique*

Soutenance le 15 mai 2006 devant le jury composé de :

Rapporteurs :	Roger J.-B. Wets	University of California at Davis
	Jean-Pierre Quadrat	I.N.R.I.A., Rocquencourt
Examinateurs :	Guy Cohen	E.N.P.C. et I.N.R.I.A., Rocquencourt
	Jean-Baptiste Hiriart-Urruty	Université Paul Sabatier, Toulouse
Directeurs de thèse :	Pierre Carpentier	E.N.S.T.A., Paris
	Agnès Sulem	I.N.R.I.A., Rocquencourt
Invités :	René Aïd	E.D.F. R&D, Clamart
	René Henrion	Weierstrass Institut, Berlin



Cette thèse a été effectuée conjointement au Laboratoire de Mathématiques Appliquées de l'ENSTA, au CERMICS de l'ENPC, et dans le département OSIRIS d'EDF R&D sous la forme d'un contrat CIFRE.



Table des matières

Remerciements viii

Avant-Propos x

Résumé xii

Extended Abstract xiv
Dual effect free Stochastic Systems xiv
Stability of Multistage Stochastic Programs xv
Functional Stochastic Gradient Algorithms xv
Decomposition for Multistage Stochastic Programs xv
Objective Functions which are nonlinear functions of the Expectation xvi

Chapitre I. Introduction aux problèmes d'optimisation stochastique 2
I.1. Problème général et typologie 2
I.2. Problèmes à plusieurs niveaux et information 4
I.3. Application aux problèmes à plusieurs pas de temps 5
I.4. Enjeux d'une telle classification et plan du mémoire 6

Chapitre II. Effet dual 8
II.1. Résumé 8
II.2. Introduction 9
II.3. Recalls on dual effect and stochastic systems 9
II.3.1. Discrete-time stochastic input-output systems 9
II.3.2. How to measure information ? 10
II.3.3. Dual effect, feedback sets, and change of variables 11
II.4. Characterization of dual effect free scalar systems 13
II.5. Proof of the Characterization Theorem 14
II.5.1. Lemmas 15
II.6. Conclusion 20

Chapitre III. Stabilité 22
III.1. Résumé 22
III.2. Distance de Fortet-Mourier en optimisation stochastique 23
III.2.1. Motivation 23
III.2.2. Cadre général et résultat de stabilité classique 23
III.2.3. Convergence de la distance de Fortet-Mourier 25
III.2.4. Contre-exemple 26
III.2.5. Méthode alternative 28
III.2.6. Conclusion 30
III.3. Stabilité des problèmes stochastiques linéaires 30
III.3.1. Motivation 30
III.3.2. Distances entre tribus et filtrations 31
III.3.3. Un premier résultat de stabilité 36
III.3.4. Critères séparés 37
III.4. Perspectives 40

Chapitre IV. Algorithmes stochastiques et boucle fermée 44

IV.1. Résumé 44
IV.2. Introduction 45
IV.2.1. Ecriture naïve 45
IV.2.2. Idée fondamentale et enjeux 47
IV.3. Résolution de problèmes stochastiques en boucle fermée 47
IV.3.1. Remarques préliminaires 47
IV.3.2. Résultat de convergence 49
IV.3.3. Résultat de grandes déviations 53
IV.3.4. Remarques et illustrations diverses 54
IV.3.5. Application à un problème de gestion de réservoir 59
IV.4. Cadre général du gradient perturbé 64
IV.4.1. Approximation stochastique généralisée 65
IV.4.2. Algorithme de Arrow-Hurwicz hilbertien 70
IV.4.3. Illustrations 73
IV.5. Liens avec des idées existantes 76
IV.5.1. Régles de décision linéaires 76
IV.5.2. Espaces de noyaux reproduisants (RKHS) 79
IV.6. Conclusion et perspectives 83

Chapitre V. Décomposition des grands systèmes stochastiques 86
V.1. Résumé 86
V.2. État de l'art, ou ce qui se fait, et ce qui pose problème 87
V.2.1. Pour se fixer les idées 87
V.2.2. Problèmes stochastiques en boucle ouverte 87
V.2.3. Problèmes stochastiques en boucle fermée 88
V.3. Autour de la programmation dynamique 90
V.3.1. Dualité et cas linéaire quadratique 90
V.3.2. Commandes décentralisées 95
V.4. Principe du problème auxiliaire stochastique 100
V.4.1. Introduction et algorithme 100
V.4.2. Convergence 101
V.5. Conclusion 105

Chapitre VI. Fonctions non linéaires de l'espérance et boucle ouverte 108
VI.1. Résumé 108
VI.2. Introduction 108
VI.2.1. Motivation 108
VI.2.2. f quadratique 109
VI.3. Deux algorithmes voisins 110
VI.3.1. Approche par estimateur 110
VI.3.2. Approche par dualité de Fenchel 111
VI.4. Approche par dualité lagrangienne 112
VI.4.1. Lagrangien simple et Arrow-Hurwicz stochastique 112
VI.4.2. Lagrangien augmenté dans le cas non-convexe 113
VI.5. Résumé – Typologie 114
VI.6. Contraintes en espérance convexes 115
VI.6.1. Dualité, introduction d'un Lagrangien augmenté 115
VI.6.2. Travail à λ fixé 116
VI.6.3. Mélange des itérations de décomposition et de résolution interne 116
VI.6.4. Convergence 117
VI.6.5. Application numérique 117
VI.7. Contraintes d'égalité en espérance non convexes 118
VI.7.1. Dualité et Lagrangien augmenté 118
VI.7.2. Application numérique 118
VI.8. Contraintes d'inégalité en espérance non convexes 118

VI.8.1. Utilisation d'un Lagrangien augmenté 118
VI.8.2. Algorithme 119
VI.8.3. Contrainte en probabilité mollifiée 120
VI.8.4. Application numérique 120
VI.9. Conclusion et perspectives 122

Chapitre VII. Conclusion 124
VII.1. Synthèse 124
VII.1.1. Résumé des contributions 124
VII.1.2. Situation 124
VII.2. Perspectives 125

Annexe A. Optimisation 126
A.1. Analyse convexe 126
A.1.1. Ensembles convexes et projections 126
A.1.2. Fonctions convexes 128
A.2. Principe du problème auxiliaire (PPA) 131
A.3. Analyse fonctionnelle 132
A.4. Lemmes techniques 133

Annexe B. Quasimartingales 136

Annexe C. Algorithmes stochastiques en dimension finie 138
C.1. Schéma de Robbins-Monro 138
C.2. Application au gradient stochastique 140

Annexe. Bibliographie 144

Annexe. Index 148

If our statesmen were visionaries something practical might be done. If we ask for something in the abstract we might get something in the concrete.

G.K. Chesterton, *What's wrong with the World*, 1910.

Remerciements

Mes premiers remerciements vont aux personnes qui m'ont fait la joie et l'honneur de faire partie du jury de cette thèse. Ce manuscrit et tout le travail de thèse qui l'a précédé, doivent beaucoup à deux ouvrages de référence en optimisation,[**60, 84**], respectivement coécrits par Jean-Baptiste Hiriart-Urruty et Roger Wets. C'est donc avec beaucoup de respect que je les remercie d'avoir pris part à mon jury. Les nombreuses réflexions autour de la programmation dynamique stochastique et des algorithmes de gradient stochastique menées durant la thèse ont souvent fait surgir le nom d'un de leurs grands connaisseurs, Jean-Pierre Quadrat. Je suis donc très honoré de le compter dans mon jury, et j'espère que les travaux que j'ai menés feront entendre l'écho des siens. J'ai eu durant cette thèse la chance de faire la rencontre, au cours d'une conférence à Tucson en 2004, de René Henrion. C'est à cette sympathie initiale que je dois aujourd'hui de mieux connaître le domaine si passionnant des contraintes en probabilité, puisque nous avons depuis pu mener à bien un travail commun à ce sujet. C'est une grande joie de voir René, ami et collègue, dans mon jury. Cette thèse n'aurait pu voir le jour sans la persévérance et les encouragements de René Aïd, qui fut depuis mes premiers pas chez EDF R&D en février 2003, un soutien, tant technique que logistique, constant. Je tiens à profiter de ces lignes pour lui écrire toute ma reconnaissance, et lui redire mon estime et mes remerciements. Ce fut également un honneur d'avoir Agnès Sulem pour directrice. Je la remercie d'avoir accepté cette responsabilité, et d'avoir toujours prêté beaucoup d'attention à ce que je lui exposais, alors même que ce n'était que balbutiant. Il aurait été incongru de soutenir cette thèse devant un jury ne comptant pas Guy Cohen. Durant toute cette thèse, au cours de réunions hebdomadaires à l'École des Ponts et Chaussées, Guy a contesté, encouragé, et aiguillonné les recherches que je menais, et m'a par la même occasion montré ce que pouvait être un chercheur en optimisation. Je le remercie pour cela, et pour l'honneur qu'il me fait d'être dans mon jury. Enfin, c'est vers Pierre Carpentier que vont mes remerciements. En acceptant de me diriger, il m'a permis d'une part de poursuivre sur la lancée de mes études en optimisation à l'École Nationale Supérieure de Techniques Avancées, et d'autre part de bénéficier de ses conseils, de son amitié, et de sa bienveillance.

C'est dans les locaux d'EDF R&D, à Clamart, que j'ai passé la majeure partie de cette thèse. Parmi tous les membres d'OSIRIS que je remercie, je fais un signe particulier à René Aïd et Yannick Jacquemart qui m'ont accordé leur confiance respectivement pour démarrer cette thèse et poursuivre dans l'entreprise, à mes encadrants Nadia Oudjane et Léonard Bacaud, à mon chef de projet Xavier Warin, à mes compagnons de course à pied, Jérôme Collet, Gilles Deurveilher, Sébastien Finet, Pascal Martinetto, Wim van Ackooij, à mes compagnons de thèse Arnaud Lenoir, Arnaud Porchet et Babacar Seck, aux amateurs de rugby Hubert Fedry et Jérôme Quénu, et surtout à mes collègues de travail et amis Kengy Barty, Pierre Girardeau, et Jean-Sébastien Roy, avec lesquels nous avons fait beaucoup de travail, dans une atmosphère joyeuse et dynamique, et avec lesquels j'espère que nous continuerons à en faire.

J'ai aussi été amené à passer du temps à l'ENPC et à l'ENSTA. Je remercie Antonino Zanette et Jérôme Lelong avec lesquels j'ai beaucoup discuté autour des options américaines et des algorithmes stochastiques, Jean-François Delmas et Bernard Lapeyre, mes références probabilistes, et Sylvie Berte pour son efficacité quotidienne. Il règne au laboratoire de Mathématiques Appliquées de l'ENSTA une excellente ambiance, et je remercie tous les membres de ce laboratoire pour cela. C'était toujours un moment agréable que je passais à l'ENSTA. En particulier, je remercie Frédéric Jean (monsieur café), Jérôme Pérez (monsieur psaume), Patrick

Ciarlet qui m'a confié des petites classes de son cours de calcul scientifique, Christophe Hazard et ses discussions stimulantes, Kamel Berriri mon camarade de bureau, et j'en oublie.

Comme je l'ai déjà mentionné plus haut, des réunions hebdomadaires ont émaillé cette thèse, les réunions du groupe SOWG (Systems Optimization Working Group), composé de Laetitia Andrieu, Kengy Barty, Pierre Carpentier, Jean-Philippe Chancelier, Guy Cohen, Anes Dallagi, Michel de Lara, et Babacar Seck. Je remercie tous les membres du groupe pour cette écoute régulière, ces questions, ces aides, ces exposés, ces disputes, sans lesquelles il aurait été impossible d'avancer. En particulier, je remercie Jean-Philippe pour sa bienveillance, Michel pour sa précision, et Anes pour tout.

Grâce à EDF, j'ai également pu faire la connaissance de Werner Römisch. Je profite de ces lignes pour remercier Werner de son amitié, et de la collaboration que nous avons eue ensemble, avec Holger Heitsch, sur la stabilité des problèmes stochastiques à plusieurs pas de temps. J'espère que nous aurons l'occasion dans l'avenir de poursuivre cette collaboration. Je remercie également Alexander Shapiro pour les échanges que nous avons eu autour de son travail sur les contraintes en probabilité, et Frédéric Bonnans pour son oreille attentive lorsque je venais à l'INRIA présenter mes travaux à Agnès Sulem. L'idée d'une thèse en optimisation m'est venue alors que j'écoutais les cours d'optimisation que donnait Jean-Charles Gilbert à l'ENSTA. Je tiens ici à le remercier pour la qualité de son cours, et l'enthousiasme avec lequel il enseigne, sans lequel je ne me serais pas lancé dans l'aventure. Enfin, je remercie Bruno Bouchard pour l'intérêt qu'il a porté à nos travaux sur les approximations stochastiques fonctionnelles appliquées à la valorisation d'options américaines.

Mes derniers remerciements vont à ma famille et belle-famille. Ce sont mes parents qui par leur exemple m'ont donné dès mon plus jeune âge le goût d'apprendre, le goût de réfléchir, et m'ont toujours encouragé à faire des études. Je les remercie ici du fond du cœur pour le soutien qu'ils m'ont apporté durant toutes ces années, et les valeurs qu'ils m'ont transmises. J'ai aussi une pensée particulière pour mes deux frères et ma sœur, qui se préparent à de longues études avec ou sans thèse. Je leur souhaite de faire leur propre chemin et espère être pour eux un tremplin. Je dédie ce mémoire à mes grands parents, enfants d'immigrés ou ayant fait peu d'études. C'est de leur goût d'apprendre, d'avancer et de s'intégrer qu'a pu naître ce travail. Je profite également de ces lignes pour remercier ma belle-famille pour l'accueil qu'ils ont réservé à un optimiseur.

Enfin, c'est vers Lucile que je me tourne, et à qui je fais mes plus chaleureux remerciements. Ce ne sont pas des mots qui pourront suffire à la remercier pour tout ce qu'elle m'a donné et me donne toujours. Je lui dédie non seulement ce mémoire, mais surtout les nombreuses années de recherche et de découvertes qu'il nous reste ensemble.

Avant-Propos

Il est assez heureux d'apprendre que l'usage français du verbe optimiser nous est arrivé vers 1844 d'Angleterre, où *to optimize* signifiait *se comporter en optimiste*. L'acception mathématique actuelle de ce verbe ne doit pas faire oublier ce passé joyeux. L'adjectif stochastique vient quant à lui de la Grèce antique, où $\sigma\tau o\chi\alpha\sigma\tau\iota\kappa o\zeta$ qualifiait les personnes *habiles à conjecturer*. Il est donc assez drôle de constater aujourd'hui que l'optimisation stochastique caractérise étymologiquement les devins optimistes...De là à dire que le père des optimiseurs stochastiques français est Panoramix il n'y a qu'un pas, que Goscinny eût vite franchi, mais que nous nous contenterons d'esquisser. Néanmoins, il est vrai que la recherche exige persévérance et ouverture d'esprit, qui ne sont finalement qu'une forme particulière d'optimisme et de conjecture.

Que le lecteur se rassure tout de suite, le présent mémoire n'est ni un ouvrage de divination, ni un recueil d'histoires drôles.

Ce mémoire est donc rédigé comme suit : après un chapitre de préliminaires (chapitre I) dans lequel nous donnons les classifications et problématiques nécessaires pour comprendre les analyses menées plus loin, et dans lequel nous indiquons la forme générale du problème d'optimisation stochastique qui est étudié dans les autres chapitres, nous passons à l'analyse, chapitre par chapitre, des différentes caractéristiques de ce problème : au chapitre II, nous nous intéressons au problème du traitement de l'information et des influences des décisions passées sur l'information future, pour caractériser les systèmes dans lesquels ces influences sont nulles, au chapitre III, nous proposons une analyse de la stabilité du problème par rapport à l'information d'une part et à la loi de probabilité d'autre part, à travers des exemples de discrétisation, au chapitre IV, nous donnons un algorithme de résolution des problèmes dits en boucle fermée, avec structure d'information dynamique, utilisant les idées de l'estimation non paramétrique et des approximations stochastiques, et au chapitre V, nous étudions la possibilité de mener une théorie de la décomposition dans les problèmes stochastiques avant toute discrétisation de l'aléa. Enfin, le chapitre VI propose un catalogue de divers algorithmes stochastiques pour minimiser des fonctions non-linéaires de l'espérance.

Résumé

Cette thèse s'attache à l'étude des problèmes d'optimisation stochastique, en les abordant sous divers angles. Le premier thème abordé est la caractérisation des systèmes stochastiques à plusieurs pas de temps sans effet dual en boucle ouverte. Il est prouvé qu'en dimension 1, seuls les systèmes linéaires le sont. Le deuxième thème est l'étude des résultats existants de stabilité pour ces problèmes. A travers un exemple, on montre les limites des approches mettant en jeu des distances entre mesures de probabilité, et, après avoir présenté comment bâtir une topologie sur les tribus, on montre deux résultats de stabilité mettant en lumière l'importance des contraintes de non-anticipativité dans les problèmes à plusieurs pas de temps. Le troisième thème concerne les approches variationnelles pour l'optimisation stochastique fonctionnelle. On propose une nouvelle famille d'algorithmes stochastiques permettant de rechercher les commandes optimales fonctionnellement sans aucune discrétisation préalable de l'aléa, et avec une garantie asymptotique d'optimalité. Des outils théoriques sont donnés et démontrés pour permettre l'analyse de ces méthodes. Le quatrième thème s'occupe de la décomposition des grands systèmes stochastiques. La présence de contraintes informationnelles est discutée, et on montre comment la programmation dynamique stochastique peut dans une certaine mesure être associée à la décomposition. Puis, on propose dans la lignée des algorithmes stochastiques fonctionnels présentés dans la partie précédente, un principe du problème auxiliaire stochastique, ou approché, qui permet de décomposer des problèmes stochastiques de grande taille en toute généralité. Le cinquième et dernier thème est consacré à l'étude des problèmes d'optimisation sans contraintes de mesurabilité, dans lesquels le critère est une fonction non-linéaire d'une espérance. Divers algorithmes d'approximation stochastique sont proposés et démontrés dans le cas convexe, avant d'être appliqués à la résolution de problèmes non-convexes comme des problèmes sous contraintes en probabilité.

Extended Abstract

Stochastic Optimization Problems may be typically represented by :

$$\min_{u \in U^f} \mathbb{E}\left(j(u(\boldsymbol{\xi}), \boldsymbol{\xi})\right), \tag{.1}$$

where $\boldsymbol{\xi}$ is a random variable with values in a finite dimensional space Ξ, and $u : \Xi \to U$ is a function with some integrability properties, and the decision variable. U^f is whence a subset of some functional space, typically a L_p space on Ξ. Finally, j is some normal integrand with typically convexity properties with respect to its first component. Two main types of problems exist in this paradigm :

- Static Information Problems : problems where the feasible set U^f is exogeneous to the decision variable u. Typically, U^f represents some non-anticipativity constraints, bounding constraints on the decision variable, etc.
- Dynamic Information Problems : problems where U^f is implicitly defined with u. Typically, it is the case of multistage problems where the decision may affect the future available information.

In the first category enter all the problems investigated by the so-called stochastic programming. Linear or nonlinear multistage stochastic programs with non-anticipativity constraints are static information problems : the measurability constraints are given by the underlying stochastic process $\boldsymbol{\xi}$ without any further changes due to the decisions.

In the second category enter all the problems known as Markov decision Processes, or Stochastic Dynamic Programming. In such problems, the measurability of the decision is given through a state of the system which is affected by a stochastic process (typically white noise) and by past decisions : a priori, the information is dynamic, i.e. depends on the decisions.

During my PhD Thesis, I investigated issues related to those two types of problems.

Dual effect free Stochastic Systems

For dynamic information problems, the decision variable may affect both the future information, and the cost function. In such problems, there is always a tradeoff between the quantity of future information we would like to have, and the cost of a decision which would give such an information. This double effect of the decision is known in the literature as dual effect of controls (see [**9**], or [**5**]). Dynamic information problems are typically given through a pair of forward equations for $t = 1, \dots, T-1$:

$$y_{t+1} = h_t(x_{t+1}, \boldsymbol{\xi}_{t+1}), \tag{.2a}$$

$$x_{t+1} = f_t(x_t, u_t, \boldsymbol{\xi}_{t+1}), \tag{.2b}$$

where x_t is the state of the system at time step t, u_t is the decision at time step t, and y_{t+1} is the observation at time step $t+1$. $\boldsymbol{\xi}$ is a random process, and the decision u_t has to be taken measurably with respect to y_t. The question was to characterize the functions h_t and f_t such that y_t (i.e. the measurability constraint) is independent from the past decisions $(u_s)_{s \leq t-1}$. A previous work (see [**9**]) proved that a necessary (and sufficient under some other assumptions) condition for having this dual effect free property is the property of no open-loop dual effect (NOLDE). The latter property means that for all constant controls (u_t), y_t is measurably constant. I proved with M. de Lara (Ecole Nationale des Ponts et Chaussées) that all the systems (.2) satisfying this NOLDE property were linear with respect to a diffeomorphic change of variables. Since all linear systems satisfy the NOLDE property, it proves a characterization of dual effect free stochastic systems.

Stability of Multistage Stochastic Programs

After having investigated dynamic information problems, I focused on the stability of multistage stochastic programs. In the stochastic programming community, classical results describe the stability of one or two-stage stochastic programs, i.e. of stochastic programs with decision variables which are not functions subject to measurability constraints, but only finite dimensional vectors. A very interesting stability analysis shows the qualitative and quantitative properties of the optimal cost taken as a function of the probability measure corresponding to our random variable $\boldsymbol{\xi}$ (see e.g. the survey [**86**]). A first approximation said that in the case of multistage stochastic programs, such a theory can also be developped, and it was done in [**51**].

A deeper insight in these problems, provided e.g. by [**8**], enlightened the role played by measurability constraints in optimization problems. After having surveyed the main result related to distance of σ-fields and filtrations, and having extended the results of [**8**], I developped a further analysis on that topic and showed on small examples that without a *measurability term*, no stability result can be expected in the multistage context. In a joint work with H. Heitsch and W. Römisch (both at Humboldt University, Berlin, Germany), we proved a new stability result for multistage linear stochastic programs giving a stability bound depending on a sum of two terms, one term expressing a distance of probability measures, and one term turning out to be a filtration distance (see [**57**]). I proved a similar result using lagrangian duality arguments for multistage stochastic programs with separated criterion.

Functional Stochastic Gradient Algorithms

The previous stability analysis is an important prerequisite for dealing with discretized multistage stochastic programs. A disappointing feature with stochastic programming is that there is no algorithmic theory without a priori discretization of the uncertainty space, through a scenario tree (see e.g. [**91**]), a basis of functions (see e.g. [**39**]), etc.

To make profit of the uncertainty, I developped with K. Barty and J.-S. Roy (both at Electricité de France, Clamart, France) a new variational approach directly on the initial problem (.1). If you make a naive projected gradient algorithm for this problem, you obtain iterates u^k defined as

$$u^{k+1}(\cdot) = \Pi_{U^f}\left(u^k(\cdot) - \rho^k \nabla_u j(u^k(\cdot), \cdot)\right).$$

Such an algorithm is not applicable, due to the projection and the computation of the functional gradient. In many cases, our feasible set may be rewritten as the intersection of a closed convex set U_c^f (bounding constraints) and a closed vector subspace U_v^f (measurability constraints), respecting the property that the projection of the subspace on the convex subset remains on the subspace. On the basis of a general proposition on such projection, I prove that $\Pi_{U^f} = \Pi_{U_c^f} \circ \Pi_{U_v^f}$. Hence, the lonely remaining problem was to compute the functional gradient. Gathering stochastic approximation techniques and mollifying techniques, we proposed the following stochastic algorithm, and proved its convergence under some classical assumptions on the sequences ϵ^k, ρ^k, and the mollifiers K^k :

$$u^{k+1}(\cdot) = \Pi_{U_c^f}\left(u^k(\cdot) - \rho^k \nabla_u j(u^k(\boldsymbol{\xi}^{k+1}), \boldsymbol{\xi}^{k+1}) \frac{1}{\epsilon^k} \Pi_{U_v^f}\left(K^k(\boldsymbol{\xi}^{k+1}, \cdot)\right)\right). \tag{.3}$$

If the mollifiers are known with a small number of parameters, it remains to remember the i.i.d. sample $(\boldsymbol{\xi}^{k+1})$ to rebuild the solution everywhere.

I then developped a deep analysis of this stochastic procedure, and extended it to the solution of functional fixed point equations, such as Bellman equations appearing in the pricing of bermudan options in mathematical finance. Some preprints are available on this subject, see [**11**], [**10**], [**12**].

Decomposition for Multistage Stochastic Programs

The variational approach developped in the preceding chapter to solve multistage stochastic problems without any a priori discretization of the uncertainty was the first step to build

a decomposition theory for large-scale stochastic optimization problems. In the stochastic programming literature, almost all the decomposition schemes appear only after a discretization of the uncertainty through typically a scenario tree (see e.g. [**58**]). Such schemes suffer from the fact that the uncertainty may be heterogeneous between different subproblems, and whence should be discretized differently in each subproblem. Such different discretizations are possible only if the decomposition-coordination scheme is made before, i.e. on the uncertain optimization problem. On the basis of the so-called Auxiliary Problem Principle (see [**29**]), I developped a new decomposition theory suitable for stochastic problems, where the coordination tools are allowed to be approximations of the true coordination tools (due to the infinite dimensional nature of the whole problem). I then applied this Stochastic Auxiliary Problem Principle to solve large-scale decentralized optimization problems.

In another part of this chapter, I also investigated the links between Dynamic Programming and Decomposition and proved on examples the incompatibility between those two principles, enlightening the need of a variational approach to solve the decomposed subproblems.

Objective Functions which are nonlinear functions of the Expectation

Finally, on the last part of my dissertation, I analyzed the different possible algorithms (stochastic or not) able to deal with stochastic optimization problems where the objective function is a nonlinear function of the expectation instead of being the expectation itself. This last work was motivated by chance constraints optimization problems for which one may want to use augmented lagrangian techniques, turning the usual cost function into a nonlinear function of the expectation. These algorithms were tested on small examples presenting nonconvexities.

CHAPITRE I

Introduction aux problèmes d'optimisation stochastique

Nous allons donner ici le modèle typique de tous les problèmes d'optimisation stochastique que nous aborderons dans la suite du mémoire. Ce chapitre se veut donc être une clé de lecture du mémoire qui suit. Son objet est de donner une classification pertinente des problèmes d'optimisation stochastique, issue notamment de [**8**], et plus largement du travail mené depuis plusieurs années par le SOWG [1].

En prenant les notations mises au point dans [**8**], une très large classe de problème d'optimisation stochastique peuvent être écrits sous la forme :

$$\min_{\boldsymbol{u}\in U^f} J(\boldsymbol{u}) := \mathbb{E}\left(j(\boldsymbol{u},\boldsymbol{\xi})\right) \tag{I.1}$$

On se donne alors un espace de probabilité $(\Omega,\mathcal{F},\mathbb{P})$. La variable de commande sera notée $\boldsymbol{u}$, l'aléa $\boldsymbol{\xi}$, variables aléatoires sur $(\Omega,\mathcal{F},\mathbb{P})$, à valeurs dans des espaces de dimension finie. La fonction de coût sera notée j et l'ensemble admissible U^f. Nous préciserons selon les cas à quoi correspondent chacun de ces ingrédients dans les problèmes considérés. De manière très générale, la commande $\boldsymbol{u}$ est recherchée dans un sous-ensemble U^f de l'ensemble des variables aléatoires sur $(\Omega,\mathcal{F},\mathbb{P})$. Les contraintes éventuelles pesant sur $\boldsymbol{u}$ sont donc formulées grâce à l'ensemble U^f.

Selon les spécifications des composantes du problème (I.1), on se trouvera face à des difficultés de nature et de profondeur différentes. C'est pourquoi avant toute chose, il est nécessaire d'opérer une typologie des problèmes d'optimisation stochastique. On distingue principalement deux types de problèmes :

– les problèmes en *boucle ouverte*,

– et les problèmes en *boucle fermée*.

Pour donner une première idée de ce à quoi correspond cette typologie, disons que dans un problème en boucle ouverte, la variable de commande est recherchée comme une constante ne dépendant pas des réalisations de l'aléa $\boldsymbol{\xi}$, c'est à dire comme une variable aléatoire dégénérée, tandis que dans un problème en boucle fermée, la variable de commande est recherchée comme une véritable fonction de l'aléa, permettant ainsi d'agir de façon différenciée sur des aléas donnant des informations différentes.

Afin de pouvoir donner des définitions de ces problèmes, nous introduisons maintenant le modèle mathématique.

I.1. Problème général et typologie

Soit U et Ξ deux espaces de Banach séparables, munis respectivement de leurs tribus boréliennes notées $\mathcal{B}_U$ et $\mathcal{B}_\Xi$, et de leurs normes notées $\|\cdot\|_U$ et $\|\cdot\|_\Xi$. Soit $(\Omega,\mathcal{F},\mathbb{P})$ un espace de probabilité.

Soit $\boldsymbol{\xi}:\Omega\to\Xi$ une variable aléatoire, de loi μ. Soit $j:U\times\Xi\to\mathbb{R}$ une intégrande normale (cf. Définition A.23). On définit alors $\mathcal{U} := \{\boldsymbol{u}:\Omega\to U \ : \ \boldsymbol{u} \text{ mesurable}\}$, et l'application $J:\mathcal{U}\to\mathbb{R}\cup\{\infty\}$ telle que :

$$\forall \boldsymbol{u}\in\mathcal{U},\ J(\boldsymbol{u}) := \mathbb{E}\left(j(\boldsymbol{u},\boldsymbol{\xi})\right).$$

L'ensemble U^f est donc naturellement un sous-ensemble de $\mathcal{U}$. Dorénavant, nous considérerons l'ensemble U^f défini comme l'intersection de deux sous-ensembles de $\mathcal{U}$, notés U^{info} et U^{ponct},

[1] Systems Optimization Working Group, abrité par l'Ecole Nationale des Ponts et Chaussées, et composé de Laetitia Andrieu, Kengy Barty, Pierre Carpentier, Jean-Philippe Chancelier, Guy Cohen, et Michel de Lara, auxquels se sont ajoutés en 2003 Anes Dallagi et moi-même.

i.e., $U^f = U^{\text{info}} \cap U^{\text{ponct}}$. On précisera plus loin la forme de U^{info}. En revanche, U^{ponct} aura dans la suite la forme suivante :

$$U^{\text{ponct}} = \{\boldsymbol{u} \in \mathcal{U},\ :\ \boldsymbol{u}(\omega) \in \boldsymbol{\Gamma}(\omega),\ \text{pour } \mathbb{P}\text{-presque tout } \omega \in \Omega\}\,, \tag{I.2}$$

avec $\boldsymbol{\Gamma}$ une multi-application de Ω dans U.

Nous pouvons maintenant opérer une première typologie :

Définition I.1 (Boucle ouverte, Boucle fermée). *Le problème* (I.1) *est dit en boucle ouverte (open loop) si et seulement si*

$$U^{\text{info}} = \{\boldsymbol{u} \in \mathcal{U}\ :\ \boldsymbol{u} \text{ est } \sigma(\{\Omega, \emptyset\}) - \text{mesurable}\}.$$

Dans le cas contraire, le problème est dit en boucle fermée (closed loop).

Ainsi, le problème (I.1) est dit en boucle ouverte si et seulement si $U^{\text{info}} = \{\boldsymbol{u} \in \mathcal{U}\ :\ \exists v \in U,\ \boldsymbol{u} = v,\ \text{p.s.}\}$. La définition I.1 fait donc une distinction en terme de mesurabilité de la commande. Cette distinction s'inscrit tout à fait dans l'idée qu'une large classe des difficultés inhérentes aux problèmes d'optimisation stochastique trouve sa source dans la mauvaise compréhension des contraintes de mesurabilité pesant sur la commande, comme nous tenterons de l'expliquer dans la suite.
La distinction boucle ouverte - boucle fermée est de plus essentielle d'un point de vue pratique, car elle conditionne l'utilisation des algorithmes de résolution du problème considéré. En effet, comme nous le détaillerons dans la suite du mémoire, les problèmes en boucle ouverte peuvent être résolus numériquement par des méthodes proches des méthodes de l'optimisation déterministe, comme la programmation linéaire, ou les algorithmes de gradient stochastique (cf. par exemple [**72**]). En revanche, les problèmes en boucle fermée posent une importante difficulté numérique. A l'heure actuelle, principalement deux familles de méthodes sont disponibles pour les traiter : d'une part, les méthodes par chroniques arborescentes (cf. par exemple [**47**]), et d'autre part les techniques de programmation dynamique stochastique (cf. par exemple [**18, 19**]).

En spécifiant davantage U^f, nous allons maintenant faire une remarque de modélisation utile pour la suite. Soit $\boldsymbol{h} : \Omega \to Y$ une variable aléatoire de loi ν, et supposons que

$$U^{\text{info}} = \{\boldsymbol{u} \in \mathcal{U}\ :\ \boldsymbol{u} \text{ est } \sigma(\boldsymbol{h}) - \text{mesurable}\}\,,$$

et

$$U^{\text{ponct}} = \{\boldsymbol{u} \in \mathcal{U},\ :\ \boldsymbol{u}(\omega) \in K(\boldsymbol{\xi}(\omega)),\ \text{pour } \mathbb{P}\text{-presque tout } \omega \in \Omega\}\,,$$

avec K une multi-application de Ξ dans U. Posons alors $\tilde{\boldsymbol{\xi}} = (\boldsymbol{\xi}, \boldsymbol{h})$, et $\tilde{\Xi} = \Xi \times Y$. En définissant maintenant

$$\Psi^f = \left\{\psi : Y \to U\ :\ \psi \text{ est mesurable, et } \psi(y) \in K(\xi),\ \text{pour } \mu \otimes \nu - \text{presque tous } (\xi, y) \in \tilde{\Xi}\right\},$$

le problème (I.1) est équivalent au problème de feedback suivant :

$$\min_{\psi \in \Psi^f} \hat{J}(\psi) := \mathbb{E}\left(j(\psi(\boldsymbol{h}), \boldsymbol{\xi})\right). \tag{I.3}$$

L'équivalence est une conséquence directe de l'égalité $U^f = \{\psi(\boldsymbol{h})\ :\ \psi \in \Psi^f\}$. Enfin, en définissant

$$\Phi^f = \left\{\phi : \tilde{\Xi} \to U\ :\ \phi(\xi, \cdot) \in \Psi^f,\ \text{pour } \mu\text{-presque tout } \xi \in \Xi\right\},$$

on obtient finalement le problème en feedback équivalent aux problèmes (I.1) et (I.3) :

$$\min_{\phi \in \Phi^f} \tilde{J}(\phi) := \mathbb{E}\left(\tilde{j}(\phi(\tilde{\boldsymbol{\xi}}), \tilde{\boldsymbol{\xi}})\right), \tag{I.4}$$

avec pour tout $u \in U$, et tout $\tilde{\xi} = (\xi, y) \in \tilde{\Xi}$, $\tilde{j}(u, \tilde{\xi}) = j(u, \xi)$. Plus tard, il arrivera dans le reste du mémoire que l'on écrive u (resp. U^f), en lieu et place de ϕ (resp. Φ^f), comme par exemple dans le chapitre IV. La notation dans le problème (I.4) a l'avantage de mettre en évidence dans le cas de la boucle ouverte et de la boucle fermée l'indépendance ou la dépendance du contrôle par rapport à l'aléa.

REMARQUE I.2 (Choix de modélisation). *Dans la thèse* [**8**]*, c'est souvent les notations du problème* (I.1) *qui sont choisies, mais leur coexistence avec des notations du type* (I.4) *rend parfois difficile la compréhension et la représentation des espaces dans lesquels on travaille. Bien entendu, ces deux représentations sont équivalentes dans le cadre décrit ci-dessus, et présentent selon les points que l'on souhaite mettre en évidence plus ou moins d'intérêt.*

Nous allons maintenant entrer plus avant dans la problèmatique des contraintes d'information.

I.2. Problèmes à plusieurs niveaux et information

On parle souvent pour les problèmes stochastiques de contraintes d'information : en un mot, U^f est souvent inclus dans un ensemble défini à l'aide de contraintes de mesurabilité : il est demandé aux commandes $\boldsymbol{u} \in U^f$ d'être mesurables (éventuellement composante par composante) par rapport à la tribu (ou un ensemble de tribus correspondant aux composantes de la commande) engendrée par une application donnée.

En restant dans le cadre général donné par (I.1), considérons les problèmes stochastiques dits *multistage*, c'est à dire avec plusieurs niveaux de décision. Ces niveaux de décision peuvent correspondre à un écoulement du temps. Ils peuvent aussi correspondre à une répartition de décisions simultanées entre plusieurs acteurs. Dès qu'une telle structure se présente, il convient de définir clairement l'information disponible à chaque niveau et les effets des décisions sur ces informations.

Pour ce faire, supposons qu'il y ait $n \in \mathbb{N}$ niveaux de décisions, et notons $\boldsymbol{u}_i : \Omega \to U_i$ la décision de niveau i, variable aléatoire à valeurs dans un espace de commande noté U_i, pour tout $i \in \{1, \dots, n\}$, et $U := \Pi_{j=1}^n U_j$. De façon analogue, on notera $\mathcal{U}_i$ l'ensemble des variables aléatoires mesurables de Ω dans U_i, et $\mathcal{U} := \Pi_{j=1}^n \mathcal{U}_j$. Définissons également pour tout $i \in \{1, \dots, n\}$ les applications $h_i : U \times \Xi \to Y_i$, avec Y_i un espace de Hilbert. On appellera Y_i l'espace d'information associé au $i$ème niveau de décision. Les applications (h_i) représentent les quantités d'information disponibles pour les commandes $(\boldsymbol{u}_i)$. On notera $Y := \Pi_{j=1}^n Y_j$.
Dès lors, une façon de demander que chacune des étapes de décision respecte sa contrainte d'information est d'écrire :

$$\forall i \in \{1, \dots, n\},\ \boldsymbol{u}_i \text{ est mesurable par rapport à } h_i(\boldsymbol{u}, \boldsymbol{\xi}). \tag{I.5}$$

De façon plus compacte, on peut définir $h : U \times \Xi \to Y$ par :

$$\forall u \in U,\ \xi \in \Xi,\ h(u, \xi) := (h_i(u, \xi))_{1 \le i \le n}\,.$$

On peut remarquer qu'alors, (I.5) implique la proposition suivante (mais ne lui est pas équivalente) :

$$\boldsymbol{u} \text{ est } \sigma(h(\boldsymbol{u}, \boldsymbol{\xi})) - \text{mesurable}. \tag{I.6}$$

On définit maintenant U^{info}, l'ensemble des commandes *informationnellement admissibles* :

DÉFINITION I.3. *Posons* $U^{\text{info}} := \{\boldsymbol{u} \in \mathcal{U} \ :\ \forall i = 1, \dots, n,\ \boldsymbol{u}_i \text{ est } \sigma(h_i(\boldsymbol{u}, \boldsymbol{\xi})) - \text{mesurable}\}$. *$\boldsymbol{u} \in \mathcal{U}$ est dite informationnellement admissible si et seulement si $\boldsymbol{u} \in U^{\text{info}}$.*

Avec la définition I.3, nous sommes maintenant en mesure de donner une autre typologie des problèmes d'optimisation stochastique à plusieurs niveaux. C'est la typologie d'*information statique* ou *dynamique*. Considérons le problème :

$$\min_{\boldsymbol{u} \in U^f = U^{\text{info}} \cap U^{\text{ponct}}} J(\boldsymbol{u}) \tag{I.7}$$

DÉFINITION I.4 (Information statique/dynamique). *Le problème* (I.7) *est dit à structure d'information statique si et seulement s'il existe une application $\overline{h} : \Xi \to Y$ telle que*

$$U^{\text{info}} = \left\{\boldsymbol{u} \in \mathcal{U} \ :\ \forall i = 1, \dots, n,\ \boldsymbol{u}_i \text{ est } \sigma(\overline{h}_i(\boldsymbol{\xi})) - \text{mesurable}\right\}.$$

Dans le cas contraire, le problème (I.7) *est dit à structure d'information dynamique.*

Cela revient donc à dire que les applications h_i sont indépendantes des commandes $\boldsymbol{u}$, ou, en d'autres termes, que la structure d'information du problème n'est pas affectée par les décisions prises aux divers niveaux. On pourrait imaginer des situations plus compliquées, dans lesquelles une partie seulement des niveaux de décisions influencerait l'information.

Une autre terminologie utilisée dans le cadre des structures d'information dynamique est l'*effet dual*, ou *double effet* des commandes. L'idée de ce terme est de désigner l'action à deux niveaux qu'ont les commandes lorsque l'information est dynamique : les commandes peuvent à la fois affecter la qualité de l'information disponible, et la fonction de coût du problème. Cette question sera abordée plus en détails au chapitre II.

Dans la suite du mémoire, nous nous intéresserons à la caractérisation des systèmes ayant une structure d'information statique. Pour les besoins des preuves, et des raisons historiques (datant notamment des travaux [**25**]), nous introduirons un ordre sur les fonctions mesurables, dont les parentés avec la mesurabilité classique seront assez claires, mais de manipulation plus simple.

Pour la résolution numérique des problèmes d'optimisation stochastique, cette distinction est extrêmement importante. En effet, on utilise souvent des méthodes postulant a priori une certaine forme de la commande, et donc une certaine dépendance fonctionnelle de celle-ci par rapport à l'aléa... et à elle-même. Or, si l'ensemble admissible est défini par une équation de mesurabilité du type (I.6), ces méthodes sont inapplicables. Il est donc tout à fait nécessaire quand on entreprend la résolution numérique d'un problème, de savoir s'il est ou non à structure d'information statique.

I.3. Application aux problèmes à plusieurs pas de temps

Afin de rendre plus concrets les discours sur la boucle fermée d'une part, et les décisions à plusieurs niveaux d'autre part, appliquons ces concepts à un problème stochastique à plusieurs pas de temps, communément appelé problème *multistage*.

Considérons le problème suivant :

$$\min_{\boldsymbol{u}} \mathbb{E}\left(\sum_{t=1}^{T} L_t(\boldsymbol{u}_1,\dots,\boldsymbol{u}_t,\boldsymbol{\xi}_1,\dots,\boldsymbol{\xi}_t)\right) \quad \text{s.c. } \forall 1\le t\le T,\ \boldsymbol{u}_t \text{ est } \sigma(h_t(\boldsymbol{u},\boldsymbol{\xi}))-\text{mesurable}, \tag{I.8}$$

avec pour tout $t \in \{1,\dots,T\}$, et pour $\mathbb{P}$-presque tout ω, $\boldsymbol{u}_t(\omega) \in U_t$, et $\boldsymbol{u}_t \in \mathcal{U}_t$, $\boldsymbol{\xi}_t \in \Xi$. On a de plus les fonctions de coût $L_t : \left(\Pi_{s=1}^t U_s\right) \times (\Xi)^t \to \mathbb{R}$, et les fonctions d'observation $h_t : \left(\Pi_{s=1}^T U_s\right) \times (\Xi)^T \to Y_t$.

Classiquement, si la commande $\boldsymbol{u}_t$ représente la t^{ime} décision dans l'ordre chronologique, il est raisonnable de lui demander de ne dépendre que du passé du système. Distinguons trois cas généraux :

- $\boldsymbol{u}_t$ ne dépend de rien du tout :
  On aura par exemple $h_t(u,\xi) = 0$ pour tous u, ξ, ce qui imposera à $\boldsymbol{u}_t$ d'être une constante : c'est le cas de commandes en boucle ouverte.
- $\boldsymbol{u}_t$ dépend de tout le passé :
  On aura par exemple $h_t(u,\xi) := (u_1,\dots,u_{t-1},\xi_1,\dots,\xi_t)$. Ainsi, récursivement, on aura :

  $$\boldsymbol{u}_1\ \sigma(\boldsymbol{\xi}_1)-\text{mesurable},\dots\boldsymbol{u}_t\ \sigma(\boldsymbol{\xi}^t)-\text{mesurable},\dots$$

  et on sera bien dans un cas d'information statique en boucle fermée. Communément, on désignera de plus cette structure d'information comme une information en mémoire parfaite.
- $\boldsymbol{u}_t$ oublie une partie du passé :
  C'est le cas par exemple avec $h_t(u,\xi) = (u_{t-1},\xi_{t-1})$. Dans ce cas, le système a oublié les aléas précédents, et ne voit plus que sa commande et le bruit à l'instant précédent. Dans

ce cas, les applications h_t ne peuvent être décrites indépendemment des commandes : c'est la situation de commandes en boucle fermée, avec structure d'information dynamique.

On s'aperçoit donc aisément sur cet exemple que l'introduction des concepts de boucle ouverte et boucle fermée, puis d'information statique et dynamique, sont d'une grande importance. Dans la suite du mémoire, nous aborderons surtout les cas en information statique.

I.4. Enjeux d'une telle classification et plan du mémoire

La classification entreprise dans les pages précédentes entre commandes en boucle ouverte, commandes en boucle fermée, information statique et information dynamique, est bien plus profonde qu'elle peut le paraître de prime abord. Historiquement, les problèmes les mieux étudiés sont les problèmes en boucle ouverte, et avec structure d'information statique. On peut les représenter génériquement par :

$$\min_{u \in U^f} \mathbb{E}\left(f(u, \boldsymbol{\xi})\right), \tag{I.9}$$

avec U^f un convexe fermé d'un espace de Hilbert U, souvent de dimension finie, $\boldsymbol{\xi}$ une variable aléatoire définie comme avant, à valeurs dans Ξ, et $f : U \times \Xi \to \mathbb{R}$ une application typiquement convexe en u et mesurable en ξ.

On dispose pour ce type de problèmes de nombreux résultats de discrétisation, de stabilité de la solution par rapport à des perturbations de l'ensemble admissible U^f, permettant des résolutions numériques efficaces et bien contrôlées. A titre d'exemple, nous pouvons citer les travaux [**90**], donnant des résultats de convergence du type théorème central limite ou principe de grandes déviations pour des schémas de Monte-Carlo, ou les résultats de stabilités développés par [**86, 48, 79**]. En remontant davantage dans le temps, les problématiques abordées par exemple dans [**81, 82**] autour des problèmes dits *two-stage* entrent dans le formalisme de (I.9), en définissant la fonction f comme étant la fonction de recours.

Plus récemment, les préoccupations se sont tournées davantage vers les problèmes *multistage* en information statique complète et boucle fermée, pour reprendre notre terminologie. Assez systématiquement, comme cela sera exposé dans la suite de ce mémoire, en particuliers aux chapitres III et V, les approches proposées sont de nature arborescente ou par scénarios, comme en témoignent les travaux et recueils [**26, 91, 58**] qui offrent un large panel des possibilités offertes pour la résolution de ce type de problèmes discrétisés de la sorte. Cependant, en discrétisant ainsi aussitôt ces problèmes à plusieurs niveaux de décision, on s'aliène du même coup nombre de possibilités en appauvrissant d'emblée la représentation de l'aléa qui fait a priori la richesse et la chance des problèmes stochastiques. D'autre part, en appliquant un peu trop vite aux problèmes *multistage* les résultats obtenus pour les problèmes en boucle ouverte, on passe à côté de difficultés essentielles, et l'on peut même se tromper.

C'est précisément pour cette raison que la classification qui précède est importante. Trop souvent, le passage de la boucle ouverte vers la boucle fermée est sous-estimé, et on peut avoir tendance à penser que les résultats puissants valables en boucle ouverte vont se transmettre à la boucle fermée sans coup férir. Au contraire, le passage de la boucle ouverte vers la boucle fermée n'est pas simplement analogue au passage de la dimension finie vers la dimension infinie. Principalement pour une raison fort simple à comprendre, mais malaisée à circonvenir : la présence des contraintes d'information évoquées auparavant. D'autres raisons rendent ce passage difficile, mais elles regardent plus les méthodes numériques utilisées lors de la résolution.

Ce mémoire est donc construit autour des raisons qui rendent importantes la classification informationnelle définie dans les pages qui précédent :

(1) Dans le chapitre II, nous nous intéressons, dans la lignée de [**9**], au problème de caractériser (en dimension 1) les systèmes stochastiques sans effet dual. Ce chapitre est le seul du mémoire à traiter d'information dynamique, afin, en quelque sorte, de tordre le cou pour la suite à cette question.

(2) Dans le chapitre III, nous illustrons à travers un exemple simple puis des résultats de stabilité d'un nouveau type, l'importance dans les problèmes *multistage* des contraintes de mesurabilité. Ce chapitre fait ensuite le lien avec la thèse [**8**] qui propose des schémas pratiques de résolution.

(3) Le chapitre IV, le plus volumineux du mémoire, s'attache à proposer une nouvelle méthode d'essence variationnelle pour résoudre sans discrétisation de l'aléa les problèmes stochastiques en boucle fermée. Il prend donc totalement le contrepied des approches arborescentes discutées dans le chapitre précédent pour se concentrer sur une nouvelle manière d'envisager la résolution pratique des mêmes problèmes. L'efficacité des algorithmes proposés (et dont des preuves de convergence sont fournies), et leur généralité sont ensuite discutées à travers des exemples d'applications.

(4) La problématique abordée dans le chapitre V concerne la décomposition des problèmes d'optimisation stochastique en boucle fermée de grande taille. Ce chapitre montre les particularités de la boucle fermée pour la décomposition, et détaille les liens de la programmation dynamique stochastique avec la décomposition. Les limites mises en valeur ouvrent ensuite la voie à un principe du problème auxiliaire stochastique (héritier de [**29, 32**]) dont une démonstration est donnée, ainsi qu'une application aux problèmes en information décentralisée. L'intérêt de ce principe du problème auxiliaire est de donner un cadre théorique général pour décomposer des problèmes en boucle fermée. Ce principe fait par ailleurs écho au chapitre précédent dont il est d'un certain point de vue une généralisation.

(5) Le chapitre VI est quant à lui un peu à part, puisqu'il s'attache aux problèmes en boucle ouverte dont le critère est une fonction non-linéaire de l'espérance. Après avoir proposé diverses approches stochastiques pour résoudre cette famille de problèmes, ce chapitre applique ces méthodes à la résolution de problèmes d'optimisation stochastique non-convexes comme certains problèmes sous contraintes en probabilité, résolus en mêlant une approche stochastique à l'usage de lagrangiens augmentés.

S'il est une chose que l'auteur de ce mémoire souhaite, c'est de rendre à travers la diversité des problématiques abordées, le caractère foisonnant et passionnant de l'optimisation stochastique, ses enjeux, et les moyens donnés aux optimiseurs pour l'aborder.

CHAPITRE II

Effet dual

REMARQUE II.1. *Ce chapitre est à peu de choses près conforme à l'article intitulé* A Theorem on dual effect free stochastic scalar state space systems*, coécrit avec Michel de Lara (École Nationale des Ponts et Chaussées), soumis à Annals of Operations Research en décembre 2004, et actuellement en révision. Il représente la poursuite des recherches engagées depuis plusieurs années par le groupe SOWG.*

II.1. Résumé

Beaucoup de problèmes d'optimisation stochastique se posent comme des problèmes de contrôle d'un système stochastique entrée-sortie en temps discret sur un horizon $\{0, \dots, T\}$, sous un critère de minimisation. Les sorties d'un tel système sont les observations, et les entrées les contrôles. Un exemple typique de système stochastique entrée-sortie contrôlé est donné par le problème (I.8).

Typiquement, les entrées à l'instant t ne peuvent dépendre que des observations passées (i.e. pour $s \leq t$), tandis que l'observation à l'instant t dépend des décisions (et donc des entrées) prises auparavant. De telles contraintes sur les entrées sont appelées *contraintes de non-anticipativité*.

Dans un tel contexte, les contrôles peuvent donc avoir une double influence sur le système : à la fois sur le critère à minimiser et sur les observations. Ce double effet est connu dans la littérature sous le nom de double effet, ou *dual effect* en anglais (voir par exemple le travail de [**95**]). Pour reprendre la typologie donnée dans le chapitre I, de tels problèmes sont en toute généralité des problèmes en boucle fermée avec structure d'information dynamique.

D'un point de vue numérique, dès lors que les contrôles ont un double effet, une discrétisation a priori de la structure d'information du problème n'a plus de sens, puisque cette discrétisation devrait au contraire être faite a posteriori, selon les influences des contrôles sur les observations. Il est donc d'une grande importance de savoir si un contrôle va produire ou non un double effet, i.e. s'il va ou non modifier les observations futures.

La première difficulté dans cette classification a été récemment franchie par SOWG (voir [**8**] et l'article [**9**]), qui a montré que l'ensemble des contrôles ne provoquant pas de double effet, c'est à dire ne modifiant pas la structure d'information, peut être décrit explicitement, sous les hypothèses que :

- les contrôles en boucle ouverte (c'est à dire constants par rapport aux observations) sont eux-mêmes libres de tout double effet, et
- le système conserve une mémoire parfaite des observations passées, et
- les observations sont causales.

Les deux dernières hypothèses étant assez faciles à vérifier sur un système, et même assez naturelles, un dernier obstacle subsistait néanmoins dans cette caractérisation , relatif à la première hypothèse.

En effet, la propriété pour les contrôles constants de ne pas modifier la structure d'information est une propriété structurelle du système entrée-sortie considéré, en d'autres termes, c'est une propriété qui dépend directement de la fonction d'observation du système, et de la fonction de transition entrée vers observation (i.e. de la dynamique du système). En appelant NOLDE (no open loop dual effect) les systèmes vérifiant la première hypothèse, la dernière difficulté est donc de caractériser les systèmes NOLDE.

Ce chapitre, après quelques rappels techniques de définitions et de propriétés des systèmes entrée-sortie et de la structure de mesurabilité choisie, donne la démonstration de deux résultats principaux : les systèmes à dynamique et observation linéaires sont NOLDE, et la propriété d'être

NOLDE est invariante par changement de variables. Puis, sous des hypothèses supplémentaires d'injectivité, une réciproque partielle est donnée : on montre en dimension 1 que tout système NOLDE peut être rendu linéaire par changement de variables.

Le résultat général en dimension supérieure nécessiterait sans doute encore beaucoup de travail, et ce chapitre prétend seulement apporter une première pierre à l'édifice. Moralement, on peut cependant estimer que s'attaquer à un système stochastique entrée-sortie fortement non-linéaire est à peu près désespéré !

II.2. Introduction

Discrete-time stochastic input-output systems classically appear in stochastic programming or stochastic control, through for example respectively multistage stochastic programs and Markov decision processes.
Such systems are characterized by an instantaneous cost or reward function, stochastic dynamics and observation functions, and a time horizon $T \in \mathbb{N}$. At a time step t, given an observation y_t and a current state x_t, a control decision u_t has to be made, such that it minimizes the expected future cost. This decision u_t generates a future state x_{t+1} and a future observation y_{t+1}. The decision has therefore a double influence : on the one hand, it affects the instantaneous cost, and on the other hand, it changes the future observations. There is hence a tradeoff between the current cost function and the future available information. This double effect of controls is referred in literature (see [**95**]) as *dual effect* of controls. Excepted a few very particular cases like Markov decision processes, dual effect renders discrete-time stochastic systems practically unsolvable. Extensive explanations about this phenomenon are given in [**3, 4, 5**], where it appears that dual effect unables all current numerical techniques.

It is therefore very important to characterize dual effect free stochastic systems or at least dual effect free controls. The recent works [**9**] characterizes the set of dual effect free controls for a given discrete-time stochastic input-output system. Their characterization only holds under the assumption that open-loop (i.e. constant) controls are dual effect free. This assumption will be recorded in the following as No Open-Loop Dual Effect property, or simply NOLDE property, and systems satisfying this assumption will be called NOLDE systems.

We provide in this paper a characterization of one-dimensional NOLDE systems. Under injectivity and differentiability assumptions, we prove that a discrete-time stochastic input-output system is NOLDE if and only if it is linear up to a change of variables.

The paper is composed as follows. After recalling some important facts and definitions related to dual effect and discrete-time stochastic input-output systems (section II.3), we state our main characterization result in Theorem II.17. After this statement, we illustrate on an example the interest of our result. In section II.5, we provide the proof of our theorem.

II.3. Recalls on dual effect and stochastic systems

There are different ways to study discrete-time stochastic input-output systems. The main frameworks are the set-theoretical framework, the measurability framework, the topological framework, and the linear framework. Our study will be made in the set-theoretical framework, which can be considered as the most general one. In the following subsections, we will propose set-theoretical definitions for such systems, especially for the aspects related to information constraints, and feedbacks.

II.3.1. Discrete-time stochastic input-output systems. Discrete-time stochastic systems are characterized by controls, states, observations and noises. The controls will be denoted by u, the states by x and the observations by y. The noises will be denoted by ω, and the corresponding capital letters will denote the sets in which these variables take values.

Let us now be more formal. Let $T \in \mathbb{N}$ be the time horizon, and $A = \{0, 1, \dots, T-1\}$ the time description. For all $t \in A$, define the control set U_t, the observation set Y_{t+1} and the state set X_{t+1}, as abstract sets. Let $U = \times_{t\in A} U_t$, $Y = \times_{t\in A} Y_{t+1}$ and $X = \times_{t\in A} X_{t+1}$. Let also V_{t+1} and W_{t+1} denote noise states for all $t \in A$ and analogously $V = \times_{t\in A} V_{t+1}$, $W = \times_{t\in A} W_{t+1}$, and $\Omega = V \times W$, the noise sets of the system.

Let for all $t \in A$, $H_{t+1} : X_{t+1} \times W_{t+1} \to Y_{t+1}$ denote the observation function, and $F_{t+1} : X_t \times U_t \times V_{t+1} \to X_{t+1}$ denote the dynamics function. We then define our discrete-time stochastic input-output system as :

$$\text{(II.1)} \qquad \left\{ \begin{array}{rcl} x_{t+1} & = & F_{t+1}(x_t, u_t, v_{t+1}), \text{ for all } t \in \{0, \dots, T-1\}, \\ y_{t+1} & = & H_{t+1}(x_{t+1}, w_{t+1}), \text{ for all } t \in \{0, \dots, T-1\}. \end{array} \right.$$

The first state of the system (II.1) is given, and denoted by x_0. For the simplicity of notations, we define $y_0 = H_0(x_0)$. Scheme 1 sums up what a stochastic input-output system is.

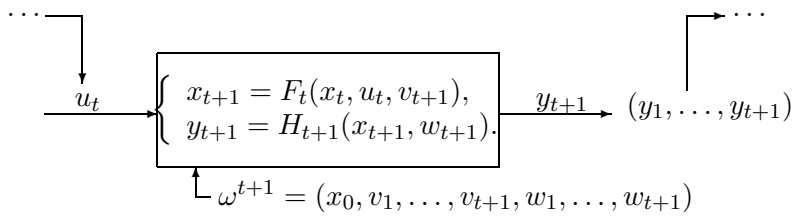


FIG. 1. Discrete-time stochastic input-output system

Each observation variable y_t, for $t \in \{0, \dots, T-1\}$, is a function of past noises $x_0, v_1, \dots, v_t, w_t$ and controls $u_0, \dots, u_{t-1}$, through the sequence of states. We define therefore observation functions of system (II.1) directly on the basis of past controls and noises, and without the states.

– We define instantaneous observation functions for all $t \in \{1, \dots, T\}$ by $h_t^+ : U \times \Omega \to Y_t$ with :

$$h_t^+(u, \omega) := H_t(F_t(F_{t-1}(\dots(F_1(x_0, u_0, v_1), u_1, v_2), \dots)u_{t-1}, v_t), w_t), \ \forall u \in U, \ \omega \in \Omega,$$

and the first observation by :

$$h_0^+(u, \omega) := x_0, \ \forall u \in U, \ \omega \in \Omega.$$

– We define global observation functions for all $t \in \{0, \dots, T\}$ by :

$$\begin{array}{rcl} h_t : U \times \Omega & \longrightarrow & \times_{s \le t} Y_s \\ (u, \omega) & \longmapsto & (h_0^+(u, \omega), \dots, h_t^+(u, \omega)). \end{array}$$

The set of all feedbacks for the system (II.1) is then given by $\Gamma := \{u : \Omega \to U\}$. Finally, we provide the definition of scalar system :

DÉFINITION II.2 (Scalar systems). *System (II.1) will be said to be scalar if for all* $t \in A$, $U_t = X_{t+1} = Y_{t+1} = V_{t+1} = W_{t+1} = \mathbb{R}$, *i.e., if state, control, observation and noise ate each time step are real-valued.*

As we recorded in the introduction, the decision u_t have to be made with respect to the past observations $(y_s)_{s \le t}$. We will give in the following subsection a definition of this *measurability* with respect to observations.

II.3.2. How to measure information ? The notions we define in this section are only based on sets, and are hence algebraic. The first issue is to characterize the information carried by a mapping.

DÉFINITION II.3 (Information order on mappings). *Let* $f_i : \Omega \to \mathcal{F}_i, i = 1, 2$ *be two functions on some abstract spaces. We will say that* f_1 *is less informative than* f_2, *and will write* $f_1 \preceq f_2$ *if :*

$$\forall \omega, \omega' \in \Omega, \ f_2(\omega) = f_2(\omega') \Rightarrow f_1(\omega) = f_1(\omega').$$

One can alternatively propose the following definition for this order on functions sharing the same domain :

PROPOSITION II.4. *Let $f_i : \Omega \to \mathcal{F}_i, i = 1, 2$ be two functions on some abstract spaces. The two following statements are equivalent :*

(1) $f_1 \preceq f_2$,

(2) $\exists\, g : Im(f_2) \to Im(f_1)$ s.t. $f_1 = g \circ f_2$.

Preuve : cf. [**9**]. □

DÉFINITION II.5. *Let $f_i : \Omega \to \mathcal{F}_i, i = 1, 2$ be two functions on some abstract spaces. We will say that f_1 and f_2 are equivalent and write $f_1 \equiv f_2$ if $f_1 \preceq f_2$ and $f_2 \preceq f_1$.*

A straightforward application of the preceding definitions and proposition is the following :

PROPOSITION II.6. *Let $f_i : \Omega \to \mathcal{F}_i, i = 1, 2$ be two functions on some abstract spaces. The three following statements are equivalent :*

(1) $f_1 \equiv f_2$,

(2) there exists a bijection $g : Im(f_2) \to Im(f_1)$ s.t. $f_1 = g \circ f_2$.

(3) there exists a one-to-one mapping $g : Im(f_2) \to \mathcal{F}_1$ s.t. $f_1 = g \circ f_2$.

Preuve : cf. [**9**]. □

REMARQUE II.7 (Measurability framework). *Proposition II.4 underlines the links between classical measurability, and our order relation on mappings. Indeed, we can see that under some additional assumption on our domains (they should be measurable spaces), we could say that if f_1 is measurable with respect to f_2, then $f_1 \preceq f_2$. It is but not exactly the same notion, and we will avoid to use abusively measurability theory words in our context.The links between these concepts are especially adressed in* [**9**].

II.3.3. Dual effect, feedback sets, and change of variables. We are now able to define the information requirements on the controls for the system (II.1).

DÉFINITION II.8 (Information amount). *For all control variable $u \in \Gamma$, let us define :*

$$\forall t \in \{0, \dots, T\},\ \eta_t^u := h_t(u(.), .) : \Omega \to \times_{s \leq t} Y_s.$$

Let u, u' be two such control variables. We will say that :

– u amounts to more information than u' if $\forall t \in \{0, \dots, T\}$, $\eta_t^{u'} \preceq \eta_t^u$;

– u amounts to the same information as u' if $\forall t \in \{0, \dots, T\}$, $\eta_t^{u'} \equiv \eta_t^u$.

DÉFINITION II.9 (Feasible set). *We can hence define the set of feasible feedbacks, denoted by $\mathbb{F}^{ad}$ as :*

$$\mathbb{F}^{ad} := \{u \in \Gamma \ :\ u_t \preceq \eta_t^u(.),\ \forall t \in \{0, \dots, T-1\}\}.$$

The feasible feedbacks are therefore subject to information constraints. The information structure described by the mappings (η_t^u) is causal (this concept is introduced in [**9**]), in the sense that $\eta_t^u \preceq \eta_{t+1}^u$ for all $u \in \mathbb{F}^{ad}$, i.e. the information grows with the time steps. Moreover, the feasible feedbacks may be taken with respect to all past observations, as reflected by the global observation mappings. This property is referred to in [**9**] as perfect memory of the system.

DÉFINITION II.10 (Open-loop controls). *The set of open-loop feedbacks (constant mappings) $\perp$ is defined by :*

$$\perp := \left\{u \in \Gamma \ :\ u(\omega) = u(\omega'),\ \forall \omega, \omega' \in \Omega\right\}.$$

The dual effect of control variables on the system is defined as follows :

DÉFINITION II.11 (NOLDE). *We will say that the system (II.1) is NOLDE (no open-loop dual effect) if any open-loop control amounts to the same information. In that case, some spaces Z_t exist, as well as some mappings $\zeta_t : \Omega \to Z_t$, for all $t \in \{0, \dots, T-1\}$, such that :*

$$\forall u \in \perp,\ \eta_t^u \equiv \zeta_t.$$

REMARQUE II.12. *The definition of NOLDE systems introduces mappings* $(\zeta_t)_{t-1\in A}$ *which can be taken to be surjective. According to proposition II.6, one can thus say that the system* (II.1) *is NOLDE if and only if :*

$$\forall t \in \{0,\dots,T\},\ \exists \tilde{h}_t : U \times Z_t \to \times_{s\le t} Y_s, \quad s.t. \quad \forall u \in U,\ \forall \omega \in \Omega,\ h_t(u,\omega) = \tilde{h}_t(u,\zeta_t(\omega)),\ \textit{and} \tag{II.2}$$

$$\forall u \in U,\ \tilde{h}_t(u,.)\ \textit{is one-to-one.} \tag{II.3}$$

We are interested from the beginning on dual effect free controls. It is very natural to assume that the system is NOLDE, if one want to characterize all dual effect free feedbacks, since the open-loop controls are the simpliest ones. We go therefore to the following straightforward definition of dual effect free feedbacks :

DÉFINITION II.13 (Dual effect free feedbacks). *If the system* (II.1) *is NOLDE, the set of dual effect free feedbacks* $\mathbb{F}^{nde}$ *is defined by :*

$$\mathbb{F}^{nde} := \left\{ u \in \mathbb{F}^{ad} \ :\ \eta_t^u(.) \equiv \zeta_t(.),\ \forall t \in \{0,\dots,T-1\} \right\}.$$

The latter definition of dual effect free controls is not efficient at all, since it involves the mappings $(\eta_t^u)_{0\le t\le T-1}$ which depends on all the past observations and feedbacks. Here comes the main result of [**9**], namely the fact that $\mathbb{F}^{nde}$ may be characterized directly on the control variables, without any reference to the observation functions, beside the NOLDE property of the system. The main statement of [**9**] is therefore :

THÉORÈME II.14. *Assume that* (II.1) *is NOLDE. Then,*

$$\mathbb{F}^{nde} = \{ u \in \Gamma \ :\ u_t \preceq \zeta_t,\ \forall t \in \{0,\dots,T-1\} \}.$$

Preuve : cf. [**9**]. □

The important claim of Theorem II.14 assumes that the discrete-time stochastic system is NOLDE. To complete the study of dual effect free controls, it is thus necessary to characterize NOLDE systems. In this set-theoretic framework, it is clear that bijective change of variables cannot affect the properties of the system. The following definition and proposition formalize this claim.

DÉFINITION II.15 (Change of variables). *An independent change of variables in the system* (II.1) *consists in bijective mappings* $\beta_t, \rho_t, \alpha_t, \gamma_t, \theta_t$ *on respectively* Y,U,X,V *and* W, *for all* $t \in \{1,\dots,T\}$, *defining new variables as :*

$$\begin{cases} \hat{y}_t := \beta_t(y_t), & \textit{(observation)},\\ \hat{u}_t := \rho_t(u_t), & \textit{(control)},\\ \hat{x}_t := \alpha_t(x_t), & \textit{(state)},\\ \hat{v}_t := \gamma_t(v_t), & \textit{(dynamics noise)},\\ \hat{w}_t := \theta_t(w_t), & \textit{(observation noise)}. \end{cases}$$

PROPOSITION II.16. *If the system* (II.1) *is NOLDE, then it remains NOLDE under any independent change of variables.*

Preuve : Assume that (II.1) is NOLDE. We make a change of variables at each time step, i.e. let $\alpha_{U,t} : U_t \to U_t$, $\alpha_{Y,t} : \times_{s\le t} Y_s \to \times_{s\le t} Y_s$, $\alpha_{\Omega,t} : \Omega_t \to \Omega_t$, $\alpha_{X,t} : X_t \to X_t$ be bijections. We make now the corresponding change of variables and wonder whether the new system is still NOLDE.
We note $\alpha_U : U \to U$ and $\alpha_\Omega : \Omega \to \Omega$ the bijections defined by :

$$\begin{aligned} \forall u \in U,\ \alpha_U(u) &= (\alpha_{U,t}(u_t))_t,\\ \forall \omega \in \Omega,\ \alpha_\Omega(\omega) &= (\alpha_{\Omega,t}(\omega_t))_t. \end{aligned}$$

Since (II.1) is NOLDE, we have with usual notations :

$$\forall t, \forall u \in U,\ h_t(u,.) \equiv \zeta_t(.) \text{ on } \Omega.$$

Using Lemma II.21, and the fact that α_U is a bijection, it is then clear that :

$$\forall t, \forall u \in U,\ \hat{h}_t(\hat{u},\hat{.}) = \alpha_{Y,t} \circ h_t(\alpha_U(u), \alpha_\Omega(.)) \equiv \zeta_t(.) \text{ on } \Omega.$$

We now have just to consider the change of variables for the state variable x. By definition of h_t, it is clear that it does not affect the NOLDE property of the system, since each change of variables is independent of the other ones. Indeed, to make a change of variables on x is equivalent to deal with new dynamics having the same properties as the previous ones. □

As a consequence of Proposition II.16, any characterization result of the NOLDE property of stochastic systems can only be stated up to a change of variable. Next section provides precisely such a characterization result.

II.4. Characterization of dual effect free scalar systems

We give here the main result of this paper. It characterizes NOLDE systems directly on the dynamics and observation functions.

THÉORÈME II.17. *Let us consider system* (II.1). *Assume that it is scalar and stationary, and that* $T \geq 2$. *Assume also that :*

$$\forall x \in \mathbb{R}, \quad H(x,.) \text{ is one-to-one,} \tag{II.4a}$$

$$\forall w \in \mathbb{R}, \quad H(.,w) \text{ is one-to-one,} \tag{II.4b}$$

$$\forall (x,u) \in \mathbb{R}^2, \quad F(x,u,.) \text{ is one-to-one,} \tag{II.4c}$$

$$\forall (x,u,v) \in \mathbb{R}^3, \quad \partial F/\partial u(x,u,v) \neq 0, \tag{II.4d}$$

$$\forall (u,v) \in \mathbb{R}^2, \quad F(.,u,v) \text{ is one-to-one.} \tag{II.4e}$$

Then, (II.1) *is NOLDE if and only if there is a change of variables making* F *linear in its three variables and* H *linear in its two variables.*

It follows therefore from Theorems II.17 and II.14 :

COROLLAIRE II.18. *Under the assumptions of Theorem II.17, if the system* (II.1) *is equal to the linear system* (II.7), *up to an independent change of variables, the set of dual effect free feedbacks is given by*

$$\mathbb{F}^{nde} = \{u \in \Gamma \ : \ u_t \preceq \zeta_t, \ \forall t \in \{0,\dots,T-1\}\},$$

with $\zeta_t(\omega) = w_t + C_t v_t + C_t A_{t-1} v_{t-1} + \cdots + C_t A_{t-1} \cdots A_0 v_0 + C_t A_{t-1} \cdots A_0 x_0$.

All the concepts and definitions related to stochastic systems were given for the sake of generality in the set-theoretic framework. They are of course also correct in the linear spaces framework. Since Theorem II.17 is oriented towards numerical considerations, it is stated in the linear spaces framework, especially in the scalar case. The limitation to the scalar case essentially comes from the tools used in the proof, i.e., differential arguments and open-mapping theorem. Our conjecture is that the result holds true in general spaces with a complete order which can be embedded in the real line, but we fail for the moment to prove this generalization.

Let us now discuss on two examples the assumptions (II.4) of Theorem II.17. One often consider stochastic systems of the type :

$$\begin{cases} y_t = v_t, \\ x_{t+1} = F_t(x_t, u_t, v_{t+1}), \end{cases} \tag{II.5}$$

$$\begin{cases} y_t = x_t, \\ x_{t+1} = F_t(x_t, u_t, v_{t+1}), \end{cases} \tag{II.6}$$

with some given non-linear dynamic functions. We can analyse the behaviour of these systems :

- The first case (II.5) is somehow degenerate, since controls do not appear in the observation function. Systems of type (II.5) are obviously NOLDE, and assumption (II.4b) allow us to forget such systems in our analysis. This assumption can therefore be seen as a necessary restriction to interesting cases.
- The second case (II.6) reveals a weakeness of our assumptions. Due to assumption (II.4a), it is out of our field. Indeed, if we try to adapt the proof to this type of system, we see quickly that we are not able to build a suitable change of variables. On the other hand, if F satisfy the injectivity assumptions (II.4c)–(II.4e), the knowledge of the state is equivalent to the

knowledge of the noise (thanks also to the perfect memory of the system). Hence, such systems are NOLDE. Moreover, such systems are practically of great importance, since they correspond to Markov decision processes. Our theorem fails therefore to characterize the NOLDE systems when the systems are somehow degenerate.

Our result is therefore not completely sufficient to understand the behaviour of NOLDE systems. It has to be considered as a first step into this direction. Its main interest, on the contrary, is to provide an easy criterion to give up the study of nonlinear stochastic systems. In fact, Theorem II.17 claims that open-loop controls in deeply nonlinear systems cause dual effect. Since numerical schemes like scenario trees or other prescribed discretizations of the noises are not compatible with dual effect, it claims that deeply nonlinear systems cannot be numerically solved.

We end this section with an example showing the interest of our result.

II.5. Proof of the Characterization Theorem

Theorem II.17 is based on Proposition II.19 (*if* sense) and Proposition II.20 (*only if* sense).

PROPOSITION II.19. *Assume that for all $t \in \{0,\dots,T-1\}$, the sets U_t, X_{t+1}, Y_{t+1}, V_{t+1} and W_{t+1} are linear spaces, and assume that there are linear operators A_t, B_t, C_t, such that :*

$$\text{(II.7)} \qquad \left\{ \begin{array}{c} H_t(x,w) = C_t x + w, \\ F_t(x,u,v) = A_t x + B_t u + v. \end{array} \right.$$

Then, the system (II.1) *is NOLDE.*

Preuve : The proof is straightforward. We use the observation functions h_t^+ and h_t. We have to show that for all t, the observation amount is independent from the open-loop controls.
It holds under our linearity assumptions that :

$$\begin{array}{rcl} \forall u \in U,\ \omega \in \Omega,\ h_t^+(u,\omega) & = & w_t + C_t v_t + C_t A_{t-1} v_{t-1} + \cdots + C_t A_{t-1} \cdots A_0 v_0 + C_t A_{t-1} \cdots A_0 x_0 \\ & & + \underbrace{C_t B_{t-1} u_{t-1} + C_t A_{t-1} B_{t-2} u_{t-2} + \cdots C_t A_{t-1} \cdots A_2 B_1 u_1}_{:=g(u)}, \\ & = & w_t + C_t v_t + C_t A_{t-1} v_{t-1} + \cdots + C_t A_{t-1} \cdots A_0 v_0 + C_t A_{t-1} \cdots A_0 x_0 + g(u). \end{array}$$

It is then obvious that : $h_t^+(u,.) \equiv h_t^+(u',.)$, since any mapping is equivalent with its translated by some translation, since translations are here bijections, thanks to the underlying linear structure. Hence, $h_t(u,.) \equiv h_t(u',.)$, which completes the proof. □

PROPOSITION II.20. *Consider system (II.1). Assume that the system is scalar, and homogeneous, i.e., that there are mappings $F \in C^1(\mathbb{R}^3,\mathbb{R})$ and $H \in C^1(\mathbb{R}^2,\mathbb{R})$ such that :*

$$\forall t \in \{1,\dots,T\},\ H_t = H,\ F_t = F.$$

Assume that following injectivity assumptions hold :

$$\text{(II.8)} \qquad \left\{ \begin{array}{ccc} \forall x \in \mathbb{R}, & H(x,.) & \text{is one-to-one,} \\ \forall w \in \mathbb{R}, & H(.,w) & \text{is one-to-one,} \\ \forall (x,u) \in \mathbb{R}^2, & F(x,u,.) & \text{is one-to-one,} \\ \forall (x,u,v) \in \mathbb{R}^3, & \partial F/\partial u(x,u,v) & \neq 0, \\ \forall (u,v) \in \mathbb{R}^2, & F(.,u,v) & \text{is one-to-one.} \end{array} \right.$$

Assume also that $T \geq 2$.
Then, if the system is NOLDE, there is a diffeomorphic change of variables under which F and H are such that :

$$\begin{array}{rcl} \hat{F}(\hat{x},\hat{u},\hat{v}) & = & \hat{x} + \hat{u} + \hat{v}, \\ \hat{H}(\hat{x},\hat{w}) & = & \hat{x} + \hat{w}. \end{array}$$

Preuve : The proof is based on lemmas II.23 and II.25. Indeed,using the definition of a NOLDE system, we get some mappings ζ_t and some one-to-one functions $\tilde{h}_t$ such that :

$$\forall u \in U,\ \forall \omega \in \Omega,\ h_t(u,\omega) = \tilde{h}_t(u,\zeta_t(\omega)).$$

Using the remark about causality, we can write the previous equation for $t = 1$, which is :

$$H(F(x_0,u_0,v_1),w_1) = \tilde{h}_1(u_0,\zeta_1(x_0,v_1,w_1)).$$

This equation is exactly the assumption made in lemma II.23. Therefore, as we verify all assumptions of this lemma, we shall write that there is a diffeomorphic change of variables such that :

$$\begin{aligned}\hat{H}(\hat{x},\hat{w}) &= \hat{x}+\hat{w},\\ \hat{F}(\hat{x},\hat{u},v) &= \hat{u}+\bar{F}(\hat{x},v).\end{aligned}$$

$\hat{F}$ is not yet linear with respect to the state variable.
Nevertheless, we know that NOLDE properties of any system are stable with respect to any change of variables. We now consider system (II.1) in which we have made the preceding change of variables making $\hat{H}$ linear and $\hat{F}$ quasi linear.
For the simplicity of notations, we will write in the sequel $\hat{H}=H$ and $\hat{F}=F$. We now write at the following time step the NOLDE property of the new system. It means that it holds, for all $(x_0,v_0,v_1,u_0,u_1,v_2,w_2,w_1)\in\mathbb{R}^8$:

$$\begin{aligned}\begin{pmatrix} H(F(x_0,u_0,v_1),w_1)\\ H(F(F(x_0,u_0,v_1),u_1,v_2),w_2)\end{pmatrix} &= \begin{pmatrix}\tilde{h}_{2,1}((u_0,u_1),\zeta_{2,1}(x_0,v_1,v_2,w_1,w_2))\\ \tilde{h}_{2,2}((u_0,u_1),\zeta_{2,2}(x_0,v_1,v_2,w_1,w_2))\end{pmatrix}, \text{i.e.,}\\ \begin{pmatrix}\bar{F}(x_0,v_1)+u_0+w_1\\ \bar{F}(\bar{F}(x_0,v_1)+u_0,v_2)+u_1+w_2\end{pmatrix} &= \begin{pmatrix}\tilde{h}_{2,1}((u_0,u_1),\zeta_{2,1}(x_0,v_1,v_2,w_1,w_2))\\ \tilde{h}_{2,2}((u_0,u_1),\zeta_{2,2}(x_0,v_1,v_2,w_1,w_2))\end{pmatrix}.\end{aligned}$$

We will now consider only the second part of the couple. It is there clear that u_1 and w_1 do not play any role. We can therefore write that for all $(x_0,u_0,v_1,v_2,w_2)\in\mathbb{R}^5$,

$$\bar{F}(\bar{F}(x_0,v_1)+u_0,v_2)+w_2=\tilde{\tilde{h}}_{2,2}(u_0,\tilde{\zeta}_{2,2}(x_0,v_1,v_2,w_2))$$

with given mappings $\tilde{\zeta}_{2,2}$ and $\tilde{\tilde{h}}_{2,2}$ such that for all $u_0\in\mathbb{R}$, $\tilde{\tilde{h}}_{2,2}(u_0,.)$ is one-to-one. This last equation allows us to apply lemma II.25, since injectivity assumptions are satisfied. We can therefore conclude that there exists a coordinates change for v such that one has :

$$\bar{F}(x,\hat{v})=x+\hat{v},$$

and it ends the proof. □

II.5.1. Lemmas. The first lemma is dedicated to some obvious properties of our order relation on functions.

LEMME II.21. *Let $f:\Omega\to F$, $g:\Omega\to G$, and $\alpha:E\to\Omega$ be some mappings on some abstract spaces. Then :*
(i) $f\equiv g\Rightarrow f\circ\alpha\equiv g\circ\alpha$.
(ii) If $F=G$, let $\beta:F\to F$ be a bijection. Then :
$f\equiv g\Rightarrow\beta\circ f\equiv g$.

Preuve : (i) $f\equiv g\Leftrightarrow\exists p:G\to F$, one-to-one, s.t. $f=p\circ g$. Then, it is clear that $f\circ\alpha=p\circ g\circ\alpha$ which is the result.
(ii) Suppose that $f\equiv g$. Then, there is some one-to-one mapping $p:F\to F$ such that $f=p\circ g$, i.e., $\beta\circ f=(\beta\circ p)\circ g$. Since β is a bijection, and p is one-to-one, $(\beta\circ p):F\to F$ is one-to-one, and hence, $\beta\circ f\equiv g$. □

LEMME II.22. *Let $f\in C^0(\mathbb{R},\mathbb{R})$ be one-to-one, and $g\in C^0(\mathbb{R}^p,\mathbb{R})$, $h\in C^0(\mathbb{R}\times\mathbb{R}^p,\mathbb{R})$, $k\in C^0(\mathbb{R}^2,\mathbb{R})$, and assume that :*

$$\forall(x,a,b)\in\mathbb{R}^p\times\mathbb{R}^2,\ f(b+g(x))+h(a,x)=k(a,b). \tag{II.9}$$

Then for all $a\in\mathbb{R}$, $h(a,.)\equiv g$, i.e., there is some one-to-one mapping $\rho_a\in C^0(\mathbb{R},\mathbb{R})$ such that $h(a,.)=\rho_a\circ g$.

Preuve : From equation (II.9), we get that :

$$\forall(x,a,b)\in\mathbb{R}^p\times\mathbb{R}\times\mathbb{R},\ h(a,x)=k(a,b)-f(b+g(x)),$$

the right hand side of this equation does not depend on b, and we can hence take $b=0$, which leads to :

$$\forall(x,a)\in\mathbb{R}^p\times\mathbb{R},\ h(a,x)=k(a,0)-f(g(x)).$$

Let us define for all $a\in\mathbb{R}$ the mapping $\rho_a:\mathbb{R}\to\mathbb{R}$ by :

$$\forall y\in\mathbb{R},\ \rho_a(y)=k(a,0)-f(y).$$

Since f is one-to-one, ρ_a is also one-to-one, and we have that :

$$\forall x \in \mathbb{R}^p,\ h(a,x) = \rho_a(g(x)),$$

which is the result. □

We give here some useful lemmas which are used in the proof of our main result.

LEMME II.23. *Let $F \in C^1(\mathbb{R}^3,\mathbb{R})$ and $H \in C^1(\mathbb{R}^2,\mathbb{R})$. Assume that :*

$$\forall (x,u,v) \in \mathbb{R}^3,\ \frac{\partial F}{\partial u}(x,u,v) \neq 0. \tag{II.10}$$

Assume also that there are some mappings $W, X \in C^1(\mathbb{R}^2,\mathbb{R})$ and $G \in C^1(\mathbb{R}^3,\mathbb{R})$ such that :

$$\forall (x,w,y) \in \mathbb{R}^3,\ y = H(x,w) \Leftrightarrow w = W(x,y), \tag{II.11}$$

$$\forall (x,w,y) \in \mathbb{R}^3,\ y = H(x,w) \Leftrightarrow x = X(w,y), \tag{II.12}$$

$$\forall (x,u,v,z) \in \mathbb{R}^4,\ z = F(x,u,v) \Leftrightarrow x = G(z,u,v). \tag{II.13}$$

and assume that there are some mappings $\tilde{h} \in C^1(\mathbb{R}^2,\mathbb{R})$ and $\zeta \in C^1(\mathbb{R}^3,\mathbb{R})$ such that :

$$\forall (x,u,v,w) \in \mathbb{R}^4,\ H(F(x,u,v),w) = \tilde{h}(u,\zeta(x,v,w)),$$

and for all $u \in \mathbb{R}$, $\tilde{h}(u,.)$ is injective.
Then there is a coordinate system given by four diffeomorphisms of $\mathbb{R}$

$$\left\{ \begin{array}{rcl} \hat{y} &=& \beta(y), \\ \hat{x} &=& -\alpha(x), \\ \hat{w} &=& \theta(w), \\ \hat{u} &=& \rho(u), \end{array} \right. \tag{II.14}$$

such that H and F can be rewritten in the following simple form :

$$\forall (x,u,v,w) \in \mathbb{R}^4,\ \left\{ \begin{array}{rcccl} \beta(H(x,w)) &=& \hat{H}(\hat{x},\hat{w}) &:=& \hat{x}+\hat{w}, \\ -\alpha(F(x,u,v)) &=& \hat{F}(\hat{x},\hat{u},v) &:=& \bar{F}(\hat{x},v)+\hat{u}. \end{array} \right. \tag{II.15}$$

Preuve :

(1) The idea of the proof is roughly the following : Given some real bijections α and β, it is always true that we can express a function of two variables $(x,y) \in \mathbb{R}^2$ as a function of the two changed variables $(\alpha(x)+\beta(y), \alpha(x)-\beta(y)) \in \mathbb{R}^2$. Then, we can express with differential arguments that our mysterious function depends in fact only on $(\alpha(x)+\beta(y))$, by writing that its derivative with respect to its second variable is null. That is exactly our goal :
we use the equation (II.11) and say that W does not depend on $(\alpha(x)-\beta(y))$. If such is the case, we can then write that there is some bijection θ such that :

$$W(x,y) = \theta(\alpha(x)+\beta(y)),$$

which yields to the linearization of H. Saying that W does not depend on $(\alpha(x)-\beta(y))$ can be expressed by writing :

$$\frac{1}{\alpha'(x)}\frac{\partial W}{\partial x}(x,y) - \frac{1}{\beta'(y)}\frac{\partial W}{\partial y}(x,y) = 0,\ \forall (x,y) \in \mathbb{R}^2.$$

The goal is then to identify the bijections α and β verifying the last equation. To do that, we use the injectivity properties (II.11-II.13) and the injectivity of $\tilde{h}(u,.)$.
We then apply same techniques to linearize the function F. Let's do the proof :

(2) We first use equation (II.11). It means that :

$$\forall (x,y) \in \mathbb{R}^2,\ H(x,W(x,y)) = y.$$

We can differentiate this equation, and we get the following two equations :

$$\begin{array}{rcl} \forall (x,y) \in \mathbb{R}^2,\ \dfrac{\partial H}{\partial x}(x,W(x,y)) + \dfrac{\partial H}{\partial w}(x,W(x,y))\dfrac{\partial W}{\partial x}(x,y) &=& 0, \\ \forall (x,y) \in \mathbb{R}^2,\ \dfrac{\partial H}{\partial w}(x,W(x,y))\dfrac{\partial W}{\partial y}(x,y) &=& 1. \end{array}$$

It means clearly that $\partial W/\partial y \neq 0$ everywhere, and hence we get :

$$\forall (x,y) \in \mathbb{R}^2,\ \frac{\frac{\partial W}{\partial x}}{\frac{\partial W}{\partial y}}(x,y) = -\frac{\partial H}{\partial x}(x, W(x,y)). \tag{II.16}$$

Our goal is now to express the x-derivative of H as the quotient of one function depending on x over one function depending on y.

(3) We now use the injectivity of $\tilde{h}(u,.)$. This implies (together with the differentiability assumptions on H and F) that there is some mapping $T \in C^1(\mathbb{R}^2, \mathbb{R})$ such that :

$$\forall (y,u,\nu) \in \mathbb{R}^3,\ y = \tilde{h}(u,\nu) \Leftrightarrow \nu = T(u,y),$$

which yields to the following equation :

$$\forall (u,\nu) \in \mathbb{R}^2,\ T(u, \tilde{h}(u,\nu)) = \nu.$$

This equation shows, by differentiation w.r.t. ν, that $\partial T/\partial y \neq 0$. We use now our last assumption, and get the following equation :

$$\forall (x,u,v,w) \in \mathbb{R}^4,\ T(u, H(F(x,u,v),w)) = \zeta(x,v,w).$$

We compute the u-derivative of this equation and get :

$$\forall (x,u,v,w) \in \mathbb{R}^4,\ \frac{\partial T}{\partial u}(u, H(F(x,u,v),w)) + \frac{\partial T}{\partial y}(u, H(F(x,u,v),w))\frac{\partial H}{\partial x}(F(x,u,v),w)\frac{\partial F}{\partial u}(x,u,v) = 0, \tag{II.17}$$

which can be rewritten, since by injectivity, $\partial T/\partial y \neq 0$:

$$\forall (x,u,v,w) \in \mathbb{R}^4,\ \frac{\frac{\partial T}{\partial u}}{\frac{\partial T}{\partial y}}(u, H(F(x,u,v),w)) = -\frac{\partial H}{\partial x}(F(x,u,v),w)\frac{\partial F}{\partial u}(x,u,v). \tag{II.18}$$

$$\forall (x,u,v,y) \in \mathbb{R}^4,\ \frac{\frac{\partial T}{\partial u}}{\frac{\partial T}{\partial y}}(u,y) = -\frac{\partial H}{\partial x}(F(x,u,v), W(F(x,u,v),y))\frac{\partial F}{\partial u}(x,u,v). \tag{II.19}$$

$$\forall (z,u,v,y) \in \mathbb{R}^4,\ \frac{\frac{\partial T}{\partial u}}{\frac{\partial T}{\partial y}}(u,y) = -\frac{\partial H}{\partial x}(z, W(z,y))\frac{\partial F}{\partial u}(G(z,u,v),u,v). \tag{II.20}$$

We now define for all $(u,y) \in \mathbb{R}^2$ the mapping $\varphi : \mathbb{R}^2 \to \mathbb{R}$ by :

$$\varphi(u,y) := \frac{\frac{\partial T}{\partial u}}{\frac{\partial T}{\partial y}}(u,y).$$

We also define for convenience the mapping $\psi : \mathbb{R}^3 \to \mathbb{R}$, using G defined in (II.13), by :

$$\forall (z,u,v) \in \mathbb{R}^3,\ \psi(z,u,v) := \frac{\partial F}{\partial u}(G(z,u,v),u,v). \tag{II.21}$$

(4) We have

$$\begin{aligned}
\forall (z,u,v,y) \in \mathbb{R}^4,\ \varphi(u,y) &= \frac{\frac{\partial T}{\partial u}}{\frac{\partial T}{\partial y}}(u,y), \\
&= -\frac{\partial H}{\partial x}(z, W(z,y))\frac{\partial F}{\partial u}(G(z,u,v),u,v) \quad \text{by (II.20)}, \\
&= \frac{\frac{\partial W}{\partial x}}{\frac{\partial W}{\partial y}}(z,y)\frac{\partial F}{\partial u}(G(z,u,v),u,v) \quad \text{by (II.16)}, \\
&= \psi(z,u,v)\frac{\frac{\partial W}{\partial x}}{\frac{\partial W}{\partial y}}(z,y) \quad \text{by (II.21)}.
\end{aligned} \tag{II.22}$$

We write this latter equality for $(u,v) = (0,0)$ and we get, since $\partial F/\partial u \neq 0$ and thus $\psi(z,u,v) \neq 0$:

$$\forall (z,y) \in \mathbb{R}^4,\ \frac{\varphi(0,y)}{\psi(z,0,0)} = \frac{\frac{\partial W}{\partial x}}{\frac{\partial W}{\partial y}}(z,y). \tag{II.23}$$

(5) Since $\psi(z,u,v) \neq 0$, there exists a mapping α such that

$$\alpha'(x) = \frac{1}{\psi(x,0,0)}\,. \tag{II.24}$$

We have $\varphi(u,y) := \frac{\frac{\partial T}{\partial u}}{\frac{\partial T}{\partial y}}(u,y)$ where $\partial T/\partial u \neq 0$ by equation (II.20). Thus, there exists a mapping β such that

$$\beta'(y) = \frac{1}{\varphi(0,y)}. \tag{II.25}$$

The mappings $x \to -\alpha(x)$ and $y \to \beta(y)$ define coordinate change, since α' and β' are everywhere non-null. From (II.23), we now get :

$$\forall (x,y) \in \mathbb{R}^2,\ \frac{1}{\alpha'(x)}\frac{\partial W}{\partial x}(x,y) - \frac{1}{\beta'(y)}\frac{\partial W}{\partial y}(x,y) = 0\,. \tag{II.26}$$

This easily implies that W does not depend on $(\alpha(x) - \beta(y))$. By the fact that H is a diffeomorphism when restricted to any of its arguments, there is hence some diffeomorphism θ such that :

$$\forall (x,y) \in \mathbb{R}^2,\ W(x,y) = \theta(\alpha(x) + \beta(y)). \tag{II.27}$$

Equation (II.27) achieves the goal for H. Let us define the following change of coordinates :

$$\left\{ \begin{array}{rcl} \hat{x} & = & -\alpha(x), \\ \hat{y} & = & \beta(y), \\ \hat{w} & = & \theta^{-1}(w), \end{array} \right. \tag{II.28}$$

With equations (II.28), we can now write :

$$\forall (x,w) \in \mathbb{R}^2,\ \beta(H(x,w)) = \theta^{-1}(w) - \alpha(x) = \hat{w} + \hat{x},$$

which is the "linearization" of H.

(6) We now want to linearize F in u. By definitions, we can rewrite equations (II.22) and (II.23) as follows :

$$\forall (z,u,v,y) \in \mathbb{R}^4,\ \frac{\partial F}{\partial u}(G(z,u,v),u,v)\alpha'(z) = \varphi(u,y)\beta'(y),$$

and hence :

$$\forall (x,u,v) \in \mathbb{R}^3,\ \frac{\partial F}{\partial u}(x,u,v)\alpha'(F(x,u,v)) = \varphi(u,0)\beta'(0). \tag{II.29}$$

We integrate equation (II.29) in u, and we get :

$$\forall (x,u,v) \in \mathbb{R}^3,\ \alpha(F(x,u,v)) = \alpha(F(x,0,v)) + \beta'(0)\int_0^u \varphi(\nu,y)d\nu. \tag{II.30}$$

Let us now set :

$$\hat{u} = \rho(u) := -\beta'(0)\int_0^u \varphi(\nu,0)d\nu. \tag{II.31}$$

This defines a change of variables since $\varphi(\nu,y) \neq 0$ everywhere. We get then :

$$\forall (x,u,v) \in \mathbb{R}^3,\ -\alpha(F(x,u,v)) \quad = \quad \bar{F}(\hat{x},v) + \hat{u},$$

with some given function $\bar{F} \in C^1(\mathbb{R}^2,\mathbb{R})$:

$$\forall (x,v) \in \mathbb{R}^2,\ \bar{F}(x,v) = -\alpha(F(\alpha^{-1}(-x),0,v)).$$

It hence proves the lemma. Moreover, we get that the mappings $\bar{F}(x,.)$ and $\bar{F}(.,v)$ are one-to-one for all $x,v \in \mathbb{R}$, because of the properties of F and α. □

Remarque II.24. *If one has that $F(x,u,v) = u + \bar{F}(x,v)$ in the previous lemma, the proof shows that $\alpha(x) = x$, and hence, that the change of variables on the first variable of H, on F and on its first variable is equal to identity.*

The previous lemma can be seen as a linearization result. We show now another linearization result :

LEMME II.25. *Let $\bar{F} \in C^1(\mathbb{R}^2, \mathbb{R})$ be such that for all $(x, v) \in \mathbb{R}^2$, $\bar{F}(x, .)$ and $\bar{F}(., v)$ are one-to-one respectively. Let also $\tilde{h} \in C^0(\mathbb{R}^2, \mathbb{R})$ and $\zeta \in C^0(\mathbb{R}^4, \mathbb{R})$ be such that for all $u \in \mathbb{R}$, $\tilde{h}(u, .)$ is one-to-one, and such that it holds :*

$$\forall (x_0, u_0, v_1, v_2, w_2) \in \mathbb{R}^5,\ \bar{F}(\bar{F}(x_0, v_1) + u_0, v_2) + w_2 = \tilde{h}(u_0, \zeta(x_0, v_1, v_2, w_2)), \tag{II.32}$$

Then there is some bijection $\beta \in \mathcal{C}^1(\mathbb{R}, \mathbb{R})$ such that :

$$\forall (x, v) \in \mathbb{R}^2,\ \bar{F}(x, v) = x + \beta(v).$$

Preuve :

(1) The first step of the proof is to show that there are some mappings θ and β such that :

$$\forall (x, v) \in \mathbb{R}^2,\ \bar{F}(x, v) = \theta(x + \beta(v)).$$

To show that, we define the mapping $\bar{\zeta} : \mathbb{R}^3 \to \mathbb{R}$ by $\bar{\zeta}(x_0, v_1, v_2) = \zeta(x_0, v_1, v_2, 0)$ for all $(x_0, v_1, v_2) \in \mathbb{R}^3$. We can now rewrite equation (II.32) by taking $w_2 = 0$, which means :

$$\forall (x_0, u_0, v_1, v_2) \in \mathbb{R}^4,\ \bar{F}(\bar{F}(x_0, v_1) + u_0, v_2) = \tilde{h}(u_0, \bar{\zeta}(x_0, v_1, v_2)).$$

This equation allows to apply lemma II.23, since all invertibility assumptions are satisfied. $\bar{F}$ plays here exactly the role of H in this previous lemma. By the remark II.24, the change of variable on the first variable of $\bar{F}$ is in fact the identity.
It appears hence that there are two bijections $\theta, \beta \in C^1(\mathbb{R}, \mathbb{R})$, such that :

$$\forall (x, v) \in \mathbb{R}^2,\ \bar{F}(x, v) = \theta(x + \beta(v)).$$

Since we deal with bijections, we can choose θ such that $\theta(0) = 0$, without loss of generality (we can of course translate θ).

(2) We show now that θ is linear. To do that, we use one more time equation (II.32). It holds, thanks to our injectivity assumptions, that there is some mapping $\psi \in C^1(\mathbb{R}^4, \mathbb{R})$ s.t. :

$$\forall (z, w_2, x_0, v_1, v_2) \in \mathbb{R}^5,\ z = \zeta(x_0, v_1, v_2, w_2) \Leftrightarrow w_2 = \psi(z, x_0, v_1, v_2).$$

Hence, we can write equation (II.32) while replacing $\bar{F}$ by θ, and writing $v_.$ instead of $\beta(v_.)$ (which is equivalent since β is a bijection), i.e.,

$$\forall (z, u_0, x_0, v_1, v_2) \in \mathbb{R}^5,\ \theta(\theta(x_0 + v_1) + v_2 + u_0) + \psi(z, x_0, v_1, v_2) = \tilde{h}(u_0, z).$$

It implies that the mapping $(x_0, v_1, v_2) \mapsto \theta(\theta(x_0 + v_1) + v_2 + u_0) + \psi(z, x_0, v_1, v_2)$ is constant. By appealing to lemma II.22, it shows that for all $z \in \mathbb{R}$, there is some mapping $\rho(z, .) \in C^1(\mathbb{R}, \mathbb{R})$ which is one-to-one, and such that :

$$\forall (z, x_0, v_1, v_2) \in \mathbb{R}^3,\ \psi(z, x_0, v_1, v_2) = \rho(z, \theta(x_0 + v_1) + v_2).$$

Therefore, it holds that :

$$\forall (y, u_0, z) \in \mathbb{R}^3,\ \theta(y + u_0) + \rho(z, y) = \tilde{h}(u_0, z). \tag{II.33}$$

We differentiate now equation (II.33) in variable y, and it leads to :

$$\forall (y, u_0, z) \in \mathbb{R}^3,\ \theta'(y + u_0) = -\frac{\partial \rho}{\partial y}(z, y),$$

which means :

$$\forall (y, u_0) \in \mathbb{R}^2,\ \theta'(y + u_0) = \theta'(y).$$

This last equation shows that the mapping $\theta'(.)$ is constant, and therefore θ is linear, since $\theta(0) = 0$. Therefore, we can write without loss of generality that θ is identity (since we make only change of variables), and it proves the lemma.

□

II.6. Conclusion

Coming from a previous result on stochastic dual effect free controls, to the question to characterize NOLDE systems, we have only considered the scalar systems for which one has open-loop perfect memory, and observation causality. In that case, we have shown that any linear system is NOLDE, and that any dual effect property is invariant under a coordinates change.
From this point, we have proven that any NOLDE scalar system which has some injectivity properties can be transformed into a system with linear dynamics and observation through a coordinates change.
It would then be great to understand better why one can prove such an algebraic result with such analytical methods (inversion theorem, etc.).
Our result has been proved in the scalar case, with differential arguments. We are not able today to prove it in a more general statement for the multi-dimensional case. The point would be to understand how injectivity arguments could be translated and used in the multi-dimensional case. This seems to be related with usual concepts of controllability and observability of systems.

CHAPITRE III

Stabilité

REMARQUE III.1. *La première partie de ce chapitre a fait l'objet d'une note publiée en décembre 2004 sur le site* http ://hera.rz.hu-berlin.de/speps//, *sous le titre* On the Fortet-Mourier metric for the stability of Stochastic Optimization Problems, an example. *La seconde partie de ce chapitre est un travail commun réalisé avec Werner Römisch et Holger Heitsch (Humboldt Universität zu Berlin), soumis sous le titre* Stability of Multistage Stochastic Programs *à SIAM Journal on Optimization en mai 2005, et maintenant accepté pour publication.*

III.1. Résumé

Dès lors que l'on souhaite résoudre numériquement un problème du type (I.1) ou (I.4), il est important de connaître les éventuelles propriétés de stabilité de ce problème, c'est-à-dire la façon dont la commande optimale et le coût optimal se comportent lorsque les paramètres du problème changent. Cette branche importante de l'optimisation s'est beaucoup développée, notamment par les travaux [**79**] ou [**23**].

Son intérêt numérique est immédiat : si l'on parvient à quantifier les variations de la commande optimale en fonction des variations des paramètres, on sera assuré qu'une discrétisation des paramètres dont on connaît l'erreur conduira à une solution approchée dont on saura la distance au véritable optimum.

Pour les problèmes stochastiques comme ceux qui nous préoccupent, la première chose est de déterminer les paramètres par rapport auxquels on aimerait perturber le problème, en un mot, ce que l'on souhaiterait discrétiser ! Dans notre cas, il assez clair qu'il s'agit de l'aléa, i.e. de la variable aléatoire $\boldsymbol{\xi}$ introduite dans le chapitre I. C'est exactement dans cette voie que les travaux [**86**], [**73**], [**51**], [**71**] se sont développés : se concentrant sur des résultats propres à la boucle ouverte, dans lesquels on peut montrer que l'erreur sur le coût optimal lorsque l'on change la loi de $\boldsymbol{\xi}$ est gouvernée par une distance entre ces lois du type distance de Fortet-Mourier (voir [**52**]), ces travaux ont déduit des techniques de discrétisation de l'aléa pour les problèmes à plusieurs pas de temps comme le problème (I.8). Ces techniques produisent des arbres d'aléas représentant des lois discrètes dont la distance à la loi véritable est connue, ou du moins, minimisée dans un certain sens (voir [**55**],[**47**]) ; la structure d'arbre de l'aléa discrétisé est là pour assurer que les commandes déterminées dans ce cadre discret ne sont pas anticipatives. Ainsi, on pourrait a priori attendre de ce type de techniques un traitement à la fois des espérances intervenant dans la fonction de coût, et des contraintes probabilistes (mesurabilité, ou non-anticipativité) pesant sur la commande.

A l'opposé de ces soucis de nature quantitative, on trouve l'approche développée dans [**8**] pour discrétiser les problèmes d'optimisation stochastique tout en respectant les contraintes de mesurabilité pesant sur la commande. Cette approche, très qualitative, commence par traiter à l'aide de techniques de quantification la mesurabilité pesant sur les variables de commande, puis procède à une discrétisation des espérances restant dans la formulation du problème. Un résultat de convergence de cette méthode est donné, et même, pour des fonctions de coût particulières, un résultat de Lipschitz, i.e. de stabilité quantitative.

Le but de ce chapitre est de montrer dans un premier temps comment l'usage exclusif de distances sur des lois ne permet pas de tenir compte des contraintes de mesurabilité, et peut conduire à des discrétisations produisant des solutions approchées arbitrairement mauvaises de la vraie solution. Dans un second temps, pour des problèmes stochastiques linéaires, on donne un résultat de stabilité faisant intervenir un terme mesurant la déviation de la loi de la variable aléatoire, et un autre terme mesurant la déviation de la structure d'information associée.

Remarque III.2 (Problème type considéré). *Dans la première section de ce chapitre, on utilisera la modélisation avec feedback, c'est à dire la modélisation du problème* (I.4)*, tandis que dans les sections suivantes du chapitre, on utilisera la modélisation du problème* (I.1)*. Cette divergence s'explique par la différence de ce que nous souhaitons mettre en avant dans chacune des sections : la première section nécessite pour une bonne compréhension une notation fonctionnelle de la contrainte de mesurabilité, tandis que les sections suivantes, de par leur recours fréquent à la théorie des probabilités et de la mesurabilité, nécessitent une notation plus probabiliste.*

III.2. Distance de Fortet-Mourier en optimisation stochastique

III.2.1. Motivation. Dans les problèmes du type (I.1), l'aléa, modélisé par la variable aléatoire $\boldsymbol{\xi}$, intervi.e.nt à deux niveaux :

– à travers le *calcul d'une espérance* dans la fonction objectif du problème,

– à travers l'existence de possibles *contraintes de mesurabilité* dans l'ensemble admissible.

Immédiatement, une considération pratique s'impose : ne connaissant pour ainsi dire jamais la véritable distribution des variables aléatoires sous-jacentes au problème, on considère toujours une version discrète ou du moins approchée du problème initial. En un mot, on substitue une *loi approchée* à la loi véritable sous l'espérance, et une *description approchée de l'espace mesurable* à la description théorique.

Deux questions se posent donc dorénavant : d'une part, comment calculer de façon approchée ces espérances, sachant que l'on ne dispose souvent que de tirages des lois des variables aléatoires, et non d'une expression de leur loi ? d'autre part, comment écrire de façon discrète des contraintes de mesurabilité ? de sorte dans les deux cas que la solution approchée du problème tende vers une solution réalisable du problème initial, ou du moins que l'on puisse reconstruire à partir d'une solution discrète une solution réalisable. De ces deux questions en naît une troisième : est-il opportun de *discrétiser* de la même façon le calcul d'une espérance et l'écriture de contraintes de mesurabilité du problème ?

Le problème de calcul approché d'une espérance est abondamment traité dans la littérature tant du côté probabiliste que du côté de la programmation stochastique. Pour fixer les idées, cette question est souvent résolue par une technique de Monte-Carlo, puisque la donnée d'entrée traditionnelle d'un problème d'optimisation stochastique est souvent un échantillon des variables aléatoires. Une autre alternative consiste en une quantification de l'espace, i.e. en un partitionnement de l'espace en sous-ensembles auxquels sont fixées des probabilités d'occurence.

Cependant, ce calcul a souvent occulté les problèmes liés à la mesurabilité, et une attitude commune est d'oublier ces contraintes de mesurabilité en optimisant sur des commandes déterministes, i.e. en boucle ouverte (voir [**71**]), comme si les résultats obtenus de cette façon, en termes de stabilité de la solution, etc, passaient facilement au problème contraint, i.e. à la boucle fermée.

La question est ici précisément d'utiliser le résultat de stabilité montré par Pflug dans [**71**] faisant intervenir une distance entre mesures de probabilité et utilisé ensuite pour construire des arbres de scénarios, pour mettre en évidence l'importance de séparer les deux discrétisations.

III.2.2. Cadre général et résultat de stabilité classique. Le cadre général posé par Pflug [**71**] est le suivant : on cherche à résoudre un problème de contrôle stochastique à l'aide d'un problème approché, et on s'intéresse donc à l'erreur d'approximation. De façon générique, on se donne un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. Soit $\boldsymbol{\xi}$ une variable aléatoire de loi μ à valeurs dans Ξ un espace métrique, muni de la distance $c(\cdot,\cdot)$. On se donne une suite de variables aléatoires $\boldsymbol{\xi}_n$ à valeurs dans $\Xi_n \subset \Xi$, et l'on note μ_n leurs lois (dont le support est Ξ_n). De façon analogue, on note $\mathcal{F}_n$ la tribu engendrée sur Ω par $\boldsymbol{\xi}_n$.

On se donne un espace de commande noté U, typiquement, $U \subset \mathbb{R}^p$, et une fonction critère $j : U \times \Xi \longrightarrow \mathbb{R}$. On suppose que $j(u,\cdot)$ est mesurable sur Ξ pour tout $u \in U$ et que pour presque tout $\xi \in \Xi, j(\cdot,\xi)$ est continue sur U. j est donc une fonction de Carathéodory, et c'est donc une intégrande normale, i.e. si $u : \Xi \to U$ est mesurable, alors $j(u(\cdot),\cdot)$ l'est aussi. On définit alors l'espace fonctionnel des commandes $\Gamma = \left\{u : \Xi \to U \text{ mesurable } : j(u(\cdot),\cdot) \in L^1(\Xi,\mathcal{F},\mu)\right\}$.On définit alors l'espace $U^f \subset \Gamma$ des commandes admissibles, et on note pour tout $u \in U^f$, $J(u) =$

$\mathbb{E}\left(j(u(\boldsymbol{\xi}),\boldsymbol{\xi})\right)$.
On définit de même pour tout $n \in \mathbb{N}$ l'espace fonctionnel de commandes :

$$\Gamma_n = \left\{u : \Xi_n \to U \text{ mesurable } : \ j(u(\cdot),\cdot) \in L^1(\Xi_n, \mathcal{F}_n, \mu_n)\right\} \ ,$$

et $U_n^f \subset \Gamma_n$ l'espace des commandes admissibles associé. On note $J_n(u_n) = \mathbb{E}\left(j(u_n(\boldsymbol{\xi}_n),\boldsymbol{\xi}_n)\right)\ \forall u_n \in U_n^{ad}$.
Afin d'assurer l'existence de solutions aux problèmes (III.1)-(III.2) ci-dessous, on suppose que U^f et pour tout $n \in \mathbb{N}$ U_n^f sont convexes fermés, et que $j(\cdot,\xi)$ est convexe coercive pour tout $\xi \in \Xi$.

$$\min_{u \in U^f} \mathbb{E}\left(j(u(\boldsymbol{\xi}),\boldsymbol{\xi})\right) \ ; \tag{III.1}$$

$$\forall n \in \mathbb{N}, \ \min_{u_n \in U_n^f} \mathbb{E}\left(j(u_n(\boldsymbol{\xi}_n),\boldsymbol{\xi}_n)\right) \ . \tag{III.2}$$

On fait maintenant pour tout $n \in \mathbb{N}$ les hypothèses suivantes :
– $\forall u \in U^f,\ u_{|\Xi_n} \in U_n^f$,
– il existe un opérateur de prolongement[1] $\phi_n : \Xi \longrightarrow \Xi_n$ mesurable et surjectif tel que $\forall u_n \in U_n^f,\ u_n \circ \phi_n \in U^f$ et $(u_n \circ \phi_n)_{|\Xi_n} = u_n$.

Il est alors possible de calculer $J(u_n) = \mathbb{E}\left(j(u_n \circ \phi_n(\boldsymbol{\xi}),\boldsymbol{\xi})\right)$ pour tout $u_n \in U_n^f$ et $J_n(u) = \mathbb{E}\left(j(u_{|\Xi_n}(\boldsymbol{\xi}_n),\boldsymbol{\xi}_n)\right)$ pour tout $u \in U^f$.

Là où nous en sommes, il faut constater que la démarche entreprise par Pflug traite la discrétisation de l'espérance et celle de la contrainte de mesurabilité (incluse dans U^f) de la même façon, i.e. avec l'unique donnée d'une loi approchée dite μ_n et de son support Ξ_n très naturellement prolongé en Ξ.

Pour se fixer les idées, $\Xi \subset \mathbb{R}^m$, Ξ_n sera une partie finie de Ξ correspondant à un n-échantillon, sur laquelle μ_n sera la loi empirique de μ associée à l'échantillon.

Le cadre donné ci-dessus est tout à fait général et permet de traiter à la fois la boucle ouverte avec des u constants, et la boucle fermée. Afin de mesurer l'erreur commise, on choisit de poser comme critère d'erreur la quantité notée $e(u,u_n) = \left|\mathbb{E}\left(j(u(\boldsymbol{\xi}),\boldsymbol{\xi})\right) - \mathbb{E}\left(j(u_n \circ \phi_n(\boldsymbol{\xi}),\boldsymbol{\xi})\right)\right|$ avec $u \in U^f,\ u_n \in U_n^f$. Cette erreur compare donc l'évaluation du coût véritable d'une solution réalisable du problème véritable avec le coût véritable d'une solution réalisable du problème approché rendue réalisable pour le problème véritable à l'aide de l'opérateur de prolongement[2].
On note u^* la solution du problème (III.1) et u_n^* la solution du problème (III.2) pour tout $n \in \mathbb{N}$.
On va alors s'intéresser à $e(u^*,u_n^*)$ et à sa limite quand n tend vers l'infini.
On est maintenant en mesure d'énoncer la proposition suivante, dûe à Pflug, mais que l'on redémontre ici pour l'exhaustivité :

Proposition III.3 (Pflug). *Soit* $\mathcal{G} = \left\{g : \Xi \to \mathbb{R} \ : \ \exists u \in U^f,\ g(\xi) = j(u(\xi),\xi)\ \forall \xi \in \Xi\right\}$. *Alors :*

$$\forall n \in \mathbb{N}, \quad e(u^*,u_n^*) \leq 2 \sup_{g \in \mathcal{G}} \left|\mathbb{E}\left(g(\boldsymbol{\xi})\right) - \mathbb{E}\left(g(\boldsymbol{\xi}_n)\right)\right| .$$

[1] À ce stade de la position du problème, on pourait remplacer cette hypothèse par la suivante :

$$\exists \psi_n : U_n^f \longrightarrow U^f, \text{ tel que } \forall u \in U^f, u_{|\Xi_n} \in U_n^f,\ \forall u_n \in U_n^f, (\psi_n \circ u_n)_{|\Xi_n} = u_n \ .$$

Et tout pourra ensuite s'écrire très correctement avec cette notation. Considérer ce type de transformation pour passer d'un problème à l'autre permet de tenir compte d'un très grand nombre de prolongements de tous styles, mais pour notre propos, l'opérateur de prolongement ϕ_n suffit, et cela ne change rien au problème et aux questions soulevées.

[2] D'un point de vue industriel, cela reviendrait à comparer le coût de gestion optimal inconnu avec le coût de gestion sur le terrain d'une stratégie calculée par approximation du problème véritable, c'est donc la quantité que l'on souhaite absolument rendre proche de 0.

Preuve : On pose $\mathcal{M} = \{u \in U^f : \mathbb{E}\left(j(u(\boldsymbol{\xi}), \boldsymbol{\xi})\right) \leq \mathbb{E}\left(j(u^*(\boldsymbol{\xi}), \boldsymbol{\xi})\right) + 2\varepsilon\}$, avec $\varepsilon = \sup_{g\in\mathcal{G}} |\mathbb{E}\left(g(\boldsymbol{\xi})\right) - \mathbb{E}\left(g(\boldsymbol{\xi}_n)\right)|$. On raisonne par l'absurde en supposant que $u_n^* \circ \phi_n \notin \mathcal{M}$. On a alors :

$$\begin{aligned} 2\varepsilon + \mathbb{E}\left(j(u^*(\boldsymbol{\xi}), \boldsymbol{\xi})\right) <& \mathbb{E}\left(j(u_n^* \circ \phi_n(\boldsymbol{\xi}), \boldsymbol{\xi})\right) \\ \leq& \varepsilon + \mathbb{E}\left(j(u_n^*(\boldsymbol{\xi}_n), \boldsymbol{\xi}_n)\right) \text{ par définition de } \varepsilon, \text{ et comme } u_n^* \circ \phi_n|_{\Xi_n} = u_n^*, \\ \leq& \varepsilon + \mathbb{E}\left(j(u^*(\boldsymbol{\xi}_n), \boldsymbol{\xi}_n)\right) \text{ par minimalité,} \\ \leq& 2\varepsilon + \mathbb{E}\left(j(u^*(\boldsymbol{\xi}), \boldsymbol{\xi})\right) \text{ par définition de } \varepsilon, \end{aligned}$$

d'où la contradiction, ce qui achève la preuve. □

On notera dans la suite par commodité

$$\Delta_{max} = \sup_{g\in\mathcal{G}} |\mathbb{E}\left(g(\boldsymbol{\xi})\right) - \mathbb{E}\left(g(\boldsymbol{\xi}_n)\right)| = \sup_{u\in U^f} |J(u) - J_n(u)|$$

.

Le majorant de l'erreur ressemble à une distance entre deux mesures de probabilité, la distance de Fortet-Mourier (introduite dans [**52**]). Elle est définie par :

$$d(\mu, \mu_n) = \sup_{f1-\text{lipschitzienne}} \left| \int_\Xi f d\mu - \int_\Xi f d\mu_n \right| .$$

Nous introduisons alors l'hypothèse suivante :

$$|j(u(\xi), \xi) - j(u(\xi'), \xi')| \leq c(\xi, \xi') \quad \forall \xi, \xi' \in \Xi, \ \forall u \in U^f \ . \tag{III.3}$$

En faisant l'hypothèse (III.3), on peut donc majorer l'erreur par deux fois la distance de Fortet-Mourier :

$$\forall n \in \mathbb{N}, \quad e(u^*, u_n^*) \leq 2 \sup_{u\in U^f} |\mathbb{E}\left(j(u(\boldsymbol{\xi}), \boldsymbol{\xi})\right) - \mathbb{E}\left(j(u(\boldsymbol{\xi}_n), \boldsymbol{\xi}_n)\right)| \leq 2d(\mu, \mu_n) \ . \tag{III.4}$$

Ceci semble très séduisant, car pour une application, on sait que l'on peut faire tendre la distance de Fortet-Mourier entre une loi et sa loi empirique vers 0, ce qui assurerait la qualité de la solution approchée déterminée par l'écriture du problème sur la loi empirique.
Cependant, c'est ici qu'apparait une disjonction forte entre la boucle fermée et la boucle ouverte. En effet, écrire l'hypothèse (III.3) avec des fonctions $u \in U^f$ constantes ou avec des fonctions variables change radicalement le problème : dans le premier cas, l'hypothèse (III.3) se lit simplement comme une propriété analytique du critère j, tandis que dans le deuxième cas, elle nécessite pour être vérifiée d'étudier j composé avec u, pour u variant dans la classe des contrôles admissibles. En d'autres termes, *en passant à la distance de Fortet-Mourier, on a perdu les notions de structure d'information sur la commande*, car si la première inégalité de (III.4) fait intervenir la structure d'information à travers U^f, la seconde les oublie car U^f est remplacé par l'ensemble des fonctions lipschitziennes.
Par exemple, dans le cas où l'aléa et la commande sont unidimensionnels, si le critère s'écrit $j(u, \xi) = u^2 + \xi$, il sera bien entendu lipschitzien en boucle ouverte, mais si le contrôle en boucle fermée est de la forme $u(\xi) = \xi$, il ne sera plus lipschitzien sur $\mathbb{R}$.
Nous allons maintenant étudier sur un exemple simple ce phénomène pour mettre en évidence l'impossibilité de faire tendre le majorant de l'erreur vers 0 en boucle fermée, et donc le caractère inopérant de la distance de Fortet-Mourier pour ce type de problèmes.

III.2.3. Convergence de la distance de Fortet-Mourier. On se place ici dans le même cadre qu'au premier paragraphe, on se donne $(\boldsymbol{\xi}_i)_{1\leq i\leq n}$ n v.a. i.i.d de loi μ, et on pose $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{\boldsymbol{\xi}_i}$ la mesure empirique associée à μ. On a alors le théorème suivant :

Théorème III.4. *Si le support de μ est compact, alors :*

$$\lim_{n\to\infty} d(\mu, \mu_n) = 0 \text{ p.s.}$$

Preuve : On sait que la loi empirique converge en loi p.s. vers la vraie loi (Glivenko-Cantelli, [**37**], Théorème 4.4.24). Cela signifie que pour toute fonction continue bornée $f : \Xi \to \mathbb{R}$, $\mu_n(f) \to \mu(f)$ quand n tend vers l'infini. On considère maintenant $\mathcal{L}_M = \{f : supp(\mu) \subset \Xi \to \mathbb{R} \ : \ |f(\xi) - f(\xi')| \leq c(\xi, \xi') \ \forall \xi, \xi' \in supp(\mu), \ ||f||_\infty \leq M\}$ l'ensemble des fonctions lipschitziennes sur $supp(\mu) \subset \Xi$ uniformément bornées par

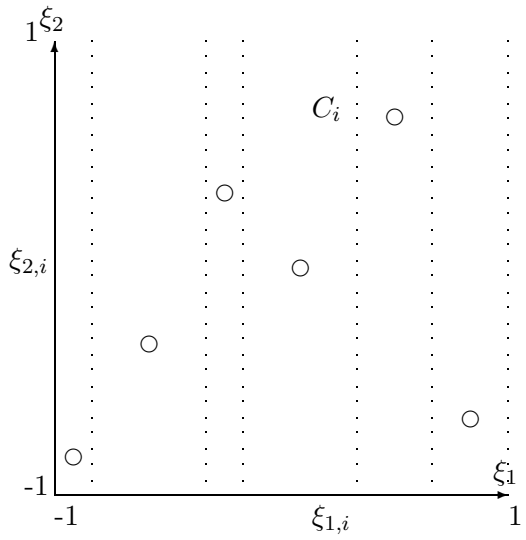


FIG. 1. Partition de $[-1,1]^2$

$M \in \mathbb{R}$. Cet ensemble est relativement compact dans l'ensemble des fonctions continues sur un compact, car c'est un ensemble équicontinu de fonctions à valeurs dans un compact (théorème d'Ascoli).
En appliquant alors le théorème 5.13.12 de [**89**], on a la convergence uniforme de μ_n vers μ sur $\mathcal{L}_M$. Enfin, prendre le sup sur les fonctions lipschitziennes sur un compact ou prendre le sup sur les fonctions lipschitziennes uniformément bornées sur un compact revient au même car vu l'expression de la distance de Fortet-Mourier, ajouter une constante aux fonctions ne change pas la valeur du sup. On a donc la convergence de la distance de Fortet-Mourier entre une loi à support compact et sa loi empirique. □

III.2.4. Contre-exemple. On considère le problème d'optimisation stochastique suivant :

$$\min_{\boldsymbol{u}\ \sigma(\boldsymbol{\xi}_1)-\text{mesurable}} \mathbb{E}\left(\varepsilon \boldsymbol{u}^2 + \beta \boldsymbol{y}^2\right) , \tag{III.5}$$

où l'on définit :

$$\begin{aligned} \varepsilon, \beta &> 0, \\ \boldsymbol{\xi}_1 &\sim \mathcal{U}_{[-1,1]}, \text{ l'état initial observé,} \\ \boldsymbol{\xi}_2 &\sim \mathcal{U}_{[-1,1]}, \text{ le bruit, indépendant de } \boldsymbol{\xi}_1, \\ \boldsymbol{y} &= \boldsymbol{\xi}_1 + \boldsymbol{u} + \boldsymbol{\xi}_2 \text{ l'état final du système,} \\ \boldsymbol{u} &= \phi(\boldsymbol{\xi}_1) \text{ une commande réelle en boucle fermée sur l'état initial.} \end{aligned}$$

L'espace de probabilité Ξ correspondant est donc $[-1,1]^2$ que l'on munit de ses boréliens, et de la mesure de Lebesgue par rapport à laquelle la loi uniforme produit sur $[-1,1]^2$ est absolument continue.

On résout explicitement le problème par programmation dynamique stochastique, en introduisant les fonctions de Bellman $V_1(y) = \beta y^2$ et $V_0(\xi_1) = \min_u \mathbb{E}\left(\varepsilon u^2 + V_1(\xi_1 + u + \boldsymbol{\xi}_2)\right)$. On obtient alors la commande optimale en boucle fermée (notée abusivement u^*) et le coût optimal :

$$u^*(\xi_1) = -\frac{\beta \xi_1}{\varepsilon + \beta} \ \forall \xi_1 \in [-1,1] \text{ et } J^* = \mathbb{E}\left(V_0(\boldsymbol{\xi}_1)\right) = \frac{1}{3}\left(\frac{\varepsilon\beta}{\varepsilon+\beta} + \beta\right) .$$

On prend maintenant un N-échantillon de la loi de $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$, noté $(\boldsymbol{\xi}_{1,i}, \boldsymbol{\xi}_{2,i})_{1\le i\le n}$. On est donc exactement dans le cas général évoqué avant, avec une probabilité μ_n égale à la loi empirique de μ : $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{(\boldsymbol{\xi}_{1,i}, \boldsymbol{\xi}_{2,i})}$. On peut alors résoudre le problème sur les N trajectoires indépendantes, ce qui nous donne les commandes optimales u_i^* correspondant à chaque trajectoire $(\boldsymbol{\xi}_{1,i}, \boldsymbol{\xi}_{2,i})$. En effet, les $\boldsymbol{\xi}_{1,i}$ sont génériquement tous différents, et la contrainte de boucle fermée

$\boldsymbol{u} = \phi(\boldsymbol{\xi}_1)$ sur l'état initial nous autorise à considérer une commande par trajectoire, ce qui fournit, tout calcul fait :

$$u_i^* = -\frac{\beta}{\varepsilon+\beta}(\boldsymbol{\xi}_{1,i} + \boldsymbol{\xi}_{2,i}) \ \forall i \in \{1,\dots,n\} \ .$$

On réalise un prolongement de cette solution ponctuelle pour la connaître sur tout l'espace. Il faut donc réaliser une partition de $[-1,1]^2$ qui à tout $x \in [-1,1]$ associe un unique couple $(\boldsymbol{\xi}_{1,i}, \boldsymbol{\xi}_{2,i})$, pour respecter la causalité du problème, i.e. le fait que la commande soit en boucle fermée sur l'état initial. Génériquement, tous les couples $(\boldsymbol{\xi}_{1,i}, \boldsymbol{\xi}_{2,i})$ sont différents avec probabilité 1. On appelle cette opération une quantification de l'espace. On réalise une partition $C = \{C_1,\dots,C_n\}$ de $[-1,1]$ telle que $\forall i \in \{1,\dots,n\}, \boldsymbol{\xi}_{1,i} \in C_i, \forall j \neq i, \boldsymbol{\xi}_{1,j} \notin C_i$. Ainsi, on aura l'opérateur ϕ du premier paragraphe donné par : $\forall (\xi_1,\xi_2) \in [-1,1]^2, \exists! i \in \{1,\dots,n\}, \xi_1 \in C_i$, qui définit $\phi(\xi_1,\xi_2) = (\boldsymbol{\xi}_{1,i}, \boldsymbol{\xi}_{2,i})$. Typiquement, si l'on classe les $\boldsymbol{\xi}_{1,i}$, on découpe l'intervalle $[-1,1]$ en bandes du type $\left[\frac{\boldsymbol{\xi}_{1,i}+\boldsymbol{\xi}_{1,i-1}}{2}, \frac{\boldsymbol{\xi}_{1,i}+\boldsymbol{\xi}_{1,i-1}}{2}\right]$. Cette partition de $[-1,1]$ nous donne une partition de $[-1,1]\times[-1,1]$ comme l'indique la figure 1, qui nous permet d'associer à tout $\xi_1 \in [-1,1]$ un unique couple $(\boldsymbol{\xi}_{1,i}, \boldsymbol{\xi}_{2,i})$. Ainsi, on peut prolonger la solution ponctuelle (u_i^*) en u_n^* définie par :

$$u_n^*(\xi_1) = -\frac{\beta}{\varepsilon+\beta}\sum_{i=1}^n 1_{C_i}(\xi_1)(\boldsymbol{\xi}_{1,i} + \boldsymbol{\xi}_{2,i}) \quad , \forall \xi_1 \in [-1,1] \ . \tag{III.6}$$

On peut dès lors s'intéresser au terme d'erreur précédent e et à sa majoration comme dans (III.4) : ici, il est bien clair que u_n^* défini par (III.6) est admissible pour le problème initial. On a la minoration :

(III.7)

$$|\mathbb{E}\left(\varepsilon u_n^{*2}(\boldsymbol{\xi}_1) + \beta(\boldsymbol{\xi}_1 + u_n^*(\boldsymbol{\xi}_1) + \boldsymbol{\xi}_2)^2\right) - \frac{1}{n}\sum_{i=1}^n \left(\varepsilon u_n^{*2}(\boldsymbol{\xi}_{1,i}) + \beta(\boldsymbol{\xi}_{1,i} + u_n^*(\boldsymbol{\xi}_{1,i}) + \boldsymbol{\xi}_{2,i})^2\right)| \leq \Delta_{max} \ .$$

Car ici :

$$\Delta_{max} = \sup_u |\mathbb{E}\left(\varepsilon u^2(\boldsymbol{\xi}_1) + \beta(\boldsymbol{\xi}_1 + u(\boldsymbol{\xi}_1) + \boldsymbol{\xi}_2)^2\right) - \frac{1}{n}\sum_{i=1}^n \left(\varepsilon u^2(\boldsymbol{\xi}_{1,i}) + \beta(\boldsymbol{\xi}_{1,i} + u(\boldsymbol{\xi}_{1,i}) + \boldsymbol{\xi}_{2,i})^2\right)| \ .$$

On calcule alors séparément les deux termes du membre de gauche de (III.7) :

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \left(\varepsilon u_n^{*2}(\boldsymbol{\xi}_{1,i}) + \beta(\boldsymbol{\xi}_1, + u_n^*(\boldsymbol{\xi}_{1,i}) + w)^2\right) =& \frac{\varepsilon\beta^2}{(\varepsilon+\beta)^2}\frac{1}{n}\sum_{i=1}^n(\boldsymbol{\xi}_{1,i} + \boldsymbol{\xi}_{2,i})^2 + \frac{\beta\varepsilon^2}{(\varepsilon+\beta)^2}\frac{1}{n}\sum_{i=1}^n(\boldsymbol{\xi}_{1,i} + \boldsymbol{\xi}_{2,i})^2 \ , \\ =& \frac{\varepsilon\beta}{\varepsilon+\beta}\frac{1}{n}\sum_{i=1}^n(\boldsymbol{\xi}_{1,i} + \boldsymbol{\xi}_{2,i})^2 \longrightarrow \frac{2}{3}\frac{\varepsilon\beta}{\varepsilon+\beta} \text{ quand } n \to \infty \ .\end{aligned}$$

Puis le second terme :

$$\begin{aligned}\mathbb{E}\left(\varepsilon u_n^{*2}(\boldsymbol{\xi}_1) + \beta(x + u_n^*(\boldsymbol{\xi}_1) + w)^2\right) =& \frac{\beta^2\varepsilon}{(\varepsilon+\beta)^2}\sum_{i=1}^n \mathbb{P}(\boldsymbol{\xi}_1 \in C_i)(\boldsymbol{\xi}_{1,i} + \boldsymbol{\xi}_{2,i})^2 \\ &+ \frac{\beta^3}{(\varepsilon+\beta)^2}\sum_{i=1}^n \mathbb{P}(\boldsymbol{\xi}_1 \in C_i)(\boldsymbol{\xi}_{1,i} + \boldsymbol{\xi}_{2,i})^2 \\ &+ \frac{2\beta}{3} - \frac{2\beta^2}{\varepsilon+\beta}\sum_{i=1}^n(\boldsymbol{\xi}_{1,i} + \boldsymbol{\xi}_{2,i})\mathbb{E}\left(\boldsymbol{\xi}_1 1_{C_i}(\boldsymbol{\xi}_1)\right) \ , \\ =& \frac{\beta^2}{\varepsilon+\beta}\sum_{i=1}^n \mathbb{P}(\boldsymbol{\xi}_1 \in C_i)(\boldsymbol{\xi}_{1,i} + \boldsymbol{\xi}_{2,i})^2 + \frac{2\beta}{3} \\ &- \frac{2\beta^2}{\varepsilon+\beta}\sum_{i=1}^n(\boldsymbol{\xi}_{1,i} + \boldsymbol{\xi}_{2,i}) \underbrace{\mathbb{E}\left(\boldsymbol{\xi}_1 1_{C_i}(\boldsymbol{\xi}_1)\right)}_{\frac{1}{n}\sum_{j=1}^n \boldsymbol{\xi}_{1,j}1_{C_i}(\boldsymbol{\xi}_{1,j}) = \frac{\boldsymbol{\xi}_{1,i}}{n}} \ .\end{aligned}$$

Ce qui nous importe ici est le calcul asymptotique, i.e. quand $n \to \infty$. Le premier terme tend clairement vers $\frac{2\beta^2}{3(\varepsilon+\beta)}$ par la loi forte des grands nombres. Le dernier terme peut aussi être traité par la loi forte des grands nombres en le séparant en $\boldsymbol{\xi}_{1,i}$ et $\boldsymbol{\xi}_{2,i}$, et en le traitant par indépendance : il tend vers $-\frac{2\beta^2}{3(\varepsilon+\beta)}$. Finalement, on a donc la limite de la différence, qui nous donne :

$$\frac{2}{3}\left(\beta - \frac{\varepsilon\beta}{\varepsilon+\beta}\right) \leq \Delta_{max} \ . \tag{III.8}$$

On voit donc que pour un ε petit, le majorant de l'erreur est minoré par un terme de l'ordre de $\frac{2\beta}{3}$. Le majorant de l'erreur en boucle fermée ne tend donc pas vers 0.
En revanche, par le théorème III.4, la distance de Fortet-Mourier entre μ et μ_n tend vers 0 quand n tend vers l'infini.
On peut dès lors apporter une conclusion : la distance de Fortet-Mourier ne peut pas majorer Δ_{max} dans cet exemple, elle ne peut donc pas nous être utile à contrôler l'erreur asymptotiquement par la borne de Pflug. On pourrait cependant se demander maintenant si l'erreur peut être contrôlée directement par la distance de Fortet-Mourier, sans utiliser la majoration de Pflug.
Observons maintenant l'erreur :

$$e(u^*, u_n^*) = \underbrace{\mathbb{E}\left(\varepsilon u_n^{*2}(\boldsymbol{\xi}_1) + \beta(\boldsymbol{\xi}_1 + u_n^*(\boldsymbol{\xi}_1) + \boldsymbol{\xi}_2)^2\right)}_{\to \frac{2\beta}{3}} - \underbrace{\mathbb{E}\left(\varepsilon u^{*2}(\boldsymbol{\xi}_1) + \beta(\boldsymbol{\xi}_1 + u^*(\boldsymbol{\xi}_1) + \boldsymbol{\xi}_2)^2\right)}_{\frac{1}{3}\left(\beta + \frac{\beta\varepsilon}{\varepsilon+\beta}\right)} . \tag{III.9}$$

On a donc d'après (III.9) que $e(u^*, u_n^*) \longrightarrow \frac{1}{3}\left(\beta - \frac{\beta\varepsilon}{\varepsilon+\beta}\right)$. Ainsi, en prenant ε petit, on aura une erreur d'ordre $\frac{\beta}{3}$. Ainsi, il est exclu de pouvoir trouver une constante $M \in \mathbb{R}$ telle que $e(u^*, u_n^*) \leq M d(\mu, \mu_n)$.
Pour compléter l'étude, il nous reste à mettre en évidence que la propriété de Lipschitz des fonctions critères en boucle fermée fait défaut. Cela se voit assez bien par rapport au partitionnement C qui a été fait de $[-1, 1]^2$. Prenons (ξ_1, ξ_2) et (ξ_1', ξ_2'), et regardons :

$$\Delta(\xi_1, \xi_1', \xi_2, \xi_2') = |j(u_n^*(\xi_1), \xi_1, \xi_2) - j(u_n^*(\xi_1'), \xi_1', \xi_2')| \ .$$

Lorsque l'on va écrire la valeur de Δ, une partie des termes sera contrôlée par $|\xi_1 - \xi_1'|$ et $|\xi_2 - \xi_2'|$. Cependant, si l'on note i l'unique indice tel que $1_{C_i}(\xi_1) > 0$ et i' l'unique indice tel que $1_{C_{i'}}(\xi_1') > 0$, on aura aussi des termes en $|\xi_{1,i} - \xi_{1,i'}|$ et $|\xi_{2,i} - \xi_{2,i'}|$. Or, il est tout à fait clair que $|\xi_{1,i} - \xi_{1,i'}|$ est contrôlé par $|\xi_1 - \xi_1'|$, mais il est également tout à fait clair que $|\xi_{2,i} - \xi_{2,i'}|$ est indépendant à la fois de $|\xi_2 - \xi_2'|$ et de $|\xi_1 - \xi_1'|$. On ne pourra pas le rendre arbitrairement petit en approchant (ξ_1, ξ_2) et (ξ_1', ξ_2'), en raison de la partition C intervenant dans u_n^* et qui doit être indépendante de ξ_2. Ainsi, les fonctions critères ne sont pas lipschitziennes, et on ne peut donc pas majorer notre sup par la distance de Fortet-Mourier.
Cependant, conclure que le problème est dans la régularité des fonctions ou dans leur caractère lipschitzien est assez illusoire et masque la difficulté majeure discernée dans l'introduction, et qui concerne la discrétisation de contraintes de mesurabilité. Le problème vient en effet ici plus du fait que nous avons appliqué à la contrainte de boucle fermée la même discrétisation que pour le calcul de l'espérance, rendant ainsi la structure d'information inexistante.

III.2.5. Méthode alternative. Comme on l'a discerné en introduction, *les termes d'aléas* apparaissent dans un problème d'optimisation stochastique à deux endroits différents. L'exemple développé dans le paragraphe précédent montre qu'un traitement conjoint de l'aléas dans ces deux endroits mène à la mise en œuvre de solutions compliquées et qui plus est fausses en boucle fermée malgré leur justesse en boucle ouverte. En effet, la distance de Fortet-Mourier invoquée par Pflug est tout à fait inopérante dans le cadre pourtant très simple de l'exemple cité avant.
Pour tenter de mieux saisir ce qui se cache derrière l'échec de l'exemple et qui va bien au-delà des fonctions lipschitziennes, on se propose ici de reposer le problème dans un cadre où la discrétisation se fera en deux temps indépendants dans l'espérance et dans les contraintes de mesurabilité.

On dispose toujours de l'espace Ξ, muni d'une métrique c, compact et séparable, et d'une suite de lois $(\mu_n)_{n\in\mathbb{N}}$, chacune des lois ayant un support noté $\Xi_n = supp(\mu_n)$, avec $\mathcal{F}_n$ la tribu associée. On notera $\boldsymbol{\xi}_n$ une variable aléatoire à valeurs dans Ξ de loi μ_n.

On utilise le même formalisme d'espace admissible U^f, et Γ, U_n^f et Γ_n, dans le cadre des intégrandes normales. La nouveauté est ici d'introduire parallèlement une suite d'espaces admissibles $(U_n^{f,K})_{K\in\mathbb{N},\, n\in\mathbb{N}}$, tels que :

$$\begin{aligned}\forall K \in \mathbb{N}, &\exists (A_k^K)_{1\le k\le K} \text{ partition de } \Xi \text{ telle que,}\\ &\forall n\in\mathbb{N},\ U_n^{f,K} \subseteq \{u:\Xi_n\to U\ :\ \exists (u_k)_{1\le k\le K}\in U,\\ &\qquad\qquad u(\cdot)=\sum_{k=1}^K u_k 1_{A_k}(\cdot),\ \text{et } j(u(\cdot),\cdot)\in L^1(\Xi_n,\mathcal{F}_n,\mu_n)\Big\}\ .\end{aligned}$$

En d'autres termes, ces $U_n^{f,K}$ quantifient les contraintes sur la commande en prenant une commande égale à une fonction étagée. De façon analogue, on se donne des espaces $(U^{f,K})_{K\in\mathbb{N}}$ tels que pour tout $K\in\mathbb{N}$:

$$U^{f,K} \subseteq \left\{u:\Xi\to U\ :\ \exists (u_k)_{1\le k\le K}\in U,\ u(\cdot)=\sum_{k=1}^K u_k 1_{A_k}(\cdot),\ \text{et } j(u(\cdot),\cdot)\in L^1(\Xi,\mathcal{F},\mu)\right\}\ .$$

Une telle définition permet de conserver la propriété d'inclusion $U^{f,K}\subset U^f$.

Enfin, on se donne des opérateurs de prolongement pour tout $n\in\mathbb{N}$, notés $\phi_n:\Xi\to\Xi_n$ tels que :

– $\forall K\in\mathbb{N},\ \forall u\in U^{f,K},\ u_{|\Xi_n}\in U_n^{f,K}$,
– $\forall u_n\in U_n^{f,K}$ alors $u_n\circ\phi_n\in U^{f,K}$ et $u_n\circ\phi_n|_{\Xi_n}=u_n$.

On utilise ensuite les notations :

$$\forall n\in\mathbb{N},\forall u\in U_n^f,\ J_n(u)=\mathbb{E}\left(j(u_n(\boldsymbol{\xi}_n),\boldsymbol{\xi}_n)\ \text{et}\ ,\right.$$

$$\forall K\in\mathbb{N},\ \forall n\in\mathbb{N},\ \forall u\in U_n^{f,K},\ J_n^K(u)=\sum_{k=1}^K \mathbb{E}\left(j(u_k,\boldsymbol{\xi}_n)1_{A_k}(\boldsymbol{\xi}_n)\right)\ .$$

On a donc ici tenu compte à la fois de la discrétisation de l'espérance, en $n\in\mathbb{N}$, et de la discrétisation de la contrainte de mesurabilité, en $K\in\mathbb{N}$.
Avant d'en finir avec les notations, on introduit pour des raisons de légèreté :

$$V(\nu,X)=\inf_{u\in X}\int_\Xi j(u(\xi),\xi)\nu(d\xi),\quad \text{pour toute loi } \nu \text{ sur } \Xi, \text{ et tout ensemble } X \text{ admissible.}$$

L'erreur considérée par Pflug s'écrit désormais comme :

$$e(u^*,u_n^{*,K})=|V(\mu,U^f)-J(u_n^{*,K})|\ .$$

On peut donc la majorer de la façon suivante :

$$\begin{aligned}e(u^*,u_n^{*,K}) \le &|V(\mu,U^f)-V(\mu,U^{f,K})| \text{ (erreur de quantification),}\\ &+|V(\mu,U^{f,K})-V(\mu_n,U_n^{f,K})| \text{ (erreur de quadrature),}\\ &+|V(\mu_n,U_n^{f,K})-J(u_n^{*,K})| \text{ (erreur de prolongement).}\end{aligned}$$

Cette décomposition fait apparaître plusieurs choses :

– L'utilisation d'une distance de Fortet-Mourier pourrait se justifier pour mesurer l'erreur de quadrature, mais ici, une simple loi forte des grands nombres devrait suffire, à K fixé quand n tend vers l'infini, si l'on prend pour les μ_n les lois empiriques.
– L'erreur de quantification, isolée de toute discrétisation de l'espérance, pourra tendre vers 0 pour un peu que les quantifications successives tendent vers l'identité quand K tend vers l'infini.

– L'erreur de prolongement va tendre quant à elle vers 0 pour des raisons encore une fois de loi forte des grands nombres quand n tend vers l'infini à K fixé, si l'on a pris comme dit avant les discrétisations en n et K.

En faisant tendre d'abord n, puis K vers l'infini, on pourra donc rendre l'erreur de Pflug aussi petite que l'on voudra, sous certaines hypothèse assurément moins fortes que l'hypothèse de Lipschitz dont il a été question un peu avant. Pour d'autres développements sur la question, on pourra se référer à la thèse [**8**] qui s'occupe entre autres de ce problème, et est à l'origine de ces réflexions.

Enfin, il est très facile de voir que si l'on procède ainsi, en quantifiant d'abord l'espace, puis en tirant des trajectoires, l'exemple choisi pour montrer la vanité de la distance de Fortet-Mourier verra bien son erreur tendre vers 0.

III.2.6. Conclusion. Ne pas isoler les sources potentielles d'erreur dans la discrétisation de problèmes d'optimisation stochastique peut donc conduire à des raisonnements incomplets dont la mise en oeuvre pratique s'avère même fausse, comme dans le cas de l'utilisation de distances de Fortet-Mourier pour mesurer l'erreur commise par approximation de l'espérance et de la contrainte de mesurabilité dans un problème d'optimisation stochastique en boucle fermée.

L'approche au contraire évitant ces désagréments, et ne nécessitant par ailleurs que des hypothèses moins fortes sur le critère, consiste en une séparation des termes touchant à la mesurabilité, et des termes touchant aux variables aléatoires. L'idée est bien entendu ensuite de donner des discrétisations certes indépendantes, mais cohérentes pour chacun de ces types de termes.

Les résultats qui permettent de conclure asymptotiquement en la validité de l'approche consistent d'une part en des applications de la loi forte des grands nombres, et d'autre part en des résultats de convergence sur les tribus, comme ceux dont il sera question dans la section suivante.

III.3. Stabilité des problèmes stochastiques linéaires

III.3.1. Motivation. La section précédente a montré qu'un résultat de stabilité basé uniquement sur des considérations de convergence en loi des variables aléatoires sous-jacentes ne permettait pas de rendre compte correctement du comportement des problèmes d'optimisation stochastique sous contraintes de mesurabilité. Si le cas général semble difficile, on peut néanmoins espérer obtenir dans certains cas un résultat de stabilité du type de la proposition III.3 dans lequel le majorant comporterait un terme propre à gouverner l'erreur en loi, et un terme propre à gouverner l'erreur en mesurabilité.

Dans la continuité de [**51**], nous sommes parvenus en collaboration avec H. Heitsch et W. Römisch de l'université Humboldt de Berlin, à énoncer un résultat de stabilité dans le cas de problèmes stochastiques linéaires, basé sur des hypothèses d'ensembles de niveaux localement bornés. Ce résultat est énoncé dans la sous-section III.3.3.

Dans la thèse [**8**], le théorème IV.14 donne un résultat de ce type, dans le cadre des problèmes avec contrainte d'information à deux pas de temps, avec une fonction de coût s'exprimant comme le produit d'un terme d'aléa et d'un terme de commande. Dans la continuité de ce résultat, nous proposons un autre résultat de stabilité pour le cas de problèmes stochastiques à critère séparant aléa et commande. Ce résultat, énoncé dans la sous-section III.3.4 peut être vu comme une généralisation de la sous-section III.3.3.

Le problème d'optimisation (I.8) peut s'écrire de façon légèrement différente sous la forme :

$$v(\boldsymbol{\xi}) := \min_{\boldsymbol{x}} F(\boldsymbol{\xi}, \boldsymbol{x}) := \mathbb{E}\left(\sum_{t=1}^{T} L_t(\boldsymbol{x}^t, \boldsymbol{\xi}^t)\right)$$

$$\forall 1 \leq t \leq T,\ \boldsymbol{x}_t \text{ est } \sigma(\boldsymbol{\xi}^t) - \text{mesurable}, \tag{III.10a}$$

$$\boldsymbol{x}_t \in X_t,\ \text{p.s.} \tag{III.10b}$$

$$\boldsymbol{x} \in \mathcal{X},\ \text{p.s.}, \tag{III.10c}$$

où l'on cherche désormais le contrôle non comme une fonction, mais directement comme une variable aléatoire. Bien entendu, entre la notation fonctionnelle (u_t) et la notation aléatoire $(\boldsymbol{x}_t)$, il y a le lien suivant pour tout t, à savoir $u_t(\boldsymbol{\xi}) = \boldsymbol{x}_t$. Ce type de formalisme est classiquement employé dans la communauté *stochastic programming* (http ://www.stoprog.org), c'est la raison pour laquelle nous énoncerons nos résultats de stabilité dans ce cadre. Ce formalisme conduit à rechercher le contrôle $\boldsymbol{x} = (\boldsymbol{x}_t)_{1\leq t\leq T}$ comme un élement d'un espace de Banach du type $L^{r'}(\Omega, \mathbb{R}^m, \mathbb{P})$. On notera $\boldsymbol{x} \in \mathcal{X}(\boldsymbol{\xi})$ lorsque $\boldsymbol{x}$ vérifiera les contraintes (III.10a)–(III.10b)–(III.10c).

Les contraintes (III.10b) sont les contraintes de borne sur le contrôle instantané, les contraintes (III.10c) sont les éventuelles contraintes dynamiques sur le contrôle, et les contraintes (III.10a) sont les contraintes de mesurabilité.

Comme la section précédente l'a illustré, un résultat de stabilité pour les problèmes stochastiques comme (III.10) nécessite a priori l'introduction de termes aptes à mesurer les variations de l'information. Bref, de termes fondés sur des distances entre les tribus et les filtrations impliquées dans la contrainte de mesurabilité (III.10a). La sous-section III.3.2 s'applique donc, dans le sillage de [**8**], à déterminer les outils nécessaires.

III.3.2. Distances entre tribus et filtrations. U prérequis pour la convergence de schémas de discrétisation, ou l'obtention de résultats de stabilité est une notion de convergence ou de métrique de convergence sur l'espace des tribus et des filtrations. Nous allons dans cette sous-section donner quelques outils mathématiques pour cela. Cette sous-section est donc à voir comme une revue de la littérature existante utilisable pour bâtir des notions de convergence appropriées à l'étude de la stabilité des problèmes d'optimisation stochastique.

En suivant le travail de [**8**], et en s'appuyant sur les travaux [**34, 62**], on peut définir une topologie sur l'ensemble des filtrations sur Ω à partir d'une topologie sur l'ensemble des tribus sur Ω. En effet, on dira qu'une suite de filtrations $(\mathcal{F}_t^n)_{1\leq t\leq T}$ converge vers une filtration $(\mathcal{F}_t)_{1\leq t\leq T}$ si et seulement si pour tout $t \in \{1, \ldots, T\}$, la suite de tribus $(\mathcal{F}_t^n)$ converge vers la tribu $\mathcal{F}_t$.

On peut donc se restreindre à l'étude des topologies sur l'ensemble des tribus. On distingue principalement deux topologies, une topologie dite uniforme, et une topologie dite forte. Pour commencer, nous allons montrer comment construire ces deux topologies, puis nous les métriserons et donnerons en les comparant quelques propriétés à leur sujet.

III.3.2.1. *Topologies, construction générale.* On rappelle que $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de probabilité séparable et métrique, de métrique notée c. On note $\mathcal{F}^{**}$ l'ensemble des sous-tribus de $\mathcal{F}$.

Définition III.5. *On définit l'opérateur* $\vee : \mathcal{F}^{**} \times \mathcal{F}^{**} \to \mathcal{F}^{**}$ *dit de jointure par :*

$$\forall \mathcal{B}_1, \mathcal{B}_2 \in \mathcal{F}^{**},\ \mathcal{B}_1 \vee \mathcal{B}_2 = \sigma\left(\mathcal{B}_1 \cup \mathcal{B}_2\right)\ .$$

Définition III.6. *Soit* $\mathcal{B}_1, \mathcal{B}_2 \in \mathcal{F}^{**}$*, on dira que* $\mathcal{B}_1$ *et* $\mathcal{B}_2$ *sont équivalentes et on notera* $\mathcal{B}_1 \sim_{\mathbb{P}} \mathcal{B}_2$ *si* $\mathcal{B}_1$ *et* $\mathcal{B}_2$ *diffèrent d'un ensemble d'ensembles de mesure* $\mathbb{P}$*-négligeable.*

Munis de cette relation d'équivalence, on peut alors définir $\mathcal{F}^*$ comme étant l'ensemble quotient de $\mathcal{F}^{**}$ par $\sim_{\mathbb{P}}$, ie l'ensemble des classes d'équivalence pour $\sim_{\mathbb{P}}$ de sous-tribus de $\mathcal{F}$. Pour $\mathcal{B} \in \mathcal{F}^{**}$, on notera par abus $\mathcal{B} \in \mathcal{F}^*$ la classe d'équivalence correspondante.

On note $\mathcal{V}$ l'ensemble des variables aléatoires réelles sur $(\Omega, \mathcal{F}, \mathbb{P})$. On équipe $\mathcal{V}$ de la métrique de la convergence en probabilité, i.e. :

$$\forall f, g \in \mathcal{V},\ \theta(f, g) = \inf\{\varepsilon\ :\ \varepsilon > 0,\ \mathbb{P}(|f - g| > \varepsilon) < \varepsilon\}\ .$$

On définit de facon standard la tribu engendrée par une variable aléatoire $f \in \mathcal{V}$ par :

$$\sigma(f) = \sigma\left\{\mathcal{B} \in \mathcal{F}\ :\ \exists B \in \mathcal{B}(\mathbb{R}),\ \mathcal{B} = f^{-1}(B)\right\}\ .$$

On note $L_1 = L^1(\Omega, \mathcal{F}, \mathbb{P})$ l'espace vectoriel des classes d'équivalence des fonctions de Ω dans $\mathbb{R}$ boréliennes intégrables, muni de la norme $||\cdot||$ définie par :

$$\forall f \in L_1,\ ||f|| = \int_\Omega |f(\omega)| d\mathbb{P}(\omega)\ .$$

On note alors $L_1^b \subset L_1$ l'ensemble des fonctions bornées de L_1. L_1 et L_1^b sont donc des espaces de Banach, et on a la propriété suivante :

Proposition III.7.

$$\forall \mathcal{B}_1, \mathcal{B}_2 \in \mathcal{F}^{**},\ (\mathbb{E}(f|\mathcal{B}_1) = \mathbb{E}(f|\mathcal{B}_2)\ p.s.\ \forall f \in L_1) \Leftrightarrow \mathcal{B}_1 \sim_{\mathbb{P}} \mathcal{B}_2\ .$$

Preuve : $\Leftarrow$: trivial avec la définition de l'espérance conditionnelle.
$\Rightarrow$: On veut montrer que $\mathcal{B}_1 \sim_{\mathbb{P}} \mathcal{B}_2$, i.e., que :

$$\forall B_2 \in \mathcal{B}_2,\ \exists C \in \mathcal{F}\ \exists B_1 \in \mathcal{B}_1,\ \mathbb{P}(C) = 0,\ B_1 \cap C = \emptyset,\ B_2 = B_1 \cup C\ .$$

On prend donc $B_2 \in \mathcal{B}_2$, et on pose $B_1 = \{\mathbb{E}(1_{B2}|\mathcal{B}_1) > 0\} \in \mathcal{B}_1$, et $C = B_2 \setminus B_1$. Il reste à montrer que $\mathbb{P}(C) = 0$, ie $\mathbb{P}(B_2 \setminus B_1) = \mathbb{P}(B_2 \cap B_1^c) = 0$.

$$\mathbb{P}(B_2 \cap B_1^c) = \mathbb{E}\left(1_{B_2} 1_{B_1^c}\right) = \mathbb{E}\left(\mathbb{E}\left(1_{B2} 1_{\mathbb{E}(1_{B2}|\mathcal{B}_1)=0}|\mathcal{B}_2\right)\right) = \mathbb{E}\left(\mathbb{E}\left(1_{B2} 1_{(\mathbb{E}(1_{B2}|\mathcal{B}_1)=0)}|\mathcal{B}_1\right)\right) = 0\ ,$$

par hypothèse. □

Par conséquent, on pourra écrire sans confusion $\mathbb{E}(\cdot|\mathcal{B})$ pour $\mathcal{B} \in \mathcal{F}^*$.

Nous souhaitons maintenant étudier l'espace $\mathcal{F}^*$ d'un point de vue topologique. Pour introduire les opérateurs de topologie habituels, considérons l'espace $\mathcal{L}(L_1^b) = \left\{T : L_1^b \to L_1^b\ :\ T \text{ linéaire continue }\right\}$. Par conséquent, pour tout $\mathcal{B} \in \mathcal{F}^*$, $\mathbb{E}(\cdot|\mathcal{B}) \in \mathcal{L}(L_1^b)$. Or, il est clair que l'opérateur $(\mathcal{B} \mapsto \mathbb{E}(\cdot|\mathcal{B}))$ est injectif. Par conséquent, on peut voir $\mathcal{F}^*$ comme un sous-ensemble de $\mathcal{L}(L_1^b)$, ce qui signifie que toute topologie sur $\mathcal{L}(L_1^b)$ induit une topologie sur $\mathcal{F}^*$.
Il est naturel d'équiper $\mathcal{L}(L_1^b)$ de l'opérateur de topologie forte s et de l'opérateur de topologie uniforme u. On pose $\Phi = \left\{f \in L_1^b\ :\ |f(\omega)| \leq 1 \text{ pour presque tout } \omega \in \Omega\right\}$. Soit $(T_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\mathcal{L}(L_1^b)$ et $T \in \mathcal{L}(L_1^b)$. Alors :

$$T_n \xrightarrow{s} T\ \Leftrightarrow \forall f \in L_1^b,\ \lim_{n\to\infty} ||T_n(f) - T(f)|| = 0\ ; \tag{III.11}$$

$$T_n \xrightarrow{u} T\ \Leftrightarrow \lim_{n\to\infty} \sup_{f\in\Phi} ||T_n(f) - T(f)|| = 0\ . \tag{III.12}$$

La question est maintenant d'étudier ces deux topologies induites sur $\mathcal{F}^*$, et d'essayer de les métriser. C'est ce à quoi nous allons nous employer dans la suite. On écrit les définitions de convergence dans $\mathcal{F}^*$ induites par (III.11) et (III.12). On se donne une suite de tribus $(\mathcal{B}_n)_{n\geq 1}$ et $\mathcal{B}_0 \in \mathcal{F}^*$. Alors :

$$\mathcal{B}_n \xrightarrow{s} \mathcal{B}_0 \Leftrightarrow \forall f \in L_1^b,\ \lim_{n\to\infty} \int_\Omega |\mathbb{E}(f|\mathcal{B}_n) - \mathbb{E}(f|\mathcal{B}_0)| d\mathbb{P} = 0\ ; \tag{III.13}$$

$$\mathcal{B}_n \xrightarrow{u} \mathcal{B}_0 \Leftrightarrow \lim_{n\to\infty} \sup_{f\in\Phi} \int_\Omega |\mathbb{E}(f|\mathcal{B}_n) - \mathbb{E}(f|\mathcal{B}_0)| d\mathbb{P} = 0\ . \tag{III.14}$$

III.3.2.2. *Topologie de convergence uniforme.* La topologie de convergence uniforme dans l'espace des tribus a été étudiée par Boylan ([**24**]), Neveu ([**68**]), Rogge ([**85**]), Allen([**1**]) et Cotter ([**34, 35**]). Chacun a apporté sa contribution à l'étude de cette notion de convergence, en l'attaquant sous des angles assez différents. Ainsi, Boylan et Neveu, en cherchant à caractériser la convergence uniforme de martingales, ont exhibé une distance sur $\mathcal{F}^*$ qui s'est avérée métriser comme l'a montré Rogge, la topologie uniforme sur $\mathcal{F}^*$, tandis qu'Allen et Cotter se sont intéressés directement à la notion de distance sur $\mathcal{F}^*$ pour en déduire des résultats de continuité entre des problèmes de minimisation sous contraintes de mesurabilité de la commande.

Il y a deux résultats importants pour la topologie uniforme sur $\mathcal{F}^*$:
– elle est métrisable, et on connaît la métrique associée,
– elle permet d'étudier la continuité d'opérateurs sur des tribus.

Commençons par définir la distance adéquate : Pour deux parties A et B de Ω, on définit $A\Delta B = (A \setminus B) \cup (B \setminus A)$, la différence symétrique entre A et B.

DÉFINITION III.8. *On définit l'opérateur $d : \mathcal{F}^{**} \times \mathcal{F}^{**} \to \mathbb{R}^+$ par :*

$$\forall \mathcal{B}_1, \mathcal{B}_2 \in \mathcal{F}^{**},\ d(\mathcal{B}_1, \mathcal{B}_2) = \sup_{B_1 \in \mathcal{B}_1} \inf_{B_2 \in \mathcal{B}_2} \mathbb{P}(B_1 \Delta B_2) + \sup_{B_2 \in \mathcal{B}_2} \inf_{B_1 \in \mathcal{B}_1} \mathbb{P}(B_1 \Delta B_2). \tag{III.15}$$

On a alors

THÉORÈME III.9 (Boylan). *d est une pseudo-distance sur $\mathcal{F}^{**}$. De plus,*

$$\forall \mathcal{B}_1, \mathcal{B}_2 \in \mathcal{F}^{**},\ d(\mathcal{B}_1, \mathcal{B}_2) = 0 \Leftrightarrow (\forall f \in L_1,\ \mathbb{E}(f|\mathcal{B}_1) = \mathbb{E}(f|\mathcal{B}_2)\ \text{p.s.}).$$

d est donc une distance sur $\mathcal{F}^*$.

Preuve : cf. [**24**]. □

PROPOSITION III.10 (Allen). *On a les deux propriétés suivantes :*

(i) Si $\mathbb{P}$ est purement atomique, $(\mathcal{F}^*, d)$ est un espace compact séparable.

(ii) Si $\Omega = [0,1]$ et $\mathcal{F} = \mathcal{B}([0,1])$, et $\mathbb{P}$ est la mesure de Lebesgue, alors l'ensemble des partitions finies de Ω n'est pas dense dans $\mathcal{F}^*$.

Preuve : cf. [**1**]. □

Dans l'article [**85**], une autre distance équivalente à la distance de Boylan est introduite :

PROPOSITION III.11 (Rogge). *Soient $\mathcal{B}_1, \mathcal{B}_2 \in \mathcal{F}^*$, on définit la distance de Rogge par :*

$$\delta(\mathcal{B}_1, \mathcal{B}_2) = \max\left(\sup_{B_1 \in \mathcal{B}_1} \inf_{B_2 \in \mathcal{B}_2} \mathbb{P}(B_1 \Delta B_2), \sup_{B_2 \in \mathcal{B}_2} \inf_{B_1 \in \mathcal{B}_1} \mathbb{P}(B_1 \Delta B_2 \right).$$

δ est une distance sur $\mathcal{F}^*$ équivalente à la distance de Boylan d.

Preuve : cf. [**85**]. □

Nous allons maintenant donner les résultats montrant que ces deux distances d et δ métrisent bien la topologie de convergence uniforme introduite plus haut :

LEMME III.12 (Rogge). *Soient $\mathcal{B}_1$ et $\mathcal{B}_2$ deux sous-tribus de $\mathcal{F}$. Soit $f : \Omega \to [0,1]$ une fonction $\mathcal{B}_2$-mesurable. Alors :*

(i) $\mathbb{E}(|\mathbb{E}(f|\mathcal{B}_1) - f|) \leq 2\delta(\mathcal{B}_1, \mathcal{B}_2)(1 - \delta(\mathcal{B}_1, \mathcal{B}_2))$,

(ii) $\left\{\mathbb{E}\left(|\mathbb{E}(f|\mathcal{B}_1) - f|^2\right)\right\}^{\frac{1}{2}} \leq \{\delta(\mathcal{B}_1, \mathcal{B}_2)(1 - \delta(\mathcal{B}_1, \mathcal{B}_2))\}^{\frac{1}{2}}$,

(iii) $\delta(\mathcal{B}_1, \mathcal{B}_2) \leq \sup_{g \in \Phi} \|\mathbb{E}(g|\mathcal{B}_1) - \mathbb{E}(g|\mathcal{B}_2)\|$.

Preuve : cf. [**85**]. □

Enfin, on a le théorème général suivant :

THÉORÈME III.13 (Rogge). *Soit $H \subset L^q(\Omega, \mathcal{F}, \mathbb{P})$, soit $a > 0$. On définit alors :*

$$\delta_{H,q}(a) = \sup_{f \in H} \left\{ \left(\mathbb{E}\left(|f 1_{\{|f| > a\}}|^q\right)\right)^{\frac{1}{q}} \right\}.$$

Alors, pour tous $\mathcal{B}_1, \mathcal{B}_2 \in \mathcal{F}^*$:

$$\sup_{f \in H} \left\{ \left[\mathbb{E}\left(|\mathbb{E}(f|\mathcal{B}_1) - \mathbb{E}(f|\mathcal{B}_2)|^q\right)\right]^{\frac{1}{q}} \right\} \leq C_q a \left[\delta(\mathcal{B}_1, \mathcal{B}_2)(1 - \delta(\mathcal{B}_1, \mathcal{B}_2))\right]^{\frac{1}{q}} + 2\delta_{H,q}(a), \tag{III.16}$$

où :
(i) $C_q = 2^{1+\frac{1}{q}}$ si $\mathcal{B}_1 \subset \mathcal{B}_2$ et $1 \leq q < 2$,
(ii) $C_q = 2$ si $\mathcal{B}_1 \subset \mathcal{B}_2$ et $q \geq 2$,
(iii) $C_q = 2^{1+\frac{1}{q}}$ si $\mathcal{B}_1$ et $\mathcal{B}_2$ sont arbitraires et $q \geq 2$.

En résumé, d et δ métrisent donc la topologie de convergence uniforme sur $\mathcal{F}^*$. Néanmoins, cette topologie n'a pas toutes les qualités. En effet, les partitions finies ne sont pas denses pour cette topologie dans l'ensemble des tribus. Or, du point de vue de l'approximation de tribus, il est naturel de songer aux partitions finies. L'absence de densité les écarte cependant si l'on se restreint à cette topologie. De plus, l'opérateur σ qui engendre des tribus à partir de variables aléatoires n'est pas continu pour cette topologie. Cette topologie est donc a priori trop forte pour convenir à la discrétisation des problèmes d'optimisation stochastique.

III.3.2.3. *Topologie de convergence forte.* Pour les raisons négatives évoquées avant, on se concentre donc sur une topologie plus faible sur $\mathcal{F}^*$, la topologie de convergence forte.
Pour commencer, introduisons la distance. On a la théorème et la définition suivants :

Théorème III.14 (Cotter). *Si L_1 est séparable, et si (f_j) est une famille dénombrable dense dans L_1, alors la métrique ρ définie ci-dessous génère la topologie de convergence forte sur $\mathcal{F}^*$. De plus, deux métriques utilisant des familles denses dans L_1 différentes sont uniformément équivalentes.*

$$\rho(\mathcal{B},\mathcal{B}') = \sum_{j=1}^{\infty} \frac{1}{2^j} \min\left(\|\mathbb{E}\left(f_j|\mathcal{B}\right) - \mathbb{E}\left(f_j|\mathcal{B}'\right)\|, 1\right) \quad \forall \mathcal{B},\mathcal{B}' \in \mathcal{F}^* \ .$$

Dans ce cas, $(\mathcal{F}^*, \rho)$ est un espace métrique complet séparable.

Preuve : cf. [**34**]. □

Proposition III.15. *Soit $\mathcal{B} \in \mathcal{F}^*$, $\varepsilon > 0$. Il existe une partition finie mesurable $\mathcal{B}' \subset \mathcal{B}$ telle que $\rho(\mathcal{B}', \mathcal{B}) < \varepsilon$.*

En d'autres termes, l'ensemble des partitions finies mesurables de Ω est dense dans $\mathcal{F}^*$ pour la distance de Cotter.

Dans un travail antérieur, on trouve cette équivalence entre divers concepts de convergence du type convergence ponctuelle :

Théorème III.16 (Kudo). *Soit $(\mathcal{B}_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\mathcal{F}^*$.*
Les trois propositions sont équivalentes :

(i) $\mathbb{E}\left(1_A|\mathcal{B}_n\right) \xrightarrow{\mathbb{P}} \mathbb{E}\left(1_A|\mathcal{B}_0\right)$ *quand* $n \to \infty \quad \forall A \in \mathcal{F}$.
(ii) $\lim_{n\to\infty} \int_\Omega |\mathbb{E}\left(f|\mathcal{B}_n\right)| d\mathbb{P} = \int_\Omega |\mathbb{E}\left(f|\mathcal{B}_0\right)| d\mathbb{P} \quad \forall f \in L_1^b$.
(iii) $\forall u \in (0,1),\ \forall A \in \mathcal{F},\ \int_\Omega |u - \mathbb{E}\left(1_A|\mathcal{B}_n\right)| d\mathbb{P} \longrightarrow \int_\Omega |u - \mathbb{E}\left(1_A|\mathcal{B}_0\right)| d\mathbb{P}$ *quand* $n \to \infty$.

Preuve : cf. [**64**]. □

Cotter nous permet maintenant de relier cette notion de convergence ponctuelle à la convergence forte métrisée par la métrique ρ :

Proposition III.17 (Cotter). *$(\mathcal{B}_n)$ converge au sens de ρ (fortement) si l'on a l'une des conditions suivantes :*
(i) $\forall A \in \mathcal{F}$, $(\mathbb{E}\left(1_A|\mathcal{B}_n\right))$ converge en probabilité ou en norme L_1,
(ii) $\forall f$ dans un ensemble dense de L_1, $(\mathbb{E}\left(f|\mathcal{B}_n\right))$ converge en probabilité ou en norme L^p, pour un certain $p \geq 1$,
(iii) $\liminf_{n\to\infty} \mathcal{B}_n := \bigvee_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \mathcal{B}_n = \bigcap_{m=1}^{\infty} \bigvee_{n=m}^{\infty} \mathcal{B}_n =: \limsup_{n\to\infty} \mathcal{B}_n$.

On s'intéresse désormais aux propriétés de continuité dans cette topologie forte :

Théorème III.18 (Cotter). *Soit $\mathcal{A}$ une partition finie de Ω. Alors l'application $\mathcal{B} \longmapsto \mathcal{A} \vee \mathcal{B}$ est continue sur $\mathcal{F}^*$.*

Preuve : cf [**34**], Théorème 3.3. □

On continue avec un résultat négatif pour la distance de Cotter :

EXEMPLE III.19. *L'opérateur de jointure n'est pas continu pour la distance de Cotter : On pose Ω la boule unité de $\mathbb{R}^2$ munie de ses boréliens et de la mesure uniforme de Lebesgue. On pose les applications : $h_0(x,y) = y$ et $h_n(x,y) = y + \frac{x}{n}$, et $\mathcal{B}_0 = \sigma(h_0)$, $\mathcal{B}_n = \sigma(h_n)$ les tribus associées. Le couple $(\mathcal{B}_0, \mathcal{B}_n)$ partitionne l'espace Ω. De plus, pour $f \in L_1$: $\mathbb{E}(f|\mathcal{B}_0)(x,y) = \int_0^1 f(t,y)dt$ et $\mathbb{E}(f|\mathcal{B}_n)(x,y) = \int_0^1 f(t, y + \frac{x-t}{n})dt$. Donc, si f est continue, on peut appliquer le théorème de convergence dominée, et on a la convergence simple : $\mathbb{E}(f|\mathcal{B}_n) \to \mathbb{E}(f|\mathcal{B}_0)$ puis à nouveau par convergence dominée la même convergence en norme L_1. Par densité de l'ensemble des fonctions continues sur Ω dans L_1, avec le théorème de caractérisation de la convergence ponctuelle, on a $\mathcal{B}_n \to \mathcal{B}_0$. Cependant $\mathcal{B}_n \vee \mathcal{B}_0 = \mathcal{F}$ pour tout n, d'où la non-continuité de l'opérateur de jointure, car $\mathcal{B}_n \vee \mathcal{B}_0$ ne converge pas vers $\mathcal{B}_0$.*

THÉORÈME III.20 (Cotter). *σ est continue pour ρ en $f \in \mathcal{V}$ si et seulement si $\sigma(f) = \mathcal{F}$. On dira alors que f est d'*information complète.

Preuve : cf. [**34**]. □

Ces résultats sont donc a priori plutôt négatifs, car on ne parvient pas à caractériser la convergence des tribus engendrées lorsque la famille de v.a. sous-jacente ne converge pas vers une v.a. d'information complète. Cependant, on trouve dans [**8**] le résultat suivant :

PROPOSITION III.21 (Barty). *Soit $\mathcal{G} \in \mathcal{F}^*$ et $(f_j)_{j\in\mathbb{N}}$ une famille dense de fonctions de $L^1(\Omega, \mathcal{F}, \mathbb{P})$. Alors la famille de fonctions $(g_j := \mathbb{E}(f_j|\mathcal{G}))_{j\in\mathbb{N}}$ est dense dans $L^1(\Omega, \mathcal{G}, \mathbb{P})$, et en posant $\rho_{\mathcal{F},f}$ et $\rho_{\mathcal{G},g}$ les distances de Cotter associées, on a :*

$$\forall \mathcal{B}_1, \mathcal{B}_2 \in \mathcal{F}^*,\ \mathcal{B}_1, \mathcal{B}_2 \subset \mathcal{G} \ \Rightarrow \rho_{\mathcal{F},f}(\mathcal{B}_1, \mathcal{B}_2) = \rho_{\mathcal{G},g}(\mathcal{B}_1, \mathcal{B}_2)\ .$$

Preuve : [**8**]. □

De la proposition III.21 et de la proposition III.20, on déduit le corollaire suivant :

COROLLAIRE III.22 (Barty). *Soit $(h_n) \in \mathcal{V}$ une suite qui converge en probabilité vers $h \in \mathcal{V}$. Si pour tout $n \in \mathbb{N}, \sigma(h_n) \subset \sigma(h)$, alors $(\sigma(h_n))$ converge fortement vers $\sigma(h)$.*

Preuve : cf. [**8**]. □

Ces dernières propriétés de continuité pour la topologie de convergence forte en font donc une candidate intéressante pour les discrétisations. Le théorème III.23 dû à Barty achève d'illustrer l'intérêt de cette topologie pour les problèmes qui nous préoccupent :

THÉORÈME III.23 (Barty). *Soit $\mathcal{X}$ un ensemble convexe fermé de $\mathbb{R}^p$, Ξ un espace métrique de dimension finie, et $\boldsymbol{\xi}$ une variable aléatoire à valeurs dans Ξ. Soit $j : \mathcal{U} \times \Xi \to \mathbb{R}$ une intégrande normale, telle qu'il existe $h : \Xi \to \mathbb{R}$ tel que d'une part $\mathbb{E}(|h(\boldsymbol{\xi})|) < +\infty$ et :*

$$h(\boldsymbol{\xi}) \leq j(x, \boldsymbol{\xi}) \text{ p.s. } \forall x \in \mathcal{X}.$$

Soit $(\mathcal{B}_n)_{n\in\mathbb{N}}$ une suite de sous-tribus de $\mathcal{F}$, et $\mathcal{B} \in \mathcal{F}^{**}$, telles que $\mathcal{B}_n \xrightarrow{s} \mathcal{B}$ quand $n \to \infty$ et $\mathcal{B}_n \subset \mathcal{B}$ pour tout $n \in \mathbb{N}$. On définit alors :

$$V(\mathbb{P}, \mathcal{B}) = \min_{\boldsymbol{x}:\Omega\to\mathcal{U},\ \boldsymbol{x}\mathcal{B}-mes} J(\boldsymbol{x}) = \mathbb{E}(j(\boldsymbol{x}, \boldsymbol{\xi}))\ .$$

Si J est continue sur l'ensemble des v.a. mesurables munies de la convergence en probabilité, alors on a :

$$\lim_{n\to\infty} |V(\mathbb{P}, \mathcal{B}_n) - V(\mathbb{P}, \mathcal{B})| = 0.$$

Preuve : cf. [**8**]. □

Espaces de tribus	$\mathcal{F}^{**}/\sim_{\mathbb{P}}=\mathcal{F}^*$, $\mathcal{F}^* \subset \mathcal{L}(L_1) := \{T : L_1 \to L_1 \ : \ T \text{ linéaire continu}\}$, car $(\mathcal{B} \in \mathcal{F}^* \mapsto \mathbb{E}(\cdot\|\mathcal{B}) \in \mathcal{L}(L_1))$ est injectif.	
Topologies	convergence uniforme	convergence forte
Travail	Rogge ([**85**]) sur les inégalités, Boylan ([**24**]) sur les martingales.	Kudo ([**64**]) sur les $\mathbb{P}-\lim \begin{pmatrix} \sup \\ \inf \end{pmatrix}$ Cotter ([**34, 35**]).
Distances sur $\mathcal{F}^*$	Boylan, d, Rogge, δ.	Cotter, ρ.
Propriétés	Densité des partitions finies dans $\mathcal{F}^*$	
	Non.	Oui.
	Continuité de l'opérateur σ :	
	Non.	En tout point d'information complète.

TAB. 1. Résumé

III.3.2.4. *Résumé.* Le tableau 1 faisant office de survol des résultats cités dans les pages précédentes, je me contenterai ici de dire qu'a priori, chacune des deux distances présente un intérêt. Typiquement, la distance de Boylan, peu efficace quant à la continuité de l'opérateur σ ou à la densité des partitions finies, présente malgré tout l'avantage d'avoir une valeur numérique intrinsèque, je veux dire par là intimement liée à la probabilité utilisée, sans coefficient arbitraire, alors que la distance de Cotter, intéressante à l'inverse pour la continuité de σ ou la densité des partitions finies, ne saurait sans doute avoir un intérêt quantitatif, puisque les coefficients en 2^j présents devant chacun des termes de la sommes sont arbitraires, puisque l'on ne doit même pas spécifier l'ordre dans lequel on prend les fonctions de la famille dénombrable dense utilisée. Aussi la topologie la plus appropriée pour la plus grande gamme de problèmes se situe-t-elle sans doute entre les deux topologies introduites ici. Une voie d'investigation future est sans doute la recherche d'une nouvelle norme sur $\mathcal{L}(L_1)$ que la norme uniforme utilisée, afin de considérer une nouvelle topologie *un peu moins forte* que la topologie uniforme, mais *plus forte* que notre topologie forte.

III.3.3. Un premier résultat de stabilité. Avec H. Heitsch et W. Römisch (voir [**57**]), nous avons considéré le problème suivant :

$$v(\boldsymbol{\xi}) = \min_{\boldsymbol{x}} F(\boldsymbol{\xi}, \boldsymbol{x}) := \mathbb{E}\left(\sum_{t=1}^{T} \langle b_t(\boldsymbol{\xi}_t), \boldsymbol{x}_t\rangle_{\mathbb{R}^{m_t}}\right)$$

$$\forall 1 \leq t \leq T, \ \boldsymbol{x}_t \text{ est } \sigma(\boldsymbol{\xi}^t) - \text{mesurable}, \tag{III.17a}$$

$$\boldsymbol{x}_t \in X_t, \text{ p.s.} \tag{III.17b}$$

$$A_{t,0}\boldsymbol{x}_t + A_{t,1}(\boldsymbol{\xi}_t)\boldsymbol{x}_{t-1} = h_t(\boldsymbol{\xi}_t), \text{ p.s.}, \tag{III.17c}$$

avec pour tout t, $\boldsymbol{\xi}_t$ une variable aléatoire à valeurs dans $\mathbb{R}^d$, $\boldsymbol{x}_t$ une variable aléatoire à valeurs dans $\mathbb{R}^{m_t}$, et donc $X_t \subset \mathbb{R}^{m_t}$, et $A_{t,0}$ (resp. $A_{t,1}(\xi_t)$) une matrice de $\mathbb{R}^{m_t}$ (resp. $\mathbb{R}^{m_{t-1}}$) dans $\mathbb{R}^{n_t}$ et h_t (resp. b_t) une application de $\mathbb{R}^d$ dans $\mathbb{R}^{n_t}$ (resp. $\mathbb{R}^{m_t}$). On suppose de plus que b_t, h_t, et $A_{t,1}$ dépendent linéairement de $\boldsymbol{\xi}_t$. Finalement, en posant $m = \sum_{t=1}^{T} m_t$, on recherche le contrôle $\boldsymbol{x} = (\boldsymbol{x}_t)_{1\leq t\leq T}$ comme un élément de l'espace de Banach $L^{r'}(\Omega, \mathbb{R}^m, \mathbb{P})$, avec r' dépendant du problème et de l'intégrabilité de $\boldsymbol{\xi}$. On rappelle qu'on notera $x \in \mathcal{X}(\boldsymbol{\xi})$ pour tout élément x vérifiant (III.17a)-(III.17b)–(III.17c). Dans ce cadre, nous avons montré :

THÉORÈME III.24. *Supposons que :*

(i) $\boldsymbol{\xi} \in L^r(\Omega, \mathbb{R}^{dT}, \mathbb{P})$,

(ii) $\boldsymbol{x} \in L^{r'}(\Omega, \mathbb{R}^m, \mathbb{P})$, *avec :*

$$r' = \begin{cases} \frac{r}{r-1} & , \quad \textit{si seuls les coûts } (b_t) \textit{ sont aléatoires,} \\ r & , \quad \textit{si seuls les seconds membres } (h_t) \textit{ sont aléatoires,} \\ r = 2 & , \quad \textit{si seuls les coûts et les seconds membres sont aléatoires,} \\ \infty & , \quad \textit{si toutes les matrices } A_{t,1} \textit{ sont aléatoires et } r = T. \end{cases}$$

Alors, la fonction critère du problème (III.17) *est bien définie.*

(iii) Pour tout $t \in \{1, \dots, T\}$, X_t *est un cône polyhédral fermé non vide, et il existe* $\delta > 0$ *tels que pour tout* $\tilde{\boldsymbol{\xi}} \in L^r(\Omega, \mathbb{R}^{dT}, \mathbb{P})$ *tel que* $\|\boldsymbol{\xi} - \tilde{\boldsymbol{\xi}}\|_r \leq \delta$, *et tout* $t \in \{1, \dots, T\}$, *les ensembles admissibles des contrôles sont non vides quelles que soient les décisions passées (*relatively complete recourse around $\boldsymbol{\xi}$*).*

(iv) $v(\boldsymbol{\xi}) < +\infty$, *et pour tout* $\epsilon > 0$, *il existe* $B \subset L^{r'}(\Omega, \mathbb{R}^m, \mathbb{P})$ *borné, et* $\delta > 0$ *tels que :*

$$\forall \tilde{\boldsymbol{\xi}} \text{ t.q. } \|\boldsymbol{\xi} - \tilde{\boldsymbol{\xi}}\|_r \leq \delta, \quad l_\alpha(F(\tilde{\boldsymbol{\xi}}, \cdot)) := \left\{ \tilde{\boldsymbol{x}} \in \mathcal{X}(\tilde{\boldsymbol{\xi}}) \ : \ F(\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{x}}) \leq v(\boldsymbol{\xi}) + \alpha \right\} \subset B. \tag{III.18}$$

Alors, il existe $L, \alpha, \delta > 0$ *tels que :*

$$|v(\boldsymbol{\xi}) - v(\tilde{\boldsymbol{\xi}})| \leq L \left(\|\boldsymbol{\xi} - \tilde{\boldsymbol{\xi}}\|_r + D_f((\mathcal{F}_t), (\tilde{\mathcal{F}}_t)) \right), \tag{III.19}$$

pour tout $\tilde{\boldsymbol{\xi}} \in L^r(\Omega, \mathbb{R}^{dT}, \mathbb{P})$ *tel que* $\|\boldsymbol{\xi} - \tilde{\boldsymbol{\xi}}\|_r \leq \delta$ *et* $v(\tilde{\boldsymbol{\xi}})$ *fini. De plus,* $(\mathcal{F}_t)$ *(resp.* $(\tilde{\mathcal{F}}_t)$*) représente la filtration naturelle engendrée par* $\boldsymbol{\xi}$ *(resp.* $\tilde{\boldsymbol{\xi}}$*), et* D_t *est définie pour tout* $t \in \{1, \dots, T\}$ *par :*

(III.20)

$$D_f((\mathcal{F}_t), (\tilde{\mathcal{F}}_t)) := \sup_{\epsilon \in (0, \alpha]} \inf_{(\boldsymbol{x}, \tilde{\boldsymbol{x}}) \in l_\alpha(F(\tilde{\boldsymbol{\xi}}, \cdot)) \times l_\alpha(F(\boldsymbol{\xi}, \cdot))} \sum_{t=2}^{T-1} \max \left(\|\boldsymbol{x}_t - \mathbb{E}\left(\boldsymbol{x}_t | \mathcal{F}_t\right) \|_{r'}, \|\boldsymbol{x}_t - \mathbb{E}\left(\boldsymbol{x}_t | \tilde{\mathcal{F}}_t\right) \|_{r'} \right).$$

Preuve : cf. [**57**], Théorème 2.1. □

Ce résultat présente le principal mérite de mettre en lumière que la variation du coût optimal v d'un problème du type (III.10) vu comme une fonction de l'aléa, est bornée par la somme d'un terme du type distance des lois (i.e. $\|\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}\|_r$) et d'un terme du type distance entre filtrations. La démonstration de ce résultat, de nature assez technique, et basée sur principalement sur des arguments de continuité et de sélections mesurables provenant de l'ouvrage [**84**] n'est pas donnée ici pour une plus grande facilité de lecture. Le lecteur intéressé pourra se référer à l'article [**57**] pour y trouver tous les détails nécessaires.

III.3.4. Critères séparés. Néanmoins, au vu du Théorème IV.14 de [**8**], il apparaît possible de généraliser le résultat précédent (i.e. le Théorème III.24) au cas de critères F non-linéaires en $\boldsymbol{x}$. L'idée du chapitre IV de [**8**] est de *pénaliser* la contrainte de mesurabilité du problème (III.21) suivant :

$$v(\mathcal{B}) := \min_{\boldsymbol{x}} \mathbb{E}\left(\langle f(\boldsymbol{x}), b(\boldsymbol{\xi}) \rangle_{\mathbb{R}^n} \right)$$

$$\text{s.c.} \quad \boldsymbol{x} \in X, \text{ p.s.} \tag{III.21a}$$

$$\boldsymbol{x} \text{ est } \mathcal{B} - \text{mesurable}, \tag{III.21b}$$

avec $\boldsymbol{\xi}$ (resp. $\boldsymbol{x}$) une variable aléatoire à valeurs dans $\mathbb{R}^d$ (resp. $\mathbb{R}^m$), X un compact de $\mathbb{R}^m$, et $\mathcal{B}$ une sous-tribu de $\mathcal{F}$ sur Ω. $f : \mathbb{R}^m \to \mathbb{R}^n$, et $b : \mathbb{R}^d \to \mathbb{R}^n$ sont telles que l'application $g : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$ définie par $g(x, \xi) = \langle f(x), b(\xi) \rangle_{\mathbb{R}^n}$ est convexe en x pour tout ξ. Barty obtient alors le résultat suivant :

Lemme III.25 (Barty). *Posons* $g_{\mathcal{B}} : \mathbb{R}^m \times \Omega \to \mathbb{R}$ *définie pour presque tout* $\omega \in \Omega$ *par*

$$\forall x \in \mathbb{R}^m, \ g_{\mathcal{B}}(x, \omega) = \langle f(x), \mathbb{E}\left(b(\boldsymbol{\xi}) | \mathcal{B}\right)(\omega) \rangle_{\mathbb{R}^n}.$$

Le problème (III.21) *admet la même solution que le problème* (III.22) *:*

$$\min_{\boldsymbol{x}} \mathbb{E}\left(g_{\mathcal{B}}(\boldsymbol{x})\right) \quad \text{s.c.} \quad \boldsymbol{x} \in X. \tag{III.22}$$

Preuve : cf. [**8**], Lemme IV.13. □

Ainsi, on a pu pénaliser la contrainte de mesurabilité dans la fonction critère du problème d'optimisation. A partir du lemme III.25, on peut obtenir le théorème suivant :

THÉORÈME III.26 (Barty). *Supposons que :*

(i) f *est continue,*

(ii) $\mathbb{E}\left(\|b(\boldsymbol{\xi})\|_{\mathbb{R}^n}\right) < +\infty$,

(iii) X *est un compact de* $\mathbb{R}^m$.

Alors, la fonction $\|f(\cdot)\|$ *atteint son supremum sur* X*, noté* $M > 0$*, et on a pour toute tribu* $\mathcal{B}'$ *sur* Ω*,*

$$|v(\mathcal{B}) - v(\mathcal{B}')| \leq 3M\mathbb{E}\left(\|\mathbb{E}\left(b(\boldsymbol{\xi})|\mathcal{B}\right) - \mathbb{E}\left(b(\boldsymbol{\xi})|\mathcal{B}'\right)\|_{\mathbb{R}^n}\right).$$

Preuve : cf. [**8**], Théorème IV.14. □

Bien entendu, dans le cadre d'un problème stochastique dynamique du type (III.10), pénaliser les contraintes de mesurabilité n'est pas aussi aisé, du fait principalement des contraintes couplantes entre les pas de temps, données par (III.10c). Nous nous plaçons ici dans le cadre du problème stochastique dynamique suivant :

$$v(\boldsymbol{\xi}) = \min_{\boldsymbol{x}} F(\boldsymbol{\xi}, \boldsymbol{x}) := \mathbb{E}\left(\sum_{t=1}^{T} \langle \boldsymbol{b}_t, f_t(\boldsymbol{x}_t)\rangle_{\mathbb{R}^{p_t}}\right)$$

$$\forall 1 \leq t \leq T,\ \boldsymbol{x}_t \text{ est } \sigma(\boldsymbol{\xi}^t) - \text{mesurable}, \tag{III.23a}$$

$$\boldsymbol{x}_t \in X_t,\ \text{p.s.} \tag{III.23b}$$

$$A_{t,0}\boldsymbol{x}_t + A_{t,1}\boldsymbol{x}_{t-1} = \boldsymbol{h}_t,\ \text{p.s.}, \tag{III.23c}$$

avec pour tout t, $\boldsymbol{\xi}_t$ une variable aléatoire à valeurs dans $\mathbb{R}^d$, $\boldsymbol{x}_t$ une variable aléatoire à valeurs dans $\mathbb{R}^{m_t}$, et donc $X_t \subset \mathbb{R}^{m_t}$, et $A_{t,0}$ (resp. $A_{t,1}$) une matrice de $\mathbb{R}^{m_t}$ (resp. $\mathbb{R}^{m_{t-1}}$ dans $\mathbb{R}^{n_t}$ et $\boldsymbol{h}_t$ (resp. $\boldsymbol{b}_t$) une variable aléatoire à valeurs dans $\mathbb{R}^{n_t}$ (resp. $\mathbb{R}^{p_t}$). f_t est une application de $\mathbb{R}^{m_t}$ dans $\mathbb{R}^{p_t}$. Finalement, en posant $m = \sum_{t=1}^{T} m_t$, et $n = \sum_{t=1}^{T} n_t$, on recherche le contrôle $\boldsymbol{x} = (\boldsymbol{x}_t)_{1\leq t\leq T}$ comme un élément de l'espace de Banach $L^{r'}(\Omega, \mathbb{R}^m, \mathbb{P})$.

Afin de pouvoir obtenir un résultat du type du théorème III.26, on a recours à la théorie de la dualité pour l'optimisation convexe[3]. Dans ce cas, en dualisant les contraintes (III.23c), on forme le lagrangien $L : L^{r'}(\Omega, \mathbb{R}^m, \mathbb{P}) \times L^{r}(\Omega, \mathbb{R}^n, \mathbb{P})$ du problème (III.23) qui s'écrit :

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = \mathbb{E}\left(\sum_{t=1}^{T} \langle \boldsymbol{b}_t, f_t(\boldsymbol{x}_t)\rangle_{\mathbb{R}^{p_t}} + \sum_{t=2}^{T} \langle \boldsymbol{\lambda}_t, A_{t,0}\boldsymbol{x}_t + A_{t,1}\boldsymbol{x}_{t-1} - \boldsymbol{h}_t\rangle_{\mathbb{R}^{n_t}}\right).$$

Sous des hypothèses adéquates assurant l'existence d'un point selle (ici, il s'agira de qualifier les contraintes, autrement dit de faire une hypothèse du type *relatively complete recourse*), on peut dès lors introduire le problème dual équivalent suivant :

$$\max_{\boldsymbol{\lambda}\in L^{r}(\Omega,\mathbb{R}^n,\mathbb{P})} \min_{\boldsymbol{x}\in L^{r'}(\Omega,\mathbb{R}^m,\mathbb{P})} L(\boldsymbol{x}, \boldsymbol{\lambda}) \tag{III.24}$$

$$\text{s.c.} \forall 1 \leq t \leq T,\ \boldsymbol{x}_t \text{ est } \sigma(\boldsymbol{\xi}^t) - \text{mesurable},$$

$$\boldsymbol{x}_t \in X_t,\ \text{p.s.}$$

On peut alors faire le raisonnement suivant : à $\boldsymbol{\lambda}$ fixé, l'application $L(\cdot, \boldsymbol{\lambda})$ s'écrit comme une somme sur les pas de temps, bref, elle est décomposable en temps, tandis que les contraintes n'ayant pas été dualisées (i.e. (III.23a)–(III.23b)) sont décomposées elles aussi en temps. Bref, le problème de minimisation (III.24) est totalement décomposé en temps. On peut donc appliquer sur chaque problème temporel le théorème III.26, ce qui donnerait un résultat de stabilité. Cependant, il est bien clair que le multiplicateur de Lagrange optimal $\boldsymbol{\lambda}^*$ dépend de $\boldsymbol{\xi}$. Il faudra

[3]Les premiers usages de la dualité pour l'étude des problèmes stochastiques convexes remontent aux articles fondateurs [**81, 82**]. Ces articles travallaient cependant entre les espaces de Banach L^1 et L^∞. Nous nous restreindrons ici au cas de Banach réflexifs afin d'éviter l'apparition de parties singulières dans les multiplicateurs.

donc être prudent par rapport à cette dépendance.

On peut montrer le résultat général de stabilité suivant, que l'on énonce pour plus de simplicité dans le cas L^2 :

THÉORÈME III.27. *Supposons que $r = r' = 2$, et que pour tout $t \in \{1, \dots, T\}$,*

(i) $\boldsymbol{\xi}_t \in L^2(\Omega, \mathbb{R}^d, \mathbb{P})$,

(ii) pour presque tout $\omega \in \Omega$, l'application $x \mapsto \langle \boldsymbol{b}_t(\omega), f_t(x)\rangle_{\mathbb{R}^{p_t}}$ est convexe sur $\mathbb{R}^{m_t}$,

(iii) f_t est différentiable et de gradient continu sur X_t,

(iv) $\mathbb{E}\left(\|\boldsymbol{b}_t\|_{\mathbb{R}^{p_t}}\right) < +\infty$

(v) X_t est compact, et il existe $\delta > 0$ tel que pour tout $\tilde{\boldsymbol{\xi}} \in L^2(\Omega, \mathbb{R}^{dT}, \mathbb{P})$ tel que $\|\boldsymbol{\xi} - \tilde{\boldsymbol{\xi}}\|_{L^2} \leq \delta$, et pour $t \geq 2$, l'ensemble des contrôles admissibles pour la filtration engendrée par $\tilde{\boldsymbol{\xi}}$ est non vide, quelles que soient les décisions passées.

Alors, il existe $M > 0$ tel que pour tout $\tilde{\boldsymbol{\xi}} \in L^2(\Omega, \mathbb{R}^{dT}, \mathbb{P})$ vérifiant $\|\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}\|_{L^2} \leq \delta$, il existe deux applications $\boldsymbol{\lambda}, \tilde{\boldsymbol{\lambda}} \in L^2(\Omega, \mathbb{R}^n, \mathbb{P})$ telles que :

(III.25)

$$
\begin{aligned}
|v(\boldsymbol{\xi}) - v(\tilde{\boldsymbol{\xi}})| \leq & M \sum_{t=1}^{T} \mathbb{E}\left(\|\mathbb{E}\left(\boldsymbol{b}_t | \sigma(\tilde{\boldsymbol{\xi}}^t)\right) - \mathbb{E}\left(\boldsymbol{b}_t | \sigma(\boldsymbol{\xi}^t)\right)\|_{\mathbb{R}^{p_t}}\right) \\
& + M' \sum_{t=1}^{T} \max\left(\mathbb{E}\left(\|\boldsymbol{\lambda}_t - \mathbb{E}\left(\boldsymbol{\lambda}_t | \sigma(\tilde{\boldsymbol{\xi}}^t)\right)\|_{\mathbb{R}^{m_t}}\right), \mathbb{E}\left(\|\tilde{\boldsymbol{\lambda}}_t - \mathbb{E}\left(\tilde{\boldsymbol{\lambda}}_t | \sigma(\boldsymbol{\xi}^t)\right)\|_{\mathbb{R}^{m_t}}\right)\right) \\
& + M' \sum_{t=1}^{T-1} \max\left(\mathbb{E}\left(\|\boldsymbol{\lambda}_{t+1} - \mathbb{E}\left(\boldsymbol{\lambda}_{t+1} | \sigma(\tilde{\boldsymbol{\xi}}^t)\right)\|_{\mathbb{R}^{m_t}}\right), \mathbb{E}\left(\|\tilde{\boldsymbol{\lambda}}_{t+1} - \mathbb{E}\left(\tilde{\boldsymbol{\lambda}}_{t+1} | \sigma(\boldsymbol{\xi}^t)\right)\|_{\mathbb{R}^{m_t}}\right)\right) \qquad \text{(III.26)}
\end{aligned}
$$

Preuve : Le lagrangien L s'écrit sous la forme :

$$
\begin{aligned}
L(\boldsymbol{x}, \boldsymbol{\lambda}) = & \sum_{t=1}^{T} \mathbb{E}\left(\langle \boldsymbol{b}_t, f_t(\boldsymbol{x}_t)\rangle_{\mathbb{R}^{p_t}}\right) \\
& + \sum_{t=2}^{T} \mathbb{E}\left(\langle A_{t,0}^* \boldsymbol{\lambda}_t, \boldsymbol{x}_t\rangle_{\mathbb{R}^{m_t}} + \langle A_{t,1}^* \boldsymbol{\lambda}_t, \boldsymbol{x}_{t-1}\rangle_{\mathbb{R}^{m_{t-1}}} - \langle \boldsymbol{\lambda}_t, \boldsymbol{h}_t\rangle_{\mathbb{R}^{n_t}}\right), \\
= & \mathbb{E}\left(\sum_{t=1}^{T} \langle \boldsymbol{b}_t, f_t(\boldsymbol{x}_t)\rangle_{\mathbb{R}^{p_t}} + \langle A_{t,0}^* \boldsymbol{\lambda}_t + A_{t+1,1}^* \boldsymbol{\lambda}_{t+1}, \boldsymbol{x}_t\rangle_{\mathbb{R}^{m_t}} - \langle \boldsymbol{\lambda}_t, \boldsymbol{h}_t\rangle_{\mathbb{R}^{n_t}}\right),
\end{aligned}
$$

en posant $A_{1,0} = 0$ et $A_{T+1,1} = 0$. Ainsi, on peut introduire la fonction duale suivante :

$$
\begin{aligned}
V(\boldsymbol{\xi}, \boldsymbol{\lambda}) := \sum_{t=1}^{T} -\mathbb{E}\left(\langle \boldsymbol{\lambda}_t, \boldsymbol{h}_t\rangle_{\mathbb{R}^{n_t}}\right) + \min_{\boldsymbol{x}_t} & \mathbb{E}\left(\langle \boldsymbol{b}_t, f_t(\boldsymbol{x}_t)\rangle_{\mathbb{R}^{p_t}} + \langle A_{t,0}^* \boldsymbol{\lambda}_t + A_{t+1,1}^* \boldsymbol{\lambda}_{t+1}, \boldsymbol{x}_t\rangle_{\mathbb{R}^{m_t}}\right) \\
& \forall 1 \leq t \leq T,\ \boldsymbol{x}_t \text{ est } \sigma(\boldsymbol{\xi}^t) - \text{mesurable}, \\
& \boldsymbol{x}_t \in X_t,\ \text{p.s.} \qquad \text{(III.27)}
\end{aligned}
$$

En utilisant le lemme III.25, pour tout λ, on a :

(III.28)

$$
\begin{aligned}
V(\boldsymbol{\xi}, \boldsymbol{\lambda}) = \sum_{t=1}^{T} -\mathbb{E}\left(\langle \boldsymbol{\lambda}_t, \boldsymbol{h}_t\rangle_{\mathbb{R}^{n_t}}\right) + \min_{\boldsymbol{x}_t} & \mathbb{E}\left(\langle \mathbb{E}\left(\boldsymbol{b}_t | \sigma(\boldsymbol{\xi}^t)\right), f_t(\boldsymbol{x}_t)\rangle_{\mathbb{R}^{p_t}} + \langle \mathbb{E}\left(A_{t,0}^* \boldsymbol{\lambda}_t + A_{t+1,1}^* \boldsymbol{\lambda}_{t+1} | \sigma(\boldsymbol{\xi}^t)\right), \boldsymbol{x}_t\rangle_{\mathbb{R}^{m_t}}\right) \\
& \boldsymbol{x}_t \in X_t,\ \text{p.s.}
\end{aligned}
$$

On utilise maintenant un argument de dualité. En effet, la condition de recours relativement complet autour de $\boldsymbol{\xi}$ assure la qualification des contraintes d'égalité (cf. par exemple [**7**], Chapitre 3). Cette condition de qualification des contraintes assure donc l'existence d'un point-selle pour le problème considéré,

d'où l'égalité suivante entre le problème primal et le problème dual :

(III.29)

$$v(\boldsymbol{\xi}) = \max_{\boldsymbol{\lambda}} \min_{\boldsymbol{x}} \mathbb{E}\left(\sum_{t=1}^{T} -\langle \boldsymbol{\lambda}_t, \boldsymbol{h}_t\rangle_{\mathbb{R}^{n_t}} + \langle \mathbb{E}\left(\boldsymbol{b}_t|\sigma(\boldsymbol{\xi}^t)\right), f_t(\boldsymbol{x}_t)\rangle_{\mathbb{R}^{p_t}} + \langle\langle \mathbb{E}\left(A^*_{t,0}\boldsymbol{\lambda}_t + A^*_{t+1,1}\boldsymbol{\lambda}_{t+1}|\sigma(\boldsymbol{\xi}^t)\right), \boldsymbol{x}_t\rangle_{\mathbb{R}^{m_t}}\right)$$
$$\text{s.c. } \lambda \in L^2(\Omega, \mathbb{R}^n, \mathbb{P}),\ \boldsymbol{x} \in L^2(\Omega, \mathbb{R}^m, \mathbb{P})$$
$$\forall t \in \{1, \dots, T\},\ \boldsymbol{x}_t \in X_t \text{ p.s.}$$

On notera alors

$$F_{\boldsymbol{\xi}}(\boldsymbol{x}, \boldsymbol{\lambda}) := \mathbb{E}\left(\sum_{t=1}^{T} -\langle \boldsymbol{\lambda}_t, \boldsymbol{h}_t\rangle_{\mathbb{R}^{n_t}} + \langle \mathbb{E}\left(\boldsymbol{b}_t|\sigma(\boldsymbol{\xi}^t)\right), f_t(\boldsymbol{x}_t)\rangle_{\mathbb{R}^{p_t}} + \langle \mathbb{E}\left(A^*_{t,0}\boldsymbol{\lambda}_t + A^*_{t+1,1}\boldsymbol{\lambda}_{t+1}|\sigma(\boldsymbol{\xi}^t)\right), \boldsymbol{x}_t\rangle_{\mathbb{R}^{m_t}}\right).$$

Pour tout $\boldsymbol{\xi} \in L^2(\Omega, \mathbb{R}^{dT}, \mathbb{P})$, on notera $(\boldsymbol{x}^*_{\boldsymbol{\xi}}, \boldsymbol{\lambda}^*_{\boldsymbol{\xi}})$ le point-selle du problème (III.29), dont on a montré qu'il existe.

On se donne maintenant $\tilde{\boldsymbol{\xi}} \in L^2(\Omega, \mathbb{R}^{dT}, \mathbb{P})$ dans un voisinage de rayon δ de $\boldsymbol{\xi}$. Les inégalités de point-selle écrites à l'optimum donnent :

$$|v(\boldsymbol{\xi}) - v(\tilde{\boldsymbol{\xi}})| \leq \max\left(|F_{\boldsymbol{\xi}}(\boldsymbol{x}^*_{\tilde{\boldsymbol{\xi}}}, \boldsymbol{\lambda}^*_{\boldsymbol{\xi}}) - F_{\tilde{\boldsymbol{\xi}}}(\boldsymbol{x}^*_{\tilde{\boldsymbol{\xi}}}, \boldsymbol{\lambda}^*_{\boldsymbol{\xi}})|, |F_{\boldsymbol{\xi}}(\boldsymbol{x}^*_{\boldsymbol{\xi}}, \boldsymbol{\lambda}^*_{\tilde{\boldsymbol{\xi}}}) - F_{\tilde{\boldsymbol{\xi}}}(\boldsymbol{x}^*_{\boldsymbol{\xi}}, \boldsymbol{\lambda}^*_{\boldsymbol{\xi}})|\right). \tag{III.30}$$

En injectant dans (III.30) toutes ces valeurs, et appliquant l'inégalité de Cauchy-Schwartz, on obtient, en notant $M_t > 0$ le supremum sur X_t de l'application $(x \mapsto \|f(\cdot)\|_{\mathbb{R}^{p_t}} + \|\cdot\|_{\mathbb{R}^{m_t}})$, et $M := \sum_{t=1}^{T} M_t$:

$$\begin{aligned}
|v(\boldsymbol{\xi}) - v(\tilde{\boldsymbol{\xi}})| \leq & M \sum_{t=1}^{T} \mathbb{E}\left(\|\mathbb{E}\left(\boldsymbol{b}_t|\sigma(\tilde{\boldsymbol{\xi}}^t)\right) - \mathbb{E}\left(\boldsymbol{b}_t|\sigma(\boldsymbol{\xi}^t)\right)\|_{\mathbb{R}^{p_t}}\right) \\
& + M \max\left(\mathbb{E}\left(\sum_{t=1}^{T} \|A^*_{t,0}(\boldsymbol{\lambda}^*_{\boldsymbol{\xi},t} - \mathbb{E}\left(\boldsymbol{\lambda}^*_{\boldsymbol{\xi},t}|\sigma(\tilde{\boldsymbol{\xi}}^t)\right)) + A^*_{t+1,1}(\boldsymbol{\lambda}^*_{\boldsymbol{\xi},t+1} - \mathbb{E}\left(\boldsymbol{\lambda}^*_{\boldsymbol{\xi},t+1}|\sigma(\tilde{\boldsymbol{\xi}}^t)\right))\|_{\mathbb{R}^{m_t}}\right),\right. \\
& \left.\mathbb{E}\left(\sum_{t=1}^{T} \|A^*_{t,0}(\boldsymbol{\lambda}^*_{\tilde{\boldsymbol{\xi}},t} - \mathbb{E}\left(\boldsymbol{\lambda}^*_{\tilde{\boldsymbol{\xi}},t}|\sigma(\boldsymbol{\xi}^t)\right)) + A^*_{t+1,1}(\boldsymbol{\lambda}^*_{\tilde{\boldsymbol{\xi}},t+1} - \mathbb{E}\left(\boldsymbol{\lambda}^*_{\tilde{\boldsymbol{\xi}},t+1}|\sigma(\boldsymbol{\xi}^t)\right))\|_{\mathbb{R}^{m_t}}\right)\right)
\end{aligned} \tag{III.31}$$

Finalement, il existe donc une constante M' liée aux opérateurs A telle que :

$$\begin{aligned}
|v(\boldsymbol{\xi}) - v(\tilde{\boldsymbol{\xi}})| \leq & M \sum_{t=1}^{T} \mathbb{E}\left(\|\mathbb{E}\left(\boldsymbol{b}_t|\sigma(\tilde{\boldsymbol{\xi}}^t)\right) - \mathbb{E}\left(\boldsymbol{b}_t|\sigma(\boldsymbol{\xi}^t)\right)\|_{\mathbb{R}^{p_t}}\right) \\
& + M' \sum_{t=1}^{T} \max\left(\mathbb{E}\left(\|\boldsymbol{\lambda}^*_{\boldsymbol{\xi},t} - \mathbb{E}\left(\boldsymbol{\lambda}^*_{\boldsymbol{\xi},t}|\sigma(\tilde{\boldsymbol{\xi}}^t)\right)\|_{\mathbb{R}^{m_t}}\right), \mathbb{E}\left(\|\boldsymbol{\lambda}^*_{\tilde{\boldsymbol{\xi}},t} - \mathbb{E}\left(\boldsymbol{\lambda}^*_{\tilde{\boldsymbol{\xi}},t}|\sigma(\boldsymbol{\xi}^t)\right)\|_{\mathbb{R}^{m_t}}\right)\right) \\
& + M' \sum_{t=1}^{T-1} \max\left(\mathbb{E}\left(\|\boldsymbol{\lambda}^*_{\boldsymbol{\xi},t+1} - \mathbb{E}\left(\boldsymbol{\lambda}^*_{\boldsymbol{\xi},t+1}|\sigma(\tilde{\boldsymbol{\xi}}^t)\right)\|_{\mathbb{R}^{m_t}}\right), \mathbb{E}\left(\|\boldsymbol{\lambda}^*_{\tilde{\boldsymbol{\xi}},t+1} - \mathbb{E}\left(\boldsymbol{\lambda}^*_{\tilde{\boldsymbol{\xi}},t+1}|\sigma(\boldsymbol{\xi}^t)\right)\|_{\mathbb{R}^{m_t}}\right)\right),
\end{aligned} \tag{III.32}$$

ce qui donne le résultat recherché. □

Remarque III.28. *Le terme de droite de l'équation* (III.26) *tend vers 0 lorsque la filtration engendrée par* $(\tilde{\boldsymbol{\xi}}^t)$ *converge au sens fort vers la filtration engendrée par* $(\boldsymbol{\xi}^t)$.

III.4. Perspectives

Dans la continuité des travaux réalisés notamment par [**8, 86, 73, 51, 71**], nous avons montré dans ce chapitre la nécessité de tenir compte de la distance entre les quantités d'information pour la stabilité des problèmes stochastiques en boucle fermée. Puis, dans le cadre particulier des critères séparés ou linéaires, avec contraintes d'égalité linéaires, nous avons donné deux résultats de stabilité nouveaux qui illustrent cette nécessité de tenir compte des filtrations engendrées.

Il apparaît assez difficile d'obtenir des résultats plus précis dans le cadre de critères non séparés. D'autre part, il ne faut pas attendre, comme la sous-section sur les distances entre tribus et filtrations le montre, d'information quantitative de tels résultats de stabilité. La seule information valable qu'ils fournissent est de nature qualitative. Cela nous conduit à donner le cadre dans lequel ces résultats doivent être considérés : comme évoqué dans la section sur la

distance de Fortet-Mourier, l'important est *in fine* de proposer des schémas de discrétisation convergents pour les problèmes stochastiques à plusieurs pas de temps. Les résultats de stabilité donnés ici peuvent être utilisés à cette fin : à partir d'une suite de processus $(\boldsymbol{\xi}_t^n)$ discrétisant un problème du type (III.17) ou (III.23), si l'on montre que pour tout $t \in \{1, \dots, T\}$,

$$\mathbb{E}\left(\|\mathbb{E}\left(\boldsymbol{b}_t|\sigma(\boldsymbol{\xi}_1^n, \dots, \boldsymbol{\xi}_t^n)\right) - \mathbb{E}\left(\boldsymbol{b}_t|\sigma(\boldsymbol{\xi}^t)\right)\|_{\mathbb{R}^{p_t}}\right) \longrightarrow 0, \quad \text{quand } n \to \infty,$$

on sera assuré d'avoir la convergence des coûts approchés vers le coût véritable[4]. Il reste cependant un travail important à fournir pour proposer de tels schémas de discrétisation. C'est précisément l'objet des recherches actuelles de W. Römisch et H. Heitsch (voir [**56**]) visant à réutiliser dans le nouveau cadre de stabilité décrit par le théorème III.24, les résultats qu'ils avaient obtenus sur la distance de Fortet-Mourier et la construction d'arbres de scénarios (par exemple [**55**]).

[4]Par exemple, la méthode de quantification proposée par [**8**] remplit cette condition.

$<$

CHAPITRE IV

Algorithmes stochastiques et boucle fermée

REMARQUE IV.1. *Le travail présenté dans ce chapitre est pour l'essentiel le fruit d'une collaboration avec Kengy Barty, alors post-doctorant chez EdF R&D, et Jean-Sébastien Roy, alors ingénieur-chercheur chez EdF R&D, et doctorant à l'université de Paris 6. Le travail présenté ici a été fait entre novembre 2004 et juin 2005, et a fait l'objet de trois articles :* Closed-Loop Stochastic Gradient *publié sur le site* speps*, et soumis à Mathematical Programming en juin 2005,* Temporal Difference Learning for pricing American style options*, accessible sur le site* http ://www.optimization-online.org *et soumis à IEEE Transactions on Automatic Control en juin 2005, et* Hilbert valued Perturbed Gradient Algorithms*, accessible sur le site* http ://www.optimization-online.org *et soumis à Mathematics of Operations Research en janvier 2006.*

IV.1. Résumé

Pour entreprendre numériquement la résolution de problèmes en boucle fermée du type (I.4), il est courant de commencer par discrétiser les variables aléatoires affectant le problème, puis de résoudre le problème déterministe résultant, c'est dans cette optique qu'ont été développés les travaux [**25, 26**], puis [**8**] sur la quantification de l'information, ceux de [**70**] sur une autre quantification des variables, ou encore tous ceux désignés par l'expression *stochastic programming* qui débutent par une construction d'arbre d'aléa (voir par exemple le handbook [**91**]). Un autre type d'approche, dérivé du contexte de la programmation dynamique stochastique (dont les fondements remontent à l'ouvrage [**13**]), est l'utilisation d'un paramétrage a priori du contrôle (resp. de la fonction de Bellman) sous la forme d'une combinaison linéaire de fonctions prédéterminées. On trouve la déclinaison de cette approche dans le cas de la programmation dynamique notamment dans le livre [**19**], ou encore dans [**39**]. On appellera cette seconde approche *règles de décision linéaires*.

Outre les problèmes de stabilité relevés dans le chapitre III, les approches discrétisant l'aléa comportent d'autres faiblesses, notamment en terme de réinterprétation : que faire avec le contrôle trouvé sur la structure discrétisée, hormis une interpolation certes naturelle mais plus difficilement justifiable ? comment l'étendre à l'espace d'aléa tout entier ? quelle garantie d'optimalité a ce prolongement ? L'approche par règles de décision linéaires, quant à elle, comporte beaucoup d'avantages, notamment, elle évite l'étape risquée d'un prolongement d'une solution discrète. En revanche, elle perd d'emblée toute optimalité globale : l'ensemble de fonctions étant fixé, on ne peut prétendre à l'optimum que dans l'espace vectoriel engendré par ces fonctions. Si l'optimum véritable du problème initial est à l'extérieur, il est définitvement hors d'atteinte dans cette formulation.

Le présent chapitre se veut être une alternative au dilemme présenté avant entre arbres d'aléas et règles de décision linéaires. En effet, nous présentons ici une nouvelle approche essentiellement non-paramétrique, qui permet d'approcher l'optimum global sans jamais avoir discrétisé a priori l'espace d'aléa ni avoir prédéterminé une base de fonctions dont on resterait prisonnier. La méthode proposée est d'essence variationnelle et non-paramétrique, et provient de la conjonction des idées d'approximation fonctionnelle et de gradient stochastique.

Ce chapitre est construit comme suit : dans un premier temps, nous introduisons dans la section IV.2 notre algorithme à l'aide de remarques naïves sur le problème (I.4), en écrivant un algorithme de gradient dans l'espace fonctionnel des contrôles. Ensuite, nous montrons dans la section IV.3 la convergence de notre algorithme vers la solution du problème (I.4), prouvons un

résultat de grande déviation, et donnons son utilité et originalité, puis quelques exemples numériques. Puis, nous proposons dans la section IV.4 un cadre théorique plus général dans lequel rentre notre algorithme : les algorithmes de gradient perturbé à valeurs dans un espace de Hilbert. Dans ce cadre, nous donnons des théorèmes de convergence pour la résolution de problèmes de minimisation et de problèmes de point-selle à l'aide d'algorithmes de gradient perturbé, et nous donnons quelques exemples d'application de cette théorie, notamment pour la preuve de convergence d'algorithmes d'estimation de densité en statistique. Enfin, en section IV.5, nous montrons les similitudes et les différences de notre méthode avec notamment les méthodes de règles de décision linéaires et les méthodes d'espaces de noyaux reproduisants.

IV.2. Introduction

Les algorithmes stochastiques, introduits pour la première fois avec l'article fondateur [**77**], sont destinés à rechercher les zéros d'une fonction satisfaisant un certain nombre de propriétés et s'exprimant comme une espérance. Leur principale force est de n'utiliser itérativement que des réalisations des variables aléatoires sous-jacentes pour calculer ce zéro. Ainsi, les algorithmes stochastiques combinent avec succès les idées de Monte-Carlo (estimation d'espérance) et d'algorithmes de descente ou de point fixe (recherche du zéro d'une application monotone).

Traditionnellement, la théorie abondante sur ces algorithmes a été établie en dimension finie, pour la raison principale que nombre d'applications étaient purement finies dimensionnelles (voir le livre [**46**] pour un exposé général dans sa première partie, et pour quelques applications dans les parties suivantes). Une adaptation naturelle de ces algorithmes finis dimensionnels à l'optimisation stochastique a été faite dans le cadre des problèmes en boucle ouverte. C'est ce qu'on a appelé les techniques de gradient stochastique (cf. [**72**], ou [**67**]), ou de pseudogradient (cf. [**50**] ou [**49**]). Une théorie en dimension infinie dans le cadre hilbertien a également été réalisée à travers les travaux de [**74, 75**] pour l'approximation stochastique générale, ou de [**59, 32**] pour l'application à l'optimisation convexe.

Cependant, dans le cadre des problèmes d'optimisation stochastique en boucle fermée, il n'est pas clair du tout que ce type d'approximations puisse être utilisé. Nous avons commencé par tenter d'écrire ce que signifierait dans ce contexte un algorithme de gradient stochastique. Puis, guidés par ces tentatives intuitives de généralisation, nous avons pu mettre en place un cadre général d'algorithmes stochastiques fonctionnels, permettant la résolution de problèmes d'optimisation stochastique en boucle fermée, et de problèmes de points-fixes fonctionnels. Enfin, dans le souci de disposer d'outils généraux et puissants pour les preuves de convergence de ces algorithmes, nous avons montré des théorèmes de convergence pour des algorithmes de gradient perturbé dans un espace de Hilbert.

Il est important de mettre au clair ici une distinction essentielle pour les problèmes d'optimisation stochastique. Dans la typologie des problèmes présentée au chapitre I, on insiste sur la différence entre problèmes en boucle ouverte et problèmes en boucle fermée. Généralement dans un problème en boucle ouverte, la variable de commande est un objet de dimension finie, mais pourrait sans perturber la classification être de dimension infinie, ce qui conduirait certes à des problèmes numériques mais resterait théoriquement traitable. Cependant dans les problèmes en boucle fermée, la variable de commande est naturellement recherchée comme une fonction, et donc comme un objet de dimension infinie, ce qui justifie l'intérêt apporté dans ce chapitre en section IV.4 à la généralisation en dimension infinie de résultats existant traditionnellement en dimension finie.

IV.2.1. Ecriture naïve. Afin de donner l'intuition de ce qui va suivre, nous allons écrire dans cette sous-section le cheminement qui nous a conduits à introduire les algorithmes stochastiques fonctionnels. Les mathématiques ne seront donc pas toujours d'une extrême rigueur, mais le but est seulement pour l'instant de faire comprendre les enjeux et les possibilités

de notre formulation.

Considérons le problème en boucle fermée usuel (I.4), qui pour mémoire s'écrit comme suit :

$$\min_{u\in U^f} J(u) := \mathbb{E}\left(j(u(\boldsymbol{\xi}),\boldsymbol{\xi})\right).$$

On supposera ici que $\Xi = \mathbb{R}^m$, que $\boldsymbol{\xi}$ est une variable aléatoire de loi μ à valeur dans Ξ et que

$$U^f = \{u : \Xi \to \mathbb{R}^p \ : \ u(\boldsymbol{\xi}) \text{ est } \mathcal{B}-\text{mesurable, et } \forall \xi \in \Xi,\ u(\xi) \in K_U\} \cap L^2(\Xi, \mathbb{R}^p, \mu),$$

avec K_U un compact de $\mathbb{R}^p$, et $\mathcal{B}$ une sous-tribu de la tribu $\mathcal{F}$ correspondant à l'espace de probabilité sous-jacent $(\Omega, \mathcal{F}, \mathbb{P})$. On se retrouve donc dans un contexte hilbertien. Calculons le gradient de J. Par définition de la dérivée directionnelle, on a pour tous $u, h \in L^2(\Xi, \mathbb{R}^p, \mu)$ et tout $t \in \mathbb{R}$,

$$\begin{aligned} J(u+th) =& \mathbb{E}\left(j(u(\boldsymbol{\xi}) + th(\boldsymbol{\xi}), \boldsymbol{\xi})\right), \\ =& \mathbb{E}\left(\langle \nabla_u j(u(\boldsymbol{\xi}), \boldsymbol{\xi}), th(\boldsymbol{\xi})\rangle_{\mathbb{R}^p}\right) + J(u) + O(t^2), \\ =& J(u) + t\langle \nabla_u j(u(\cdot),\cdot), h(\cdot)\rangle + O(t^2). \end{aligned}$$

On obtient donc le gradient de J, qui est lui-même un élément de $L^2(\Xi, \mathbb{R}^p, \mu)$:

$$\forall u \in L^2(\Xi, \mathbb{R}^p, \mu),\ \nabla J(u)(\cdot) = \nabla_u j(u(\cdot),\cdot).$$

A priori, *le gradient n'étant pas une espérance*, il n'y a aucune raison d'appliquer un algorithme de gradient stochastique. En revanche, il est toujours loisible sinon pratiquement utile d'écrire un algorithme de gradient projeté pour résoudre ce problème. On choisit $u^0 \in L^2(\Xi, \mathbb{R}^p, \mu)$ arbitrairement, puis on met à jour suivant la formule :

$$u^{k+1}(\cdot) = \Pi_{U^f}\left(u^k - \gamma^k \nabla J(u^k)\right)(\cdot). \tag{IV.1}$$

On peut déjà noter ici que s'il n'y avait pas de projection à effectuer, on pourrait dérouler l'algorithme (IV.1) pour tout $\xi \in \Xi$, et calculer ainsi le contrôle optimal point par point dans tous les points d'intérêt. Néanmoins, il est clair que c'est la projection (en mélangeant les $\xi \in \Xi$) qui rend l'algorithme (IV.1) impossible à effectuer.
Dans le cas non-contraint, le problème se réécrit simplement comme trouver l'application u^* qui annule le gradient en tout $\xi \in \Xi$, tandis que dans le cas contraint, le problème se réécrit comme trouver l'application $u^* \in U^f$ qui est telle que :

$$\forall u \in U^f,\ \langle \nabla J(u^*), u - u^*\rangle \geq 0, \tag{IV.2}$$

ce qui ramène une espérance dans la formulation à travers le produit scalaire.

Néanmoins, on peut toujours artificiellement introduire une espérance dans l'expression du gradient : en notant pour tout $\xi \in \Xi$, δ_ξ la masse de Dirac en ξ, on obtient :

$$\forall u \in L^2(\Xi, \mathbb{R}^p, \mu),\ \nabla J(u)(\cdot) = \mathbb{E}\left(\nabla_u j(u(\boldsymbol{\xi}), \boldsymbol{\xi})\delta_{\boldsymbol{\xi}}(\cdot)\right),$$

ce qui dès lors donne l'idée d'un algorithme de gradient stochastique s'écrivant :

$$\begin{aligned} &\text{Soit } \boldsymbol{\xi}^{k+1} \text{ indépendante de } (\boldsymbol{\xi}^1, \dots, \boldsymbol{\xi}^k), \text{ et répartie selon } \mu, \\ &\forall \xi \in \Xi,\ \boldsymbol{u}^{k+1}(\xi) = \Pi_{U^f}\left(\boldsymbol{u}^k - \gamma^k \nabla J(\boldsymbol{u}^k)(\boldsymbol{\xi}^{k+1})\delta_{\boldsymbol{\xi}^{k+1}}\right)(\xi) \end{aligned} \tag{IV.3}$$

Bien entendu, si la loi μ est continue sur Ξ, cet algorithme est complètement inopérant, et même faux, car une somme dénombrable de masses de Dirac reste toujours presque partout nulle... ou plus simplement, on a pour presque tout $\xi \in \Xi$,

$$\boldsymbol{u}^k(\xi) - \rho^k \nabla J(\boldsymbol{u}^k)(\boldsymbol{\xi}^{k+1})\delta_{\boldsymbol{\xi}^{k+1}}(\xi) = \boldsymbol{u}^k(\xi),\ \text{p.s.}$$

Remarque IV.2 (Loi atomique). *En revanche, dès que la loi μ est atomique, l'algorithme que nous avons écrit peut converger vers la solution, qui est elle-même finie-dimensionnelle, et la masse de Dirac est formellement remplacée par la fonction indicatrice. En effet, le contrôle est alors recherché naturellement comme un vecteur correspondant au contrôle pour chaque atome de μ, et notre algorithme s'interpréte comme un algorithme de gradient chaotique sur ce vecteur.*

REMARQUE IV.3 (Suite infinie de variables aléatoires). *On a considéré dans l'algorithme* (IV.3)*, et ce sera également le cas dans le reste de ce chapitre, une suite infinie de variables aléatoires i.i.d.* $(\boldsymbol{\xi}^k)$*. On peut se poser la question de l'existence d'une telle suite sur notre espace de probabilité* $(\Omega, \mathcal{F}, \mathbb{P})$*. Une telle suite n'existe que si l'espace de probabilité est suffisamment riche. En d'autres termes, si l'on dispose d'un espace de probabilité primitif* $(E, \mathcal{F}_E, \mathbb{P}_E)$ *sur lequel vivrait* $\boldsymbol{\xi}$*, on peut alors construire* $(\Omega, \mathcal{F}, \mathbb{P}) = (E^{\mathbb{N}}, \mathcal{F}_E^{\otimes\mathbb{N}}, \mathbb{P}_E^{\otimes\mathbb{N}})$*, qui sera suffisamment gros pour supporter des suites infinies comme celles que nous regarderons.*

IV.2.2. Idée fondamentale et enjeux. C'est à ce niveau *intuitif* qu'est apparue l'idée fondamentale pour tout ce qui va suivre de mollification (voir par exemple [**49**]) : dans le cas d'une loi μ continue, si la masse de Dirac n'est pas dans $L^2(\Xi, \mathbb{R}^p, \mu)$, on peut du moins l'approcher fonctionnellement autant qu'on le souhaite à l'aide d'une suite de fonctions dans $L^2(\Xi, \mathbb{R}^p, \mu)$ (cf. Figure 1). L'idée est donc de combiner les tirages $\boldsymbol{\xi}^{k+1}$ avec des fonctions *tendant* dans un certain sens vers la masse de Dirac.

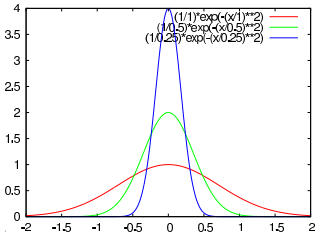


FIG. 1. Approximations de la masse de Dirac en 0

Dit autrement, on peut imaginer que si le contrôle recherché est suffisamment régulier, le déformer localement entraîne une déformation du voisinage. Toute la suite de ce chapitre est un développement et une formalisation de cette idée initiale de *déformation dans un voisinage*. En effet, nombre de questions se posent encore à ce stade, outre la convergence de tels schémas : sortira-t-on de l'obstacle d'implémentabilité soulevé par l'algorithme (IV.1), que choisir pour ces voisinages, quelles hypothèses faire pour obtenir la convergence, etc. De plus, les algorithmes stochastiques reposent sur l'idée que des tirages successifs peuvent suffire, alors que nous avons vu dans le chapitre III qu'une approximation ponctuelle de la solution pouvait mener dans le cadre des problèmes en boucle fermée à des solutions totalement fausses : il semble donc difficile a priori d'utiliser des informations locales pour obtenir une convergence globale. C'est à ce type de questions que les sections suivantes se proposent de répondre.

Bien entendu, les enjeux de telles questions sont importants. En effet, obtenir des algorithmes de nature variationnelle pour les problèmes en boucle fermée est extrêmement important du point de vue par exemple des théories de décomposition de tels problèmes, comme cela sera souligné dans le chapitre V. De plus, utiliser plus habilement l'aléa sous-jacent du problème, au lieu de s'en débarasser au plus vite par une quelconque discrétisation est sans doute plus efficace. Enfin, trouver une méthode qui par nature donnerait un contrôle bien défini sur tout l'espace d'aléa et éviterait ainsi une coûteuse phase d'extrapolation serait une avancée significative.

IV.3. Résolution de problèmes stochastiques en boucle fermée

IV.3.1. Remarques préliminaires. On s'intéresse ici à la résolution numérique du problème (I.1). Comme il a été dit dans la section précédente, on prend $j : \mathbb{R}^p \times \Xi \to \mathbb{R}$, avec $\Xi = \mathbb{R}^m$, telle que j soit une intégrande normale (cf. Définition A.23). On considère μ une probabilité sur Ξ absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}^m$, c'est à dire telle qu'il existe une densité par rapport à la mesure de Lebesgue. On note $\boldsymbol{\xi}$ une variable aléatoire à valeurs dans Ξ et de loi μ.
Sous cette hypothèse, on introduit l'espace $\mathcal{L}^2(\Xi, \mathbb{R}^p, \mu)$ des applications $u : \Xi \to \mathbb{R}^p$ de carré intégrable, i.e., telles que $\mathbb{E}\left(\|u(\boldsymbol{\xi})\|^2_{\mathbb{R}^p}\right) < +\infty$. $L^2(\Xi, \mathbb{R}^p, \mu)$ est alors défini comme l'espace quotient de $\mathcal{L}^2(\Xi, \mathbb{R}^p, \mu)$ par rapport à la relation d'équivalence μ-négligeable. C'est un espace

de Hilbert et on notera $\langle u, v\rangle := \mathbb{E}\left(\langle u(\boldsymbol{\xi}), v(\boldsymbol{\xi})\rangle_{\mathbb{R}^p}\right)$ son produit scalaire, et $\|\cdot\| := \sqrt{\langle\cdot,\cdot\rangle}$ sa norme.
On considère alors U^f comme étant un sous-ensemble de $L^2(\Xi, \mathbb{R}^p, \mu)$. Pratiquement, U^f sera un convexe fermé, obtenu la plupart du temps comme l'intersection d'un sous-espace vectoriel et d'un convexe. On notera génériquement $\Pi_{U^f} : L^2(\Xi, \mathbb{R}^p, \mu) \to U^f$ la projection sur U^f (cf. Proposition A.4).

On s'intéresse dès lors au problème :

$$\min_{u\in U^f} J(u) = \mathbb{E}\left(j(u(\boldsymbol{\xi}), \boldsymbol{\xi})\right). \tag{IV.4}$$

A supposer que j soit convexe par rapport à sa première composante (on notera alors $\nabla_u j$ un élément du sous-différentiel $\partial_u j$), on peut calculer un sous-gradient ∇J de J comme étant :

$$\forall u \in L^2(\Xi, \mathbb{R}^p, \mu),\ \nabla J(u)(\cdot) = \nabla_u j(u(\cdot), \cdot) \in L^2(\Xi, \mathbb{R}^p, \mu).$$

L'aléa du problème (I.1) n'est donc plus présent dans le gradient du critère qu'à travers la relation de μ-équivalence, c'est à dire à travers le support du gradient. Dans le cadre d'une caractérisation de l'optimalité d'un élément $u^* \in U^f$, comme l'équation (IV.2), l'aléa apparaît donc éventuellement à deux endroits :
- à travers le support de μ,
- à travers l'ensemble admissible U^f.

Ainsi, lorsque l'ensemble admissible U^f sera défini par exemple par des contraintes ponctuelles (du type $\theta(u(\xi), \xi) = 0$ pour μ-presque tout ξ, ou à l'image de l'ensemble U^{ponct} défini dans le chapitre I), seul le support de μ interviendra dans la définition de la commande optimale u^*, tandis que la loi complète μ interviendra dans le coût. Il convient donc dans la suite de faire attention à l'objet d'intérêt : dans bien des cas, si seule la commande présente un intérêt, il sera possible d'ignorer la loi de $\boldsymbol{\xi}$, pour peu que l'on connaisse son support, et d'obtenir tout de même la commande optimale pour (IV.4).

Ce qui nous préoccupe ici est de trouver un algorithme de descente pour résoudre le problème (IV.4), qui soit implémentable et qui converge vers la solution du problème. En effet, le problème étant naturellement posé dans l'espace fonctionnel $L^2(\Xi, \mathbb{R}^p, \mu)$, un algorithme de descente usuel du type (IV.1) ou même utilisant plus astucieusement l'aléa à l'image de l'algorithme (IV.3), respectivement convergera sans être implémentable dès lors que l'ensemble admissible n'est pas l'espace de Hilbert tout entier, ou sera implémentable sans converger...
Sur la base de la théorie de l'approximation fonctionnelle, nous proposons donc ici l'algorithme suivant :

Algorithme IV.4. *Étape k :*
- *Soit $\boldsymbol{\xi}^{k+1}$ indépendante des v.a. passées, et répartie selon la loi μ,*
- *Mettre à jour :*

$$\text{Soit } \boldsymbol{r}^k(\cdot) \in \partial J(\boldsymbol{u}^k)(\cdot),$$
$$\boldsymbol{u}^{k+1}(\cdot) = \Pi_{U^f}\left(\boldsymbol{u}^k(\cdot) - \rho^k \epsilon^k \boldsymbol{r}^k(\boldsymbol{\xi}^{k+1})\frac{1}{\epsilon^k} K^k(\boldsymbol{\xi}^{k+1}, \cdot)\right),$$

avec pour tout $k \in \mathbb{N}$, $K^k : \Xi \times \Xi \to \mathbb{R}$ une application bornée, et $\epsilon^k > 0$.
On peut remarquer, comme les notations le mettent en lumière, qu'à l'itération k, le contrôle $\boldsymbol{u}^k(\cdot)$ est doublement fonctionnel : d'une part, c'est un élément de $L^2(\Xi, \mathbb{R}^p, \mu)$, ce que souligne le "$\cdot$" dans la notation, et d'autre part c'est une variable aléatoire sur ω à travers les variables aléatoires $(\boldsymbol{\xi}^l)_{1\le l\le k+1}$ qui le définissent itérativement, d'où la notation en gras de ce contrôle.
Quelques remarques s'imposent immédiatement :
- L'algorithme IV.4, de par la présence de variables aléatoires i.i.d., est par essence un algorithme stochastique. Dans l'étude des algorithmes stochastiques, on considère classiquement la filtration $(\mathcal{F}^k)$ associée au processus $(\boldsymbol{\xi}^k)$. On regarde ensuite les espérances

conditionnelles des directions de descente successives par rapport à cette filtration. Usuellement, ces espérances conditionnelles sont les directions de descente théoriques, et l'on en déduit l'existence d'une martingale sous-jacente (la différence des termes précédents forme un incrément de martingale). C'est sur la base de cette martingale que l'on montre alors la convergence de l'algorithme. Ici, on a

$$\begin{aligned}\mathbb{E}\left(\boldsymbol{r}^k(\boldsymbol{\xi}^{k+1})\frac{1}{\epsilon^k}K^k(\boldsymbol{\xi}^{k+1},\cdot)\middle|\mathcal{F}^k\right)=&\mathbb{E}\left(\boldsymbol{r}^k(\boldsymbol{\xi})\frac{1}{\epsilon^k}K^k(\boldsymbol{\xi},\cdot)\middle|\boldsymbol{u}^k\right),\\ \neq&\nabla_u j(\boldsymbol{u}^k(\cdot),\cdot).\end{aligned}$$

Ces égalités sont bien entendu à prendre presque sûrement, comme égalités entre variables aléatoires. Ainsi, nos directions de descente successives ne forment pas des approximations non-biaisées de la véritable direction de descente, ce qui va nécessiter des conditions supplémentaires pour obtenir la convergence de l'algorithme.

– Par rapport à l'algorithme intuitif (IV.3), le terme $\frac{1}{\epsilon^k}K^k(\boldsymbol{\xi}^{k+1},\cdot)$ est donc venu en lieu et place de la masse de Dirac $\delta_{\boldsymbol{\xi}^{k+1}}(\cdot)$. Par analogie avec la théorie de l'estimation fonctionnelle, on appellera donc dans la suite les applications K^k des noyaux.
– Si chaque noyau K^k est connu par un nombre fini de paramètres (par exemple, si l'on prend K^k un noyau gaussien, déterminé entièrement par sa moyenne et sa variance), l'algorithme IV.4, tout en restant en dimension infinie, ne nécessite de mémoriser qu'un nombre fini de paramètres : à l'étape k, le contrôle $\boldsymbol{u}^k$ est connu comme une somme (intercalée éventuellement de projections) de noyaux, et chacun de ces noyaux est connu par un nombre fini de paramètres.
– Il reste encore à résoudre la question de l'implémentabilité, notamment en raison de la présence de la projection sur U^f dans l'algorithme IV.4.

IV.3.2. Résultat de convergence. Nous allons donner ici un résultat général de convergence de l'algorithme IV.4, recouvrant les cas d'ensemble admissible égal à un espace vectoriel ou tout simplement égal à un convexe fermé.

THÉORÈME IV.5. *(1) Supposons que pour μ-presque tout $\xi\in\mathbb{R}^m$, $u\mapsto j(u,\xi)$ est convexe, coercive, semi continue inférieurement, et à valeurs réelles. J est alors convexe et semi continue inférieurement, et pour tout $u\in U^f$, $\partial J(u)\neq\emptyset$. Supposons de plus que j est une intégrande normale sur U^f qui est un sous-ensemble convexe fermé de $L^2(\mathbb{R}^m,\mathbb{R}^p,\mu)$. Alors le problème* (IV.4) *admet un ensemble de solutions noté S.*

(2) En posant $(\mathcal{F}^k)$ la filtration définie pour tout $k\in\mathbb{N}$ par $\mathcal{F}^k=\sigma(u^0,\boldsymbol{\xi}^1\ldots,\boldsymbol{\xi}^k)$, alors l'algorithme IV.4 construit deux suites $(\boldsymbol{u}^k)$ et $(\boldsymbol{r}^k)$ adaptées à $(\mathcal{F}^k)$.

(3) Supposons que j est à sous-gradient en u linéairement borné, i.e. il existe deux réels $c,d>0$, tels que pour tout $u\in\mathbb{R}^p$,

$$\forall\xi\in\mathbb{R}^m,\ \forall v\in\partial_u j(u,\xi),\ \|v\|_{\mathbb{R}^p}\leq c\|u\|_{\mathbb{R}^p}+d, \tag{IV.5}$$

(4) Supposons qu'il existe deux réels $b_1,b_2>0$ tels que :

$$\forall k\in\mathbb{N},\ \left\|\boldsymbol{r}^k-\mathbb{E}\left(\boldsymbol{r}^k(\boldsymbol{\xi})\frac{1}{\epsilon^k}K^k(\boldsymbol{\xi},\cdot)\middle|\mathcal{F}^k\right)\right\|\leq b_1(\epsilon^k)^{1/m}\left(1+\|\boldsymbol{r}^k\|\right), \tag{IV.6a}$$

$$\forall x\in\mathbb{R}^m,\ \mathbb{E}\left((\frac{1}{\epsilon^k}K^k(x,\boldsymbol{\xi}))^2\right)\leq\frac{b_2}{\epsilon^k}, \tag{IV.6b}$$

Si U^f est un convexe fermé qui n'est pas un sous-espace vectoriel, supposons qu'il existe une application $g:\mathbb{R}\to\mathbb{R}$ continue ou bornée, telle que pour tout $k\in\mathbb{N}$,

$$\mathbb{E}\left(\left\|\boldsymbol{r}^k-\boldsymbol{r}^k(\boldsymbol{\xi}^{k+1})\frac{1}{\epsilon^k}K^k(\boldsymbol{\xi}^{k+1},\cdot)\right\|\middle|\mathcal{F}^k\right)\leq g(\|\boldsymbol{r}^k\|) \tag{IV.6c}$$

(5) Supposons que les suites (ϵ^k) et (ρ^k) soient telles que :

$$\epsilon^k, \rho^k > 0, \quad \sum_{k\in\mathbb{N}} \epsilon^k \rho^k = +\infty, \quad \sum_{k\in\mathbb{N}} \rho^k (\epsilon^k)^{1+\frac{1}{m}} < +\infty, \quad \sum_{k\in\mathbb{N}} (\rho^k)^2 \epsilon^k < +\infty. \tag{IV.7}$$

(6) Alors la suite $(\boldsymbol{u}^k)$ générée par l'algorithme IV.4 est telle que :

$$\lim_{k\to\infty} J(\boldsymbol{u}^k) = J(u^*), \text{ presque sûrement,}$$

avec $u^* \in U^*$. De plus, tout point d'accumulation de $(\boldsymbol{u}^k)$ dans la topologie faible de $L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)$ est dans S.

(7) En outre, si j est fortement convexe (en u) de module $B > 0$, alors S se réduit à un singleton, et $(\boldsymbol{u}^k)$ converge presque sûrement dans $L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)$ vers l'unique solution de (IV.4).

Preuve : La preuve suit le schéma introduit par Culioli dans sa thèse [**36**], et repris dans [**32**]. Le schéma est le suivant :

- Choisir une fonction de Lyapunov traduisant la convergence de l'algorithme par sa décroissance ;
- Majorer la variation de cette fonction de Lyapunov le long des itérations de l'algorithme ;
- Grâce à cette majoration et un argument de quasi-martingale, montrer la bornitude des itérés de l'algorithme ;
- Achever par des lemmes usuels de montrer la convergence des coûts et des argmins.

Notons u^* un élément de S. On définit alors une fonction de Lyapunov $\Lambda : L^2(\mathbb{R}^m, \mathbb{R}^p, \mu) \to \mathbb{R}$ par :

$$\forall u \in L^2(\mathbb{R}^m, \mathbb{R}^p, \mu),\ \Lambda(u) := \frac{1}{2}\|u - u^*\|^2.$$

On va maintenant étudier la variation de cette fonction de Lyapunov le long des itérations, et de là déduire la convergence de l'algorithme. Soit $k \in \mathbb{N}$. En posant

$$\begin{aligned}\delta^{k+1} := \Lambda(\boldsymbol{u}^{k+1}) - \Lambda(\boldsymbol{u}^k) =& \frac{1}{2}\|\boldsymbol{u}^{k+1} - u^*\|^2 - \frac{1}{2}\|\boldsymbol{u}^k - u^*\|^2,\\ =& \frac{1}{2}\|\boldsymbol{u}^{k+1} - \boldsymbol{u}^k\|^2 + \langle \boldsymbol{u}^{k+1} - \boldsymbol{u}^k, \boldsymbol{u}^k - u^*\rangle,\end{aligned}$$

on obtient cela par le théorème de Pythagore. De par la propriété de contraction de la projection, on obtient par définition de $\boldsymbol{u}^{k+1}$:

$$\|\boldsymbol{u}^{k+1} - \boldsymbol{u}^k\| \le \rho^k \|\boldsymbol{r}(\boldsymbol{\xi}^{k+1}) K^k(\boldsymbol{\xi}^{k+1}, \cdot)\|.$$

Soient $G^k(\cdot,\cdot) = \frac{1}{\epsilon^k} K^k(\cdot,\cdot)$, et $\boldsymbol{f}^k(\cdot) = \epsilon^k \boldsymbol{r}^k(\boldsymbol{\xi}^{k+1}) G^k(\boldsymbol{\xi}^{k+1}, \cdot)$. On étudie $\langle \boldsymbol{u}^{k+1} - \boldsymbol{u}^k, \boldsymbol{u}^k - u^*\rangle$. Pour simplifier les notations, on écrira Π la projection sur U^f. On distingue alors deux cas :

- Si U^f est un sous-espace vectoriel, Π est linéaire et auto-adjoint. On obtient donc :

$$\begin{aligned}\langle \boldsymbol{u}^{k+1} - \boldsymbol{u}^k, \boldsymbol{u}^k - u^*\rangle =& \langle \Pi\left(\boldsymbol{u}^k - \rho^k \boldsymbol{f}^k\right) - \boldsymbol{u}^k, \boldsymbol{u}^k - u^*\rangle,\\ =& -\rho^k \epsilon^k \langle \Pi\left(\boldsymbol{r}^k(\boldsymbol{\xi}^{k+1}) G^k(\boldsymbol{\xi}^{k+1}, \cdot)\right), \boldsymbol{u}^k - u^*\rangle,\\ =& -\rho^k \epsilon^k \langle \boldsymbol{r}^k(\boldsymbol{\xi}^{k+1}) G^k(\boldsymbol{\xi}^{k+1}, \cdot), \boldsymbol{u}^k - u^*\rangle.\end{aligned}$$

- Si U^f est un convexe fermé, le travail est un peu plus long. En utilisant la caractérisation de la projection (cf. Proposition A.4), on obtient :

$$\begin{aligned}\langle \boldsymbol{u}^{k+1} - \boldsymbol{u}^k, \boldsymbol{u}^k - u^*\rangle =& \langle \Pi\left(\boldsymbol{u}^k - \rho^k \boldsymbol{f}^k\right) - \boldsymbol{u}^k, \boldsymbol{u}^k - u^*\rangle,\\ =& \langle \Pi\left(\boldsymbol{u}^k - \rho^k \boldsymbol{f}^k\right) - \left(\boldsymbol{u}^k - \rho^k \boldsymbol{f}^k\right), \boldsymbol{u}^k - u^*\rangle - \rho^k \langle \boldsymbol{f}^k, \boldsymbol{u}^k - u^*\rangle,\\ =& \underbrace{\langle \Pi\left(\boldsymbol{u}^k - \rho^k \boldsymbol{f}^k\right) - \left(\boldsymbol{u}^k - \rho^k \boldsymbol{f}^k\right), \boldsymbol{u}^k - \rho^k \boldsymbol{f}^k - u^*\rangle}_{\le 0, \text{ par propriété de la projection,}}\\ & + \rho^k \langle \boldsymbol{f}^k, \Pi\left(\boldsymbol{u}^k - \rho^k \boldsymbol{f}^k\right) - \left(\boldsymbol{u}^k - \rho^k \boldsymbol{f}^k\right)\rangle - \rho^k \langle \boldsymbol{f}^k, \boldsymbol{u}^k - u^*\rangle\\ \le& (\rho^k \epsilon^k)^2 \|\boldsymbol{r}^k(\boldsymbol{\xi}^{k+1}) G^k(\boldsymbol{\xi}^{k+1}, \cdot)\|^2 - \rho^k \langle \boldsymbol{f}^k, \boldsymbol{u}^k - u^*\rangle\end{aligned}$$

Par conséquent, on a dans tous les cas

$$\begin{aligned}\delta^{k+1} \leq & \frac{3(\rho^k\epsilon^k)^2}{2}\|\boldsymbol{r}^k(\boldsymbol{\xi}^{k+1})G^k(\boldsymbol{\xi}^{k+1},\cdot)\|^2 - \rho^k\epsilon^k\langle \boldsymbol{r}^k(\boldsymbol{\xi}^{k+1})G^k(\boldsymbol{\xi}^{k+1},\cdot), \boldsymbol{u}^k - u^*\rangle,\\ \leq & \frac{3b_2(\rho^k)^2\epsilon^k}{2}\|\boldsymbol{r}^k(\boldsymbol{\xi}^{k+1})\|^2_{\mathbb{R}^p} + \rho^k\epsilon^k\langle \boldsymbol{r}^k - \boldsymbol{r}^k(\boldsymbol{\xi}^{k+1})G^k(\boldsymbol{\xi}^{k+1},\cdot), \boldsymbol{u}^k - u^*\rangle\\ & + \rho^k\epsilon^k\langle \boldsymbol{r}^k, u^* - \boldsymbol{u}^k\rangle. \end{aligned} \tag{IV.8}$$

La deuxième inégalité vient de l'hypothèse (IV.6b) sur les noyaux K^k. En utilisant la convexité de J, on a :

$$\langle \boldsymbol{r}^k, u^* - \boldsymbol{u}^k\rangle \leq J(u^*) - J(\boldsymbol{u}^k) \leq 0. \tag{IV.9}$$

En rassemblant alors les équations (IV.8) et (IV.9), on trouve :

$$\begin{aligned}\Lambda(\boldsymbol{u}^{k+1}) - \Lambda(\boldsymbol{u}^k) \leq & \frac{3b_2(\rho^k)^2\epsilon^k}{2}\|\boldsymbol{r}^k(\boldsymbol{\xi}^{k+1})\|^2_{\mathbb{R}^p} + \rho^k\epsilon^k\langle \boldsymbol{r}^k - \boldsymbol{r}^k(\boldsymbol{\xi}^{k+1})G^k(\boldsymbol{\xi}^{k+1},\cdot), \boldsymbol{u}^k - u^*\rangle\\ & + \rho^k\epsilon^k\left(J(u^*) - J(\boldsymbol{u}^k)\right). \end{aligned} \tag{IV.10}$$

En conditionnant l'équation (IV.10) par rapport à la tribu $\mathcal{F}^k := \sigma(\boldsymbol{\xi}^1, \dots, \boldsymbol{\xi}^k)$, engendrée par les tirages passés, comme $\boldsymbol{\xi}^{k+1}$ est indépendant des tirages passés, on obtient

$$\begin{aligned}\mathbb{E}\left(\Lambda(\boldsymbol{u}^{k+1}) - \Lambda(\boldsymbol{u}^k)|\mathcal{F}^k\right) \leq & \frac{3b_2(\rho^k)^2\epsilon^k}{2}\|\boldsymbol{r}^k\|^2 + \rho^k\epsilon^k\langle \boldsymbol{r}^k - \mathbb{E}\left(\boldsymbol{r}^k(\boldsymbol{\xi})G^k(\boldsymbol{\xi},\cdot)|\mathcal{F}^k\right), \boldsymbol{u}^k - u^*\rangle\\ & + \rho^k\epsilon^k\left(J(u^*) - J(\boldsymbol{u}^k)\right). \end{aligned} \tag{IV.11}$$

Par l'hypothèse de sous-gradient linéairement borné sur J, on obtient également, en utilisant l'inégalité scalaire classique $(a+b)^2 \leq 2a^2 + 2b^2$, que :

$$\|\boldsymbol{r}^k\|^2 \leq c_1\|\boldsymbol{u}^k - u^*\|^2 + c_2,$$

avec c_1, c_2 deux réels strictement positifs. Ainsi,

$$\begin{aligned}\mathbb{E}\left(\Lambda(\boldsymbol{u}^{k+1}) - \Lambda(\boldsymbol{u}^k)|\mathcal{F}^k\right) \leq & \frac{3b_2c_1(\rho^k)^2\epsilon^k}{2}\|\boldsymbol{u}^k - u^*\|^2 + \frac{3b_2c_2(\rho^k)^2\epsilon^k}{2}\\ & + \rho^k\epsilon^k\langle \boldsymbol{r}^k - \mathbb{E}\left(\boldsymbol{r}^k(\boldsymbol{\xi})G^k(\boldsymbol{\xi},\cdot)|\mathcal{F}^k\right), \boldsymbol{u}^k - u^*\rangle\\ & + \rho^k\epsilon^k\left(J(u^*) - J(\boldsymbol{u}^k)\right). \end{aligned} \tag{IV.12}$$

On utilise maintenant l'inégalité de Cauchy-Schwarz dans l'équation (IV.12) :

$$\begin{aligned}\mathbb{E}\left(\Lambda(\boldsymbol{u}^{k+1}) - \Lambda(\boldsymbol{u}^k)|\mathcal{F}^k\right) \leq & \frac{3b_2c_1(\rho^k)^2\epsilon^k}{2}\|\boldsymbol{u}^k - u^*\|^2 + \frac{3b_2c_2(\rho^k)^2\epsilon^k}{2}\\ & + \rho^k\epsilon^k\|\boldsymbol{r}^k - \mathbb{E}\left(\boldsymbol{r}^k(\boldsymbol{\xi})G^k(\boldsymbol{\xi},\cdot)|\mathcal{F}^k\right)\|\|\boldsymbol{u}^k - u^*\|\\ & + \rho^k\epsilon^k\left(J(u^*) - J(\boldsymbol{u}^k)\right). \end{aligned} \tag{IV.13}$$

Par l'hypothèse (IV.6a), il vient :

$$\begin{aligned}\mathbb{E}\left(\Lambda(\boldsymbol{u}^{k+1}) - \Lambda(\boldsymbol{u}^k)|\mathcal{F}^k\right) \leq & \frac{3b_2c_1(\rho^k)^2\epsilon^k}{2}\|\boldsymbol{u}^k - u^*\|^2 + \frac{3b_2c_2(\rho^k)^2\epsilon^k}{2}\\ & + b_1\rho^k(\epsilon^k)^{1+\frac{1}{m}}\left(1 + \|\boldsymbol{r}^k\|\right)\|\boldsymbol{u}^k - u^*\|\\ & + \rho^k\epsilon^k\left(J(u^*) - J(\boldsymbol{u}^k)\right). \end{aligned} \tag{IV.14}$$

L'hypothèse (IV.5) implique qu'il existe deux réels $c_3, c_4 > 0$ tels que :

$$\forall u \in L^2(\mathbb{R}^m, \mathbb{R}^p, \mu),\ \forall v \in \partial J(u),\ \|v\| \leq c_3\|u\| + c_4.$$

Cette dernière inégalité, avec l'inégalité scalaire classique $x \leq x^2 + 1$, donne :

$$\begin{aligned}\mathbb{E}\left(\Lambda(\boldsymbol{u}^{k+1}) - \Lambda(\boldsymbol{u}^k)|\mathcal{F}^k\right) \leq & \frac{3b_2c_1(\rho^k)^2\epsilon^k}{2}\|\boldsymbol{u}^k - u^*\|^2 + \frac{3b_2c_2(\rho^k)^2\epsilon^k}{2}\\ & + b_1\rho^k(\epsilon^k)^{1+\frac{1}{m}}(1 + c_3 + c_4)\|\boldsymbol{u}^k - u^*\|^2 + b_1\rho^k(\epsilon^k)^{1+\frac{1}{m}}(1 + c_4)\\ & + \rho^k\epsilon^k\left(J(u^*) - J(\boldsymbol{u}^k)\right). \end{aligned}$$

Par définition de Λ, on trouve finalement :

$$\mathbb{E}\left(\Lambda(\boldsymbol{u}^{k+1}) - \Lambda(\boldsymbol{u}^k)|\mathcal{F}^k\right) \leq \alpha^k\Lambda(\boldsymbol{u}^k) + \beta^k + \rho^k\epsilon^k\left(J(u^*) - J(\boldsymbol{u}^k)\right), \tag{IV.15}$$

avec $\alpha^k := 3b_2c_1(\rho^k)^2\epsilon^k + 2b_1\rho^k(\epsilon^k)^{1+\frac{1}{m}}(1+c_3+c_4)$ et $\beta^k := \frac{3b_2c_2(\rho^k)^2\epsilon^k}{2} + b_1\rho^k(\epsilon^k)^{1+\frac{1}{m}}(1+c_4)$. (α^k) et (β^k) sont par hypothèse deux séries sommables. En prenant l'espérance dans l'équation (IV.15), et définissant $y^k := \mathbb{E}\left(\Lambda(\boldsymbol{u}^k)\right)$, on trouve, grâce à l'optimalité de u^*,

$$y^{k+1} - y^k \leq \alpha^k y^k + \beta^k. \tag{IV.16}$$

Le lemme A.33 montre alors la suite (y^k) est bornée, par une constante $M > 0$. On prouve maintenant que $(\Lambda(\boldsymbol{u}^k))$ est une quasi-martingale convergente (cf. Définition B.4) :

- $(\Lambda(\boldsymbol{u}^k))$ est par définition adaptée à $(\mathcal{F}^k)$.
- Par définition, $\Lambda(\boldsymbol{u}^k) \geq 0$ pour tout $k \in \mathbb{N}$, i.e., $\inf_{k\in\mathbb{N}} \mathbb{E}\left(\Lambda(\boldsymbol{u}^k)\right) > -\infty$.
- Posons $C_k := \{\mathbb{E}\left(\Lambda(\boldsymbol{u}^{k+1}) - \Lambda(\boldsymbol{u}^k)|\mathcal{F}^k\right) > 0\}$. Il est clair que 1_{C_k} est $\mathcal{F}^k$-mesurable. En utilisant l'inégalité (IV.15), on obtient :

$$\begin{aligned}
\sum_{k\in\mathbb{N}} \mathbb{E}\left(1_{C_k} \times (\Lambda(\boldsymbol{u}^{k+1}) - \Lambda(\boldsymbol{u}^k))\right) &\leq \sum_{k\in\mathbb{N}} \mathbb{E}\left(1_{C_k} \times \mathbb{E}\left(\Lambda(\boldsymbol{u}^{k+1}) - \Lambda(\boldsymbol{u}^k)|\mathcal{F}^k\right)\right), \\
&\leq \sum_{k\in\mathbb{N}} \mathbb{E}\left(1_{C_k} \times (\alpha^k\Lambda(\boldsymbol{u}^k) + \beta^k)\right), \\
&\leq \sum_{k\in\mathbb{N}} (\alpha^k M + \beta^k), \\
&< +\infty,
\end{aligned}$$

  car les séries sont sommables.
- Il est également clair que $\sup_{k\in\mathbb{N}} \mathbb{E}\left(\Lambda(\boldsymbol{u}^k)^-\right) < \infty$, et par conséquent, en utilisant le théorème B.6, la suite de variables aléatoires $(\Lambda(\boldsymbol{u}^k))$ est une quasi-martingale et converge presque sûrement vers une variable aléatoire intégrable. Cette suite est donc presque sûrement bornée, et par définition, $(\boldsymbol{u}^k)$ et $(\boldsymbol{r}^k)$ sont donc deux suites presque sûrement bornées dans $L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)$.

On prouve maintenant que $(J(\boldsymbol{u}^k))$ converge presque sûrement vers $J(u^*)$. En revenant à l'équation (IV.15) et prenant l'espérance, on obtient :

$$\rho^k\epsilon^k\mathbb{E}\left(J(\boldsymbol{u}^k) - J(u^*)\right) \leq \alpha^k y^k + \beta^k + y^k - y^{k+1}.$$

On ajoute ces inégalités pour $k = 0, \ldots, n$:

$$\begin{aligned}
\sum_{k=0}^{n} \rho^k\epsilon^k\mathbb{E}\left(J(\boldsymbol{u}^k) - J(u^*)\right) &\leq y^0 - y^{n+1} + \sum_{k=0}^{n}(\alpha^k M + \beta^k), \\
&\leq M + M\sum_{k=0}^{n}\alpha^k + \sum_{k=0}^{n}\beta^k.
\end{aligned} \tag{IV.17}$$

En faisant alors $n \to \infty$ on trouve :

$$\sum_{k\in\mathbb{N}} \rho^k\epsilon^k\mathbb{E}\left(J(\boldsymbol{u}^k) - J(u^*)\right) < \infty.$$

Par optimalité, chacun des termes de la série sous l'espérance est positif. Dès lors :

$$\sum_{k\in\mathbb{N}} \rho^k\epsilon^k\left(J(\boldsymbol{u}^k) - J(u^*)\right) < \infty. \tag{IV.18}$$

On va maintenant utiliser le lemme A.35. Soit $l \in \mathbb{N}$. Par convexité, on a

$$J(\boldsymbol{u}^l) - J(\boldsymbol{u}^{l+1}) \leq \langle \boldsymbol{r}^l, \boldsymbol{u}^l - \boldsymbol{u}^{l+1}\rangle. \tag{IV.19}$$

Nous allons maintenant distinguer selon les cas

- U^f est un sous-espace vectoriel. On a alors

$$\begin{aligned}
J(\boldsymbol{u}^l) - J(\boldsymbol{u}^{l+1}) &\leq \langle \boldsymbol{r}^l, \boldsymbol{u}^l - \boldsymbol{u}^{l+1}\rangle, \\
&= \rho^l\epsilon^l\langle \boldsymbol{r}^l, \Pi_{U^f}\left(\boldsymbol{r}^l(\boldsymbol{\xi}^{l+1})G^l(\boldsymbol{\xi}^{l+1}, \cdot)\right)\rangle.
\end{aligned} \tag{IV.20}$$

En conditionnant par rapport à $\mathcal{F}^l$, il vient :

$$\begin{aligned} J(\boldsymbol{u}^l) - \mathbb{E}\left(J(\boldsymbol{u}^{l+1})|\mathcal{F}^l\right) \leq & \rho^l\epsilon^l\langle \boldsymbol{r}^l, \Pi_{U^f}\left(\mathbb{E}\left(\boldsymbol{r}^l(\boldsymbol{\xi})\frac{1}{\epsilon^l}K^l(\boldsymbol{\xi},\cdot)\middle|\mathcal{F}^l\right)\right)\rangle, \\ \leq & \rho^l\epsilon^l\|\boldsymbol{r}^l\|\left(\|\mathbb{E}\left(\boldsymbol{r}^l(\boldsymbol{\xi})\frac{1}{\epsilon^l}K^l(\boldsymbol{\xi},\cdot)\middle|\mathcal{F}^l\right) - \boldsymbol{r}^l\| + \|\boldsymbol{r}^l\|\right), \\ \leq & \rho^l\epsilon^l R\left(b_1(\epsilon^l)^{1/m}(1+R) + R\right), \\ \leq & \rho^l\epsilon^l\delta, \end{aligned} \tag{IV.21}$$

avec $\delta > 0$, car l'on sait déjà que la suite $(\|\boldsymbol{r}^k\|)$ est bornée par un réel $R > 0$.

– U^f est un convexe fermé et non un sous-espace vectoriel. Par l'inégalité de Cauchy-Schwarz et la propriété de contraction de la projection, l'équation (IV.19) implique

$$\begin{aligned} J(\boldsymbol{u}^l) - J(\boldsymbol{u}^{l+1}) \leq & \rho^l\epsilon^l\|\boldsymbol{r}^l\| \times \|\boldsymbol{r}^l(\boldsymbol{\xi}^{l+1})\frac{1}{\epsilon^l}K^l(\boldsymbol{\xi}^{l+1},\cdot)\|, \\ \leq & \rho^l\epsilon^l\|\boldsymbol{r}^l\|\left(\|\boldsymbol{r}^l\| + \|\boldsymbol{r}^l - \boldsymbol{r}^l(\boldsymbol{\xi}^{l+1})\frac{1}{\epsilon^l}K^l(\boldsymbol{\xi}^{l+1},\cdot)\|\right) \end{aligned}$$

D'où, en prenant l'espérance conditionnelle par rapport à $\mathcal{F}^l$, et en utilisant l'hypothèse (IV.6c) :

$$J(\boldsymbol{u}^l) - \mathbb{E}\left(J(\boldsymbol{u}^{l+1})|\mathcal{F}^l\right) \leq \rho^l\epsilon^l\delta \tag{IV.22}$$

avec $\delta > 0$, car l'on sait que la suite $(\|\boldsymbol{r}^k\|)$ est presque sûrement bornée.

On peut donc appliquer le lemme A.35, grâce aux inégalités (IV.18) et (IV.21)–(IV.22), en posant $\gamma^k = \epsilon^k\rho^k$. On en déduit :

$$\lim_{k\to\infty} J(\boldsymbol{u}^k) = J(u^*) \tag{IV.23}$$

Comme $(\boldsymbol{u}^k)$ est presque sûrement bornée, l'ensemble des points d'accumulation de cette suite dans la topologie faible est non vide. Soit $\bar{\boldsymbol{u}}$ un tel point d'accumulation dans la topologie faible. Il existe donc une sous-suite $(\boldsymbol{u}^{\phi(k)})$ qui converge faiblement vers $\bar{\boldsymbol{u}}$. Comme U^f est un sous-espace fermé, $\bar{\boldsymbol{u}} \in U^f$, et par semi continuité inférieure de J dans la topologie faible (ce qui provient de la convexité et de la semi continuité forte), il vient :

$$J(\bar{\boldsymbol{u}}) \leq \liminf_{k\to\infty} J(\boldsymbol{u}^{\phi(k)}) = J(u^*),$$

ainsi, $\bar{\boldsymbol{u}} \in S$.

Si maintenant j est fortement convexe de module $B > 0$, S est réduit alors à un singleton $\{u^*\}$. Par définition, en notant $r^* \in \partial J(u^*)$,

$$J(\boldsymbol{u}^k) - J(u^*) \geq \langle r^*, \boldsymbol{u}^k - u^*\rangle + \frac{B}{2}\|u^* - \boldsymbol{u}^k\|^2 \tag{IV.24}$$

Par optimalité $\langle r^*, \boldsymbol{u}^k - u^*\rangle \geq 0$. (IV.23) donne donc la convergence dans $L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)$ de $(\boldsymbol{u}^k)$ vers u^*, ce qui complète la preuve. □

Remarque IV.6 (Kiefer-Wolfowitz). *Dans le cadre des algorithmes stochastiques en dimension finie, Kiefer et Wolfowitz (cf.* [**63**]*) ont introduit un algorithme de gradient stochastique bruité, dans le sens où le vrai gradient, évalué en un tirage, était remplacé par deux évaluations du type différences finies. Ce bruitage du gradient nécessitait alors l'introduction, en plus du pas de descente usuel de l'algorithme (ici $\rho^k\epsilon^k$), d'un pas d'approximation par différences finies, analogue en un certain sens à notre pas de convolution ϵ^k. Les hypothèses conjointes faites sur le pas de descente et le pas de différences finies par Kiefer et Wolfowitz coïncident avec les nôtres.*

IV.3.3. Résultat de grandes déviations. Sur la base des résultats de [**74**], nous donnons maintenant un résultat de grande déviation pour l'algorithme IV.4 en l'absence de projections et dans le cas d'un critère fortement convexe.

Proposition IV.7. *(i) Nous nous plaçons sous les mêmes hypothèses que le théorème IV.5, et supposons en plus que pour tout $\xi \in \mathbb{R}^m$, $u \mapsto j(u,\xi)$ est fortement convexe de module B, différentiable et que $U^f = L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)$. Le problème* (IV.4) *admet donc une solution unique u^* vers laquelle la suite $(\boldsymbol{u}^k)$ générée par l'algorithme IV.4 converge fortement presque sûrement.*

(ii) Supposons aussi que pour tout $k \in \mathbb{N}$, $\rho^k \epsilon^k = 1/k$.
(iii) Enfin, supposons à la place de (IV.6a), *que*

$$\forall u \in L^2(\mathbb{R}^m, \mathbb{R}^p, \mu),\ \forall r \in \partial J(u),\ \left\| r - \mathbb{E}\left(r(\boldsymbol{\xi}) \frac{1}{\epsilon^k} K^k(\boldsymbol{\xi}, \cdot) \right) \right\| \leq b_1 (\epsilon^k)^{1/m} \left(1 + \|r\|\right). \tag{IV.25}$$

(iii) Alors, pour tout $\epsilon > 0$, *il existe* $\alpha > 0, C > 0$ *et* $k_0 \in \mathbb{N}$ *tels que*

$$\forall k \geq k_0,\ \mathbb{P}\left(\|\boldsymbol{u}^k - u^*\| \geq \epsilon \right) \leq \exp\left(-Ck\right).$$

Preuve : La preuve est une application du Théorème 1 de [**74**]. Sans perte de généralité, supposons que $u^* = 0$. Pour tout $k \in \mathbb{N}$, définissons l'opérateur $M_k : L^2(\mathbb{R}^m, \mathbb{R}^p, \mu) \to L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)$ par

$$\forall u \in L^2(\mathbb{R}^m, \mathbb{R}^p, \mu),\ M_k(u)(\cdot) = \mathbb{E}\left(\nabla_u j(u(\boldsymbol{\xi}^{k+1}), \boldsymbol{\xi}^{k+1}) \frac{1}{\epsilon^k} K^k(\boldsymbol{\xi}^{k+1}, \cdot) \middle| \mathcal{F}^k \right),$$

et la suite de variables aléatoires $\boldsymbol{y}_k(\cdot) = \nabla_u j(\boldsymbol{u}^k(\boldsymbol{\xi}^{k+1}), \boldsymbol{\xi}^{k+1}) \frac{1}{\epsilon^k} K^k(\boldsymbol{\xi}^{k+1}, \cdot) - M_k(\boldsymbol{u}^k)(\cdot)$. Avec ces notations, l'algorithme IV.4 se réécrit simplement :

$$\boldsymbol{u}^{k+1} = \boldsymbol{u}^k - \frac{M_k(\boldsymbol{u}^k)}{k} - \frac{\boldsymbol{y}_k}{k}.$$

Par construction, il est clair que pour tout $k \in \mathbb{N}$, $\mathbb{E}\left(\boldsymbol{y}_k | \mathcal{F}^k\right) = 0$. De même, il existe deux réels $A, B \geq 0$ tels que $\|M_k(u)\| \leq A\|u\| + B$. En effet :

$$\begin{aligned}
\|M_k(u)\| \leq & \|M_k(u) - \nabla J(u)\| + \|\nabla J(u)\|, \\
\leq & b_1(\epsilon^k)^{1/m} + \|\nabla J(u)\| \left(1 + b_1(\epsilon^k)^{1/m}\right), \text{ par l'hypothèse (IV.25)} \\
\leq & b_1(\epsilon^k)^{1/m} + \left(1 + b_1(\epsilon^k)^{1/m}\right)(c\|u\| + d), \text{ par l'hypothèse (IV.5)} \\
\leq & A\|u\| + B, \text{ car } \epsilon^k \text{ tend vers 0.}
\end{aligned}$$

Enfin, en suivant la preuve du théorème IV.5, on sait qu'il existe $R > 0$ tel que $\|\boldsymbol{y}_k\| \leq R$ presque sûrement.
Soit alors $u \in L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)$.

$$\begin{aligned}
\langle M_k(u), u \rangle = & \langle \nabla_u j(u(\cdot), \cdot), u \rangle - \langle \nabla_u j(u(\cdot), \cdot) - M_k(u), u \rangle, \\
\geq & \frac{B}{2}\|u\|^2 - \|\nabla_u j(u(\cdot), \cdot) - M_k(u)\| \|u\|, \text{ par forte convexité et Cauchy-Schwarz„} \\
\geq & \frac{B}{2}\|u\|^2 - b_1(\epsilon^k)^{1/m} \left(\|\nabla_u j(u(\cdot), \cdot)\| + 1\right) \|u\|, \text{ par l'hypothèse (IV.25),} \\
\geq & \frac{B}{2}\|u\|^2 - b_1(\epsilon^k)^{1/m} \left(c\|u\| + d + 1\right) \|u\|, \text{ par l'hypothèse (IV.5),} \\
\geq & \left(\frac{B}{2} - c b_1 (\epsilon^k)^{1/m} \right) \|u\|^2 - b_1(d+1)(\epsilon^k)^{1/m} \|u\|. \qquad \text{(IV.26)}
\end{aligned}$$

Les hypothèses (IV.7) impliquent que $\left((\epsilon^k)^{1/m}\right)$ tend vers 0. Ainsi, pour tout $\beta > 0$, il existe k_0 tel que pour tout $k \geq k_0$,

$$\langle M_k(u), u \rangle \geq \left(\frac{B}{2} - c b_1 \beta \right) \|u\|^2 - b_1(d+1)\beta \|u\|. \tag{IV.27}$$

Soit alors $\epsilon > 0$, tel que $\frac{B}{2} - c b_1 \epsilon^2 - 2 b_1 (d+1)\epsilon > 0$. En posant $\beta = \sqrt{\epsilon}$ et $\alpha = \frac{B}{2} - c b_1 \epsilon^2 - 2 b_1 (d+1)\epsilon$, il vient alors de l'équation (IV.27) que

$$\forall u \in L^2(\mathbb{R}^m, \mathbb{R}^p, \mu),\ \text{ tel que } \|u\| \geq \epsilon/2,\ \langle M_k(u), u \rangle \geq \alpha \|u\|^2,$$

ce qui correspond à l'hypothèse de coercivité du théorème 1 de [**74**]. On peut donc appliquer ce théorème, ce qui achève la preuve. □

Nous allons donner dans la section suivante diverses illustrations des hypothèses du théorème et de l'intérêt de l'algorithme IV.4.

IV.3.4. Remarques et illustrations diverses.

IV.3.4.1. *Pas décroissants.* Pour commencer, on peut faire quelques remarques sur les pas. Il y a ici deux pas de natures différentes, notés (ρ^k) et (ϵ^k). Le pas $(\rho^k\epsilon^k)$ peut être interprété comme le pas de descente usuel d'un algorithme de gradient : de par la nature stochastique de notre algorithme, il est donc naturel de le prendre décroissant, et en série divergente. Les pas (ϵ^k) sont quant à eux moins usuels. Théoriquement, ce sont les pas d'approximation fonctionnelle de la masse de Dirac par les noyaux successifs. De manière plus imagée, ils représentent le rayon du voisinage dans lequel l'information ponctuelle de gradient est étendue. Intuitivement, les pas (ϵ^k) doivent donc décroître pour que les noyaux tendent vers la masse de Dirac, mais pas trop vite pour que les voisinages successifs d'actualisation recouvrent bien tout l'espace. Enfin, on s'attend à avoir des relations liant les pas de descente et les pas d'approximation, et ce sont par exemple les hypothèses (IV.7).

Bien entendu, l'ensemble des suites vérifiant ces hypothèses (IV.7) n'est pas vide : prenons par exemple la famille de pas $\rho^k = k^{-\alpha}$ et $\epsilon^k = k^{-\beta}$, avec $\beta, \alpha > 0$. Les hypothèses (IV.7) s'écrivent alors :

$$\alpha + \beta \leq 1, \quad 2\alpha + \beta > 1, \quad \alpha + \beta\left(1 + \frac{1}{m}\right) > 1,$$

et il est donc clair qu'une infinité de suites les satisfont. Par exemple, si $\alpha = \frac{1}{2}$, on peut prendre $\beta \in \left(\frac{1/2}{1+1/m}, \frac{1}{2}\right]$ et les hypothèses (IV.7) seront vérifiées. Ce qui apparaît clairement est l'impact de la dimension : plus la dimension de l'espace d'aléa est grande, et moins l'on peut jouer avec le choix des pas de l'algorithme. La figure 2 permet de visualiser dans le plan (α, β) les choix possibles pour ces valeurs. Il est clair que lorsque m croît, le triangle des choix possibles tend à s'écraser contre le segment $\alpha + \beta = 1$.

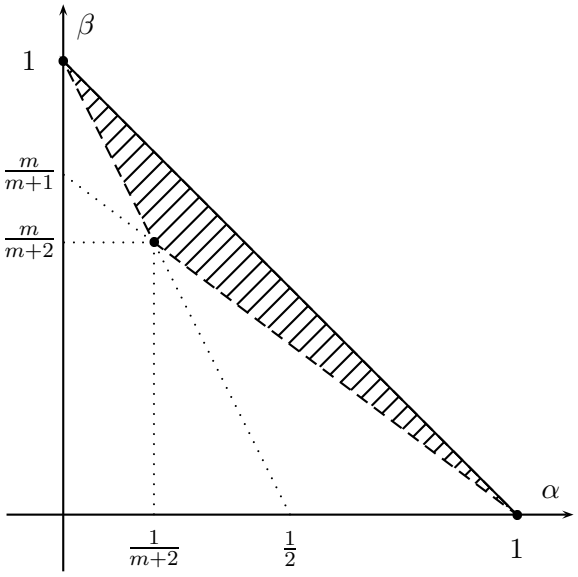


FIG. 2. Ensemble admissible des exposants des suites décroissantes en fonction de la dimension de l'espace d'aléas.

Afin d'illustrer une fois de plus les hypothèses (IV.7), nous donnons maintenant un exemple dans lequel nous fixons la manière de faire décroître les supports des noyaux K^k vers 0 par un choix de la suite (ϵ^k), et observons les conséquences de ce choix sur la suite (ρ^k). Prenons $m = 1$, et μ la loi uniforme sur $[0, 1]$. De par la présence de μ-équivalence partout, les contrôles recherchés sont donc recherchés comme des fonctions sur $[0, 1]$. Soit $\epsilon^k = 1/(k+1)$, et définissons les noyaux comme étant $K^k(x, y) = 1_{|x-y|\leq\epsilon^k/2}$ pour tout $x, y \in [0, 1]$.

Définissons maintenant les indices j_n tels que $j_0 = 0$, et pour $n \geq 1$, j_n est tel que :

$$\sum_{k=j_{n-1}}^{j_n-1} \epsilon^k \leq 1, \quad \sum_{k=j_{n-1}}^{j_n} \epsilon^k > 1.$$

Comme $\sum_{k\in\mathbb{N}} \epsilon^k = +\infty$, cette suite d'indices est bien définie. Soit alors (ξ^k) une suite de réels dans $[0,1]$ telle que pour tout $k \in \mathbb{N}$, $\xi^{j_k+1} = \frac{\epsilon^{j_k}}{2}$, $\xi^{j_k+2} = \frac{\epsilon^{j_k+1}+\epsilon^{j_k}}{2} + \xi^{j_k+1}$, $\xi^{j_k+3} = \frac{\epsilon^{j_k+2}+\epsilon^{j_k+1}}{2} + \xi^{j_k+2}$, $\ldots$, $\xi^{j_{k+1}} = \sum_{n=j_k}^{j_{k+1}-2} \epsilon^n + \frac{\epsilon^{j_{k+1}-1}}{2}$, $\xi^{j_{k+1}+1} = \frac{\epsilon^{j_{k+1}}}{2}$, et ainsi de suite. Cette suite est également bien définie. La figure 3 illustre cette construction entre les indices j_k et j_{k+1}. Finalement, nous avons construit ici un pavage du type Quasi-Monte Carlo uniforme de l'intervalle $[0,1]$.

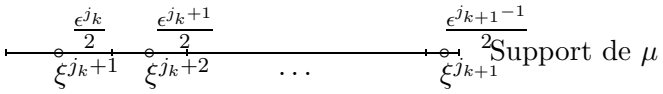


FIG. 3. Couverture de type quasi-Monte Carlo de $[0,1]$

Soit (ρ^k) une suite de réels positifs. Considérons l'algorithme :

$$\forall \xi \in \Xi,\ u^{k+1}(\xi) = u^k(\xi) - \rho^k f(\xi^{k+1}) K^k(\xi^{k+1}, \xi),$$

avec $f : [0,1] \to \mathbb{R}$ une application lipschitzienne de module L. Par définition des noyaux K^k, cet algorithme ne modifie donc l'itéré courant u^k que dans une boule de rayon $\epsilon^k/2$ centrée sur ξ^{k+1}. Cet algorithme est une version Quasi-Monte Carlo de l'algorithme IV.4.
Nous étudions maintenant cet algorithme au point $\xi = 0$. Pour simplifier les notations, on appelera $u^k(0)$, v^k. L'algorithme s'écrit donc en 0 :

$$v^{j_{k+1}} = v^{j_k} - \rho^{j_k} r^{j_k}, \tag{IV.28}$$

avec $r^{j_k} = f(\xi^{j_{k+1}}) K^{j_{k+1}-1}(\xi^{j_{k+1}}, 0)$. En effet, v^l ne bouge pas entre les indices j_k et j_{k+1} grâce à notre construction. Remarquons pour commencer que r^{j_k} est en fait une perturbation de la fonction f en 0. Afin de faire converger l'algorithme (IV.28), nous allons utiliser un théorème général de convergence des algorithmes stochastiques (par exemple, l'un des théorèmes prouvés dans [**19**]). Une condition usuelle sur les pas est la suivante :

$$\sum_{k\in\mathbb{N}} \rho^{j_k} = +\infty, \quad \sum_{k\in\mathbb{N}} (\rho^{j_k})^2 < +\infty. \tag{IV.29}$$

Le choix $\rho^{j_k} = \frac{1}{k}$, serait donc approprié. Par définition, on a $\sum_{n=j_k}^{j_{k+1}} \epsilon^k \simeq 1$.

$$\begin{aligned}\sum_{n=j_k}^{j_{k+1}} \epsilon^n &\simeq \int_{j_k}^{j_{k+1}} \frac{1}{x} dx, \\ &= [\log(x)]_{j_k}^{j_{k+1}}, \\ &= \log\left(\frac{j_{k+1}}{j_k}\right).\end{aligned}$$

Par conséquent, on voudrait choisir $\log(\frac{j_{k+1}}{j_k}) = 1$, i.e., $j_{k+1} = j_k e$. Avec $j_0 = 1$, cela donne donc $j_k = e^k$. Ainsi, on pourra prendre $\rho^n = \frac{1}{\log(n)}$, pour tout $n \in \mathbb{N}$, ce qui satisfera les conditions (IV.29). Avec $\epsilon^n = \frac{1}{n}$ pour tout $n \in \mathbb{N}$, les hypothèses(IV.7) sont donc satisfaites : elles expriment le fait que le support de la loi de la variable aléatoire sous-jacente est suffisamment exploré.

IV.3.4.2. *Choix des noyaux.* Les noyaux $K^k : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ peuvent se construire de la manière suivante. Soit $K : \mathbb{R} \to \mathbb{R}$ une densité de probabilité par rapport à la mesure de Lebesgue. Supposons par simplicité que μ soit la loi uniforme sur l'hypercube $[0,1]^m$. On pose alors

$$\forall x, y \in \mathbb{R}^m,\ K^k(x,y) = K\left(\frac{\|x-y\|_{\mathbb{R}^m}}{(\epsilon^k)^{1/m}}\right).$$

On peut alors vérifier sous quelques hypothèses d'intégrabilité de K les hypothèses (IV.6a) et (IV.6b) en utilisant par exemple le Théorème 1.3.2 de [**21**] sur les approximations fonctionnelles. Ce résultat donne la convergence de l'approximation par convolution lorsque la fenêtre du noyau tend vers 0. De ce théorème, si l'on effectue des hypothèses supplémentaires de continuité du gradient en ξ uniformément en u, on peut prouver les hypothèses (IV.6a) et (IV.6b).

IV.3.4.3. *Contraintes de mesurabilité et ordre des projections.* Tout d'abord, il est intéressant de constater que l'algorithme IV.4 est particulièrement adapté à la prise en compte de contraintes de mesurabilité sur la commande. En effet, imaginons que U^f soit défini comme le sous-ensemble de $L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)$ des applications $\sigma(h)$-mesurables, avec $h : \mathbb{R}^m \to \mathbb{R}^m$ une fonction d'observation du bruit $\boldsymbol{\xi}$. La projection sur U^f n'est dans ce cas que l'espérance conditionnelle sachant $\sigma(h)$. Par linéarité de l'espérance conditionnelle, et mesurabilité des itérés précédents, la mise à jour dans l'algorithme s'écrit donc simplement :

$$\boldsymbol{u}^{k+1}(\cdot) = \boldsymbol{u}^k(\cdot) - \rho^k \nabla_u j(\boldsymbol{u}^k(\boldsymbol{\xi}^{k+1}), \boldsymbol{\xi}^{k+1}) \mathbb{E}\left(K^k(\boldsymbol{\xi}^{k+1}, \boldsymbol{\xi}) \middle| \sigma(h)\right)(\cdot).$$

Si l'on prend pour tout $\xi \in \mathbb{R}^m$ des applications noyaux $K^k(\xi, \cdot)$ qui sont $\sigma(h)$-mesurables, l'algorithme IV.4 ne comporte donc plus de projection. En d'autres termes, il est envisageable de pré-traiter un certain nombre de contraintes du type mesurabilité par un choix adéquat de noyaux. On verra dans la sous-section suivante des exemples d'application dans lesquels de tels choix s'opèrent.
Plus généralement, nous avons la proposition suivante :

PROPOSITION IV.8. *Considérons U^f défini par :*

$$U^f = \left\{u \in L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)\ :\ u \text{ est } \sigma(h) - \text{mesurable, et } u(\xi) \in \Gamma(\xi)\ \mu - \text{p.s.}\right\}.$$

Si Γ, comme multi-application de Ξ dans $\mathbb{R}^p$ est $\sigma(h)$-mesurable, et à valeurs convexes fermées, alors

$$\forall v \in L^2(\mathbb{R}^m, \mathbb{R}^p, \mu),\ \Pi_{Uf}(v)(\cdot) = \Pi_{\Gamma(\cdot)}\left(\mathbb{E}\left(v(\boldsymbol{\xi})|\sigma(h)\right)(\cdot)\right),$$

avec $\Pi_{\Gamma(\cdot)}$ la projection dans $\mathbb{R}^p$ sur $\Gamma(\cdot)$.

Preuve : On prouve cette proposition en utilisant le lemme A.8. Soit

$$U_v^f = \left\{u \in L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)\ :\ u \text{ est } \sigma(h) - \text{mesurable}\right\},$$

et

$$U_c^f = \left\{u \in L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)\ :\ u(\xi) \in \Gamma(\xi),\ \mu - \text{presque partout}\right\}.$$

Par définition, $U^f = U_v^f \cap U_c^f$, et U_v^f est un sous-espace vectoriel fermé, tandis que U_c^f est un convexe fermé. La preuve s'effectue maintenant en deux étapes :

- Montrons tout d'abord que $\Pi_{U_c^f}(u)(\cdot) = \Pi_{\Gamma(\cdot)}(u(\cdot))$ pour tout $u \in L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)$. Il est clair que $\Pi_{\Gamma(\cdot)}(u(\cdot)) \in U_c^f$. Soit $v \in U_c^f$. Pour μ-presque tout $\xi \in \mathbb{R}^m$, comme $\Gamma(\xi)$ est convexe fermé, on a par caractérisation de la projection dans $\mathbb{R}^p$, que

  $$\langle u(\xi) - v(\xi), \Pi_{\Gamma(\xi)}(u(\xi)) - v(\xi)\rangle_{\mathbb{R}^p} \leq 0.$$

  En intégrant par rapport à μ, on obtient donc le résultat recherché, par caractérisation de la projection dans $L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)$.
- Soit alors $u \in U_v^f$. Montrons maintenant que $\Pi_{U_c^f}(u)$ est $\sigma(h)$-mesurable. On utilise pour ce faire un théorème de mesurabilité des applications marginales. En effet, posons $f : \mathbb{R}^m \times \mathbb{R}^p :\to \mathbb{R}$ définie par $f(\xi, v) = \|u(\xi) - v\|_{\mathbb{R}^p}^2$. Pour μ-presque tout $\xi \in \mathbb{R}^m$, $f(\xi, \cdot)$ est continue, et pour tout $v \in \mathbb{R}^p$, $f(\cdot, v)$ est $\sigma(h)$-mesurable. C'est donc une fonction de Carathéodory (cf. définition A.24), si l'on munit $\mathbb{R}^m$ de la tribu $\sigma(h)$. De plus, la multi-application Γ est $\sigma(h)$-mesurable, et à valeurs fermées et non-vides. Par conséquent, le théorème 8.2.11 de [**6**] s'applique, et donc

l'application $\xi \mapsto \arg\min_{v\in\Gamma(\xi)} f(\xi, v)$ est $\sigma(h)$-mesurable, ce qui veut exactement dire que la projection recherchée est $\sigma(h)$-mesurable, i.e. que $\Pi_{\Gamma(\cdot)}(u(\cdot)) \in U_v^f$.

On peut donc appliquer le lemme A.8 qui achève la preuve. □

En définissant donc U^f comme dans la proposition IV.8, et en appliquant cette proposition à l'algorithme IV.4, on trouve une formule de mise à jour s'écrivant simplement :

$$\forall \xi \in \Xi,\ \boldsymbol{u}^{k+1}(\xi) = \Pi_{\Gamma(\xi)}\left(\boldsymbol{u}^k(\xi) - \rho^k \nabla_u j(\boldsymbol{u}^k(\boldsymbol{\xi}^{k+1}), \boldsymbol{\xi}^{k+1}) \mathbb{E}\left(K^k(\boldsymbol{\xi}^{k+1}, \boldsymbol{\xi}) \middle| \sigma(h)\right)(\xi)\right).$$

Ainsi, on a fait passer la difficulté liée à la projection sur le sous-espace vectoriel des fonctions $\sigma(h)$-mesurables directement à l'intérieur du noyau, ce qui autorise un pré-traitement rendant l'exécution de l'algorithme nettement plus simple. L'algorithme se déroule schématiquement comme suit :

- Choisir une forme de noyau $K(\cdot,\cdot)$ incluant la contrainte de mesurabilité ;
- Initialiser u^0 à 0 partout ;
- À l'étape k,
  1. tirer une réalisation ξ^{k+1},
  2. évaluer $v^{k+1} = u^k(\xi^{k+1})$,
  3. calculer le gradient $s^{k+1} = \nabla_u j(v^{k+1}, \xi^{k+1})$,
  4. stocker le quadruplet $(s^{k+1}, \xi^{k+1}, \epsilon^k, \rho^k)$.

Ainsi, à chaque étape, on a à effectuer une évaluation de u^k connu comme une somme de noyaux, et à stocker le quadruplet obtenu pour de futures évaluations. Chaque étape de l'algorithme est donc extrêmement simple, la seule difficulté numérique résidant dans un calcul rapide de sommes de noyaux croissant avec les itérataions.

IV.3.4.4. *Gradient stochastique usuel.* Une autre remarque intéressante concerne le gradient stochastique usuel. En effet, il est possible de retrouver un algorithme de gradient stochastique en boucle ouverte à partir de notre algorithme. Considérons l'ensemble admissible U^f défini par :

$$U^f = \left\{u \in L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)\ :\ u \text{ est } \sigma(\{\mathbb{R}^m, \emptyset\}) - \text{mesurable, et } u(\xi) \in \Gamma\ \mu - \text{p.s.}\right\}, \tag{IV.30}$$

avec Γ un compact convexe de $\mathbb{R}^p$. Alors, les applications u appartenant à U^f ne sont autres que les applications constantes à valeurs dans Γ. Finalement, le problème (IV.4) avec la contrainte (IV.30) est un problème en boucle ouverte, auquel on pourrait directement appliquer un algorithme de gradient stochastique en boucle ouverte. Nous allons ici faire le lien entre cet algorithme en boucle ouverte direct, et l'algorithme en boucle fermée IV.4 appliqué à ce cas particulier.
La proposition IV.8 donne la projection sur U^f :

$$\Pi_{U^f}(u)(\cdot) = \Pi_\Gamma\left(\mathbb{E}\left(u(\boldsymbol{\xi})\right)\right),$$

avec $\Pi_\Gamma : \mathbb{R}^p \to \mathbb{R}^p$ la projection sur Γ dans $\mathbb{R}^p$. La mise à jour de l'algorithme IV.4 se réécrit donc comme suit, dans $\mathbb{R}^p$ tout simplement :

$$\boldsymbol{u}^{k+1} = \Pi_\Gamma\left(\boldsymbol{u}^k - \rho^k \nabla_u j(\boldsymbol{u}^k, \boldsymbol{\xi}^{k+1}) \int_{\mathbb{R}^m} K^k(\boldsymbol{\xi}^{k+1}, \xi) d\mu(\xi)\right).$$

Ainsi, on retombe sur l'algorithme de gradient stochastique en boucle ouverte dans lequel on considère maintenant un pas modifié intégrant le noyau. Pour un choix particulier de noyau, on peut perdre la dépendance du noyau en $\boldsymbol{\xi}^{k+1}$ afin de ne pas avoir de pas de descente anticipatif. Typiquement, on fera en sorte d'avoir :

$$\forall x \in \mathbb{R}^m,\ \int_{\mathbb{R}^m} K^k(x, \xi) d\mu(\xi) = \epsilon^k.$$

Pour résumer, le gradient stochastique en boucle fermé introduit par l'algorithme IV.4 est bien une généralisation du gradient stochastique en boucle ouverte.

IV.3.5. Application à un problème de gestion de réservoir. Nous allons donner ici un exemple d'application de l'algorithme IV.4, afin d'en montrer l'intérêt pratique. Il s'agit de gérer un réservoir hydraulique, successivement sur un et deux pas de temps, avec des commandes en boucle fermée, soumises à des contraintes de mesurabilité.
De manière générale, le problème de gestion de réservoir se pose comme suit : on dispose d'un certain stock d'eau initial noté s s'il est connu, et $\boldsymbol{s}$ s'il est lui-même une variable aléatoire. On doit ensuite prendre une décision de production notée génériquement $\boldsymbol{u}$, en fonction d'un prix de marché incertain noté $\boldsymbol{\xi}$. Cette décision est prise de manière à maximiser un certain profit. Ce profit s'exprime comme la somme des ventes (grossièrement $\boldsymbol{u\xi}$) et de la valeur du stock final (donnée ici a priori). On peut ensuite à loisir compliquer ce problème en faisant dépendre la commande du stock initial, ou en mettant plusieurs temps de décision correspondant à plusieurs périodes de vente, et donc plusieurs prix de vente. Il est alors raisonnable de demander à la première décision de ne dépendre que du premier prix, tandis que la deuxième décision est autorisée à dépendre des deux premiers prix, etc. De telles contraintes sont communément appelées contraintes de non-anticipativité. Nous allons dans les paragraphes qui vont suivre résoudre de manière exacte, et avec notre algorithme, de tels problèmes.

IV.3.5.1. *Exemple avec un seul pas de temps.* Nous notons ici s le stock initial du barrage, supposé pour l'instant fixé, $\boldsymbol{\xi}$ le prix de vente de la production, variable aléatoire à loi uniforme sur $[\underline{x}, \overline{x}]$. La fonction de coût à minimiser (soit l'opposé du profit à maximiser) s'écrit ici :

$$\forall \xi \in [\underline{x}, \overline{x}]\,,\ \forall u \in [0, s],\ j\,(u, \xi) = -\xi u - \sqrt{\epsilon + s - u}. \tag{IV.31}$$

Le terme $\sqrt{\epsilon + s - u}$ est la valeur du stock à la fin du jeu lorsqu'il y reste une quantité $s - u$. Le paramètre $\epsilon > 0$ sert à régler la fonction de fin de jeu du barrage. La fonction que l'on souhaite minimiser est donc l'application J définie par :

$$\forall u \in L^2([\underline{x}, \overline{x}]\,, \mathbb{R}, \lambda),\ J(u) = \mathbb{E}\,(j(u(\boldsymbol{\xi}), \boldsymbol{\xi}))\,.$$

Son gradient se calcule comme suit :

$$\nabla J(u)(\xi) = -\xi + \frac{1}{2\sqrt{\epsilon + s - u(\xi)}}.$$

Par conséquent, le contrôle optimal u^* est donné pour tout $\xi \in [\underline{x}, \overline{x}]$ par :

$$u^*(\xi) = \begin{cases} 0 & \text{si } \xi < \frac{1}{2\sqrt{\epsilon+s}} \\ s + \epsilon - \frac{1}{4\xi^2} & \text{si } \frac{1}{2\sqrt{\epsilon+s}} \leq \xi \leq \frac{1}{2\sqrt{\epsilon}} \\ s & \text{si } \xi > \frac{1}{2\sqrt{\epsilon}} \end{cases}$$

En notant $[\cdot]_a^b = \min(a, \max(\cdot, b))$, on peut le réécrire simplement

$$u^*(\xi) = \left[s + \epsilon - \frac{1}{4\xi^2} \right]_0^s.$$

Nous proposons maintenant une résolution de ce problème par l'algorithme IV.4. En effet, il s'agit de minimiser $J(u)$ sous la contrainte que u prenne ses valeurs dans $[0, s]$.

Pour une application numérique, nous prenons $s = 1, \epsilon = 0.1, [\underline{x}, \overline{x}] = [0.4, 2]$. On choisit de plus $K^k(x, y) = \frac{1}{\epsilon^k \sqrt{2\pi}} \exp\left(\frac{(x-y)^2}{2(\epsilon^k)^2}\right)$, et $\rho^k = \epsilon^k = 1/\sqrt{k}$. La figure 4 montre l'évolution du contrôle après 500, 3000 puis 10000 itérations, ainsi que la vitesse de convergence en échelle logarithmique, i.e. l'erreur en fonction du nombre d'itérations.

Afin de rendre l'exemple un peu plus compliqué, et surtout d'augmenter la dimension de l'espace d'aléas, on considère maintenant le stock initial s comme étant une variable aléatoire $\boldsymbol{s}$ suivant une loi uniforme sur $[0, 1]$, indépendante[1] de $\boldsymbol{\xi}$. La fonction de coût j devient donc une fonction de trois variables (la troisième étant le niveau de stock $\boldsymbol{s}$), et l'on recherche le

[1]On ne suppose l'indépendance que pour la simplicité des applications numériques. En effet, avec une dépendance entre les variables aléatoires, la solution théorique du problème serait la même, et il suffirait dans l'algorithme stochastique du type IV.4 de tirer conjointement les deux variables aléatoires selon leur loi jointe pour obtenir la convergence.

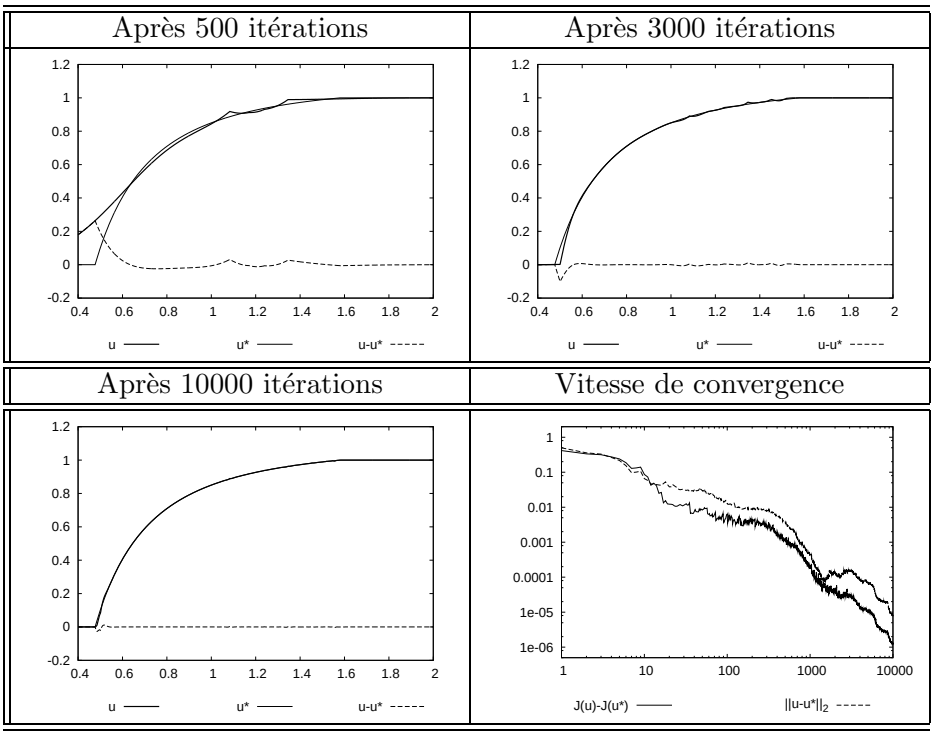


FIG. 4. Problème de gestion de réservoir à un pas de temps avec un stock initial déterministe, contrôle le long des itérations et termes d'erreurs.

contrôle u comme une application de $L^2([\underline{x}, \overline{x}] \times [0,1], \mathbb{R}, \lambda)$, i.e. pouvant dépendre du niveau de stock initial. Les calculs théoriques donnant la solution optimale sont exactement les mêmes qu'auparavant, il suffit de faire dépendre le contrôle du stock initial s. La figure 5 représente le contrôle optimal en fonction du stock initial et du prix de vente.

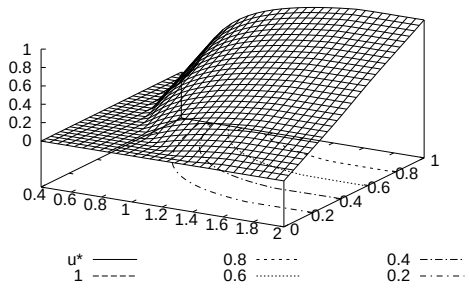


FIG. 5. Problème de réservoir à un pas de temps et deux bruits, contrôle optimal

A nouveau, on peut appliquer l'algorithme IV.4 pour résoudre ce problème, avec maintenant un tirage des deux variables aléatoires $\boldsymbol{\xi}$ et $\boldsymbol{s}$. On utilise les mêmes paramètres qu'auparavant, avec la seule distinction que les noyaux K^k deviennent maintenant définis sur $\mathbb{R}^2$ comme des produits de noyaux gaussiens. On obtient alors la figure 6 montrant le long des itérations l'évolution du contrôle (en haut) et de l'erreur sur ce contrôle (en bas).

La figure 7 donne la vitesse de convergence pour ce problème. A nouveau, il est intéressant de constater qu'il n'y a apparemment pas d'effet de seuil dans la convergence. Néanmoins, avec l'augmentation de la dimension, on a maintenant besoin d'un minimum de 10000 itérations pour obtenir une convergence correcte : ceci est tout à fait compréhensible, car le principe de notre

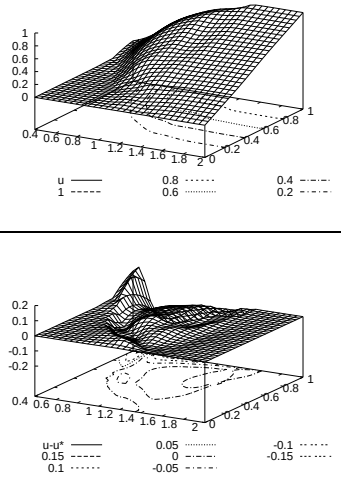
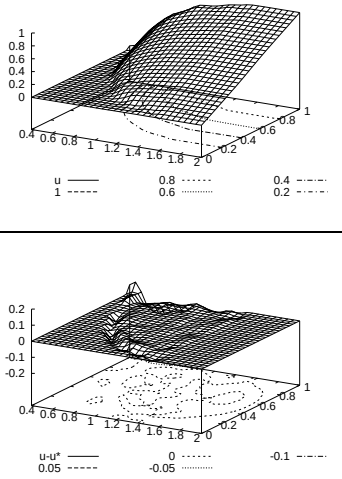
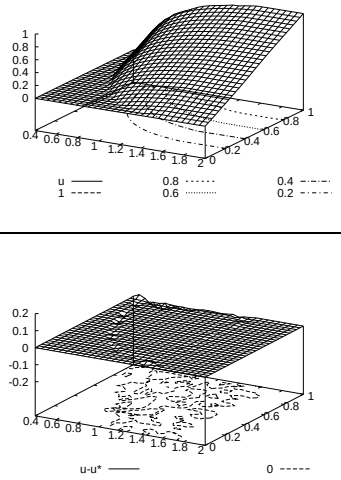
Après 100 itérations	Après 1000 itérations	Après 10000 itérations
		

FIG. 6. Problème de réservoir à un pas de temps et deux bruits, contrôles le long des itérations

algorithme est d'explorer par voisinages décroissants l'espace des aléas. Par conséquent, plus l'espace est *grand*, plus l'on a besoin de voisinages, et donc d'itérations.

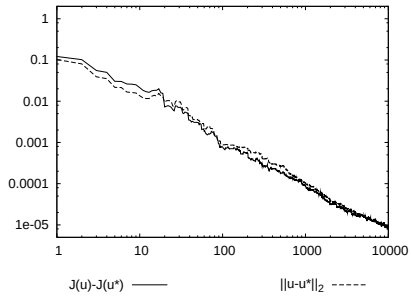


FIG. 7. Problème de réservoir à un pas de temps et deux bruits, vitesse de convergence

IV.3.5.2. *Exemple avec deux pas de temps.* Après cette mise en jambes sur le problème de gestion de réservoir à un pas de temps, nous allons le généraliser à deux pas de temps. Le principe est toujours le même, mais il y a maintenant deux prix de vente qui se dévoilent successivement, et deux décisions de production correspondantes, sur lesquelles pèse une contrainte de non-anticipativité. Nous considérons maintenant la fonction de coût

$$j(u_1, u_2, \xi_1, \xi_2) = -u_1\xi_1 - u_2\xi_2 - \sqrt{\epsilon + s - u_1 - u_2}, \tag{IV.32}$$

pour tous $(\xi_1, \xi_2) \in [\underline{x}_1, \overline{x}_1] \times [\underline{x}_2, \overline{x}_2]$, et tous $u_1 \in [0, s]$, $u_2 \in [0, s - u_1]$. Pour $i = 1, 2$, $\boldsymbol{\xi}_i$ est une variable aléatoire de loi uniforme sur $[\underline{x}_i, \overline{x}_i]$, telle que $\boldsymbol{\xi}_1$ et $\boldsymbol{\xi}_2$ sont indépendantes. On souhaite dès lors minimiser le critère

$$J(u_1, u_2) = \mathbb{E}\left(j(u_1(\boldsymbol{\xi}_1), u_2(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2), \boldsymbol{\xi}_1, \boldsymbol{\xi}_2)\right),$$

avec $u_1 \in L^2([\underline{x}_1, \overline{x}_1], \mathbb{R}, \lambda)$ et $u_2 \in L^2(\Pi_{i=1,2}[\underline{x}_i, \overline{x}_i], \mathbb{R}, \lambda \otimes \lambda)$. En exprimant le problème de cette manière, on a directement tenu compte de la contrainte de non-anticipativité sur u_1. Dans un premier temps, voici la résolution théorique du problème. On la trouve par programmation dynamique :

– On calcule d'abord le contrôle u_2^*, comme une fonction du premier contrôle u_1 et des deux prix ξ_1 et ξ_2. Cela s'effectue exactement de la même manière qu'avant, et on obtient :

$$u_2^*(\xi_1, \xi_2, u_1) = \left[\epsilon + s - u_1 - \frac{1}{4(\xi_2)^2}\right]_0^{s-u_1},$$

qui ne dépend pas directement de ξ_1. De manière plus explicite, on a :

$$u_2^*(\xi_2, u_1) = \begin{cases} s - u_1 & \text{si } \xi_2 > \frac{1}{2\sqrt{\epsilon}}, \\ \epsilon + s - u_1 - \frac{1}{4(\xi_2)^2} & \text{si } \frac{1}{2\sqrt{\epsilon+s-u_1}} \leq \xi_2 \leq \frac{1}{2\sqrt{\epsilon}}, \\ 0 & \text{si } \xi_2 < \frac{1}{2\sqrt{\epsilon+s-u_1}}. \end{cases}$$

On calcule maintenant le gradient de u_2^* par rapport à u_1 :

$$\nabla_{u_1} u_2^*(u_1, \xi_2) = \begin{cases} -1 & \text{si } \xi_2 \geq \frac{1}{2\sqrt{\epsilon+s-u_1}} \\ 0 & \text{sinon.} \end{cases}$$

– Il s'agit enfin de résoudre pour tout ξ_1 le problème suivant :

$$\min_{u_1 \in [0,s]} -u_1\xi_1 - \mathbb{E}\left(u_2^*(\boldsymbol{\xi}_2, u_1)\boldsymbol{\xi}_2 + \sqrt{\epsilon + s - u_1 - u_2^*(\boldsymbol{\xi}_2, u_1)}\right)$$

Dans notre cas, on peut résoudre ce problème en égalant le projeté du gradient à 0, ce qui donne :

$$-\xi_1 + \mathbb{E}\left(\boldsymbol{\xi}_2 1_{[\frac{1}{2\sqrt{\epsilon+s-u_1}}, \overline{x}_2]}(\boldsymbol{\xi}_2)\right) - \mathbb{E}\left(\frac{1}{2\sqrt{s+\epsilon-u_1-u_2^*(\boldsymbol{\xi}_2,u_1)}} 1_{[\frac{1}{2\sqrt{\epsilon+s-u_1}}, \overline{x}_2]}(\boldsymbol{\xi}_2)\right) + \mathbb{E}\left(\frac{1}{2\sqrt{s+\epsilon-u_1-u_2^*(\boldsymbol{\xi}_2,u_1)}}\right) = 0,$$

Supposons provisoirement que $\overline{x}_2 < \frac{1}{2\sqrt{\epsilon}}$. En injectant dans la dernière inégalité la formule explicite de u_2^*, on obtient :

$$\mathbb{E}\left(\frac{1}{2\sqrt{s+\epsilon-u_1-u_2^*(\boldsymbol{\xi}_2,u_1)}}\right) = \xi_1.$$

Il ne reste plus qu'à calculer cette espérance ($\boldsymbol{\xi}_2$ suit une loi uniforme sur $[\underline{x}_2, \overline{x}_2]$) :

$$\frac{1}{\overline{x}_2 - \underline{x}_2}\int_{\underline{x}_2}^{\overline{x}_2} \frac{d\xi_2}{2\sqrt{s+\epsilon-u_1-u_2^*(\boldsymbol{\xi}_2,u_1)}} = \xi_1, \text{ i.e.,}$$

$$\int_{\underline{x}_2}^{\frac{1}{2\sqrt{s+\epsilon-u_1}}} \frac{1}{2\sqrt{s+\epsilon-u_1}} d\xi_2 + \int_{\frac{1}{2\sqrt{s+\epsilon-u_1}}}^{\overline{x}_2} \xi_2 d\xi_2 = (\overline{x}_2 - \underline{x}_2)\xi_1,$$

Pour simplifier les calculs, posons $r = \frac{1}{2\sqrt{\epsilon+s-u_1}}$. En poursuivant, on a :

$$\begin{aligned} r(r - \underline{x}_2) + \frac{(\overline{x}_2)^2}{2} - \frac{r^2}{2} &= (\overline{x}_2 - \underline{x}_2)\xi_1, \\ (r - \underline{x}_2)^2 &= \left((\underline{x}_2)^2 - (\overline{x}_2)^2\right) + 2\xi_1(\overline{x}_2 - \underline{x}_2), \\ r &= \underline{x}_2 + \sqrt{2(\overline{x}_2 - \underline{x}_2)(\xi_1 - \frac{\overline{x}_2 + \underline{x}_2}{2})_+}. \end{aligned}$$

Finalement, on peut exprimer le contrôle optimal $u_1^*(\xi_1)$, sans davantage d'hypothèses sur $\bar{x}_2$ et ϵ :

$$u_1^*(\xi_1) = \left[\epsilon + s - \frac{1}{4\left(\underline{x}_2 + \sqrt{2(\overline{x}_2 - \underline{x}_2)(\xi_1 - \frac{\overline{x}_2 + \underline{x}_2}{2})_+}\right)^2}\right]_0^s.$$

– Le contrôle optimal u_2^{**} est ensuite donné comme $u_2^{**}(\xi_1, \xi_2) = u_2^*(\xi_2, u_1^*(\xi_1))$.

Voyons maintenant les résultats numériques et l'application de notre algorithme pour résoudre ce problème. On prend $s = 1, \epsilon = 0.1, \underline{x}_1 = \underline{x}_2 = 0.4, \overline{x}_1 = \overline{x}_2 = 2$. La figure 8 donne le contrôle optimal théorique u_2^{**} calculé précédemment avec les paramètres donnés ci-dessus.

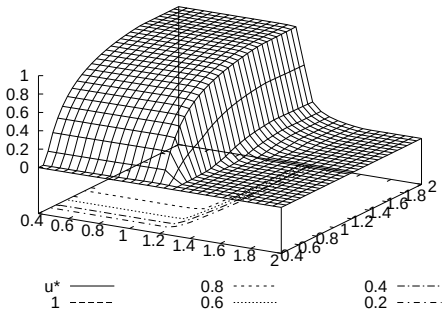


FIG. 8. Problème de gestion de réservoir à deux pas de temps, contrôle optimal au deuxième pas de temps

On choisit ici des noyaux sur $\mathbb{R}^2$ qui s'expriment comme le produit de noyaux gaussiens sur $\mathbb{R}$, ce qui rend les projections dûes à la contrainte de non-anticipativité aisées.

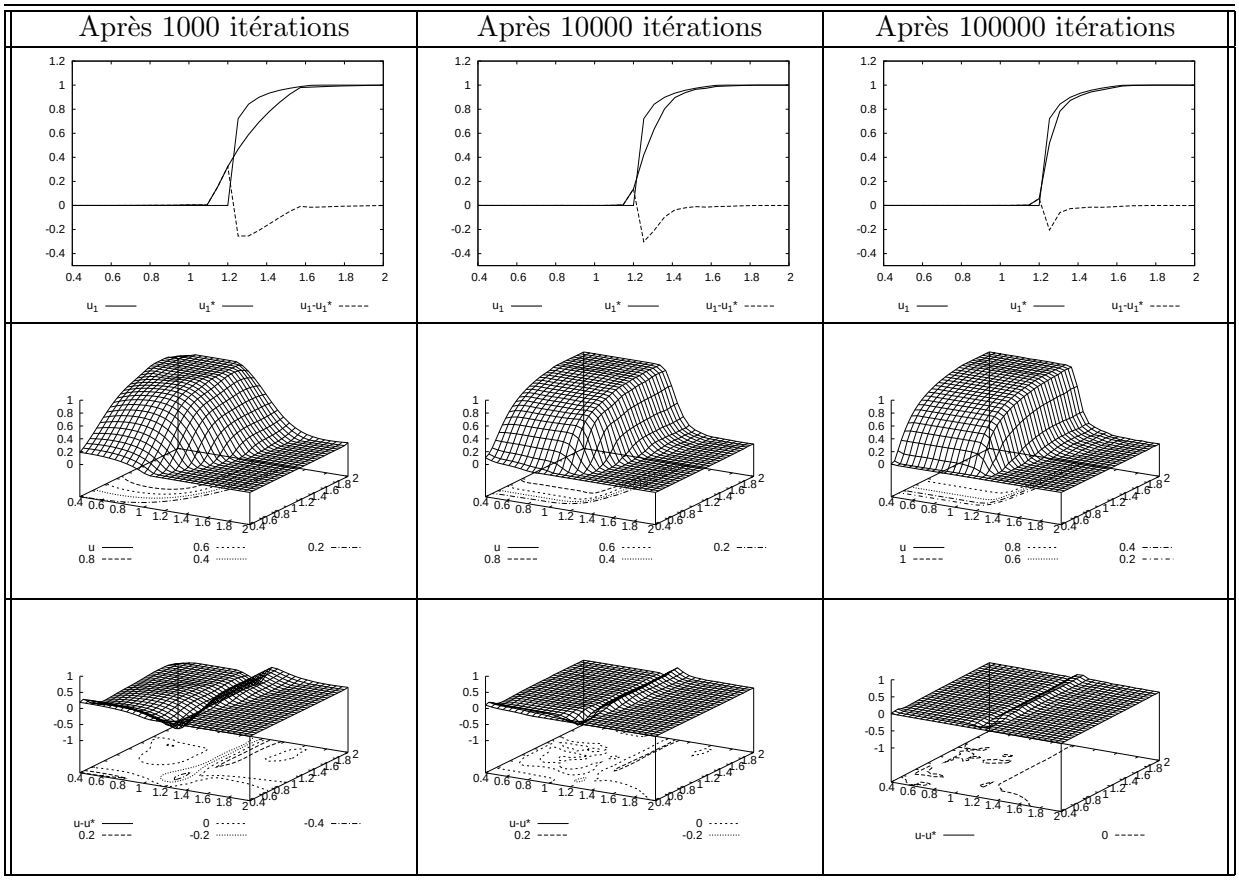


FIG. 9. Problème de gestion de réservoir à deux pas de temps, contrôle au premier pas de temps (haut), au deuxième pas de temps (milieu), et erreur du deuxième contrôle (bas) le long des itérations

On obtient alors la figure 9 qui donne l'évolution de u_1 (en haut), de u_2 (au milieu) et de l'erreur sur u_2 le long des itérations, après 1000, 10000 et 100000 itérations. Dans cet exemple, il y a une contrainte difficile à projeter, c'est la contrainte de stock liant u_1 et u_2. En effet, il faut ici que $u_2 \leq s - u_1$. Afin de rendre le problème numériquement plus sympathique, on pénalise cette contrainte sous la forme d'un terme additionnel dans le critère donné par $\lambda \times (u_1 + u_2 - s)$ avec $\lambda > 0$, et on applique l'algorithme IV.4 au problème pénalisé. On obtient comme le montre la figure 9 une convergence de l'algorithme, certes plus lente, mais toujours avérée.

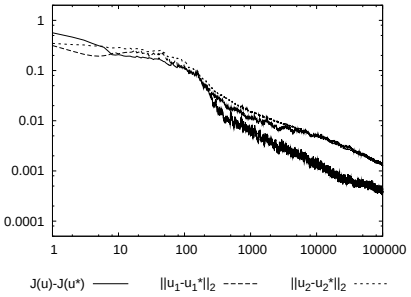


FIG. 10. Problème de gestion de réservoir à deux pas de temps, vitesse de convergence

IV.3.5.3. *Remarques numériques.* Un important travail numérique a été entrepris, notamment par Jean-Sébastien Roy que nous ne remercierons jamais assez pour son efficacité et ses excellentes idées, afin d'améliorer la vitesse de convergence pratique de l'algorithme IV.4. Ce travail s'est centré à la fois sur le choix de pas de descente et de voisinage en ligne, en fonction des tirages passés, et sur le choix de structures d'accès rapide pour les évaluations des sommes de noyaux exigées dans l'étape de mise à jour. Les principales pistes explorées sont l'utilisation de transformées gaussiennes rapides (cf. [**54, 93, 96**]).

IV.4. Cadre général du gradient perturbé

Une alternative pour l'étude des algorithmes stochastiques est de les considérer comme des algorithmes de descente perturbés, c'est à dire comme des algorithmes basés sur l'idée d'une direction de descente, qu'on ne peut connaître que de manière erronée, ou bruitée. Ces bruits peuvent ensuite être pris comme déterministes ou aléatoires. Pour l'étude générale d'algorithmes de gradient perturbés, on peut consulter l'article [**20**] qui permet de prouver la convergence des algorithmes de gradient stochastique en dimension finie, en l'absence de projection sur un convexe fermé, mais sans hypothèses de convexité sur la fonction dont on prend le gradient.

Il existe en revanche peu de travaux sur les algorithmes de gradient perturbé en dimension infinie. Le travail le plus proche du nôtre de par son approche particulièrement adaptée à l'optimisation convexe est [**59**]. Il donne en effet des conditions sous lesquelles des opérateurs stochastiques approchés du gradient ou du sous-gradient d'une fonction convexe permettent de trouver le minimum de la dite application. Du côté des approximations stochastiques dans un cadre plus probabiliste ou statistique, on peut également mentionner les travaux [**74, 75**] qui procurent des résultats en grande déviation pour des algorithmes stochastiques sans projection dans un espace de Hilbert. Enfin, dans la lignée des travaux de Révész, on peut mentionner l'important article [**97**] donnant des résultats de convergence pour des algorithmes de Robbins-Monro en dimension infinie. Leur résultat principal est assez proche du nôtre, avec une hypothèse différente sur la suite des bruits, et en fait peu applicable au cas d'approximations fonctionnelles comme celles qui nous préoccupent, et en tous les cas inadaptée à la prise en considération de projections.

C'est dans la lignée de ces travaux sur les approximations stochastiques hilbertiennes que nous nous plaçons : nous allons prouver des théorèmes de convergence pour des algorithmes généraux de gradient perturbé dans un cadre hilbertien, et montrer ensuite comment ces théorèmes

peuvent servir à démontrer la convergence de l'algorithme IV.4, et d'autres types d'algorithmes d'approximation fonctionnelle.

IV.4.1. Approximation stochastique généralisée. Nous allons donc ici nous placer dans le cadre général suivant. Les notations par rapport à la section précédente sont différentes, et ce dans le but de souligner le caractère générique de ce qui va suivre. Le problème est de résoudre :

$$\min_x f(x) \quad \text{s.c. } x \in X^f. \tag{IV.33}$$

avec :
- X un espace de Hilbert muni d'un produit scalaire et d'une norme notés $\langle\cdot,\cdot\rangle$ et $\|\cdot\|$,
- $f : X \to \mathbb{R}$ une application convexe,
- X^f un convexe fermé de X (possiblement sous-espace vectoriel) et Π_{X^f} la projection sur X^f.

On se donne également $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité, et $(\mathcal{F}_t)$ une filtration sur cet espace. On propose alors l'algorithme de gradient perturbé suivant pour résoudre le problème IV.33 :

ALGORITHME IV.9. *Étape $t \in \mathbb{N}$:*

$$\boldsymbol{x}_{t+1} = \Pi_{X^f}\left(\boldsymbol{x}_t + \gamma_t(\boldsymbol{s}_t + \boldsymbol{w}_t)\right),$$

avec $\boldsymbol{s}_t$, une direction de descente, $\boldsymbol{w}_t$ une variable aléatoire à valeurs dans X (la perturbation, le bruit sur la direction de descente), et γ_t une suite décroissante de réels positifs.

Nous allons maintenant donner un théorème de convergence de l'algorithme IV.9 dans le cas de X^f convexe fermé (pouvant être ou non un sous-espace vectoriel). Ce théorème est l'analogue dans un cadre plus général du théorème IV.5, et suit donc le même schéma de preuve.

THÉORÈME IV.10 (Projection générale). *(i) Supposons que $x \mapsto f(x)$ soit convexe, coercive et semi continue inférieurement, sur X, et que X^f soit un convexe fermé. Ainsi, pour tout $x \in X^f$, $\partial f(x) \neq \emptyset$.*

(ii) Soit $(\mathcal{F}_t)$ une filtration sur $(\Omega, \mathcal{A}, \mathbb{P})$, et supposons que pour tout $t \in \mathbb{N}$, $\boldsymbol{s}_t$ et $\boldsymbol{w}_{t-1}$ sont $\mathcal{F}_t$-mesurables.

(iii) Supposons que f est à sous-gradients linéairement bornés sur X^f, c'est à dire

$$\exists a_1, a_2 \geq 0,\ \forall x \in X^f,\ \forall v \in \partial f(x),\ \|v\| \leq a_1\|x\| + a_2. \tag{IV.34}$$

(iv) Supposons qu'il existe $\kappa > 0$ tel que

$$\forall t \in \mathbb{N},\ -\frac{1}{\kappa}\boldsymbol{s}_t \in \partial f(x_t). \tag{IV.35}$$

(v) Supposons qu'il existe $A > 0$ et deux suites de réels positifs (ϵ_t, η_t), tels que pour tout $t \in \mathbb{N}$, il existe $\boldsymbol{v}_t \in \partial f(\boldsymbol{x}_t)$, tel que,

$$\|\mathbb{E}\left(\boldsymbol{w}_t|\mathcal{F}_t\right)\| \leq \eta_t\left(1 + \|\boldsymbol{v}_t\|\right), \tag{IV.36a}$$

$$\mathbb{E}\left(\|\boldsymbol{w}_t\|^2|\mathcal{F}_t\right) \leq A\left(1 + \frac{1}{\epsilon_t}\|\boldsymbol{v}_t\|^2\right). \tag{IV.36b}$$

Si X^f n'est pas un sous-espace vectoriel, supposons de plus qu'il existe une application $g : \mathbb{R} \to \mathbb{R}$ continue ou bornée, et pour tout $t \in \mathbb{N}$ un élément $\boldsymbol{v}_t \in \partial f(\boldsymbol{x}_t)$ tels que

$$\mathbb{E}\left(\|w_t\||\mathcal{F}_t\right) \leq g(\|\boldsymbol{v}_t\|). \tag{IV.36c}$$

(vi) Supposons que les suites (γ_t), (ϵ_t) and (η_t) décroissent vers 0 et soient telles que :

$$\forall t \in \mathbb{N},\ \gamma_t, \epsilon_t > 0, \quad \sum_{t\in\mathbb{N}} \gamma_t = +\infty, \quad \sum_{t\in\mathbb{N}} \frac{(\gamma_t)^2}{\epsilon_t} < +\infty, \quad \sum_{t\in\mathbb{N}} \gamma_t\eta_t < +\infty. \tag{IV.37}$$

Alors le problème (IV.33) *admet un ensemble de solutions qu'on notera S, et une valeur optimale notée f_S. De plus, $f(\boldsymbol{x}_t) \to f_S$ p.s., quand t tend vers l'infini, et tout point d'accumulation de $(\boldsymbol{x}_t)$ dans la topologie faible appartient à S.*

(vii) De plus, si f est fortement convexe, $(\boldsymbol{x}_t)$ converge fortement vers l'unique solution x^* de (IV.33).

Preuve : On utilise le schéma de preuve introduit dans [**32**], en introduisant une fonction de Lyapunov. Soit $x^* \in S$, et pour tout $x \in X$, $\Lambda(x) := \frac{1}{2}\|x - x^*\|^2$ notre fonction de Lyapunov. On étudie son évolution le long des itérations. Pour tout $t \in \mathbb{N}$, on note $\Lambda_t = \Lambda(\boldsymbol{x}_t)$, avec $\boldsymbol{x}_t$ généré par l'algorithme IV.9. Soit $t \in \mathbb{N}$. En suivant le même raisonnement sur les projections que celui menant dans la preuve du théorème IV.5 à l'équation (IV.8), on obtient :

$$\begin{aligned}\Lambda_{t+1} - \Lambda_t =& \frac{1}{2}\|\boldsymbol{x}_{t+1} - \boldsymbol{x}_t\|^2 + \langle \boldsymbol{x}_{t+1} - \boldsymbol{x}_t, \boldsymbol{x}_t - x^*\rangle \\ \leq& \frac{3(\gamma_t)^2}{2}\|\boldsymbol{s}_t + \boldsymbol{w}_t\|^2 + \gamma_t\langle \boldsymbol{s}_t + \boldsymbol{w}_t, \boldsymbol{x}_t - x^*\rangle \end{aligned} \tag{IV.38}$$

On prend maintenant l'espérance conditionnelle sachant $\mathcal{F}_t$ dans (IV.38)

$$\begin{aligned}\mathbb{E}\left(\Lambda_{t+1}|\mathcal{F}_t\right) - \Lambda_t \leq& \frac{3(\gamma_t)^2}{2}\mathbb{E}\left(\|\boldsymbol{s}_t + \boldsymbol{w}_t\|^2|\mathcal{F}_t\right) + \gamma_t\langle \boldsymbol{s}_t, \boldsymbol{x}_t - x^*\rangle \\ &+ \gamma_t\langle \mathbb{E}\left(\boldsymbol{w}_t|\mathcal{F}_t\right), \boldsymbol{x}_t - x^*\rangle, \\ \leq& \frac{3(\gamma_t)^2}{2}\mathbb{E}\left(\|\boldsymbol{s}_t + \boldsymbol{w}_t\|^2|\mathcal{F}_t\right) + \gamma_t\kappa\left(f(x^*) - f(\boldsymbol{x}_t)\right) \\ &+ \eta_t\gamma_t\|\boldsymbol{x}_t - x^*\| + \eta_t\gamma_t\|\boldsymbol{v}_t\|\|\boldsymbol{x}_t - x^*\|, \end{aligned} \tag{IV.39}$$

de par les hypothèses (IV.35) et (IV.36a) et l'inégalité de Cauchy-Schwarz. Selon l'hypothèse (IV.34), il existe $a_1, a_3 \geq 0$ tels que

$$\|\boldsymbol{v}_t\| \leq a_1\|x_t - x^*\| + a_3.$$

On continue alors le calcul :

$$\begin{aligned}\mathbb{E}\left(\Lambda_{t+1}|\mathcal{F}_t\right) - \Lambda_t \leq& \frac{3(\gamma_t)^2}{2}\mathbb{E}\left(\|\boldsymbol{s}_t + \boldsymbol{w}_t\|^2|\mathcal{F}_t\right) + \gamma_t\kappa\left(f(x^*) - f(\boldsymbol{x}_t)\right) \\ &+ \eta_t\gamma_t(1 + a_3)\|\boldsymbol{x}_t - x^*\| + \eta_t\gamma_t a_1\|\boldsymbol{x}_t - x^*\|^2 \\ \leq& \frac{3(\gamma_t)^2}{2}\mathbb{E}\left(\|\boldsymbol{s}_t + \boldsymbol{w}_t\|^2|\mathcal{F}_t\right) + \gamma_t\kappa\left(f(x^*) - f(\boldsymbol{x}_t)\right) \\ &+ \eta_t\gamma_t(1 + a_1 + a_3)\|\boldsymbol{x}_t - x^*\|^2 + \eta_t\gamma_t\left(1 + a_3\right). \end{aligned} \tag{IV.40}$$

L'inégalité (IV.40) est obtenue grâce à l'inégalité scalaire $x \leq 1 + x^2$. On s'intéresse maintenant au premier terme du membre de droite. Selon l'hypothèse (IV.35), et l'hypothèse (IV.34), il vient

$$\|\boldsymbol{s}_t\|^2 \leq \kappa^2\left(a_1\|\boldsymbol{x}_t - x^*\| + a_3\right)^2 \leq 2\kappa^2\left(a_1^2\|\boldsymbol{x}_t - x^*\|^2 + a_3^2\right).$$

D'où :

$$\begin{aligned}\frac{(\gamma_t)^2}{2}\mathbb{E}\left(\|\boldsymbol{s}_t + \boldsymbol{w}_t\|^2|\mathcal{F}_t\right) \leq& (\gamma_t)^2\left(\|\boldsymbol{s}_t\|^2 + \mathbb{E}\left(\|\boldsymbol{w}_t\|^2|\mathcal{F}_t\right)\right), \\ \leq& (\gamma_t)^2\left(2\kappa^2(a_1^2\|\boldsymbol{x}_t - x^*\|^2 + a_3^2) + A\left(1 + \frac{1}{\epsilon_t}\|\boldsymbol{v}_t\|^2\right)\right), \\ \leq& (\gamma_t)^2\left(2\kappa^2 a_3^2 + A + \frac{2Aa_3^2}{\epsilon_t}\right) + (\gamma_t)^2\left(2\kappa^2 a_1^2 + \frac{2a_1^2 A}{\epsilon_t}\right)\|\boldsymbol{x}_t - x^*\|^2, \end{aligned} \tag{IV.41}$$

L'équation (IV.41) s'obtient en utilisant l'inégalité scalaire classique $(x+y)^2 \leq 2x^2 + 2y^2$, les hypothèses (IV.35)–(IV.36b), et la propriété de sous-gradients linéairement bornés de f. En revenant maintenant à l'équation (IV.40), on trouve :

$$\mathbb{E}\left(\Lambda_{t+1}|\mathcal{F}_t\right) - \Lambda_t \leq \alpha_t\Lambda_t + \beta_t + \gamma_t\kappa\left(f(x^*) - f(\boldsymbol{x}_t)\right) \tag{IV.42}$$

avec :

$$\begin{aligned}\alpha_t =& 12(\gamma_t)^2\left(\kappa^2 a_1^2 + \frac{a_1^2 A}{\epsilon_t}\right) + 2\gamma_t\eta_t(1 + a_1 + a_3), \\ \beta_t =& 3(\gamma_t)^2\left(2\kappa^2 a_3^2 + A + \frac{2Aa_3^2}{\epsilon_t}\right) + \gamma_t\eta_t(1 + a_3). \end{aligned}$$

Ainsi, (α_t) et (β_t) sont deux séries sommables. En prenant l'espérance dans l'équation (IV.42), et en notant $\lambda_t = \mathbb{E}(\Lambda_t)$, on obtient :

$$\lambda_{t+1} - \lambda_t \leq \alpha_t \lambda_t + \beta_t + \gamma_t \kappa \underbrace{\mathbb{E}(f(x^*) - f(\boldsymbol{x}_t))}_{\leq 0, \text{ par optimalité}}. \tag{IV.43}$$

Le lemme A.33 montre alors que (λ_t) est bornée par un réel $M > 0$. On prouve maintenant que (Λ_t) est une quasi-martingale convergente. En effet :

- Par définition, (Λ_t) est $(\mathcal{F}_t)$-adaptée.
- Par définition, pour tout $t \in \mathbb{N}$, $\Lambda_t \geq 0$, i.e., $\inf_{t\in\mathbb{N}} \mathbb{E}(\Lambda_t) \geq 0$.
- Soit pour tout $t \in \mathbb{N}$, $C_t := \{\mathbb{E}(\Lambda_{t+1} - \Lambda_t | \mathcal{F}_t) > 0\}$. Clairement, 1_{C_t} est $\mathcal{F}_t$-mesurable. Par l'équation (IV.42), on a donc :

$$\begin{aligned}
\sum_{t\in\mathbb{N}} \mathbb{E}(1_{C_t} \cdot (\Lambda_{t+1} - \Lambda_t)) &\leq \sum_{t\in\mathbb{N}} \mathbb{E}(1_{C_t} \cdot \mathbb{E}(\Lambda_{t+1} - \Lambda_t | \mathcal{F}_t)), \\
&\leq \sum_{t\in\mathbb{N}} \mathbb{E}(1_{C_t}(\alpha_t \Lambda_t + \beta_t)), \\
&\leq \sum_{t\in\mathbb{N}} (\alpha_t M + \beta_t) < +\infty.
\end{aligned}$$

- Il est clair également que $\sup_{t\in\mathbb{N}} \mathbb{E}(\min(\Lambda_t, 0)) < +\infty$. Par conséquent, en utilisant le théorème B.6, (Λ_t) est une quasi-martingale et converge p.s. vers une variable aléatoire intégrable. Par conséquent, cette suite est bornée p.s., et par définition, en utilisant l'hypothèse (IV.34), les suites $(\boldsymbol{x}_t)$ $(\boldsymbol{s}_t)$, $(\boldsymbol{v}_t)$ sont presque sûrement bornées dans X.

On prouve maintenant que $(f(\boldsymbol{x}_t))$ converge p.s. vers $f(x^*)$. L'équation (IV.43) donne pour tout $t \in \mathbb{N}$:

$$\gamma_t \mathbb{E}(f(\boldsymbol{x}_t) - f(x^*)) \leq \alpha_t \lambda_t + \beta_t + \lambda_t - \lambda_{t+1}.$$

On ajoute ces inégalités pour $t = 0, \ldots, n$:

$$\begin{aligned}
\kappa \sum_{t=0}^{n} \gamma_t \mathbb{E}(f(\boldsymbol{x}_t) - f(x^*)) &\leq \lambda_0 - \lambda_{n+1} + \sum_{t=0}^{n} (\alpha_t M + \beta_t), \\
&\leq M + M \sum_{t=0}^{n} \alpha_t + \sum_{t=0}^{n} \beta_t.
\end{aligned} \tag{IV.44}$$

En faisant tendre $n \to \infty$:

$$\sum_{t\in\mathbb{N}} \gamma_t \mathbb{E}(f(\boldsymbol{x}_t) - f(x^*)) < \infty.$$

Par optimalité, tous les termes sous l'espérance sont positifs p.s. Par conséquent :

$$\sum_{k\in\mathbb{N}} \gamma_t (f(\boldsymbol{x}_t) - f(x^*)) < \infty. \tag{IV.45}$$

On utilise maintenant le lemme A.35. Soit $l \in \mathbb{N}$. f étant convexe,

$$\begin{aligned}
f(\boldsymbol{x}_l) - f(\boldsymbol{x}_{l+1}) &\leq \langle \boldsymbol{v}_l, \boldsymbol{x}_l - \boldsymbol{x}_{l+1} \rangle, \\
&= \langle \boldsymbol{v}_l, \boldsymbol{x}_l - \Pi_{X^f}(\boldsymbol{x}_l + \gamma_l(\boldsymbol{s}_l + \boldsymbol{w}_l)) \rangle
\end{aligned}$$

Distinguons alors deux cas :

- Si X^f est un sous-espace vectoriel, la projection est linéaire et auto-adjointe, d'où

$$\begin{aligned}
f(\boldsymbol{x}_l) - f(\boldsymbol{x}_{l+1}) &\leq \langle \boldsymbol{v}_l, \boldsymbol{x}_l - \boldsymbol{x}_{l+1} \rangle, \\
&= -\gamma_l \langle \boldsymbol{v}_l, \boldsymbol{s}_l + \boldsymbol{w}_l \rangle
\end{aligned}$$

  En conditionnant par rapport à $\mathcal{F}_l$, on obtient :

$$\begin{aligned}
f(\boldsymbol{x}_l) - \mathbb{E}(f(\boldsymbol{x}_{l+1}) | \mathcal{F}_l) &\leq -\gamma_l \langle \boldsymbol{v}_l, \boldsymbol{s}_l + \mathbb{E}(\boldsymbol{w}_l | \mathcal{F}_l) \rangle \\
&\leq \gamma_l \|\boldsymbol{v}_l\| \times (\|\boldsymbol{s}_l\| + \|\mathbb{E}(\boldsymbol{w}_l | \mathcal{F}_l)\|), \\
&\leq \gamma_l \delta,
\end{aligned} \tag{IV.46}$$

  en utilisant l'hypothèse (IV.36a), car les suites $(\|\boldsymbol{s}_t\|), (\|\boldsymbol{v}_t\|)$ sont bornées presque sûrement.

– Si X^f est un convexe fermé, on a :

$$f(\boldsymbol{x}_l) - f(\boldsymbol{x}_{l+1}) \leq \gamma_l \|\boldsymbol{v}_l\| \times \|\boldsymbol{s}_l + \boldsymbol{w}_l\|.. \tag{IV.47}$$

En conditionnant par rapport à $\mathcal{F}_l$ on trouve :

$$\begin{aligned} f(\boldsymbol{x}_l) - \mathbb{E}\left(f(\boldsymbol{x}_{l+1})|\mathcal{F}_l\right) &\leq \gamma_l \|\boldsymbol{v}_l\| \times \mathbb{E}\left(\|\boldsymbol{s}_l + \boldsymbol{w}_l\||\mathcal{F}_l\right) \\ &\leq \gamma_l \|\boldsymbol{v}_l\| g(\boldsymbol{s}_l) \leq \gamma_l \delta', \end{aligned} \tag{IV.48}$$

avec $\delta' > 0$, car on sait que les suites $(\|\boldsymbol{s}_t\|)$ et $(\|\boldsymbol{v}_t\|)$ sont bornées, et on applique (IV.36c).
On peut donc appliquer le lemme A.35, grâce aux équation (IV.45) et (IV.46)–(IV.48), ce qui donne

$$\lim_{t\to\infty} f(\boldsymbol{x}_t) = f(x^*) \tag{IV.49}$$

Soit $\bar{x}$ un point d'accumulation de $(\boldsymbol{x}_t)$ dans la topologie faible. Un tel point existe car $(\boldsymbol{x}_t)$ est bornée. Il existe donc une sous-suite $(\boldsymbol{x}_{\phi(t)})$ qui converge vers $\bar{x}$. Comme X^f est convexe fermé et donc faiblement fermé, $\bar{x} \in X^f$, et par semi continuité inférieure faible de f (ce qui provient de la convexité et de la semi continuité forte), il vient :

$$f(\bar{x}) \leq \liminf_{t\to\infty} f(\boldsymbol{x}_{\phi(t)}) = f(x^*),$$

et donc $\bar{x} \in S$.
Supposons maintenant que f soit fortement convexe de module $B > 0$. Dans ce cas, S se réduit à un singleton $\{x^*\}$. Par définition,

$$f(\boldsymbol{x}_t) - f(x^*) \geq \langle \nabla f(x^*), \boldsymbol{x}_t - x^* \rangle + \frac{B}{2}\|x^* - \boldsymbol{x}_t\|^2 \tag{IV.50}$$

Par optimalité, $\langle \nabla f(x^*), \boldsymbol{x}_t - x^* \rangle \geq 0$. (IV.50) donne donc la convergence forte de $(\boldsymbol{x}_t)$ vers x^*, ce qui achève la preuve. □

Remarque IV.11 (Hypothèses (IV.36)). *Les hypothèses* (IV.36) *proviennent du caractère stochastique de l'algorithme considéré. L'hypothèse* (IV.36a) *s'interprétent comme une relaxation de l'hypothèse usuelle d'incrément de martingale qui s'écrirait* $\mathbb{E}\left(\boldsymbol{w}_t|\mathcal{F}_t\right) = 0$. *L'hypothèse* (IV.36b) *est une hypothèse sur la variance de ce qui joue le rôle d'incrément de martingale. Habituellement, la variance de l'incrément de martingale doit être bornée. Ici, nous l'autorisons à croître a priori. Enfin, l'hypothèse* (IV.36c) *peut être vue comme la concession introduite pour compenser la présence d'une projection sur un convexe fermé. Cette hypothèse est bien entendu compatible avec les autres, et n'est présente que lorsque la projection ne s'effectue pas sur un sous-espace vectoriel.*

Remarque IV.12 (Forte convexité). *En suivant le travail de* [**15**] *on peut affaiblir l'hypothèse (vii) du théorème IV.10, c'est à dire l'hypothèse de forte convexité. En effet, si la fonction f est supposée strictement convexe, la convergence forte de (x_t) vers l'unique solution x^* du problème* (IV.33) *peut également être prouvée. L'idée de* [**15**] *est la suivante : ayant $(\boldsymbol{x}_t)$ convergeant faiblement vers x^*, et $(f(\boldsymbol{x}_t))$ convergeant vers $f(x^*)$, avec f strictement convexe, on en déduit la convergence de $(\|\boldsymbol{x}_t - x^*\|)$ vers 0. C'est pour ne pas compliquer davantage la preuve que nous avons choisi ici l'hypothèse plus forte de forte convexité sur f.*

Remarque IV.13 (Direction de descente). *On peut remplacer l'hypothèse* (IV.35) *par le jeux d'hypothèses suivantes plus faibles : pour tout $x^* \in S$, pour tout $t \in \mathbb{N}$,*

$$\langle \boldsymbol{s}_t, \boldsymbol{x}_t - x^* \rangle \leq \kappa\left(f(x^*) - f(\boldsymbol{x}_t)\right), \tag{IV.51a}$$

$$\exists \boldsymbol{v}_t \in \partial f(\boldsymbol{x}_t),\ \|\boldsymbol{s}_t\| \leq c\left(1 + \|\boldsymbol{v}_t\|\right). \tag{IV.51b}$$

Nous donnons maintenant un théorème de convergence dans le cas où la fonction à optimiser est fortement convexe, et la projection se fait sur un convexe fermé. L'intérêt de ce dernier théorème est d'éviter, sous l'hypothèse de forte convexité, l'introduction de l'hypothèse sur les bruits (IV.36c). Par simplicité, nous énonçons ce théorème dans le cas différentiable, mais il est également vrai dans le cas sous-différentiable.

Théorème IV.14 (Projection sur un convexe fermé). *(i) Supposons que $x \mapsto f(x)$ soit fortement convexe de module $B > 0$, semi continue inférieurement et différentiable. Supposons de plus que X^f soit un convexe fermé de X. Alors, le problème* (IV.33) *admet une unique solution notée x^*.*

(ii) Soit $(\mathcal{F}_t)$ une filtration sur $(\Omega, \mathcal{A}, \mathbb{P})$, et supposons que pour tout $t \in \mathbb{N}$, $\boldsymbol{s}_t$ et $\boldsymbol{w}_{t-1}$ sont $\mathcal{F}_t$-mesurables.

(iii) Supposons que $\nabla f(\cdot)$ soit lipschitzienne sur X^f de module L.

(iv) Supposons qu'il existe $c, \kappa > 0$, tels que pour tout $t \in \mathbb{N}$,

$$\langle \boldsymbol{s}_t, \boldsymbol{x}_t - x^* \rangle \leq \kappa \left(f(x^*) - f(\boldsymbol{x}_t) \right), \tag{IV.52a}$$

$$\|\boldsymbol{s}_t\| \leq c \left(1 + \|\nabla f(\boldsymbol{x}_t)\|\right). \tag{IV.52b}$$

(v) Supposons qu'il existe $A > 0$ et deux suites de réels (ϵ_t, η_t), tels que pour tout $t \in \mathbb{N}$,

$$\|\mathbb{E}\left(\boldsymbol{w}_t | \mathcal{F}_t\right)\| \leq b\eta_t \left(1 + \|\nabla f(\boldsymbol{x}_t)\|\right), \tag{IV.53a}$$

$$\mathbb{E}\left(\|\boldsymbol{w}_t\|^2 | \mathcal{F}_t\right) \leq A \left(1 + \frac{1}{\epsilon_t} \|\nabla f(\boldsymbol{x}_t)\|^2\right). \tag{IV.53b}$$

(vi) Supposons que les suites (γ_t), (ϵ_t) et (η_t) décroissent vers 0, et soient telles que :

$$\forall t \in \mathbb{N},\ \gamma_t, \epsilon_t > 0, \quad \sum_{t \in \mathbb{N}} \gamma_t = +\infty, \quad \sum_{t \in \mathbb{N}} \frac{(\gamma_t)^2}{\epsilon_t} < +\infty, \quad \sum_{t \in \mathbb{N}} b\gamma_t\eta_t < +\infty. \tag{IV.54}$$

Alors $f(\boldsymbol{x}_t) \to f(x^*)$ p.s., quand t tend vers l'infini, et $(\boldsymbol{x}_t)$ converge fortement p.s. vers l'unique solution x^* du problème (IV.33).

Preuve : On suit le schéma de Robbins-Siegmund (voir [**78**]). Soit $t \in \mathbb{N}$. On a :

$$\begin{aligned}
\|\boldsymbol{x}_{t+1} - x^*\|^2 =& \|\Pi_{X^f}\left(\boldsymbol{x}_t + \gamma_t(\boldsymbol{s}_t + \boldsymbol{w}_t)\right) - x^*\|^2 \\
\leq& \|\boldsymbol{x}_t - x^* + \gamma_t(\boldsymbol{s}_t + \boldsymbol{w}_t)\|^2, \text{ par contraction de } \Pi \\
\leq& \|\boldsymbol{x}_t - x^*\|^2 + (\gamma_t)^2\|\boldsymbol{s}_t + \boldsymbol{w}_t\|^2 + 2\gamma_t\langle \boldsymbol{s}_t + \boldsymbol{w}_t, \boldsymbol{x}_t - x^*\rangle \\
\leq& \|\boldsymbol{x}_t - x^*\|^2 + 2(\gamma_t)^2\|\boldsymbol{s}_t\|^2 + 2(\gamma_t)^2\|\boldsymbol{w}_t\|^2 \\
&+ 2\gamma_t\langle \boldsymbol{w}_t, \boldsymbol{x}_t - x^*\rangle + 2\gamma_t\kappa\left(f(x^*) - f(\boldsymbol{x}_t)\right)
\end{aligned} \tag{IV.55}$$

par l'hypothèse (IV.52a), et en utilisant l'inégalité classique $(x+y)^2 \leq 2x^2 + 2y^2$. En utilisant cette inégalité une nouvelle fois, ainsi que l'hypothèse (IV.52b), on obtient :

$$\begin{aligned}
\|\boldsymbol{x}_{t+1} - x^*\|^2 \leq& \|\boldsymbol{x}_t - x^*\|^2 + 4c^2(\gamma_t)^2\left(1 + \|\nabla f(\boldsymbol{x}_t)\|^2\right) \\
&+ 2(\gamma_t)^2\|\boldsymbol{w}_t\|^2 + 2\gamma_t\langle \boldsymbol{w}_t, \boldsymbol{x}_t - x^*\rangle + 2\gamma_t\kappa\left(f(x^*) - f(\boldsymbol{x}_t)\right)
\end{aligned} \tag{IV.56}$$

La forte convexité de f s'écrit :

$$f(\boldsymbol{x}_t) - f(x^*) + \langle \nabla f(x^*), x^* - \boldsymbol{x}_t\rangle \geq \frac{B}{2}\|\boldsymbol{x}_t - x^*\|^2.$$

Par optimalité, on a $\langle \nabla f(x^*), x^* - \boldsymbol{x}_t\rangle \leq 0$. Par conséquent,

$$f(\boldsymbol{x}_t) - f(x^*) \geq \frac{B}{2}\|\boldsymbol{x}_t - x^*\|^2.$$

En prenant l'espérance conditionnelle par rapport à $\mathcal{F}_t$ dans l'équation (IV.56), et en utilisant l'inégalité de forte convexité précédente, il vient :

$$\begin{aligned}
\mathbb{E}\left(\|\boldsymbol{x}_{t+1} - x^*\|^2 | \mathcal{F}_t\right) \leq& \|\boldsymbol{x}_t - x^*\|^2 + 4c^2(\gamma_t)^2\left(1 + \|\nabla f(\boldsymbol{x}_t)\|^2\right) \\
&+ 2(\gamma_t)^2\mathbb{E}\left(\|\boldsymbol{w}_t\|^2 | \mathcal{F}_t\right) + 2\gamma_t\|\mathbb{E}\left(\boldsymbol{w}_t | \mathcal{F}_t\right)\|\|\boldsymbol{x}_t - x^*\| \\
&- B\kappa\gamma_t\|\boldsymbol{x}_t - x^*\|^2, \\
\leq& \|\boldsymbol{x}_t - x^*\|^2 + 4c^2(\gamma_t)^2 + 8c^2(\gamma_t)^2\|\nabla f(x^*)\|^2 \\
&+ 8c^2(\gamma_t)^2\|\nabla f(\boldsymbol{x}_t) - \nabla f(x^*)\|^2 + 2A(\gamma_t)^2\left(1 + \frac{1}{\epsilon_t}\|\nabla f(\boldsymbol{x}_t)\|^2\right) \\
&+ 2\gamma_t\eta_t\left(1 + \|\nabla f(\boldsymbol{x}_t)\|\right)\|\boldsymbol{x}_t - x^*\| - B\kappa\gamma_t\|\boldsymbol{x}_t - x^*\|^2
\end{aligned} \tag{IV.57}$$

graâce aux hypothèses (IV.53a)–(IV.53b). On utilise maintenant l'inégalité $x \leq 1 + x^2$ et le caractère lipschitzien de $\nabla f(\cdot)$. Ainsi,

$$\begin{aligned}\mathbb{E}\left(\|\boldsymbol{x}_{t+1} - x^*\|^2 | \mathcal{F}_t\right) \leq & \left(1 + 8c^2(\gamma_t)^2 L^2 + 4A\frac{(\gamma_t)^2}{\epsilon_t}L^2 + 2\gamma_t\eta_t(1 + L + \|\nabla f(x^*)\|\right)\|\boldsymbol{x}_t - x^*\|^2 \\ & + 2(\gamma_t)^2 A + 4c^2(\gamma_t)^2\left(1 + 2\|\nabla f(x^*)\|^2\right) + 4A\frac{(\gamma_t)^2}{\epsilon_t}\|\nabla f(x^*)\|^2 \\ & + 2\gamma_t\eta_t(1 + \|\nabla f(x^*)\|) - B\kappa\gamma_t\|\boldsymbol{x}_t - x^*\|^2 \\ = & (1 + \alpha_t)\|\boldsymbol{x}_t - x^*\|^2 + \beta_t - \delta_t, \end{aligned} \tag{IV.58}$$

avec α_t et β_t les termes de deux séries sommables par l'hypothèse (IV.54), et

$$\delta_t = B\kappa\gamma_t\|\boldsymbol{x}_t - x^*\|^2 \geq 0.$$

On peut donc appliquer le lemme de Robbins-Siegmund (cf. Lemme C.3) qui donne finalement :

$$\begin{aligned} & \|\boldsymbol{x}_t - x^*\|^2 \text{ converge p.s. quand } t \to \infty, \text{ et,} \\ & \sum_{t\in\mathbb{N}} \gamma_t\|\boldsymbol{x}_t - x^*\|^2 < +\infty. \end{aligned}$$

Ainsi, $(\|\boldsymbol{x}_t - x^*\|^2)$ converge vers 0 quand t tend vers l'infini, ce qui achève la preuve. □

Remarque IV.15 (Pas aléatoires). *Les pas (ρ_t), (ϵ_t) et (η_t) introduits dans les théorèmes IV.14 et IV.10 peuvent être pris comme des suites de variables aléatoires positives $(\mathcal{F}_t)$-adaptées. En effet, le lemme de Robbins-Siegmund, tout comme le résultat de Métivier sur les quasi-martingales peuvent toujours être appliqués dans ce cas. Cette remarque permet de choisir des pas* online*, en fonction des tirages passés, ce qui peut être intéressant numériquement en vue d'accélérer la convergence.*

Outre les hypothèses déjà discutées sur les pas de descente et pas d'approximation, ou sur la fonction f, l'intérêt majeur de ce théorème est de proposer des hypothèses faibles sur les bruits affectant la direction de descente. En effet, il est assez usuel de demander aux bruits d'être de moyenne conditionnelle nulle, i.e. d'être des incréments de martingale. Ici, on ne demande finalement qu'asymptotiquement cette propriété. De plus, nos hypothèses permettent de soulever le lien entre les pas de descente et les pas d'approximation, ce qui fournit un guide supplémentaire dans la perspective du réglage du schéma numérique d'approximation stochastique. En particulier, nous renvoyons à la figure 2, qui donnait une interprétation graphique des hypothèses sur les pas dans le cas particulier de l'algorithme IV.4. Nous verrons au paragraphe IV.4.3.1 le lien entre l'algorithme IV.9 et l'algorithme IV.4.

IV.4.2. Algorithme de Arrow-Hurwicz hilbertien. Afin d'aller plus loin dans la généralisation, nous donnons maintenant une version *à deux niveaux* de l'algorithme de gradient perturbé, utile pour les problèmes de point-selle apparaissant en théorie des jeux par exemple, ou avec la dualité lagrangienne en optimisation convexe. On propose donc de résoudre le problème :

$$\begin{aligned} & \min_{x\in X}\max_{p\in P} L(x,p), \\ & \qquad \text{s.c. } x \in X^f,\ p \in P^f, \end{aligned} \tag{IV.59}$$

avec

- X et P deux espaces de Hilbert munis respectivement des produits scalaires et normes notés $\langle\cdot,\cdot\rangle_X$, $\langle\cdot,\cdot\rangle_P$ et $\|\cdot\|_X$, $\|\cdot\|_P$,
- $L : X \times P \to \mathbb{R}$ est une application convexe-concave,
- X^f, P^f sont des convexes fermés de X et P respectivement, et $\Pi_\cdot(\cdot)$ sera l'opérateur de projection correspondant.

On se propose d'étudier le comportement de l'algorithme suivant pour la résolution du problème (IV.59) :

ALGORITHME IV.16. *Étape $t \in \mathbb{N}$:*

$$\begin{aligned}\boldsymbol{x}_{t+1} =& \Pi_{X^f}\left(\boldsymbol{x}_t + \gamma_t^x(\boldsymbol{s}_t + \boldsymbol{w}_t)\right),\\ \boldsymbol{p}_{t+1} =& \Pi_{P^f}\left(\boldsymbol{p}_t + \gamma_t^p(\boldsymbol{r}_t + \boldsymbol{v}_t)\right).\end{aligned}$$

$\boldsymbol{s}_t$ est donc une direction de descente, tandis que $\boldsymbol{r}_t$ est une direction de montée, et $\boldsymbol{w}_t, \boldsymbol{v}_t$ sont les perturbations. Les pas de gradient γ_t^x, γ_t^p seront dorénavant pris égaux.

Il serait à nouveau possible de proposer différents théorèmes de convergence avec des hypothèses de forte convexité, etc. Pour simplifier la présentation, nous ne donnons ici qu'un seul théorème regroupant les cas les plus intéressants pratiquement, et étant donc énoncé sans hypothèse de différentiabilité.

THÉORÈME IV.17 (Problème de point-selle). *(i) Supposons que $L(\cdot, p) : X \to \mathbb{R}$ soit convexe, semi continue inférieurement pour tout $p \in P$, et que $L(x, \cdot) : P \to \mathbb{R}$ soit concave, semi continue supérieurement pour tout $x \in X$. Supposons que X^f et P^f soient des convexes fermés de X et P, et qu'il existe un point selle (x^*, p^*) à L sur $X^f \times P^f$.*

(ii) Soit $(\mathcal{F}_t)$ une filtration, et supposons que pour tout $t \in \mathbb{N}$, $\boldsymbol{x}_t, \boldsymbol{s}_t, \boldsymbol{p}_t$ et $\boldsymbol{r}_t$ soient $\mathcal{F}_t$-mesurables.

(iii) Supposons que pour tout $(x, p) \in X^f \times P^f$, $\partial_x L(x, p)$ et $\partial_p L(x, p)$ soient non vides et qu'il existe deux réels $a_1, a_2 \geq 0$ tels que

$$\forall (x,p) \in X^f \times P^f,\ \forall u_x \in \partial_x L(x,p),\ \|u_x\|_X \leq a_1\|x\|_X + a_2, \tag{IV.60a}$$

$$\forall (x,p) \in X^f \times P^f,\ \forall u_p \in \partial_p L(x,p),\ \|u_p\|_P \leq a_1\|p\|_P + a_2, \tag{IV.60b}$$

(iv) Supposons qu'il existe $c, \kappa > 0$ tels que pour tout $t \in \mathbb{N}$,

$$\langle \boldsymbol{s}_t, \boldsymbol{x}_t - x^*\rangle_X \leq \kappa\left(L(x^*, \boldsymbol{p}_t) - L(\boldsymbol{x}_t, \boldsymbol{p}_t)\right), \tag{IV.61a}$$

$$\langle \boldsymbol{r}_t, \boldsymbol{p}_t - p^*\rangle_P \leq \kappa\left(L(\boldsymbol{x}_t, \boldsymbol{p}_t) - L(\boldsymbol{x}_t, p^*)\right), \tag{IV.61b}$$

$$\exists \boldsymbol{u}_t^x \in \partial_x L(\boldsymbol{x}_t, \boldsymbol{p}_t),\ \|\boldsymbol{s}_t\|_X \leq c\left(1 + \|\boldsymbol{u}_t^x\|\right), \tag{IV.61c}$$

$$\exists \boldsymbol{u}_t^p \in \partial_p L(\boldsymbol{x}_t, \boldsymbol{p}_t),\ \|\boldsymbol{r}_t\|_P \leq c\left(1 + \|\boldsymbol{u}_t^p\|\right). \tag{IV.61d}$$

(v) Supposons qu'il existe un réel $A > 0$ et des suites réelles positives (ϵ_t^x, η_t^x) et (ϵ_t^p, η_t^p) tels que pour tout $t \in \mathbb{N}$, il existe $(\boldsymbol{u}_t^x, \boldsymbol{u}_t^p) \in \partial_x L(\boldsymbol{x}_t, \boldsymbol{p}_t) \times \partial_p L(\boldsymbol{x}_t, \boldsymbol{p}_t)$ tels que

$$\|\mathbb{E}\left(\boldsymbol{w}_t|\mathcal{F}_t\right)\|_X \leq \eta_t^x\left(1 + \|\boldsymbol{u}_t^x\|_X\right), \tag{IV.62a}$$

$$\|\mathbb{E}\left(\boldsymbol{v}_t|\mathcal{F}_t\right)\|_P \leq \eta_t^p\left(1 + \|\boldsymbol{u}_t^p\|_P\right), \tag{IV.62b}$$

$$\mathbb{E}\left(\|\boldsymbol{w}_t\|_X^2|\mathcal{F}_t\right) \leq A\left(1 + \frac{1}{\epsilon_t^x}\|\boldsymbol{u}_t^x\|_X^2\right), \tag{IV.62c}$$

$$\mathbb{E}\left(\|\boldsymbol{v}_t\|_P^2|\mathcal{F}_t\right) \leq A\left(1 + \frac{1}{\epsilon_t^p}\|\boldsymbol{u}_t^p\|_P^2\right). \tag{IV.62d}$$

Si X^f (resp. P^f) n'est pas un sous-espace vectoriel, supposons aussi qu'il existe une application continue ou bornée $g_x : \mathbb{R} \to \mathbb{R}$ (resp. g_t), et pour tout $t \in \mathbb{N}$ un élément $\boldsymbol{u}_t^x \in \partial_x L(\boldsymbol{x}_t, \boldsymbol{p}_t)$ (resp. $\boldsymbol{u}_t^p \in \partial_p L(\boldsymbol{x}_t, \boldsymbol{p}_t)$) tels que,

$$\mathbb{E}\left(\|\boldsymbol{w}_t\|_X|\mathcal{F}_t\right) \leq g_x(\|\boldsymbol{u}_t^x\|_X),\ (\text{resp. } \mathbb{E}\left(\|\boldsymbol{v}_t\|_P|\mathcal{F}_t\right) \leq g_p(\|\boldsymbol{u}_t^p\|_P)). \tag{IV.62e}$$

(vi) Supposons que les suites (γ_t), (ϵ_x^t), (ϵ_t^p), (η_t^x) et (η_t^p) soient toutes strictement positives et décroissent vers 0 tout en vérifiant :

$$\sum_{t\in\mathbb{N}} \gamma_t = +\infty,\ \sum_{t\in\mathbb{N}} \gamma_t\eta_t^x < +\infty,\ \sum_{t\in\mathbb{N}} \gamma_t\eta_t^p < +\infty\ \sum_{t\in\mathbb{N}} \frac{(\gamma_t)^2}{\epsilon_t^x} < +\infty,\ \sum_{t\in\mathbb{N}} \frac{(\gamma_t)^2}{\epsilon_t^p} < +\infty. \tag{IV.63a}$$

Alors, $(\boldsymbol{x}_t)$ et $(\boldsymbol{p}_t)$ sont presque sûrement bornées, et $L(\boldsymbol{x}_t, p^*) \to L(x^*, p^*)$, et $L(x^*, \boldsymbol{p}_t) \to L(x^*, p^*)$ presque sûrement, quand t tend vers l'infini. De plus, si $L(\cdot, p^*)$ est fortement convexe, $(\boldsymbol{x}_t)$ converge presque sûrement fortement vers x^*.

Preuve : Nous suivons le même schéma de preuve que pour le théorème IV.10. Définissons pour tout $t \in \mathbb{N}$, Λ_t notre fonction de Lyapunov par :

$$\Lambda_t = \|\boldsymbol{x}_t - x^*\|_X^2 + \|\boldsymbol{p}_t - p^*\|_P^2.$$

De la même manière que pour arriver à l'équation (IV.8), nous obtenons :

$$\begin{aligned}\Lambda_{t+1} \leq & \Lambda_t + 3(\gamma_t)^2 \left(\|\boldsymbol{s}_t + \boldsymbol{w}_t\|_X^2 + \|\boldsymbol{r}_t + \boldsymbol{v}_t\|_P^2\right) \\ & + 2\gamma_t \left(\langle \boldsymbol{s}_t + \boldsymbol{w}_t, \boldsymbol{x}_t - x^* \rangle_X + \langle \boldsymbol{r}_t + \boldsymbol{v}_t, \boldsymbol{p}_t - p^* \rangle_P\right). \end{aligned} \tag{IV.64}$$

Avec l'inégalité scalaire $(a+b)^2 \leq 2(a^2+b^2)$, et par les hypothèses (IV.61a)–(IV.61b), nous poursuivons (IV.64) :

$$\begin{aligned}\Lambda_{t+1} \leq & \Lambda_t + 6(\gamma_t)^2 \left(\|\boldsymbol{s}_t\|_X^2 + \|\boldsymbol{w}_t\|_X^2\right) \\ & + 6(\gamma_t)^2 \left(\|\boldsymbol{r}_t\|_P^2 + \|\boldsymbol{v}_t\|_P^2\right) \\ & + 2\gamma_t \left(L(x^*, \boldsymbol{p}_t) - L(\boldsymbol{x}_t, \boldsymbol{p}_t) + L(\boldsymbol{x}_t, \boldsymbol{p}_t) - L(\boldsymbol{x}_t, p^*)\right) \\ & + 2\gamma_t \left(\langle \boldsymbol{w}_t, \boldsymbol{x}_t - x^* \rangle_X + \langle \boldsymbol{v}_t, \boldsymbol{p}_t - p^* \rangle_P\right). \end{aligned} \tag{IV.65}$$

De plus, les hypothèses (IV.61c)–(IV.61d) impliquent que :

$$\begin{aligned}\|\boldsymbol{s}_t\|_X^2 \leq & 2c^2 \left(1 + \|\boldsymbol{u}_t^x\|_X^2\right), \\ \|\boldsymbol{r}_t\|_P^2 \leq & 2c^2 \left(1 + \|\boldsymbol{u}_t^p\|_P^2\right). \end{aligned}$$

Finalement, avec l'hypothèse (IV.60), on arrive à :

$$\begin{aligned}\|\boldsymbol{s}_t\|_X^2 \leq & 2c^2 \left(1 + 2(a_1)^2 \|\boldsymbol{x}_t - x^*\|_X^2 + 2(a_2 + a_1\|x^*\|_X)^2\right), \\ \|\boldsymbol{r}_t\|_X^2 \leq & 2c^2 \left(1 + 2(a_1)^2 \|\boldsymbol{p}_t - p^*\|_P^2 + 2(a_2 + a_1\|p^*\|_P)^2\right). \end{aligned}$$

Définissons alors $a_3 = 4c^2(a_1)^2$ et $a_4^x = 2c^2\left(1 + 2(a_2 + a_1\|x^*\|_X)^2\right)$ et de la même façon a_4^p. On peut alors écrire :

$$\|\boldsymbol{s}_t\|_X^2 \leq a_3 \|\boldsymbol{x}_t - x^*\|_X^2 + a_4^x, \tag{IV.66a}$$

$$\|\boldsymbol{r}_t\|_X^2 \leq a_3 \|\boldsymbol{p}_t - p^*\|_P^2 + a_4^p. \tag{IV.66b}$$

Par les mêmes arguments, les hypothèses (IV.60) et (IV.62c)–(IV.62d) se réécrivent comme

$$\mathbb{E}\left(\|\boldsymbol{w}_t\|_X^2 | \mathcal{F}_t\right) \leq A\left(1 + \frac{2}{\epsilon_t^x}\left((a_1)^2 \|\boldsymbol{x}_t - x^*\|_X^2 + (a_2 + a_1\|x^*\|_X)^2\right)\right) \tag{IV.67a}$$

$$\mathbb{E}\left(\|\boldsymbol{v}_t\|_P^2 | \mathcal{F}_t\right) \leq A\left(1 + \frac{2}{\epsilon_t^p}\left((a_1)^2 \|\boldsymbol{p}_t - p^*\|_P^2 + (a_2 + a_1\|p^*\|_P)^2\right)\right) \tag{IV.67b}$$

Nous conditionnons maintenant l'équation (IV.65) par rapport à $\mathcal{F}_t$ et appliquons les inégalités (IV.66)–(IV.67).

$$\begin{aligned}\mathbb{E}\left(\Lambda_{t+1} | \mathcal{F}_t\right) \leq & \Lambda_t + 6(\gamma_t)^2 \left(a_3 \|\boldsymbol{x}_t - x^*\|_X^2 + a_4^x + a_3 \|\boldsymbol{p}_t - p^*\|_P^2 + a_4^p\right) \\ & + 6(\gamma_t)^2 \left(A\left(1 + \frac{2}{\epsilon_t^x}\left((a_1)^2 \|\boldsymbol{x}_t - x^*\|_X^2 + (a_2 + a_1\|x^*\|_X)^2\right)\right)\right) \\ & + 6(\gamma_t)^2 \left(A\left(1 + \frac{2}{\epsilon_t^p}\left((a_1)^2 \|\boldsymbol{p}_t - p^*\|_P^2 + (a_2 + a_1\|p^*\|_P)^2\right)\right)\right) \\ & + 2\gamma_t \left(L(x^*, \boldsymbol{p}_t) - L(\boldsymbol{x}_t, \boldsymbol{p}_t) + L(\boldsymbol{x}_t, \boldsymbol{p}_t) - L(\boldsymbol{x}_t, p^*)\right) \\ & + 2\gamma_t \left(\|\mathbb{E}\left(\boldsymbol{w}_t | \mathcal{F}_t\right)\| \times \|\boldsymbol{x}_t - x^*\|_X + \|\mathbb{E}\left(\boldsymbol{v}_t | \mathcal{F}_t\right)\| \times \|\boldsymbol{p}_t - p^*\|_P\right) \end{aligned} \tag{IV.68}$$

Les hypothèses (IV.62a)–(IV.62b) permettent de borner les derniers termes de l'équation (IV.68), et nous obtenons finalement

$$\begin{aligned}\mathbb{E}\left(\Lambda_{t+1}|\mathcal{F}_t\right) \leq & \Lambda_t + 6(\gamma_t)^2\left(a_3\|\boldsymbol{x}_t - x^*\|_X^2 + a_4^x + a_3\|\boldsymbol{p}_t - p^*\|_P^2 + a_4^p\right)\\ &+ 10(\gamma_t)^2\left(A\left(1 + \frac{2}{\epsilon_t^x}\left((a_1)^2\|\boldsymbol{x}_t - x^*\|_X^2 + (a_2 + a_1\|x^*\|_X)^2\right)\right)\right)\\ &+ 10(\gamma_t)^2\left(A\left(1 + \frac{2}{\epsilon_t^p}\left((a_1)^2\|\boldsymbol{p}_t - p^*\|_P^2 + (a_2 + a_1\|p^*\|_P)^2\right)\right)\right)\\ &+ 2\gamma_t\left(L(x^*, \boldsymbol{p}_t) - L(\boldsymbol{x}_t, \boldsymbol{p}_t) + L(\boldsymbol{x}_t, \boldsymbol{p}_t) - L(\boldsymbol{x}_t, p^*)\right)\\ &+ 2\eta_t^x\gamma_t\left(a_1\|\boldsymbol{x}_t - x^*\|_X + a_2 + a_1\|x^*\|_X\right)\|\boldsymbol{x}_t - x^*\|_X\\ &+ 2\eta_t^p\gamma_t\left(a_1\|\boldsymbol{p}_t - p^*\|_P + a_2 + a_1\|p^*\|_P\right)\|\boldsymbol{p}_t - p^*\|_P\end{aligned} \tag{IV.69}$$

Nous utilisons alors l'inégalité scalaire $ab \leq \frac{a^2+b^2}{2}$. Ainsi, (IV.69) se réécrit :

$$\begin{aligned}\mathbb{E}\left(\Lambda_{t+1}|\mathcal{F}_t\right) \leq & \Lambda_t + \beta_t + \alpha_t\left(\|\boldsymbol{x}_t - x^*\|_X^2 + \|\boldsymbol{p}_t - p^*\|_P^2\right) + 2\gamma_t\left(L(x^*, \boldsymbol{p}_t) - L(\boldsymbol{x}_t, p^*)\right),\\ \leq & \Lambda_t\left(1 + \alpha_t\right) + \beta_t + 2\gamma_t\left(L(x^*, \boldsymbol{p}_t) - L(\boldsymbol{x}_t, p^*)\right),\end{aligned} \tag{IV.70}$$

avec (α_t) et (β_t) les termes de deux séries sommables définies de la même manière que dans la preuve du théorème IV.10.
En utilisant alors la double inégalité de point-selle pour (x^*, p^*), on obtient avec l'inégalité (IV.70) :

$$\mathbb{E}\left(\Lambda_{t+1}|\mathcal{F}_t\right) \leq \Lambda_t\left(1 + \alpha_t\right) + \beta_t + 2\gamma_t\left(L(x^*, p^*) - L(\boldsymbol{x}_t, p^*)\right) \text{ et,} \tag{IV.71a}$$

$$\mathbb{E}\left(\Lambda_{t+1}|\mathcal{F}_t\right) \leq \Lambda_t\left(1 + \alpha_t\right) + \beta_t + 2\gamma_t\left(L(x^*, \boldsymbol{p}_t) - L(x^*, p^*)\right). \tag{IV.71b}$$

Par définition,

$$L(x^*, \boldsymbol{p}_t) - L(x^*, p^*) \leq 0, \text{ et, } L(x^*, p^*) - L(\boldsymbol{x}_t, p^*) \leq 0.$$

En utilisant maintenant le même argument de quasimartingale que précédemment, on obtient que (Λ_t) est une quasimartingale et converge presque sûrement vers une variable aléatoire intégrable. Elle est donc presque sûrement bornée, et par conséquent, $(\boldsymbol{x}_t)$ and $(\boldsymbol{p}_t)$ le sont également presque sûrement dans X et P respectivement. En utilisant les hypothèses (IV.61)–(IV.60), $(\boldsymbol{s}_t)$ et $(\boldsymbol{r}_t)$ sont aussi presque sûrement bornées.
Enfin, en faisant les mêmes calculs que ceux menant à l'équation (IV.45), on obtient

$$\sum_{t\in\mathbb{N}} \gamma_t\left(L(\boldsymbol{x}_t, p^*) - L(x^*, p^*)\right) < +\infty, \tag{IV.72a}$$

$$\sum_{t\in\mathbb{N}} \gamma_t\left(L(x^*, p^*) - L(x^*, \boldsymbol{p}_t)\right) < +\infty. \tag{IV.72b}$$

Par convexité de $L(\cdot, p^*)$ et concavité de $L(x^*, \cdot)$, on peut à nouveau vérifier les hypothèses du lemme A.35 (calculs menant aux équations (IV.46) ou (IV.48) selon les ensembles admissibles), ce qui donne ici :

$$\lim_{t\to\infty} L(\boldsymbol{x}_t, p^*) = L(x^*, p^*), \text{ et,} \tag{IV.73a}$$

$$\lim_{t\to\infty} L(x^*, \boldsymbol{p}_t) = L(x^*, p^*). \tag{IV.73b}$$

La semi continuité inférieure de $L(\cdot, p^*)$ et supérieure de $L(x^*, \cdot)$ donnent la convergence faible de $(\boldsymbol{x}_t, \boldsymbol{p}_t)$ vers (x^*, p^*).
Enfin, si $L(\cdot, p^*)$ est fortement convexe, on obtient comme en (IV.50) la forte convergence de $(\boldsymbol{x}_t)$ vers x^*. □

IV.4.3. Illustrations.

IV.4.3.1. *Convergence de l'algorithme IV.4.* Afin de revenir au cadre du problème (IV.4), nous posons $X = L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)$, et pour tout $x \in X$, $f(x) = \mathbb{E}\left(j(x(\boldsymbol{\xi}), \boldsymbol{\xi})\right)$, avec $\boldsymbol{\xi}$ une variable aléatoire de loi μ à valeurs dans $\mathbb{R}^m$, et j une intégrande normale convexe (cf. Définition A.23). A supposer que les applications soient toutes différentiables, on a naturellement :

$$\forall x \in X,\ \nabla f(x)(\cdot) = \nabla_x j(x(\cdot), \cdot).$$

Enfin, on prendra X^f un convexe fermé ou un sous-espace fermé de X. Soit $(\boldsymbol{\xi}_t)$ une suite de variables aléatoires i.i.d. de loi μ, et $(K_t : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R})$ une suite d'applications. Enfin, on se

donne deux suites de réels positifs (ρ_t, ϵ_t). L'algorithme IV.4 s'écrit pour son actualisation :

$$\boldsymbol{x}_{t+1} = \Pi_{X^f}\left(\boldsymbol{x}_t - \rho_t \nabla f(\boldsymbol{x}_t)(\boldsymbol{\xi}_{t+1}) K_t(\boldsymbol{\xi}_{t+1}, \cdot)\right). \tag{IV.74}$$

Nous allons maintenant spécifier l'algorithme général IV.9 afin de retomber sur l'algorithme IV.4. On pose alors pour tout $t \in \mathbb{N}$:

- $\mathcal{F}_t := \sigma(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_t)$ la filtration naturelle associée à $(\boldsymbol{\xi}_t)$,
- $\boldsymbol{s}_t := -\nabla f(\boldsymbol{x}_t)$,
- $\boldsymbol{w}_t := \nabla f(\boldsymbol{x}_t) - \nabla_x f(\boldsymbol{x}_t)(\boldsymbol{\xi}_{t+1}) \frac{1}{\epsilon_t} K_t(\boldsymbol{\xi}_{t+1}, \cdot)$.

L'équation IV.74 se réecrit alors :

$$\boldsymbol{x}_{t+1} = \Pi_{X^f}\left(\boldsymbol{x}_t + \gamma_t(\boldsymbol{s}_t + \boldsymbol{w}_t)\right),$$

avec $\gamma_t = \rho_t \epsilon_t$. Les hypothèses du théorème IV.5 permettent de vérifier celles du théorème IV.10, ce qui montre l'utilité du cadre général du gradient perturbé. Les hypothèses sur les bruits IV.36 sont en effet impliquées, avec notre réécriture, par les hypothèses IV.6.

IV.4.3.2. *Convergence de l'estimateur de densité de Wagner-Wolverton.* Nous allons ici montrer comment le théorème IV.10 peut permettre de démontrer la convergence d'algorithmes d'approximation d'une fonction de régression : par exemple, la convergence de l'estimateur de densité récursif introduit par Wagner et Wolverton dans [**94**]. Cet article est la première introduction d'un estimateur récursif de densité, après l'estimateur de Parzen-Rosenblatt (cf. [**87, 69**]). Le but est d'estimer la densité f sur $\mathbb{R}^m$ d'une variable aléatoire, sur la base unique de tirages successifs i.i.d. de cette variable aléatoire. L'algorithme s'écrit :

$$\begin{aligned} &\text{Tirer } \boldsymbol{\xi}_{t+1} \text{ indépendamment de } (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_t) \text{ de densité } f, \\ &\forall x \in \mathbb{R}^m,\ \boldsymbol{f}_{t+1}(x) = \boldsymbol{f}_t(x)\left(1 - \frac{1}{t+1}\right) + \frac{1}{t+1}\frac{1}{h_{t+1}^m} K\left(\frac{x - \boldsymbol{\xi}_{t+1}}{h_{t+1}}\right). \end{aligned} \tag{IV.75}$$

Bien entendu, on peut réecrire la formule IV.75 plus simplement comme suit :

$$\forall x \in \mathbb{R}^m,\ \boldsymbol{f}_{t+1}(x) = \frac{1}{t+1}\sum_{i=1}^{t+1} \frac{1}{h_i^m} K\left(\frac{x - \boldsymbol{\xi}_i}{h_i}\right). \tag{IV.76}$$

On dispose ensuite d'un théorème pour la convergence de l'algorithme (IV.75), énoncé ci-après. Nous allons donner une démonstration alternative de ce théorème utilisant notre théorie du gradient perturbé.

Théorème IV.18 (Wagner et Wolverton, 1969). *Supposons que :*

(1) La suite de réels positifs (h_t) est telle que :

$$1 \geq h_1 \geq \cdots \geq 0,\ \sum_{t\in\mathbb{N}} \frac{h_t}{t} < +\infty,\ \sum_{t\in\mathbb{N}} \frac{1}{t^2 h_t^m} < +\infty. \tag{IV.77}$$

(2) L'application $K : \mathbb{R}^m \to \mathbb{R}$ est une densité de probabilité satisfaisant

$$\forall x \in \mathbb{R}^m,\ K(x) \leq A < +\infty, \tag{IV.78a}$$

$$\int \|x\|_{\mathbb{R}^m} K(x) dx = B < +\infty. \tag{IV.78b}$$

(3) La densité de probabilité f est uniformément lipschitzienne de constante L.

Alors, la suite de fonctions $(\boldsymbol{f}_t)$ générée par l'algorithme (IV.76) *est telle que*

$$\int_{\mathbb{R}^m} \left(\boldsymbol{f}_t(x) - f(x)\right)^2 dx \to 0,\ \text{p.s. } \textit{quand } t \to +\infty.$$

Preuve : Remarquons tout d'abord que comme f est une densité lipschitzienne, f est de carré intégrable. Par simplicité, on notera $\langle \cdot, \cdot \rangle$ et $\| \cdot \|$ le produit scalaire et la norme dans $L^2(\mathbb{R}^m, \mathbb{R}, f)$, qu'on notera dans cette preuve L^2. Regardons le problème d'optimisation fonctionnel suivant :

$$\min_{g \in L^2} J(g) := \frac{1}{2}\|g - f\|^2. \tag{IV.79}$$

Comme f est dans L^2, il est clair que la solution unique de (IV.79) est f. Un algorithme de gradient perturbé sur le problème (IV.79) peut s'écrire :

$$\boldsymbol{g}_{t+1} = \boldsymbol{g}_t + \gamma_t \left(\boldsymbol{s}_t + \boldsymbol{w}_t\right), \tag{IV.80}$$

avec $\boldsymbol{g}_t = \boldsymbol{f}_t(\cdot)$, $\boldsymbol{s}_t = f(\cdot) - \boldsymbol{f}_t(\cdot)$ et $w_t = K_{t+1}(\cdot, \boldsymbol{\xi}_{t+1}) - f(\cdot)$, avec $K_s(x,y) = \frac{1}{h_s^m} K\left(\frac{x-y}{h_s}\right)$. On définit également $\gamma_t = 1/(t+1)$. Finalement, l'algorithme (IV.80) et l'algorithme (IV.75) sont identiques. En utilisant le théorème IV.10, nous allons montrer la convergence de l'algorithme (IV.80) vers la solution du problème (IV.79), ce qui prouvera le théorème.

(i) L'application J est fortement convexe, et f est la solution du problème (IV.79) ;

(ii) En définissant pour tout $t \in \mathbb{N}$, $\mathcal{F}_t$ la tribu engendrée par $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_t$, $\boldsymbol{s}_t$ et $\boldsymbol{f}_t$ sont bien $\mathcal{F}_t$-mesurables ;

(iii) Le gradient de J est donné par $\nabla J(g) = g - f$ pour tout $g \in L^2$. Il est donc lipschitzien ;

(iv) Par définition, $\langle \boldsymbol{s}_t, \boldsymbol{g}_t - f \rangle = -\|\boldsymbol{g}_t - f\|^2$ et $\|\boldsymbol{s}_t\| = \|\nabla J(\boldsymbol{g}_t)\|$; En d'autres termes, $\boldsymbol{s}_t$ est bien une direction de descente ;

(v) Il ne reste plus qu'à vérifier les hypothèses sur les bruits $\boldsymbol{w}_t$.

$$\begin{aligned}
\mathbb{E}\left(\boldsymbol{w}_t|\mathcal{F}_t\right)(x) &= \int_{\mathbb{R}^m} K_{t+1}(x,y) f(y) dy - f(x), \\
&= \int_{\mathbb{R}^m} \frac{1}{h_{t+1}^m} K\left(\frac{x-y}{h_{t+1}}\right) f(y) dy - f(x), \\
&= \int_{\mathbb{R}^m} K(z) f(x - z h_{t+1}) dz - f(x).
\end{aligned}$$

D'où, par le caractère lipschitzien de f,

$$\begin{aligned}
\|\mathbb{E}\left(\boldsymbol{w}_t|\mathcal{F}_t\right)\| =& \sqrt{\int_{\mathbb{R}^m} \|\mathbb{E}\left(\boldsymbol{w}_t|\mathcal{F}_t\right)\|_{\mathbb{R}^p}^2 (x) f(x) dx}, \\
\leq& h_{t+1} L B,
\end{aligned}$$

grâce à l'hypothèse (IV.78b). Cela donne l'hypothèse (IV.36a), avec $\eta_t = h_{t+1}$. Pour l'hypothèse sur la variance (IV.36b), on a :

$$\begin{aligned}
\mathbb{E}\left(\|\boldsymbol{w}_t\|^2|\mathcal{F}_t\right) &= \int_{\mathbb{R}^m \times \mathbb{R}^m} \left(\frac{1}{h_{t+1}^m} K\left(\frac{x-y}{h_{t+1}}\right) - f(x)\right)^2 f(x) f(y) dx dy, \\
&= \int_{\mathbb{R}^m \times \mathbb{R}^m} \frac{1}{h_{t+1}^{2m}} K\left(\frac{x-y}{h_{t+1}}\right)^2 f(x) f(y) dx dy \\
&\quad + \int_{\mathbb{R}^m} f(x)^3 dx - \frac{2}{h_{t+1}^m} \int_{\mathbb{R}^m \times \mathbb{R}^m} K\left(\frac{x-y}{h_{t+1}}\right) f(x)^2 f(y) dx dy, \\
&\leq \int_{\mathbb{R}^m} f(x)^3 dx + \frac{1}{h_{t+1}^m} \int_{\mathbb{R}^m \times \mathbb{R}^m} K(z)^2 f(x) f(x - z h_{t+1}) dx dz, \\
&\leq \int_{\mathbb{R}^m} f(x)^3 dx + \frac{A}{h_{t+1}^m} \int_{\mathbb{R}^m \times \mathbb{R}^m} K(z) f(x) \left(f(x) + L h_{t+1} \|z\|_{\mathbb{R}^m}\right) dx dz
\end{aligned}$$

Dès lors, comme f est une densité de carré intégrable, et grâce aux hypothèses (IV.78b)–(IV.78a), il existe deux constantes $C, C' > 0$ telles que :

$$\mathbb{E}\left(\|\boldsymbol{w}_t\|^2|\mathcal{F}_t\right) \leq C + \frac{C'}{h_{t+1}^m},$$

ce qui donne l'hypothèse (IV.36b) du théorème IV.10 ;

(vi) Enfin, les hypothèses sur les suites (IV.77) impliquent dans notre cas les hypothèses (IV.37).

On peut donc appliquer le théorème IV.10, qui prouve le théorème de Wagner et Wolverton d'une autre manière. □

En faissant les mêmes hypothèses sur le noyau $K(\cdot)$, Révész démontre dans [**74**] un résultat de grande déviation pour l'estimateur de Wolverton-Wagner. Dans [**76**], avec des hypothèses légèrement plus restrictives sur la densité f (du type support compact) et sur le noyau $K(\cdot)$ (de forme créneau), il donne des théorèmes de la limite centrale pour l'erreur en norme L^2 et en norme infinie. Son travail part d'un *découpage* de l'algorithme assez fondamentalement différent de celui

que nous proposons ici : chez lui, l'erreur (notre $\boldsymbol{w}_t$) est toujours un incrément de martingale, mais c'est l'opérateur de descente (notre $\boldsymbol{s}_t$) qui est bruité. De plus, ses résultats asymptotiques de vitesse nécessitent la bornitude des itérés de l'algorithme qui doit donc être présupposée, alors qu'elle est pour nous une conséquence du théorème général de gradient perturbé IV.10. Les travaux [**40, 41**] ont également exploré ce problème, et donné des conditions nécessaires et suffisantes sur la suite (h_t) pour assurer la convergence de l'estimateur de Wagner-Wolverton. Pour un aperçu général de l'approximation non-paramétrique de densité, nous renvoyons à [**43**].

IV.5. Liens avec des idées existantes

Nous allons ici évoquer quelques autres alternatives utilisées pour résoudre les problèmes en boucle fermée, et qui peuvent comporter des analogies avec notre approche : les règles de décision linéaires, et les espaces de noyaux reproduisants. Au passage, nous proposerons dans chacun des deux cas une version chaotique d'algorithmes existants, afin de montrer une fois encore l'utilité du cadre général que nous avons dressé dans ce chapitre.

IV.5.1. Régles de décision linéaires.

IV.5.1.1. *Paramétrage de la commande.* Classiquement, dans le cadre de problèmes en boucle fermée comme le problème (I.1), afin de rendre la résolution possible, on recourt à des règles de décision prédéterminées. Historiquement, de telles démarches remontent notamment à [**61**]. De façon générale, la démarche est la suivante :

(1) Se donner une base de fonctions $\Phi^n = (\phi_1, \dots, \phi_n)$ avec pour tout $i \in \{1, \dots, n\}$, $\phi_i \in L^2(\Xi, \mathbb{R}^p, \mu)$,
(2) Rechercher u solution du problème (I.1) sous la contrainte supplémentaire que $u(\cdot) = \sum_{i=1}^n \alpha_i \phi_i(\cdot)$.

Abusivement, nous noterons Φ^n l'opérateur linéaire de $\mathbb{R}^n$ dans $L^2(\Xi, \mathbb{R}^p, \mu)$ défini par $\Phi^n \alpha = \sum_{i=1}^n \alpha_i \phi_i$ pour tout $\alpha \in \mathbb{R}^n$.

Un tel schéma de résolution transporte par l'opérateur Φ^n le problème de minimisation de l'espace fonctionnel $L^2(\Xi, \mathbb{R}^p, \mu)$ dans l'espace de dimension finie $\mathbb{R}^n$, composé des coefficients des fonctions de base. Finalement, il s'agit de résoudre :

$$\min_{\alpha \in \mathbb{R}^n} \mathbb{E}\left(j(\Phi^n \alpha(\boldsymbol{\xi}), \boldsymbol{\xi})\right), \tag{IV.81a}$$

$$\text{s.c.} \Phi^n \alpha \in U^f. \tag{IV.81b}$$

Bien entendu, le critère du problème (IV.81a) reste une fonction convexe de $\alpha \in \mathbb{R}^n$ de par la linéarité de Φ^n. Cependant, il faut faire attention à la contrainte d'admissibilité qui en toute généralité, selon la forme de l'ensemble U^f peut devenir vide dans $\mathbb{R}^n$. Des études en cours (cf. [**38**], ou [**70**]) travaillent précisément sur des règles linéaires conjointement dans le dual et le primal pour éviter ce genre d'écueils. Sans parler de règles linéaires dans le dual, on peut en effet réécrire le problème (IV.81) directement dans l'espace $L^2(\Xi, \mathbb{R}^p, \mu)$ de la manière équivalente suivante :

$$\min_{u \in L^2(\Xi, \mathbb{R}^p, \mu)} \mathbb{E}\left(j(u(\boldsymbol{\xi}), \boldsymbol{\xi})\right), \tag{IV.82a}$$

$$\text{s.c.} \quad u \in U^f \cap \operatorname{Im}(\Phi^n). \tag{IV.82b}$$

On peut dès lors penser à des théorèmes asymptotiques, en se donnant une famille dénombrable dense de fonctions (ϕ_i) dans $L^2(\Xi, \mathbb{R}^p, \mu)$ (cela existe, on peut par exemple prendre une base hilbertienne), et en définissant pour tout $k \in \mathbb{N}$ l'application linéaire Φ^k construite comme avant avec les k premières fonctions ϕ_i. Il s'agit alors de montrer que $\left(U^f \cap \operatorname{Im}(\Phi^k)\right)$ en tant que suite d'ensembles converge au sens de Painlevé-Kuratowski vers U^f, pour obtenir, à quelques conditions près, la convergence des problèmes approchés (IV.82) vers le problème initial (I.1) (ce type de raisonnement, basé sur le concept d'épi-convergence exposé dans [**84**] est par exemple utilisé dans [**70**]).

Exemple IV.19 (Ensemble admissible et règles linéaires). *Comme dit, le problème en $\alpha \in \mathbb{R}^n$ demeure convexe : si l'ensemble convexe U^f de $L^2(\Xi, \mathbb{R}^p, \mu)$ s'écrit comme l'ensemble des applications u telles que $g(u) \leq 0$, avec $g : L^2(\Xi, \mathbb{R}^p, \mu) \to \mathbb{R}$ convexe, alors l'application $f : \mathbb{R}^n \to \mathbb{R}$ définie par $f(\alpha) = g(\Phi^n \alpha)$ sera elle aussi convexe par linéarité de Φ^n. De même, dans le cas de contraintes de mesurabilité, les contraintes de mesurabilité pourront être automatiquement incorporées dans la formulation en utilisant des fonctions de base ϕ_i ayant la bonne mesurabilité. En revanche, si U^f est l'ensemble des applications u telles que μ-presque partout, $A(\xi)u(\xi) = b(\xi)$, avec $A : \Xi \to \mathbb{R}^{d \times p}$ et $b : \Xi \to \mathbb{R}^d$ deux applications mesurables, le passage aux règles linéaires peut rendre U^f vide, car on dispose alors d'un nombre fini de paramètres (le vecteur $(\alpha_i)_{1 \leq i \leq n}$) pour satisfaire un nombre de contraintes de l'ordre du cardinal de Ξ (et donc typiquement infini).*

Comme le problème (IV.81) est plus contraint que le problème (I.1), on n'a aucune garantie de l'optimalité de la solution trouvée dans la base Φ^n : on sait seulement que l'on aura recherché la meilleure solution dans l'image de Φ^n. Ainsi, dans l'hypothèse où la base choisie est mauvaise pour le problème, on obtienda une solution approchée mauvaise. Au contraire, si Φ^n n'est réduite qu'à un seul élément qui est la solution optimale du problème (I.1), la solution de (IV.81) sera évidemment optimale. C'est pour cette raison qu'un axe intéressant de recherche dans ce domaine est l'enrichissement de la base Φ^n : dans quelle direction enrichir Φ^n pour obtenir une meilleure solution approchée ? Quelle est la garantie d'optimalité, etc. À notre connaissance, seul le résultat asymptotique évoqué plus haut existe à ce propos. Nous allons maintenant comparer l'approche par règles de décisions linéaires avec l'approche du gradient stochastique en boucle fermée.

IV.5.1.2. *Comparaison.* Ce qui est intéressant ici est de constater les ressemblances et dissemblances entre notre approche et les règles de décision linéaires :

- Dans les deux cas, la solution est obtenue comme une somme de fonctions élémentaires (les approximations de la masse de Dirac dans notre cas, la base de fonctions pour les régles linéaires).
- Dans le cas des règles linéaires, la solution obtenue est la meilleure dans l'espace vectoriel engendré par les fonctions de base, tandis que dans notre cas, on n'a aucune garantie d'optimalité à nombre d'itérations fixé, dans la base engendrée par les approximations de la masse de Dirac utilisées.
- Dans le cas des règles linéaires, il reste toujours une erreur résiduelle dûe à l'erreur d'approximation par une base finie, et l'on ne sait dans quelle direction enrichir cette base, tandis que dans notre cas, on peut montrer asymptotiquement la convergence du schéma vers l'optimum : l'enrichissement est automatique.

Pour nous résumer, sous des dehors assez semblables, l'approche de gradient stochastique en boucle fermée développée dans ce mémoire, et l'approche par règles de décision linéaires sont fondamentalement différentes :

- Côté règles de décision linéaires, la question est d'approcher directement la solution du problème, dans un sous-espace vectoriel appelé à croître vers l'espace entier. L'asymptotique se situe donc dans la dimension de l'espace d'approximation ;
- Côté gradient stochastique en boucle fermée, à partir d'un point courant (l'itéré), on approche la distance à la solution optimale dans un espace de dimension 1 dont la seule fonction de base est le noyau courant[2]. L'asymptotique se situe donc dans la succession des points courants.

Ces deux approches sont donc très différentes. Néanmoins, on pourrait imaginer que l'une serve à l'autre. Dans le cas où le choix d'une base de fonctions pour des règles de décision linéaires, n'est pas naturel, on pourrait faire tourner l'algorithme IV.4 jusqu'à un nombre donné d'itérations, et utiliser la solution courante ainsi produite comme élément de la base de fonctions, que l'on pourrait ensuite compléter par une base typiquement polynômiale.

[2]Cette interprétation est du reste la même pour tout algorithme de gradient ou plus généralement de descente : on s'y intéresse moins à la solution qu'à la distance qui nous en sépare depuis le point courant.

IV.5.1.3. *Algorithmes chaotiques pour règles de décision linéaires.* La thèse [**38**] s'intéresse particulièrement au problème (IV.81), principalement de deux points de vue : que peut-on dire des propriétés asymptotiques des contrôles obtenus par régles de décision linéaires lorsque le cardinal de la base de fonction tend vers l'infini ? comment est-il souhaitable de choisir la base primale (eu égard aux contraintes pouvant former un ensemble admissible vide une fois le problème réduit) ? Nous allons ici nous concentrer sur une instance particulière du problème (IV.81), dans lequel U^f est le sous-espace vectoriel des applications $\sigma(h)$-mesurables de $L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)$, avec $h : \mathbb{R}^m \to \mathbb{R}^m$ mesurable. Nous allons dans ce cadre proposer un nouvel algorithme, chaotique, permettant à partir de fonctions de base dont le nombre grandit, d'approcher asymptotiquement la vraie solution du problème.
Nous avons donc défini $U = L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)$ l'espace de Hilbert des commandes, et $J : U \to \mathbb{R}$ une fonction de coût convexe. Soit alors $(\phi_i)_{i\in I}$ une base orthonormale (dénombrable) de U^f qui est un espace de Hilbert muni du même produit scalaire que U. Soit $\Phi : \mathbb{R}^I \to L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)$ l'opérateur linéaire défini par :

$$\forall \alpha \in \mathbb{R}^I,\ \Phi\alpha = \sum_{i\in I} \alpha_i \phi_i,$$

On se pose le problème suivant :

$$\min_{u\in U^f = \mathrm{Im}(\Phi)} J(u). \tag{IV.83}$$

Ce problème peut se réécrire sous la forme équivalente suivante :

$$\min_{(\alpha_i)_{i\in I}} J\left(\sum_{i\in I} \alpha_i \phi_i\right). \tag{IV.84}$$

On pourrait donc écrire l'algorithme de gradient suivant sur α_i, et son analogue sur $u = \Phi\alpha$:

$$\forall i \in I,\ \alpha_i^{k+1} = \alpha_i^k - \rho^k \Phi^* \nabla J(\Phi\alpha^k), \tag{IV.85}$$

$$u^{k+1}(\cdot) = u^k(\cdot) - \rho^k \Phi\Phi^* \nabla J(u^k). \tag{IV.86}$$

De par la définition de l'opérateur Φ, l'opérateur Φ^* à valeurs dans $\mathbb{R}^I$ est défini par :

$$\forall u \in U,\ \Phi^* u = \left(\mathbb{E}\left(\langle \phi_i(\boldsymbol{\xi}), u(\boldsymbol{\xi})\rangle_{\mathbb{R}^p}\right)\right)_{i\in I},$$

et l'équation (IV.86) se réécrit

$$u^{k+1}(\cdot) = u^k(\cdot) - \rho^k \sum_{i\in I} \mathbb{E}\left(\langle \phi_i(\boldsymbol{\xi}), \nabla J(u^k)(\boldsymbol{\xi})\rangle_{\mathbb{R}^p}\right) \phi_i(\cdot). \tag{IV.87}$$

On voit donc dans l'équation de mise-à-jour (IV.87) deux *sommes* apparaître. L'une porte sur les fonctions de base (somme en $i \in I$), et l'autre est l'espérance en $\boldsymbol{\xi}$ selon la loi μ. Soit $\nu = (\nu_i)_{1\le i\le I}$ une mesure de probabilité sur I. En posant alors pour tout $i \in I$, $\psi_i = \frac{1}{\nu_i}\phi_i$, on peut réécrire l'algorithme (IV.87) comme

$$u^{k+1}(\cdot) = u^k(\cdot) - \rho^k \sum_{i\in I} \nu_i \mathbb{E}\left(\langle \phi_i(\boldsymbol{\xi}), \nabla J(u^k)(\boldsymbol{\xi})\rangle_{\mathbb{R}^p}\right) \psi_i(\cdot). \tag{IV.88}$$

De là vient alors l'idée d'un algorithme chaotique (cf. par exemple [**20**]) pour éviter ces calculs d'espérance à chaque itération :

Soit $\boldsymbol{i}^{k+1}$ variable aléatoire distribuée selon ν, indépendante des v.a. passées,

Soit $\boldsymbol{\xi}^{k+1}$ variable aléatoire distribuée selon μ, indépendante des v.a. passées,

$$\boldsymbol{u}^{k+1}(\cdot) = \boldsymbol{u}^k(\cdot) - \rho^k \langle \phi_{\boldsymbol{i}^{k+1}}(\boldsymbol{\xi}^{k+1}), \nabla J(\boldsymbol{u}^k)(\boldsymbol{\xi}^{k+1})\rangle_{\mathbb{R}^p} \psi_{\boldsymbol{i}^{k+1}}(\cdot). \tag{IV.89}$$

On peut alors montrer la proposition suivante :

Proposition IV.20. *(i) Supposons que $J : L^2(\mathbb{R}^m, \mathbb{R}^p, \mu) \to \mathbb{R}$ soit convexe, coercive, semi continue inférieurement et différentiable. Soit $U^f = \{u \in L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)\ :\ u\ \sigma(h) - \text{mesurable}\}$, avec $h : \mathbb{R}^m \to \mathbb{R}^m$ mesurable. Alors, il existe un ensemble S non vide de solutions à* (IV.83)*, et l'on notera J_S la valeur optimale du problème.*

(ii) Supposons de plus que J soit à gradient linéairement borné.
(iii) Supposons que la base orthonormale $(\phi_i)_{i\in I}$ de U^f soit telle que :

$$\exists C>0,\ \forall i\in I,\ \forall \xi\in\mathbb{R}^m,\ \|\phi_i(\xi)\|_{\mathbb{R}^p}\leq C \tag{IV.90}$$

(iv) Supposons que la suite positive (ρ^k) soit telle que

$$\sum_{k\in\mathbb{N}}\rho^k=+\infty,\quad \sum_{k\in\mathbb{N}}(\rho^k)^2<+\infty.$$

(v) Alors la suite $(\boldsymbol{u}^k)$ générée par l'algorithme (IV.89) *est telle que $\lim_{k\to\infty}J(\boldsymbol{u}^k)=J_S$ presque sûrement, et tout point d'accumulation de $(\boldsymbol{u}^k)$ est presque sûrement un élément de S.*
(vi) Si de plus J est fortement convexe, alors $(\boldsymbol{u}^k)$ converge presque sûrement fortement vers l'unique solution de (IV.83).

Preuve : La preuve de cette proposition est une application du théorème IV.10. En posant $\boldsymbol{w}^k=\nabla J(\boldsymbol{u}^k)-\langle\phi_{\boldsymbol{i}^{k+1}}(\boldsymbol{\xi}^{k+1}),\nabla J(\boldsymbol{u}^k)(\boldsymbol{\xi}^{k+1})\rangle_{\mathbb{R}^p}\psi_{\boldsymbol{i}^{k+1}}$, on peut réécrire (IV.89) comme suit :

$$\boldsymbol{u}^{k+1}=\boldsymbol{u}^k+\rho^k\left(-\nabla J(\boldsymbol{u}^k)+\boldsymbol{w}^k\right).$$

Il ne reste qu'à vérifier les hypothèses sur $\boldsymbol{w}^k$. En notant $\mathcal{F}^k=\sigma(\boldsymbol{\xi}^1,\boldsymbol{i}^1,\cdots,\boldsymbol{\xi}^k,\boldsymbol{i}^k)$, on a par indépendance des tirages :

$$\begin{aligned}\mathbb{E}\left(\boldsymbol{w}^k|\mathcal{F}^k\right)=&\nabla J(\boldsymbol{u}^k)-\sum_{i\in I}\nu_i\mathbb{E}\left(\langle\nabla J(\boldsymbol{u}^k)(\boldsymbol{\xi}),\phi_i(\boldsymbol{\xi})\rangle_{\mathbb{R}^p}|\boldsymbol{u}^k\right)\psi_i,\\ =&\nabla J(\boldsymbol{u}^k)-\sum_{i\in I}\mathbb{E}\left(\langle\nabla J(\boldsymbol{u}^k)(\boldsymbol{\xi}),\phi_i(\boldsymbol{\xi})\rangle_{\mathbb{R}^p}|\boldsymbol{u}^k\right)\phi_i=0,\end{aligned}$$

par définition d'une base hilbertienne, du fait que $\nabla J(\boldsymbol{u}^k)$ est par construction $\sigma(h)$-mesurable, ce qui permet de vérifier facilement l'hypothèse (IV.36a). On a également par indépendance des tirages, en utilisant l'inégalité de Cauchy-Schwarz, et l'inégalité scalaire classique $(a+b)^2\leq 2a^2+2b^2$:

$$\begin{aligned}\mathbb{E}\left(\|\boldsymbol{w}^k\|_{L^2}^2|\mathcal{F}^k\right)\leq&2\left(\|\nabla J(\boldsymbol{u}^k)\|_{L^2}^2+\sum_{i\in I}\nu_i\mathbb{E}\left(\|\phi_i(\boldsymbol{\xi})\|_{\mathbb{R}^p}^2\|\nabla J(\boldsymbol{u}^k)\|_{\mathbb{R}^p}^2|\boldsymbol{u}^k\right)\right),\\ \leq&2\|\nabla J(\boldsymbol{u}^k)\|_{L^2}^2\left(1+C^2\right),\end{aligned}$$

ce qui permet de vérifier l'hypothèse (IV.36b). On peut donc appliquer le théorème IV.10 qui achève la preuve. □

Les hypothèses de la proposition IV.20 sont toutes naturelles du point de vue des algorithmes chaotiques, sauf l'hypothèse (IV.90), propre au contexte des règles de décision linéaires. À titre d'illustration, si $\mathbb{R}^p=\mathbb{R}^m=\mathbb{R}$, et si $\mu(d\xi)=\frac{1}{2\pi}1_{[-\pi,\pi]}(\xi)\lambda(d\xi)$, λ étant la mesure de Lebesgue, on peut choisir pour Φ la base de Fourier $(\xi\mapsto\cos(n\xi)/\sqrt{2},\xi\mapsto\sin(n\xi)/\sqrt{2},1)_{n\in\mathbb{N}}$, qui forme une base orthonormale de $L^2([-\pi,\pi],\mathbb{R},\mu)$. Cette base vérifie bien entendu l'hypothèse (IV.90), avec $C=1$.

IV.5.2. Espaces de noyaux reproduisants (RKHS). Une autre approche peut faire penser à notre algorithme, il s'agit des espaces de noyaux reproduisants (Reproducing Kernel Hilbert Spaces, ou RKHS en anglais).

Le but de cette sous-section est de faire le lien entre notre approche par noyaux et la littérature abondante concernant les espaces de noyaux reproduisants et leur utilisation dans les procédures d'apprentissage, éventuellement stochastiques, comme par exemple dans [**92**].

IV.5.2.1. *Espaces de noyaux reproduisants.* De façon générale, les RKHS peuvent être introduits comme suit :

Définition IV.21. *Soit H un espace de Hilbert d'applications de E dans F deux espaces abstraits. H est muni du produit scalaire $\langle\cdot,\cdot\rangle_H$, et de la norme $\|\cdot\|_H=\sqrt{\langle\cdot,\cdot\rangle_H}$. H est dit être un RKHS s'il existe une application $K:E\times E\to\mathbb{R}$ telle que :*
(i) $\forall x\in E,\ K(x,\cdot)\in H$,
(ii) $\forall f\in H,\ \forall x\in E,\ \langle K(x,\cdot),f\rangle_H=f(x)$. Dans ce cas, on appelle K la fonctionnelle d'évaluation du RKHS.

On peut également introduire de tels espaces de manière constructive, et c'est ce que nous allons faire maintenant.
Soient m et p deux entiers. Considérons l'espace $L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)$ des applications u de $\mathbb{R}^m$ dans $\mathbb{R}^p$ telles que :

$$\mathbb{E}\left(\|u(\boldsymbol{\xi})\|^2_{\mathbb{R}^p}\right) < +\infty.$$

C'est un espace de Hilbert. Soit $K : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ un noyau de Mercer, c'est-à-dire une forme continue symétrique, semi-définie positive dans le sens où :

$$\forall l \in \mathbb{N},\ \forall (x_i)_{1\leq i\leq l} \in (\mathbb{R}^m)^l,\ \forall (c_i)_{1\leq i\leq l} \in \mathbb{R}^l,\ \sum_{i,j=1}^{l} c_i c_j K(x_i, x_j) \geq 0.$$

Nous allons maintenant construire le RKHS noté $\mathcal{H}_K$ associé au noyau de Mercer K.
Soit $V_K := \mathrm{Vect}\,(K(t,\cdot)\ :\ t \in \mathbb{R}^m)$. On notera par simplicité $K_t(\cdot) = K(t,\cdot)$. Par continuité et symétrie de l'application K, on peut, avec le théorème de Riesz, définir un semi-produit scalaire sur V_K, noté $\langle\cdot,\cdot\rangle_K$ par

$$\forall x, x' \in \mathbb{R}^m,\ \langle K_x, K_{x'}\rangle_K := K(x, x'),$$

et la semi-norme associée. En remarquant que $V_0 = \{f \in V_K\ :\ \|f\|_K = 0\}$ est un sous-espace, on peut définir sur l'espace quotient V_k/V_0 un produit scalaire et une norme à l'aide de $\langle\cdot,\cdot\rangle_K$. On pose alors $\mathcal{H}_K$ comme étant la complétion de V_K/V_0, ce qui nous fournit l'espace de Hilbert associé au noyau de Mercer K. Par construction, on aura ici que si $K(x,\cdot) \in L^2(\mathbb{R}^m, \mathbb{R}, \mu)$ pour tout $x \in \mathbb{R}^m$, alors $\mathcal{H}_K \subset L^2(\mathbb{R}^m, \mathbb{R}, \mu)$.

À partir de $\mathcal{H}_K$, nous allons construire un espace de Hilbert inclus dans $L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)$, qu'on notera $\mathcal{H}_{K,p}$. Posons

$$\mathcal{H}_{K,p} := \left\{u \in L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)\ :\ \forall 1 \leq i \leq p,\ u_i \in \mathcal{H}_K\right\}.$$

On le munit du produit scalaire :

$$\forall u, v \in \mathcal{H}_{K,p},\ \langle u, v\rangle_{K,p} = \sum_{i=1}^{p} \langle u_i, v_i\rangle_K,$$

qui est bien un produit scalaire de par les propriétés de $\mathcal{H}_K$. On obtient donc

$$\forall u \in \mathcal{H}_{K,p},\ \|u\|_{K,p} := \sqrt{\sum_{i=1}^{p} \|u_i\|^2_K}.$$

On notera dans la suite $C_K := \sup_{x\in\mathbb{R}^m} \sqrt{K(x,x)}$.

On a vu qu'à partir de tout noyau de Mercer $K : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$, on sait construire un espace de Hilbert $\mathcal{H}_{K,p} \subset L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)$. On peut également remarquer que $L^2(\mathbb{R}^m, \mathbb{R}, \mu)$ n'est pas un RKHS. En effet, la fonctionnelle dite d'évaluation dans L^2 serait la masse de Dirac δ, dont on sait qu'elle n'appartient pas à L^2. Dans la perspective d'approcher avec une suite de RKHS l'espace $L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)$ tout entier, il faut considérer des noyaux de Mercer du type $K(x,y) = G(\frac{x-y}{h})$, où G est une application de $\mathbb{R}^m$ dans $\mathbb{R}$ telle que :

$$\begin{aligned}\int_{\mathbb{R}^m} G(x)dx =& 1,\\ \forall x \in \mathbb{R},\ G(x) =& G(-x) \geq 0,\end{aligned}$$

avec un scaling pour tenir compte du fait que l'on n'intègre pas par rapport à la mesure de Lebesgue, mais par rapport à une loi de probabilité μ. En faisant tendre h vers 0, on peut espérer obtenir une succession d'espaces de Hilbert $\mathcal{H}_{K,p,h}$ tendant vers $L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)$. Ce dernier point est encore à étudier, afin de réfléchir au sens de cette convergence d'espace. Typiquement, on souhaiterait montrer que $\mathcal{H}_{K,p,h}$ tend vers $L^2(\mathbb{R}^m, \mathbb{R}^p, \mu)$ au sens de Painlevé-Kuratowski (cf. [**84**], Chapter 4.B).

IV.5.2.2. *Application à l'optimisation.* Soit K un noyau de Mercer comme avant. Considérons maintenant le problème d'optimisation suivant :

$$\min_{u\in\mathcal{H}_{K,p}\cap U^f} J(u) := \int_{\mathbb{R}^m} j(u(\xi),\xi)d\mu(\xi), \tag{IV.91}$$

avec $j : \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}$ une application donnée, et U^f un sous-espace vectoriel fermé de $L^2(\mathbb{R}^m,\mathbb{R}^p,\mu)$. Définissons l'algorithme suivant :

$$\begin{aligned} &\text{Soit } \boldsymbol{\xi}^{k+1} \text{ distribuée selon } \mu, \text{ indépendante des v.a. passées,} \\ &\boldsymbol{u}^{k+1}(\cdot) = \Pi_{U^f}^{K,p}\left(\boldsymbol{u}^k(\cdot) - \rho^k\nabla_u j(\boldsymbol{u}^k(\boldsymbol{\xi}^{k+1}),\boldsymbol{\xi}^{k+1})K(\boldsymbol{\xi}^{k+1},\cdot)\right), \end{aligned} \tag{IV.92}$$

où l'on a supposé que J était différentiable. La projection $\Pi_{U^f}^{K,p}$ est définie par :

$$\forall u\in\mathcal{H}_{K,p},\ \Pi_{U^f}^{K,p}(u) \in \arg\min_{v\in U^f\cap\mathcal{H}_{K,p}} \|u-v\|_{K,p}.$$

On a alors le théorème suivant :

THÉORÈME IV.22. *(i) Supposons que $j(\cdot,\xi)$ est convexe, différentiable, semi continue inférieurement uniformément en $\xi\in\mathbb{R}^m$, et que J est coercive sur $\mathcal{H}_{K,p}\cap U^f$, avec U^f un sous-espace vectoriel fermé de $L^2(\mathbb{R}^m,\mathbb{R}^p,\mu)$. Alors* (IV.91) *admet des solutions, et on notera $S_K\subset\mathcal{H}_{K,p}$ l'ensemble des solutions.*

(ii) Supposons que K est un noyau de Mercer tel que

$$C_K = \sup_{x\in\mathbb{R}^m}\sqrt{K(x,x)} < \infty.$$

(iii) Supposons que la suite (ρ^k) vérifie :

$$\rho^k > 0, \quad \sum_{k\in\mathbb{N}}\rho^k = +\infty, \quad \sum_{k\in\mathbb{N}}(\rho^k)^2 < +\infty. \tag{IV.93}$$

Si de plus, $j(\cdot,\xi)$ est à gradient linéairement borné, i.e., s'il existe $c,d>0$ tels que pour tous u,ξ,

$$\|\nabla_u j(u,\xi)\|_{\mathbb{R}^p} \le c\|u\|_{\mathbb{R}^p} + d, \tag{IV.94}$$

alors pour tout $u_K^*\in S_K$, $(\boldsymbol{u}^k)$ générée par (IV.92) *vérifie :*

$$\lim_{k\to\infty} J(\boldsymbol{u}^k) = J(u_K^*),\ p.s.$$

et tout point d'accumulation de $(\boldsymbol{u}^k)$ dans la topologie faible de $\mathcal{H}_{K,p}$ est presque sûrement élément de S_K.

(iv) Si de plus, $j(\cdot,\xi)$ est fortement convexe de module $B>0$, S_K est réduit à un singleton, et $(\boldsymbol{u}^k)$ converge fortement presque sûrement vers l'unique solution u_K^*.

Preuve : On suit le schéma de preuve introduit dans [**32**], par fonction de Lyapunov. Soit $\Lambda : \mathcal{H}_K \to \mathbb{R}$ définie par :

$$\forall u\in\mathcal{H}_{K,p},\ \Lambda(u) = \frac{1}{2}\|u-u_K^*\|_{K,p}.$$

On note $\Lambda^k = \Lambda(\boldsymbol{u}^k)$ pour tout $k \in \mathbb{N}$. D'où :

$$\begin{aligned}
\Lambda^{k+1} - \Lambda^k =& \frac{1}{2}\|\boldsymbol{u}^{k+1} - \boldsymbol{u}^k\|_{K,p}^2 + \langle \boldsymbol{u}^{k+1} - \boldsymbol{u}^k, \boldsymbol{u}^k - u_K^* \rangle_{K,p} \\
\leq& \frac{(\rho^k)^2}{2} \sum_{i=1}^{p} \nabla_u j(\boldsymbol{u}^k(\boldsymbol{\xi}^{k+1}), \boldsymbol{\xi}^{k+1})_i^2 \langle K(\boldsymbol{\xi}^{k+1}, \cdot), K(\boldsymbol{\xi}^{k+1}, \cdot) \rangle_K \\
& - \rho^k \sum_{i=1}^{p} \nabla_u j(\boldsymbol{u}^k(\boldsymbol{\xi}^{k+1}), \boldsymbol{\xi}^{k+1})_i \langle K(\boldsymbol{\xi}^{k+1}, \cdot), \boldsymbol{u}_i^k - u_{K,i}^* \rangle_K \\
=& \frac{(\rho^k)^2}{2} K(\boldsymbol{\xi}^{k+1}, \boldsymbol{\xi}^{k+1}) \|\nabla_u j(\boldsymbol{u}^k(\boldsymbol{\xi}^{k+1}), \boldsymbol{\xi}^{k+1})\|_{\mathbb{R}^p}^2 \\
& - \rho^k \sum_{i=1}^{p} \nabla_u j(\boldsymbol{u}^k(\boldsymbol{\xi}^{k+1}), \boldsymbol{\xi}^{k+1})_i \left(\boldsymbol{u}_i^k(\boldsymbol{\xi}^{k+1}) - u_{K,i}^*(\boldsymbol{\xi}^{k+1}) \right) \\
\leq& \frac{(C_K \rho^k)^2}{2} \|\nabla_u j(\boldsymbol{u}^k(\boldsymbol{\xi}^{k+1}), \boldsymbol{\xi}^{k+1})\|_{\mathbb{R}^p}^2 \\
& - \rho^k \langle \nabla_u j(\boldsymbol{u}^k(\boldsymbol{\xi}^{k+1}), \boldsymbol{\xi}^{k+1}), \boldsymbol{u}^k(\boldsymbol{\xi}^{k+1}) - u_K^*(\boldsymbol{\xi}^{k+1}) \rangle_{\mathbb{R}^p}
\end{aligned} \tag{IV.95}$$

On a utilisé dans les inégalités précédentes la contraction et la linéarité de la projection sur un sous espace vectoriel, puis la propriété de reproductibilité dans $\mathcal{H}_K$. On note maintenant $\mathcal{F}^k$ la tribu engendrée par $(\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^k)$. On prend l'espérance conditionnelle dans (IV.95), ce qui donne par indépendance entre $\boldsymbol{\xi}^{k+1}$ et $\mathcal{F}^k$:

$$\mathbb{E}\left(\Lambda^{k+1} - \Lambda^k | \mathcal{F}^k\right) \leq \frac{(C_K \rho^k)^2}{2} \|\nabla J(\boldsymbol{u}^k)\|^2 - \rho^k \langle \nabla J(\boldsymbol{u}^k), \boldsymbol{u}^k - u_K^* \rangle \tag{IV.96}$$

En, effet, par définition, $\nabla J(u)(\cdot) = \nabla_u j(u(\cdot), \cdot)$. On utilise maintenant la convexité de J, qui donne :

$$\langle \nabla J(\boldsymbol{u}^k), \boldsymbol{u}^k - u_K^* \rangle \geq \underbrace{J(\boldsymbol{u}^k) - J(u_K^*)}_{\geq 0, \text{ par optimalité}} .$$

De plus, l'hypothèse (IV.94), avec l'inégalité bien connue $(x+y)^2 \leq 2x^2 + 2y^2$ donne l'existence de deux réels $a, b > 0$ tels que :

$$\|\nabla J(\boldsymbol{u}^k)\|^2 \leq a\|\boldsymbol{u}^k - u_K^*\|^2 + b.$$

En revenant à (IV.96), nous obtenons donc :

$$\mathbb{E}\left(\Lambda^{k+1} - \Lambda^k | \mathcal{F}^k\right) \leq \alpha^k \Lambda^k + \beta^k - \rho^k \left(J(\boldsymbol{u}^k) - J(u_K^*)\right), \tag{IV.97}$$

avec $\alpha^k = a(C_K \rho^k)^2$, et $\beta^k = \frac{b(C_K \rho^k)^2}{2}$ les termes de deux séries sommables. L'utilisation d'un théorème de convergence des quasi-martingales donne alors que (Λ^k) est une quasi-martingale qui converge presque sûrement, et donc qu'elle est bornée par un réel $M > 0$. En prenant l'espérance dans (IV.97), on obtient, en notant $\lambda^k = \mathbb{E}\left(\Lambda^k\right)$:

$$\rho^k \mathbb{E}\left(J(\boldsymbol{u}^k) - J(u_K^*)\right) \leq \lambda^k - \lambda^{k+1} + \alpha^k \lambda^k + \beta^k. \tag{IV.98}$$

On somme (IV.98) de 0 à n, puis on fait tendre n vers l'infini, ce qui donne :

$$\sum_{k \in \mathbb{N}} \rho^k \mathbb{E}\left(J(\boldsymbol{u}^k) - J(u_K^*)\right) < +\infty. \tag{IV.99}$$

comme chacun des termes est positif par optimalité dans $\mathcal{H}_{K,p}$, on a même :

$$\sum_{k \in \mathbb{N}} \rho^k \left(J(\boldsymbol{u}^k) - J(u_K^*)\right) < +\infty, \quad \text{p.s.} \tag{IV.100}$$

De plus, il existe un réel $\delta > 0$ tel que :

$$\|\boldsymbol{u}^{k+1} - \boldsymbol{u}^k\|_{K,p} \leq \rho^k \delta,$$

grâce à la bornitude de Λ^k et à l'hypothèse (IV.5). On peut donc appliquer le lemme habituel qui nous donne que :

$$J(\boldsymbol{u}^k) \to J(u^*), \quad \text{quand } k \to \infty.$$

La preuve se termine grâce à la semi-continuité inférieure de J pour obtenir la convergence faible, puis grâce à l'inégalité de forte convexité pour obtenir la convergence forte. □

Nous allons maintenant ébaucher un résultat qui reste encore à prouver. Considérons une suite de noyaux de Mercer K^n tels que $K^n(x,x) \to \infty$ quand $n \to \infty$ pour tout $x \in \mathbb{R}^m$, et tels

que $\mathrm{supp} K^n(x,\cdot) \to \{x\}$ quand $n \to \infty$, pour tout $x \in \mathbb{R}^m$. Plaçons nous sous les hypothèses du point (iv) du théorème IV.22, et notons pour tout $n \in \mathbb{N}$, u_n^* la solution de (IV.91) et (u_n^k) la suite générée par (IV.92) avec le noyau de Mercer K^n. Le théorème IV.22 donne alors :

$$\begin{aligned} \forall n \in \mathbb{N},\ \|u_n^k - u_n^*\|_{K^n,p} &\to 0, \\ |J(u_n^k) - J(u_n^*)| &\to 0, \quad \text{quand } k \to \infty. \end{aligned} \tag{IV.101}$$

D'autre part, en notant u^* la solution du problème :

$$\min_{u \in U^f} J(u) = \int_{\mathbb{R}^m} j(u(\xi),\xi) d\mu(\xi), \tag{IV.102}$$

si l'on obtient la convergence au sens de Painlevé-Kuratowski de $\mathcal{H}_{K^n,p}$ vers $L^2(\mathbb{R}^m,\mathbb{R}^p,\mu)$ quand n tend vers l'infini, et sous des hypothèses adéquates sur J et U^f, on pourrait obtenir un résultat d'épi-convergence (cf. [**84**], Chapter 7.E-7.F) du type :

$$\begin{aligned} \|u_n^* - u^*\| &\to 0, \\ |J(u_n^*) - J(u^*)| &\to 0, \quad \text{quand } n \to \infty. \end{aligned} \tag{IV.103}$$

En rassemblant les équations (IV.101)–(IV.103), on obtiendrait donc :

$$\forall k,n \in \mathbb{N},\ |J(u_n^k) - J(u^*)| \leq \underbrace{|J(u_n^k) - J(u_n^*)|}_{\to 0 \text{ pour tout } n} + \underbrace{|J(u_n^*) - J(u^*)|}_{\to 0} \tag{IV.104}$$

On aimerait enfin diagonaliser le procédé et faire tendre conjointement k et n vers l'infini. On ne sait toutefois pas comment relier pratiquement la convergence en n à celle en k. Néanmoins, nous ne voyons pas à l'heure actuelle comment achever proprement cette démonstration.

IV.5.2.3. *Comparaison avec le résultat existant.* Dans ce chapitre, on s'est intéressé précisément au problème (I.4), et on a donné un théorème de convergence d'un algorithme stochastique basé sur des noyaux dont la fenêtre tend vers 0 (algorithme IV.4), c'est-à-dire des noyaux correspondants à la suite de noyaux de Mercer donnée plus haut. Ce théorème de convergence donne finalement le même résultat que la sous-section précédente. Sa force principale est de relier la convergence des noyaux vers la masse de Dirac avec le pas de descente ρ^k de l'algorithme. En ce sens, le théorème IV.5 pour l'algorithme IV.4 est plus opératoire.

IV.6. Conclusion et perspectives

Nous avons donc proposé dans ce chapitre un nouveau type d'algorithmes stochastiques en dimension infinie, adaptés aux problèmes en boucle fermée, basés sur la combinaison d'idées venant de l'approximation fonctionnelle (approximation par convolution) et de l'approximation stochastique (tirage des noyaux). Après avoir prouvé la convergence de ce type d'algorithmes pour résoudre les problèmes d'optimisation stochastique en boucle fermée, nous avons proposé quelques applications à de tels problèmes.

Dans le souci de généraliser cette idée de base et de fournir des résultats applicables à d'autres problèmes, nous avons proposé et montré la convergence d'algorithmes généraux de gradient perturbé dans un contexte hilbertien, pour des problèmes de minimisation et des problèmes de point-selle. Les hypothèses faites sur la suite de bruits affectant le gradient sont compatibles avec les hypothèses naturelles de l'approximation fonctionnelle et permettent de retrouver le résultat de convergence sur l'algorithme de gradient stochastique en boucle fermée.

Enfin, la comparaison de notre algorithme avec d'autres techniques de résolution de problèmes en boucle fermée, comme les règles de décision linéaires ou les espaces de noyaux reproduisants en montre l'originalité. Plus encore que ses capacités numériques immédiates, nous pensons que la nature variationnelle de l'approche par noyaux que nous avons présentée en fait l'intérêt : dans la perspective des techniques de décomposition par exemple, disposer d'algorithmes variationnels pour les problèmes en boucle fermée ouvre de nouvelles possibilités, comme cela sera exploré dans le chapitre V.

Une autre application de la théorie générale sur le gradient perturbé a été proposée dans [**12**] : il s'agit de résoudre des équations de point fixe fonctionnelles résultant typiquement de l'écriture du principe de programmation dynamique pour les problèmes de contrôle optimal stochastique en temps discret et horizon infini. Là encore, on peut constater la robustesse et la facilité de vérification des hypothèses de convergence proposées dans le cadre du gradient perturbé, puisqu'elles permettent de montrer la convergence d'algorithmes généraux de différences temporelles avec espace d'état non-dénombrable (possiblement continu). Une occurence particulière de ces problèmes de point fixe apparaît en mathématiques financières lors de la valorisation d'options américaines via leur discrétisation bermudéenne (voir par exemple [**53**] pour la valorisation d'options américaines). Dans ce contexte, nous travaillons à l'implémentation pratique d'algorithmes du type IV.4, et avons obtenu des résultats numériques très satisfaisants en termes de prix d'option comme en terme de couverture.

Ainsi, outre pour la décomposition des grands problèmes stochastiques, l'application des idées du gradient stochastique en boucle fermée et du gradient perturbé dans un espace de Hilbert est possible pour les problèmes de programmation dynamique, ce qui montre l'intérêt d'approches variationnelles en dimension infinie, comme alternative aux discrétisations de tous ordres.

CHAPITRE V

Décomposition des grands systèmes stochastiques

V.1. Résumé

Un problème récurrent en optimisation est la résolution de grands systèmes. Des contraintes numériques de toutes sortes incitent à décomposer de tels problèmes en une succession de problèmes auxiliaires coordonnés plus faciles à résoudre soit en raison de leur petite taille, soit en raison de leur forme particulière, etc. Dans le cadre de problèmes déterministes, la décomposition est bien connue, et, après avoir été longtemps considérée comme une branche *à part*, bi-niveaux, de l'optimisation usuelle à un seul niveau, a été réconciliée avec les approches variationnelles classiques par notamment les travaux de [**28**].
Le cadre stochastique a lui aussi été exploré, et a permis de soulever un certain nombre de questions dont un bon nombre restent en suspens. La section V.2 de ce chapitre propose une revue des techniques utilisées et des possibilités ouvertes dans le cas de problèmes stochastiques. En particulier, nous y insistons sur le caractère très original et difficile des problèmes en boucle fermée. À l'heure actuelle, les seules classes de méthodes connues pour décomposer ce type de problèmes consistent à discrétiser l'aléa sous la forme d'un arbre de scénarios, puis à utiliser des techniques déterministes sur le problème ainsi discrétisé. À notre connaissance, aucune technique ne travaille en pratique directement sur le problème initial dès que l'aléa est un continuum. Lorsque les problèmes sont de petite taille, c'est précisément ce à quoi s'emploie la programmation dynamique stochastique qui ne nécessite pas de discrétisation de l'aléa particulière pour être mise en œuvre. Dans la section V.3, nous étudions les liens de la programmation dynamique stochastique avec la décomposition. Cette étude est motivée par la résolution de problèmes à critère additif, couplés par une contrainte linéaire. En particulier, nous montrons dans le cas de problèmes linéaires quadratiques que le mutiplicateur de Lagrange optimal associé à la contrainte couplante linéaire suit une dynamique optimale, et peut ainsi rentrer dans l'état du problème. En un mot, d'un problème stochastique de programmation dynamique en dimension n, on peut passer à n problèmes de programmation dynamique stochastique en dimension 3 ou 4 selon les caractéristiques du problème initial. Ce cas particulier met en lumière les limitations inhérentes à la décomposition stochastique : on ne peut pas dans ce cadre décomposer en toute généralité puis utiliser indifféremment telle ou telle technique de résolution sur les sous-problèmes, sans faire d'autres hypothèses sur les contrôles admissibles ou la forme des contraintes et du critère. Pour illustrer cela, nous montrons ensuite comment, en se restreignant à des contrôles décentralisés, on peut proposer un schéma de décomposition compatible avec des programmations dynamiques stochastiques locales. Néanmoins, la voie de la programmation dynamique stochastique pour la décomposition ne nous semble pas suffisamment générique pour y consacrer plus de recherches. Dans la section V.4, nous proposons, forts de ce constat, un outil variationnel, dans la lignée de [**28**], pour décomposer un problème stochastique avant toute discrétisation de l'aléa pouvant être continu. L'idée est la suivante : un problème en boucle fermée se caractérise par des contrôles en dimension infinie. Ainsi, les outils de coordination de toute décomposition seront-ils par nature de dimension infinie, et donc inutilisables en pratique. Nous démontrons donc un principe du problème auxiliaire stochastique dans lequel les outils de coordination peuvent être remplacés par des estimations, fussent-elles biaisées. Ce principe du problème auxiliaire donne un cadre et une portée plus générale aux algorithmes stochastique présentés dans le chapitre IV, et procure un outil adapté à l'étude des problèmes en commandes décentralisées.

V.2. État de l'art, ou ce qui se fait, et ce qui pose problème

V.2.1. Pour se fixer les idées. Lorsque l'on parle de problèmes d'optimisation stochastique à plusieurs pas de temps, comme par exemple le problème (I.8), et que l'on souhaite recourir à une décomposition du problème, plusieurs types de décomposition peuvent être envisagés. Assez schématiquement, on peut souhaiter opérer une décomposition spatiale ou une décomposition temporelle. Pour se fixer les idées et expliciter notre pensée, considérons le problème suivant :

(V.1a)
$$\min_{\boldsymbol{u},\boldsymbol{x}} \mathbb{E}\left(\sum_{t=1}^{T-1}\sum_{i=1}^{n} L_{t,i}(\boldsymbol{u}_{i,t},\boldsymbol{x}_{i,t},\boldsymbol{\xi}^t)\right),$$

(V.1b)
$$\text{s.c.}\boldsymbol{u}=(\boldsymbol{u}_{i,t}),\ \forall 1\leq i\leq n,\ 1\leq t\leq T-1,\ \boldsymbol{u}_{i,t}\in L^2(\Omega,U_{i,t},\mathbb{P}) \text{ et est } \sigma(h_t(\boldsymbol{\xi}^t))-\text{mesurable},$$

(V.1c)
$$\forall 1\leq i\leq n,\ 1\leq t\leq T-1,\ \boldsymbol{x}_{i,t+1}=f_{i,t}(\boldsymbol{x}_{i,t},\boldsymbol{u}_{i,t},\boldsymbol{\xi}_{t+1}),\ \text{p.s.}$$

(V.1d)
$$\forall 1\leq t\leq T-1, \Theta_t(\boldsymbol{u}_t,\boldsymbol{\xi}_t)=0,\ \text{p.s.}$$

C'est l'archétype du problème en boucle fermée. Dans ce problème, $\boldsymbol{\xi}=(\boldsymbol{\xi}_1,\dots,\boldsymbol{\xi}_T)$ est le bruit, supposé blanc, $\boldsymbol{\xi}^t=(\boldsymbol{\xi}_1,\dots,\boldsymbol{\xi}_t)$ est le passé du bruit à l'instant t, et $\boldsymbol{u}$ et $\boldsymbol{x}$ sont donc des variables aléatoires soumises à des contraintes de mesurabilité par rapport au bruit. De manière équivalente, on pourrait rechercher pour tout $t\leq T$, $\boldsymbol{x}_t$ et $\boldsymbol{u}_t$ non comme des variables aléatoires, mais comme des applications de l'espace d'arrivée de $\boldsymbol{\xi}^t$ dans leurs espaces d'arrivée respectifs (cf. Remarque I.2).

Si l'on excepte un instant l'aléa, il y a dans ce problème deux dimensions qui en font un problème de grande taille : le temps $1\leq t\leq T$, et l'espace $1\leq i\leq n$. On a donc a priori un choix de décomposition à faire. La contrainte (V.1c) est couplante dans le temps, mais découplée en espace, tandis que la contrainte (V.1d) est couplante en espace mais découplée en temps. On fera également l'hypothèse que Θ_t est affine en u afin de considérer une contrainte convexe, et $\Theta_t(u,\xi_t)=\sum_{i=1}^n \Theta_{i,t}(u_{i,t},\xi_t)$. Enfin, le critère est quant à lui additif (et donc découplé).

Afin de rendre compte au mieux des possibilités connues pour décomposer le problème (V.1), et de mettre en lumière les difficultés liées à cette tâche, nous allons considérer différents cas de contraintes de mesurabilité, ou, autrement, dit, commencer par le cas où pour tout t, $h_t(\cdot)=0$, i.e. le cas où on ne recherche pas des variables aléatoires $\boldsymbol{u}$, mais uniquement des vecteurs réels $(u_{i,t})_{1\leq i\leq n,\,1\leq t\leq T}$. Ensuite, nous parlerons du cas où pour tout t, $h_t=Id$, qui correspond aux problèmes en boucle fermée avec information complète.

V.2.2. Problèmes stochastiques en boucle ouverte. On s'intéresse ici au problème suivant :

(V.2a)
$$\min_{u,x} \mathbb{E}\left(\sum_{t=1}^{T-1}\sum_{i=1}^{n} L_{t,i}(u_{i,t},x_{i,t},\boldsymbol{\xi}^t)\right),$$

(V.2b)
$$\text{s.c.}\forall 1\leq i\leq n,\ 1\leq t\leq T-1,\ x_{i,t+1}=f_{i,t}(x_{i,t},u_{i,t}),$$

(V.2c)
$$\forall 1\leq t\leq T-1, \sum_{i=1}^{n}\Theta_{i,t}(u_{i,t})=0.$$

Dans ce problème, l'aspect stochastique n'intervient que dans la fonction de coût. En particulier, les commandes optimales seront des réels, et les contraintes couplantes en temps (V.2b), et en espace (V.2c) sont purement déterministes, ce qui rend les éventuels multiplicateurs de Lagrange associés également déterministes. Dans ce cas, sous des hypothèses de qualification, on peut recourir aux usuelles techniques de décomposition, consistant principalement en des relaxations lagrangiennes du problème (V.2), ou en une décomposition par les ressources de la contrainte

couplante en espace. Typiquement, si l'on dualise la contrainte (V.2c) à l'aide du multiplicateur de Lagrange $\lambda = (\lambda_t)_{1\leq t\leq T-1}$, on arrive à la résolution du problème dual qui se sépare en n sous-problèmes du type

$$\begin{aligned} &\min_{u_i,x_i} \mathbb{E}\left(\sum_{t=1}^{T-1} L_{t,i}(u_{i,t},x_{i,t},\boldsymbol{\xi}^t) + \lambda_t \Theta_{i,t}(u_{i,t})\right), \\ &\text{s.c.} \forall 1\leq t\leq T-1,\ x_{i,t+1} = f_{i,t}(x_{i,t},u_{i,t}). \end{aligned} \tag{V.3}$$

Les sous-problèmes (V.3) pour $1 \leq i \leq n$ peuvent alors être résolus à λ fixé en utilisant la technique souhaitée (programmation dynamique, etc.), tandis que l'on itérera sur les λ jusqu'à obtenir le point-selle du Lagrangien.

D'autres décompositions peuvent également être envisagées, décomposition par les quantités ou par prédiction (cf. [**31**], ou [**17**], Chapitre 6 pour un aperçu approfondi des techniques usuelles), et un choix est fait entre ces méthodes pour diverses raisons numériques liées aux caractéristiques du problème à résoudre. Pour résumer, la boucle ouverte amène aux conclusions suivantes. On aboutit à un problème à deux niveaux : le niveau *primal* des sous-problèmes obtenus par décomposition, et le niveau *dual* de la mise-à-jour des multiplicateurs de Lagrange (ou autres outils de coordination). Bien entendu, les sous-problèmes du niveau primal sont alors de même nature que le problème initial, sans contrainte couplante avec les autres sous-problèmes, et de taille réduite (en nombre de variables et éventuellement de contraintes), et l'intérêt de la décomposition est dès lors évident, pour peu que le niveau dual de coordination, ne soit pas trop coûteux.

Tous ces algorithmes sont bien connus, et peuvent être associés, dans le cadre d'une fonction de coût ou d'une contrainte couplante s'écrivant comme une espérance, avec des algorithmes du type du gradient stochastique. Ce type de mariages à fait l'objet de la thèse [**36**] sur la décomposition des grands systèmes dans le cadre stochastique. On retrouve les résultats de cette thèse également dans [**32**], et nous en avons donné un échantillon dans l'annexe C à travers l'étude du problème (C.8).

V.2.3. Problèmes stochastiques en boucle fermée. Passons maintenant au cas de contraintes de mesurabilité effectives. On ne peut y transposer le raisonnement déterministe effectué ci-dessus. Les deux contraintes couplantes font ici intervenir des variables aléatoires et sont de plus des contraintes presque sûres. En conséquence, les multiplicateurs de Lagrange associés à ces contraintes sont eux aussi des variables aléatoires. Ces nouvelles variables aléatoires intervenant dans le problème dual décomposé entraînent une nouvelle difficulté : quelle est leur loi, leur corrélation dans le temps, entre elles, etc. ? En un mot, les variables aléatoires entraînent un autre couplage que le simple couplage *déterministe* ou fonctionnel décrit par les équations (V.1c)–(V.1d) et dans lequel on aurait oublié les $\mathbb{P}$-presque sûrement : un couplage probabiliste.

V.2.3.1. *Particularité de la boucle fermée.* Afin de mieux comprendre les enjeux et difficultés associés au cas stochastique, nous allons faire un essai intuitif de décomposition en utilisant la décomposition par les prix. Supposons que l'on introduise des multiplicateurs $(\boldsymbol{\lambda}_t)$ associés à la contrainte (V.1d) (ces multiplicateurs correspondent à une contrainte presque sûre et sont donc eux-mêmes des variables aléatoires ; plus précisément, si le contrôle est cherché dans un espace L^2, les multiplicateurs seront également dans un espace L^2, associé à l'espace d'arrivée de la fonction de contrainte Θ). Le Lagrangien associé au problème (V.1) s'écrit alors comme une somme sur les pas de temps t et les unités i. Sous des hypothèses d'existence de point-selle

et de convexité, on pourra donc s'intéresser à des sous-problèmes du type :

$$\min_{\boldsymbol{u}_i,\boldsymbol{x}_i} \mathbb{E}\left(\sum_{t=1}^{T-1} L_{t,i}(\boldsymbol{u}_{i,t},\boldsymbol{x}_{i,t},\boldsymbol{\xi}^t) + \boldsymbol{\lambda}_t \Theta_{i,t}(\boldsymbol{u}_{i,t},\boldsymbol{\xi}^t)\right), \tag{V.4a}$$

$$\text{s.c.} \boldsymbol{u} = (\boldsymbol{u}_{i,t}),\ 1 \leq t \leq T-1,\ \boldsymbol{u}_{i,t} \in L^2(\Omega, U_{i,t}, \mathbb{P}) \text{ et est } \sigma(\boldsymbol{\xi}^t) - \text{mesurable}, \tag{V.4b}$$

$$\forall 1 \leq t \leq T-1,\ \boldsymbol{x}_{i,t+1} = f_{i,t}(\boldsymbol{x}_{i,t},\boldsymbol{u}_{i,t},\boldsymbol{\xi}_{t+1}), \tag{V.4c}$$

(V.4d)

pour tout $i \in \{1,\dots,n\}$ et à $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_t)$ fixé. A première vue, le problème (V.4) est soluble par programmation dynamique stochastique. Cependant, les variables aléatoires $(\boldsymbol{\lambda}_t)$ peuvent être corrélées dans le temps, etc. ce qui rend en fait une programmation dynamique stochastique impossible sans du moins une augmentation de l'état. De plus, les itérations de mise à jour sur ces variables duales pourraient leur faire perdre toute propriété d'indépendance, ou autres. On peut donc faire ici quelques remarques importantes :

- L'aspect stochastique, dans les problèmes de grande taille, met en jeu des contraintes nouvelles en terme d'indépendance, de mesurabilité, qui excluent a priori certaines méthodes de résolution pour les sous-problèmes après décomposition.
- Les algorithmes de mise à jour des variables primales et duales doivent donner des itérés restant des variables aléatoires, et doivent donc soit être fonctionnels, soit être accompagnés d'une étape d'extension des mise-à-jour discrètes à l'espace fonctionnel.

Ces deux remarques sont tout à fait propres au cas stochastique ; aucune considération de ce genre n'intervient dans le cas déterministe dans lequel on peut, grâce notamment à l'utilisation du principe du problème auxiliaire, totalement déconnecter la décomposition-coordination, des résolutions de sous-problèmes.

V.2.3.2. *L'utilisation d'arbres d'aléas.* Afin de contourner les difficultés précitées, il est d'usage de commencer par discrétiser le problème (V.1) et de remplacer les $\mathbb{P}$-presque sûrement par un nombre fini de contraintes sur un ensemble fini de scénarios. Ce courant de méthodes est particulièrement exhaustivement exposé dans le recueil [**58**]. Le chapitre 3 de [**91**] s'occupe précisément de recenser les différentes méthodes de décomposition possibles lorsque l'on a remplacé les contraintes presque sûres de (V.1) en un nombre fini de contraintes correspondant à chaque scénario (ou chaque nœud de l'arbre). Néanmoins, cette classe de méthode doit commencer par expliquer comment la contrainte de mesurabilité est discrétisée, et cette question est loin d'être triviale, comme l'a montré le chapitre III. Cela dit, une fois la structure arborescente construite, nombre de méthodes de décomposition efficaces peuvent être mises en place pour résoudre ce problème. Parmi ces méthodes, nous pouvons citer par exemple la décomposition par scénarios de [**83**], ou la méthode dite de *dynamic splitting* de [**88**], ou le chapitre 3, dû à Wets, de [**50**]. Nous ne rentrerons pas ici dans le détail de ces méthodes, préférant concentrer notre étude sur le cas d'aléas se présentant comme un continuum et non comme des scénarios, afin de bien mettre en évidence le rôle joué pour la décomposition par les contraintes de non-anticipativité.

V.2.3.3. *Arsenal théorique, ou que faire dans un Hilbert ?* A côté des remarques que nous venons de faire, fondées sur les concepts basiques de décomposition par les prix pour les problèmes déterministes, on peut s'interroger sur la réutilisation dans le cadre stochastique des pincipes généraux de la décomposition, notamment le principe du problème auxiliaire (cf. [**30**]). Ce principe d'optimisation consiste à remplacer le problème initial par la résolution d'un problème de point fixe passant par la résolution itérative de problèmes auxiliaires plus simples et mieux conditionnés que le problème initial (cf. Appendice A).

Cette théorie existant dans un contexte hilbertien, il suffit a priori de demander aux variables de commande du problème (V.1) d'être de carré intégrable, et de se plonger dans l'espace de

Hilbert associé. En effet, réécrivons le problème (V.1) comme :

$$\min_{\boldsymbol{u}\in\mathcal{U}} \sum_{t=1}^{T}\sum_{i=1}^{n} J_{i,t}(\boldsymbol{u}_{i,t}) \tag{V.5a}$$

$$\text{s.c.} \forall 1 \leq i \leq n,\ 1 \leq t \leq T,\ \boldsymbol{u}_{i,t} \in L^2(\Omega, U_{i,t}, \mathbb{P}),\ \text{et est } \sigma(\boldsymbol{\xi}^t) - \text{mesurable}, \tag{V.5b}$$

$$\forall 1 \leq i \leq n,\ F_i(\boldsymbol{u}_i) = 0, \tag{V.5c}$$

$$\forall 1 \leq t \leq T,\ \Theta_t(\boldsymbol{u}_t, \boldsymbol{\xi}_t) = 0, \tag{V.5d}$$

Le problème (V.5) et les contraintes qui lui sont imposées correspondent donc exactement au problème (V.1) et à ses contraintes. La nouvelle formulation n'est là que pour faire apparaître le cadre essentiel dans lequel le principe du problème auxiliaire s'énonce, c'est à dire le cadre hilbertien. La différence principale entre les deux formulations est la mise en évidence dans la première de l'état $\boldsymbol{x}$, variable auxiliaire familière et indispensable à la mise en œuvre de la programmation dynamique stochastique (cf. [**18**] pour un exposé complet et général sur la programmation dynamique stochastique en temps discret). Le principe du problème auxiliaire peut ensuite exactement s'appliquer au problème de point-selle associé à la dualisation de l'une des contraintes (V.5c) ou (V.5d), sous des hypothèses suffisantes de convexité et d'existence du point-selle.

Une telle abstraction fait cependant oublier par trop de distance les troubles évoqués dans la sous-section précédente, et nous ne savons donc qu'en dire pour le moment. En particulier, il est difficile de donner ici une interprétation du *couplage probabiliste* dont nous avons parlé avant.
Dans la suite de ce chapitre, nous allons tout d'abord travailler autour de la programmation dynamique afin de montrer ce qu'il est possible de faire dans cette classe de méthodes, pour un problème linéaire quadratique, puis pour un problème en commande décentralisée. Cette étude préliminaire a pour but de montrer que si la programmation dynamique est une méthode performante en faible dimension, elle ne peut en toute généralité être associée à la décomposition, excepté dans une certain nombre de cas très particuliers. Dans un second temps, nous allons donner un cadre théorique pour la décomposition des problèmes stochastiques, à l'image du principe du problème auxiliaire.

V.3. Autour de la programmation dynamique

Cette section se fixe comme objectif de comprendre les liens entre la décomposition et la programmation dynamique stochastique. En particulier, si le cas général est désespéré comme l'a indiqué la section introductive, certaines adaptations peuvent être envisagées lorsque le problème à résoudre est mieux spécifié. C'est le cas d'un problème linéaire quadratique et d'un problème en information décentralisée.

V.3.1. Dualité et cas linéaire quadratique.

V.3.1.1. *Description du problème.* Nous allons ici donner une instance particulière du problème (V.1) connue sous le nom de problème linéaire quadratique, et qui correspond à peu de choses près à un problème simplifié de gestion de système de production électrique. On se donne les variables aléatoires suivantes :

- la demande en électricité, notée $\boldsymbol{d} = (\boldsymbol{d}_t)_{1\leq t\leq T}$, à valeurs dans $\mathbb{R}^T$,
- les apports hydrauliques de chaque unité de production $i \in \{1, \dots, n\}$ notés $\boldsymbol{a}_i = (\boldsymbol{a}_{i,t})_{1\leq t\leq T}$ à valeurs dans $\mathbb{R}^T$.

On appellera ces variables aléatoires les bruits du système. Le problème est alors de choisir les productions $\boldsymbol{u}_{i,t}$ de chaque unité i à chaque temps t comme des variables aléatoires dépendant des bruits du système, de manière non-anticipative, c'est à dire comme des variables aléatoires mesurables par rapport au passé des bruits. Afin de définir ce passé des bruits proprement, on pose $\mathcal{B}_{i,t} := \sigma(\boldsymbol{d}_s,\ s \leq t; \boldsymbol{a}_{i,s},\ s \leq t)$ la tribu représentant l'information décentralisée du sous-système i. On pose également $\mathcal{B}_t = \vee_{1\leq i\leq n}\mathcal{B}_{i,t}$ l'information complète au temps t, disponible pour choisir $\boldsymbol{u}_{i,t}$. En fait, $(\mathcal{B}_{i,t})_{1\leq t\leq T}$ est la filtration naturelle du processus $(\boldsymbol{d}, \boldsymbol{a}_i)$, et $(\mathcal{B}_t)_{1\leq t\leq T}$ est la filtration naturelle du processus $(\boldsymbol{d}, \boldsymbol{a}_1, \dots, \boldsymbol{a}_n)$. Les contraintes de non-anticipativité s'écriront

simplement : $\boldsymbol{u}_{i,t}$ $\mathcal{B}_t$-mesurable pour tous i, t.
Afin d'achever la description du problème, on introduit des coefficients de coût $(c_i)_{1\leq i\leq n} \in \mathbb{R}^n_+$, et des coefficients de pénalité $(\gamma_i)_{1\leq i\leq n} \in \mathbb{R}^n_{*+}$. Enfin, pour obtenir une formulation du type (V.1), on associe à chaque unité de production i un processus aléatoire d'états noté $\boldsymbol{x}_i = (\boldsymbol{x}_{i,t})_{1\leq t\leq T}$ et défini par une dynamique dépendant des variables de décision $\boldsymbol{u}_i$ et des bruits.
Dans ces conditions, le problème d'optimisation s'écrit :

$$\min_{\boldsymbol{u},\boldsymbol{x}} \sum_{t=1}^{T-1}\sum_{i=1}^{n} \mathbb{E}\left(c_i\frac{\boldsymbol{u}_{i,t}^2}{2}\right) + \sum_{i=1}^{n}\mathbb{E}\left(\frac{\gamma_i}{2}(\boldsymbol{x}_{i,T}-x_{i,1})^2\right), \tag{V.6a}$$

$$\text{s.c.} \forall t\in\{1,\dots,T-1\},\ \forall i\in\{1,\dots,n\},\ \boldsymbol{u}_{i,t}\in L^2(\Omega,\mathbb{R},\mathbb{P}),\ \text{et est } \mathcal{B}_t-\text{mesurable}, \tag{V.6b}$$

$$\forall t\in\{1,\dots,T-1\},\ \sum_{i=1}^{n}\boldsymbol{u}_{i,t} = \boldsymbol{d_t},\ \text{p.s.}, \tag{V.6c}$$

$$\forall t\in\{1,\dots,T-1\},\ \forall i\in\{1,\dots,n\},\ \boldsymbol{x}_{i,t+1} = \boldsymbol{x}_{i,t} + \boldsymbol{a}_{i,t+1} - \boldsymbol{u}_{i,t},\ \text{p.s.} \tag{V.6d}$$

On pourrait souhaiter résoudre directement par programmation dynamique stochastique le problème (V.6), avec des hypothèses naturelles de blancheur des bruits, mais alors, la résolution devrait être faite en dimension n, puisqu'il y a autant d'états que d'unités de production. Dès que $n \geq 4$, on serait donc bloqué.

Pour nous résumer, le but est de décomposer le problème (V.6) par unité de production. La seule contrainte couplante est la contrainte dite de satisfaction de demande (V.6c). Les autres contraintes sont naturellement découplées par unité de production. Enfin, les états initiaux des unités $(x_{i,1})_{1\leq i\leq n} \in \mathbb{R}^n$, sont supposés connus et donnés. Bien entendu, le problème (V.6) est un problème convexe. Nous allons dans la sous-section suivante utiliser les outils de la dualité lagrangienne pour caractériser l'optimum de ce problème.

V.3.1.2. *Décomposition par les prix.* Nous allons décomposer le problème (V.6) en utilisant la méthode de décomposition par les prix. On introduit les multiplicateurs $\boldsymbol{\lambda}_t \in L^2(\Omega,\mathbb{R},\mathbb{P})$ pour tout $t\in\{1,\dots,T-1\}$, associés à la contrainte couplante (V.6c), et on forme le Lagrangien associé au problème (V.6) (duquel on a retiré le terme indépendant des sous-problèmes) :

$$L(\boldsymbol{u},\boldsymbol{x},\boldsymbol{\lambda}) = \sum_{t=1}^{T-1}\mathbb{E}\left(\sum_{i=1}^{n} c_i\frac{\boldsymbol{u}_{i,t}^2}{2} - \boldsymbol{\lambda}_t\sum_{i=1}^{n}\boldsymbol{u}_{i,t}\right) + \sum_{i=1}^{n}\mathbb{E}\left(\frac{\gamma}{2}(\boldsymbol{x}_{i,T}-x_{i,1})^2\right). \tag{V.7}$$

Le Lagrangien a donc naturellement une structure décomposée, c'est à dire que l'on peut écrire $L(\boldsymbol{u},\boldsymbol{x},\boldsymbol{\lambda}) = \sum_{i=1}^{n} L_i(\boldsymbol{u}_i,\boldsymbol{x}_i,\boldsymbol{\lambda})$, avec pour tout $i\in\{1,\dots,n\}$,

$$L_i(\boldsymbol{u}_i,\boldsymbol{x}_i,\boldsymbol{\lambda}) = \mathbb{E}\left(\sum_{t=1}^{T-1} c_i\frac{\boldsymbol{u}_{i,t}^2}{2} - \boldsymbol{\lambda}_t\boldsymbol{u}_{i,t}\right) + \frac{\gamma}{2}\mathbb{E}\left((\boldsymbol{x}_{i,T}-x_{i,1})^2\right). \tag{V.8}$$

En faisant les hypothèses d'existence d'un point selle au Lagrangien (V.7), on peut donc considérer les n sous-problèmes apparaissant dans le problème dual, pour $i\in\{1,\dots,n\}$:

$$\begin{aligned}\psi_i(\boldsymbol{\lambda}) = \min_{\boldsymbol{u}_i,\boldsymbol{x}_i} & L_i(\boldsymbol{u}_i,\boldsymbol{x}_i,\boldsymbol{\lambda}),\\ \text{s.c.}\forall t\in\{1,\dots,T-1\},\ & \boldsymbol{x}_{i,t+1} = \boldsymbol{x}_{i,t} + \boldsymbol{a}_{i,t+1} - \boldsymbol{u}_{i,t},\ \text{p.s.},\\ \forall t\in\{1,\dots,T-1\}, & \boldsymbol{u}_{i,t}\in L^2(\Omega,\mathbb{R},\mathbb{P}),\ \text{et est } \mathcal{B}_t-\text{mesurable}.\end{aligned} \tag{V.9}$$

On doit ensuite résoudre le problème dual, qui s'écrit :

$$\begin{aligned}\max_{\boldsymbol{\lambda}} & \mathbb{E}\left(\sum_{t=1}^{T-1}\boldsymbol{\lambda}_t\boldsymbol{d}_t\right) + \sum_{i=1}^{n}\psi_i(\boldsymbol{\lambda}),\\ \text{s.c.}\forall t\in\{1,\dots,T-1\},\ & \boldsymbol{\lambda}_t\in L^2(\Omega,\mathbb{R},\mathbb{P}).\end{aligned}$$

On obtient alors une solution $(\boldsymbol{u}^*,\boldsymbol{x}^*,\boldsymbol{\lambda}^*)$ qui forme un point selle au Lagrangien, c'est à dire telle que $(\boldsymbol{u}^*,\boldsymbol{x}^*)$ est solution du problème (V.6). À ce stade, il est clair que l'on peut restreindre les

multiplicateurs $\boldsymbol{\lambda}$ à être adaptés à la filtration $(\mathcal{B}_t)$. En effet, remplacer partout dans le problème (V.9) et dans le problème dual, (λ_t) par $(\mathbb{E}(\boldsymbol{\lambda}_t|\mathcal{B}_t))$ ne change rien aux solutions. L'algorithme usuel de décomposition par les prix consiste à alterner les pas de descente dans le primal et dans le dual, et s'écrit

ALGORITHME V.1. *(1) Soit $(\boldsymbol{\lambda}_t^0)_{1\leq t\leq T-1} \in \Pi_{t=1}^{T-1} L^2(\Omega,\mathbb{R},\mathbb{P})$, tel que $\boldsymbol{\lambda}_t^0$ est $\vee_{i=1}^n \mathcal{B}_{i,t}$-mesurable.*

(2) – *Pour tout $i \in \{1,\dots,n\}$, on trouve $\boldsymbol{u}_i^{k+1}$ solution de :*

$$\min_{\boldsymbol{u}_i,\boldsymbol{x}_i} \mathbb{E}\left(\sum_{t=1}^{T-1} c_i \frac{\boldsymbol{u}_{i,t}^2}{2} - \boldsymbol{\lambda}_t^k \boldsymbol{u}_{i,t}\right) + \frac{\gamma}{2}\mathbb{E}\left((\boldsymbol{x}_{i,T} - x_{i,1})^2\right) \tag{V.10}$$
$$\text{s.c.} \forall t \in \{1,\dots,T-1\},\ \boldsymbol{x}_{i,t+1} = \boldsymbol{x}_{i,t} + \boldsymbol{a}_{i,t+1} - \boldsymbol{u}_{i,t},\ \text{p.s.},$$
$$\forall t \in \{1,\dots,T-1\}, \boldsymbol{u}_{i,t} \in L^2(\Omega,\mathbb{R},\mathbb{P}),\ \text{et est } \mathcal{B}_t - \text{mesurable}.$$

– *Mettre à jour :*

$$\forall t \in \{1,\dots,T-1\},\ \boldsymbol{\lambda}_t^{k+1} = \boldsymbol{\lambda}_t^k + \rho^k\left(\boldsymbol{d}_t - \sum_{i=1}^n \boldsymbol{u}_{i,t}^{k+1}\right),\ \text{p.s.} \tag{V.11}$$

(3) Si $\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|$ est suffisamment petit, on arrête, sinon, $k := k+1$ et on retourne à l'étape (2).

Comme on l'a déjà évoqué en introduction de ce chapitre, les sous-problèmes (V.10) ne peuvent être résolus tout simplement par programmation dynamique stochastique. En effet, les multiplicateurs courants $(\boldsymbol{\lambda}_t^k)$ qui y apparaissent ne forment pas en toute généralité un bruit blanc. On ne sait finalement rien à leur propos. La seule chance que nous avons de pouvoir utiliser une programmation dynamique est d'exhiber une dynamique suivie par les multiplicateurs. Nous allons donc maintenant écrire les conditions d'optimalité du problème (V.6), et travailler sur les multiplicateurs optimaux $\boldsymbol{\lambda}^*$.

V.3.1.3. *Conditions d'optimalité et multiplicateurs optimaux.* En suivant le travail de [**8**] (Chapitre VI), on écrit les conditions d'optimalité de Kuhn-Tucker du problème (V.6). Plus intuitivement, elles se trouvent en écrivant le Lagrangien complet associé au problème, c'est à dire :

$$\begin{aligned}\mathcal{L}(\boldsymbol{u},\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu}) = &\sum_{t=1}^{T-1}\mathbb{E}\left(\sum_{i=1}^n c_i\frac{\boldsymbol{u}_{i,t}^2}{2} + \boldsymbol{\lambda}_t(\boldsymbol{d}_t - \sum_{i=1}^n \boldsymbol{u}_{i,t})\right)\\ &+\sum_{i=1}^n \mathbb{E}\left(\frac{\gamma_i}{2}(\boldsymbol{x}_{i,T}-x_{i,1})^2 + \sum_{t=1}^{T-1}\boldsymbol{\mu}_{i,t+1}(\boldsymbol{x}_{i,t}+\boldsymbol{a}_{i,t+1}-\boldsymbol{u}_{i,t}-\boldsymbol{x}_{i,t+1})\right),\end{aligned}$$

et en dérivant formellement ce Lagrangien par rapport à chaque composante. Finalement, on obtient :

$$\forall t \in \{1,\dots,T-1\},\ \forall i \in \{1,\dots,n\},\ \boldsymbol{u}_{i,t} \in L^2(\Omega,\mathbb{R},\mathbb{P}),\ \text{et est } \mathcal{B}_t - \text{mesurable}, \tag{V.12a}$$
$$\forall t \in \{1,\dots,T-1\},\ \forall i \in \{1,\dots,n\},\ \mathbb{E}\left(c_i\boldsymbol{u}_{i,t} - \boldsymbol{\mu}_{i,t+1} - \boldsymbol{\lambda}_t|\mathcal{B}_t\right) = 0,\ \text{p.s.}, \tag{V.12b}$$
$$\forall t \in \{2,\dots,T-1\},\ \forall i \in \{1,\dots,n\},\ \boldsymbol{\mu}_{i,t} = \boldsymbol{\mu}_{i,t+1},\ \text{p.s.}, \tag{V.12c}$$
$$\forall i \in \{1,\dots,n\},\ \boldsymbol{\mu}_{i,2} = \gamma_i(\boldsymbol{x}_{i,T} - x_{i,1}),\ \text{p.s.}, \tag{V.12d}$$
$$\forall i \in \{1,\dots,n\},\ \boldsymbol{\mu}_{i,T} = \gamma_i(\boldsymbol{x}_{i,T} - x_{i,1}),\ \text{p.s.}, \tag{V.12e}$$
$$\forall t \in \{1,\dots,T-1\},\ \sum_{i=1}^n \boldsymbol{u}_{i,t} = \boldsymbol{d}_t,\ \text{p.s.}, \tag{V.12f}$$
$$\forall t \in \{1,\dots,T-1\},\ \forall i \in \{1,\dots,n\},\ \boldsymbol{x}_{i,t+1} = \boldsymbol{x}_{i,t} + \boldsymbol{a}_{i,t+1} - \boldsymbol{u}_{i,t},\ \text{p.s.} \tag{V.12g}$$

Le but est maintenant d'essayer de résoudre les conditions d'optimalité pour éventuellement en exhiber des propriétés du multiplicateur de Lagrange $\boldsymbol{\lambda}$ associé à la contrainte couplante

(V.6c). A cette fin, nous allons successivement faire des hypothèses sur le problème et les bruits du problème, rendant les conditions d'optimalité solubles. On définit la version adaptée des multiplicateurs $\boldsymbol{\lambda}$:

$$\forall t \in \{1, \dots, T-1\},\ \boldsymbol{\Lambda}_t = \mathbb{E}\left(\boldsymbol{\lambda}_t | \mathcal{B}_t\right),\ \text{p.s..}$$

Nous allons maintenant résoudre les conditions d'optimalité. En utilisant les équations (V.12c)–(V.12d)–(V.12e), les états adjoints $\boldsymbol{\mu}$ sont constants en fonction du temps :

$$\forall i \in \{1, \dots, n\},\ \forall t \in \{2, \dots, T\},\ \boldsymbol{\mu}_{i,t} = \gamma_i(\boldsymbol{x}_{i,T} - x_{i,1}),\ \text{p.s.} \tag{V.13}$$

L'équation (V.12b) se réécrit donc avec (V.13) :

$$\forall t \in \{1, \dots, T-1\},\ \forall i \in \{1, \dots, n\},\ \mathbb{E}\left(\boldsymbol{u}_{i,t} | \mathcal{B}_t\right) = \frac{1}{c_i}\left(\boldsymbol{\Lambda}_t + \gamma_i \mathbb{E}\left(\boldsymbol{x}_{i,T} - x_{i,1} | \mathcal{B}_t\right)\right),\ \text{p.s.}$$

En ajoutant ces équations pour $i = 1, \dots, n$, et en utilisant (V.12f), on obtient, comme $(\boldsymbol{d}_t)$ est naturellement adaptée à $(\mathcal{B}_t)$:

$$\forall t \in \{1, \dots, T-1\},\ \boldsymbol{d}_t = \left(\sum_{i=1}^{n} \frac{1}{c_i}\right) \boldsymbol{\Lambda}_t + \sum_{i=1}^{n} \frac{\gamma_i}{c_i} \mathbb{E}\left(\boldsymbol{x}_{i,T}, -x_{i,1} | \mathcal{B}_t\right),\ \text{p.s.},$$

ce qui peut encore s'écrire :

$$\forall t \in \{1, \dots, T-1\},\ \boldsymbol{\Lambda}_t = \frac{1}{\sum_{i=1}^{n} \frac{1}{c_i}} \left(\boldsymbol{d}_t - \sum_{i=1}^{n} \frac{\gamma_i}{c_i} \mathbb{E}\left(\boldsymbol{x}_{i,T} - x_{i,1} | \mathcal{B}_t\right)\right),\ \text{p.s.}$$

Le but initial étant d'exhiber une dynamique sur les multiplicateurs optimaux, on écrit l'équation précédente sur deux pas de temps successifs, ce qui donne tous calculs faits :

$$\forall t \in \{1, \dots, T-2\},\ \boldsymbol{\Lambda}_{t+1} = \boldsymbol{\Lambda}_t + \frac{1}{\sum_{i=1}^{n} \frac{1}{c_i}} \left(\boldsymbol{d}_{t+1} - \boldsymbol{d}_t - \sum_{i=1}^{n} \frac{\gamma_i}{c_i} \left(\mathbb{E}\left(\boldsymbol{x}_{i,T} | \mathcal{B}_{t+1}\right) - \mathbb{E}\left(\boldsymbol{x}_{i,T} | \mathcal{B}_t\right)\right)\right),\ \text{p.s.} \tag{V.14}$$

De la même façon, on obtient le multiplicateur initial :

$$\boldsymbol{\Lambda}_1 = \frac{1}{\sum_{j=1}^{n} \frac{1}{c_j}} \left(\boldsymbol{d}_1 - \sum_{i=1}^{n} \frac{\gamma_i}{c_i} \mathbb{E}\left(\boldsymbol{x}_{i,T} - \boldsymbol{x}_{i,1} | \mathcal{B}_1\right)\right)\ \text{p.s.} \tag{V.15}$$

Remarque V.2. *A ce niveau de calculs, on peut faire deux remarques :*
– *si $(\boldsymbol{d}_t)_{1 \le t \le T}$ est une $(\mathcal{B}_t)_{1 \le t \le T}$-martingale, alors $(\boldsymbol{\Lambda}_t)_{1 \le t \le T}$ l'est aussi,*
– *$(\mathbb{E}\left(\boldsymbol{\Lambda}_t\right))_{1 \le t \le T}$ suit une dynamique donnée par l'accroissement de la demande moyenne.*

Nous allons maintenant travailler sur le dernier terme de l'équation (V.14), faisant intervenir les états. Grâce à l'équation (V.12g), on peut écrire :

$$\forall t \in \{1, \dots, T-1\},\ \forall i \in \{1, \dots, n\},\ \boldsymbol{x}_{i,T} = \boldsymbol{x}_{i,t} + \sum_{s=t+1}^{T} \boldsymbol{a}_{i,s} - \sum_{s=t}^{T-1} \boldsymbol{u}_{i,s},\ \text{p.s.}$$

Nous allons maintenant faire une hypothèse supplémentaire sur la fonction de coût :

Hypothèse V.3 (Pénalités). $\forall i \in \{1, \dots, n\},\ \gamma_i = \alpha c_i$, *avec* $\alpha \in \mathbb{R}^+$.

On peut alors réécrire l'équation (V.14) comme :

$$\begin{aligned}\forall t \in \{1,\dots,T-2\},\ \boldsymbol{\Lambda}_{t+1} &= \boldsymbol{\Lambda}_t + \frac{1}{\sum_{i=1}^n \frac{1}{c_i}} \left(\boldsymbol{d}_{t+1} - \boldsymbol{d}_t - \alpha \sum_{i=1}^n \left(\sum_{s=t+1}^T \left(\mathbb{E}\left(\boldsymbol{a}_{i,s}|\mathcal{B}_{t+1}\right) - \mathbb{E}\left(\boldsymbol{a}_{i,s}|\mathcal{B}_t\right)\right) \right.\right. \\ &\qquad \left.\left. - \sum_{s=t}^{T-1} \left(\mathbb{E}\left(\boldsymbol{u}_{i,s}|\mathcal{B}_{t+1}\right) - \mathbb{E}\left(\boldsymbol{u}_{i,s}|\mathcal{B}_t\right)\right) \right) \right),\ \text{p.s.} \\ &= \boldsymbol{\Lambda}_t + \frac{1}{\sum_{i=1}^n \frac{1}{c_i}} \left(\boldsymbol{d}_{t+1} - \boldsymbol{d}_t - \alpha \sum_{s=t+1}^T \left(\mathbb{E}\left(\boldsymbol{a}_s|\mathcal{B}_{t+1}\right) - \mathbb{E}\left(\boldsymbol{a}_s|\mathcal{B}_t\right)\right) \right. \\ &\qquad \left. + \alpha \sum_{s=t}^{T-1} \left(\mathbb{E}\left(\boldsymbol{d}_s|\mathcal{B}_{t+1}\right) - \mathbb{E}\left(\boldsymbol{d}_s|\mathcal{B}_t\right)\right) \right),\ \text{p.s.} \end{aligned} \tag{V.16}$$

avec $\boldsymbol{a}_s = \sum_{i=1}^n \boldsymbol{a}_{i,s}$ pour tout s. Pour obtenir l'équation (V.16), on a interverti les sommes en i et t, en vertu de l'hypothèse V.3, et utilisé la contrainte (V.12f). Par les mêmes arguments, on peut expliciter (V.15) de la manière suivante :

$$\boldsymbol{\Lambda}_1 = \frac{1}{\sum_{j=1}^n \frac{1}{c_j}} \left(\boldsymbol{d_1} - \alpha \sum_{s=2}^T \mathbb{E}\left(\boldsymbol{a}_s|\mathcal{B}_1\right) - \alpha \sum_{s=1}^{T-1} \mathbb{E}\left(\boldsymbol{d}_s|\mathcal{B}_1\right) \right)\ \text{p.s.} \tag{V.17}$$

A ce stade, on a donc obtenu une expression de la dynamique optimale des $\boldsymbol{\Lambda}$ dans laquelle les commandes optimales n'interviennent pas. Seuls les bruits du système interviennent.

Dans la perspective d'utiliser une programmation dynamique stochastique pour la résolution du problème (V.6), une hypothèse naturelle est

HYPOTHÈSE V.4. *$(\boldsymbol{d}_t, \boldsymbol{a}_{1,t}, \dots, \boldsymbol{a}_{n,t})_{1\le t\le T}$ est un bruit blanc.*

Dans ce cas, l'équation de dynamique (V.16) se réécrit :

$$\forall t \in \{1,\dots,T-2\},\ \boldsymbol{\Lambda}_{t+1} = \boldsymbol{\Lambda}_t + \frac{1}{\sum_{i=1}^n \frac{1}{c_i}} \left(\boldsymbol{d}_{t+1}(1+\alpha) - \boldsymbol{d}_t - \alpha\mathbb{E}\left(\boldsymbol{d}_{t+1}\right) - \alpha(\boldsymbol{a}_{t+1} - \mathbb{E}\left(\boldsymbol{a}_{t+1}\right))\right),\ \text{p.s.}$$

$$\boldsymbol{\Lambda}_1 = \frac{1}{\sum_{j=1}^n \frac{1}{c_j}} \left(\boldsymbol{d_1}(1-\alpha) - \alpha \sum_{s=2}^T \mathbb{E}\left(\boldsymbol{a}_s\right) - \alpha \sum_{s=2}^{T-1} \mathbb{E}\left(\boldsymbol{d}_s\right) \right),\ \text{p.s.}$$

On est donc parvenu à exhiber une dynamique suivie par les multiplicateurs optimaux. Dans la perspective de la résolution du problème dual, il suffit donc de donner en entrée aux sous-problèmes (V.9) les prix optimaux. Posons $\boldsymbol{w}_{t+1} = \boldsymbol{d}_{t+1}(1+\alpha) - \boldsymbol{d}_t - \alpha\mathbb{E}\left(\boldsymbol{d}_{t+1}\right)$. Les prix optimaux sont alors déduits de l'équation de dynamique :

$$\forall t \in \{1,\dots,T-2\},\ \boldsymbol{\Lambda}_{t+1} = \boldsymbol{\Lambda}_t + \frac{1}{\sum_{i=1}^n \frac{1}{c_i}} \left(\boldsymbol{w}_{t+1} - \alpha(\boldsymbol{a}_{t+1} - \mathbb{E}\left(\boldsymbol{a}_{t+1}\right))\right),\ \text{p.s.} \tag{V.19}$$

V.3.1.4. *Programmations dynamiques locales.* On s'intéresse ici à la résolution des problèmes locaux, i.e. des problèmes :

$$\begin{aligned} \min_{\boldsymbol{u}_i, \boldsymbol{x}_i} & \mathbb{E}\left(\sum_{t=1}^{T-1} c_i \frac{\boldsymbol{u}_{i,t}^2}{2} - \boldsymbol{\Lambda}_t \boldsymbol{u}_{i,t} \right) + \frac{\alpha c_i}{2} \mathbb{E}\left((\boldsymbol{x}_{i,T} - x_{i,1})^2 \right), \\ \text{s.c.} & \forall t \in \{1,\dots,T-1\},\ \boldsymbol{x}_{i,t+1} = \boldsymbol{x}_{i,t} + \boldsymbol{a}_{i,t+1} - \boldsymbol{u}_{i,t},\ \text{p.s.}, \\ & \forall t \in \{1,\dots,T-2\},\ \boldsymbol{\Lambda}_{t+1} = \boldsymbol{\Lambda}_t + \frac{1}{\sum_{i=1}^n \frac{1}{c_i}} \left(\boldsymbol{w}_{t+1} - \alpha(\boldsymbol{a}_{t+1} - \mathbb{E}\left(\boldsymbol{a}_{t+1}\right))\right),\ \text{p.s.} \\ & \boldsymbol{\Lambda}_1 = \frac{1}{\sum_{j=1}^n \frac{1}{c_j}} \left(\boldsymbol{d_1}(1-\alpha) - \alpha \sum_{s=2}^T \mathbb{E}\left(\boldsymbol{a}_s\right) - \alpha \sum_{s=2}^{T-1} \mathbb{E}\left(\boldsymbol{d}_s\right) \right),\ \text{p.s.} \\ & \forall t \in \{1,\dots,T-1\}, \boldsymbol{u}_{i,t} \in L^2(\Omega, \mathbb{R}, \mathbb{P}),\ \text{et est } \mathcal{B}_t - \text{mesurable.} \end{aligned} \tag{V.20}$$

Dans ces sous-problèmes, il reste encore à blanchir le bruit $(\boldsymbol{w}_t)$. De par sa définition, il nécessite d'introduire deux états supplémentaires. Au final, on peut donc résoudre par programmation dynamique stochastique chacun des problèmes (V.20). Chaque programmation dynamique devra être effectuée en dimension 4 (les états sont $\boldsymbol{x}_i, \boldsymbol{\Lambda}, \boldsymbol{w}$, et ce dernier doit être dédoublé). La solution ainsi obtenue sera donc la solution optimale de par l'optimalité du multiplicateur associé.

V.3.1.5. *Conclusion.* La conclusion de cet exemple est donc que l'on peut, dans le cas linéaire quadratique, passer d'un problème de programmation dynamique en dimension n à n problèmes de programmation dynamique en dimension 4. Le prix à payer lorsqu'on décompose est donc élevé, et on ne peut s'attendre qu'à pire pour des problèmes aux contraintes convexes générales et coût non-quadratique. La décomposition par les prix ainsi proposée est donc compatible avec la programmation dynamique dans un cas particulier, mais laisse peu d'espoir dans le cas général. Le résultat présenté ici, et démontré grâce à la manipulation des conditions d'optimalité, est plus un résultat d'agrégation qu'un résultat de décomposition : il existe un état agrégé commun à tout le système, tel que le contrôle de chaque unité soit réalisé en feedback sur l'état local et cet état agrégé. L'intérêt majeur de ce résultat est de mettre en évidence la difficulté d'apporter une réponse générique à la question de marier la décomposition et la programmation dynamique.

V.3.2. Commandes décentralisées. Nous allons dans cette section regarder comment les résultats obtenus par [**32**] peuvent être étendus de la boucle ouverte au cas de la décomposition de problèmes en boucle fermée pour lesquels les commandes sont prises décentralisées. L'optimisation décentralisée se restreint d'emblée à des commandes en feedback sur des états (ou bruits, selon le formalisme choisi) locaux. La théorie de la décentralisation en contrôle est notamment exposée dans [**16**], au chapitre V.
Avant tout, nous allons montrer dans une première sous-section un exemple des possibilités offertes par la décentralisation des commandes, puis, sur la base de l'algorithme du problème auxiliaire classique, nous allons proposer un algorithme adapté au cas de commandes décentralisées, et montrer sa convergence.

V.3.2.1. *Boucle fermée et boucle ouverte partielles.* Considérons le problème suivant :

$$\begin{aligned} &\min_{\boldsymbol{u}} \mathbb{E}\left(j(\boldsymbol{u}, \boldsymbol{\xi}_2)\right) \ , \\ &\text{s.c. } \boldsymbol{u} \in U^f \subset \{\boldsymbol{u} \in L^2(\Omega, U, \mathbb{P}) \ : \ \boldsymbol{u} \ \sigma(\boldsymbol{\xi}_1) - \text{mesurable}\}, \end{aligned} \tag{V.21}$$

avec $\boldsymbol{\xi}_1$ et $\boldsymbol{\xi}_2$ deux variables aléatoires à valeurs respectivement dans Ξ_1 et Ξ_2, deux espaces métriques, et de loi conjointe μ (on notera μ_i, $i = 1, 2$ les lois marginales) ; U un espace vectoriel de dimension finie, et $j : U \times \Xi_2 \to \mathbb{R}$ une intégrande normale. Le problème (V.21) est dès lors équivalent au problème :

$$\begin{aligned} &\min_{u} J(u) := \mathbb{E}\left(j(u(\boldsymbol{\xi}_1), \boldsymbol{\xi}_2)\right), \\ &\text{s.c. } u(\boldsymbol{\xi}_1) \in U^f, \end{aligned} \tag{V.22}$$

dans lequel on a écrit la contrainte de mesurabilité directement *en dur* dans le problème. On peut voir que l'ensemble U^f est un sous-ensemble de l'espace des variables aléatoires de carré intégrable. C'est $u(\boldsymbol{\xi}_1)$ qui doit lui appartenir et non u. Il apparaît de plus que le contrôle u est recherché en boucle fermée sur $\boldsymbol{\xi}_1$ et boucle ouverte sur $\boldsymbol{\xi}_2$. On pourrait dès lors imaginer une approche du type gradient stochastique (cf. Appendice C) pour la partie concernant $\boldsymbol{\xi}_2$. Afin d'expliciter cette idée, calculons le gradient de J. En supposant j différentiable en sa première composante, on calcule la dérivée directionnelle de J. Soit $(u, h) \in L^2(\Xi_1, U, \mu_1)$:

$$\begin{aligned} J(u + th) &= \mathbb{E}\left(j(u(\boldsymbol{\xi}_1) + th(\boldsymbol{\xi}_1), \boldsymbol{\xi}_2)\right) \\ &= \mathbb{E}\left(\langle \nabla_u j(u(\boldsymbol{\xi}_1), \boldsymbol{\xi}_2), th(\boldsymbol{\xi}_1)\rangle_U\right) + J(u) + O(t^2) \\ &= J(u) + t\langle \mathbb{E}\left(\nabla_u j(u(\boldsymbol{\xi}_1), \boldsymbol{\xi}_2) | \boldsymbol{\xi}_1 = \cdot\right), h(\cdot)\rangle_{L^2(\Xi_1, U, \mu_1)} + O(t^2) \end{aligned}$$

On obtient donc :

$$\forall u \in L^2(\Xi_1, U, \mu_1), \ \nabla J(u)(\xi_1) = \mathbb{E}\left(\nabla_u j(u(\boldsymbol{\xi}_1), \boldsymbol{\xi}_2) | \boldsymbol{\xi}_1 = \xi_1\right). \tag{V.23}$$

On distingue alors principalement deux cas :

– Si $\boldsymbol{\xi}_1$ et $\boldsymbol{\xi}_2$ sont indépendantes, alors (V.23) peut se réécrire sous la forme :

$$\forall u \in L^2(\Xi_1, U, \mu_1),\ \nabla J(u)(\xi_1) = \mathbb{E}\left(\nabla_u j(u(\xi_1), \boldsymbol{\xi}_2)\right).$$

À supposer que $U^f = \{u(\boldsymbol{\xi}_1) \in L^2(\Omega, U, \mathbb{P}) \ :\ u(\xi_1) \in \Gamma(\xi_1),\ \mu_1 - \text{p.s.}\}$, avec Γ une multi-application à valeurs convexes fermées, on peut appliquer un algorithme de gradient stochastique projeté (cf. Appendice C, équation (C.3)) pour résoudre le problème (V.21), du type :

$$\begin{aligned} &\text{Soit } \boldsymbol{\xi}_2^{k+1} \text{ distribuée selon } \mu_2, \text{ indépendante des v.a. passées,} \\ &\forall \xi_1 \in \Xi_1,\ u^{k+1}(\xi_1) = \Pi_{\Gamma(\xi_1)}\left(u^k(\cdot) - \rho^k \nabla_u j(u(\cdot), \boldsymbol{\xi}_2^{k+1})\right)(\xi_1). \end{aligned} \tag{V.24}$$

Cependant, un tel algorithme ne serait de toute façon pas implémentable tel quel, vu que la variable de commande u demeure en dimension infinie, et que l'opération de projection est loin d'être triviale en toute généralité. Une possibilité pour traiter cette question est précisément l'objet du chapitre IV, avec par exemple l'algorithme IV.4.

– Si $\boldsymbol{\xi}_1$ et $\boldsymbol{\xi}_2$ sont corrélées, on ne peut a priori rien gagner à utiliser des tirages sur $\boldsymbol{\xi}_2$ (conditionnellement à ξ_1), à moins de recourir à la machinerie du chapitre IV, mais ce n'est pas ici notre objet. Il y a cependant quelques cas particuliers dans lesquels la situation d'estimer un gradient qui est désormais une espérance conditionnelle n'est pas si désespérée :

(1) Le cas d'une *corrélation explicite* : c'est le cas lorsque $\boldsymbol{\xi}_2$ peut s'écrire explicitement comme une fonction de $\boldsymbol{\xi}_1$. Supposons qu'il existe une application $g : \Xi_1 \to \Xi_2$ et une variable aléatoire $\boldsymbol{\epsilon}$ à valeurs dans Ξ_2, indépendante de $\boldsymbol{\xi}_1$, et telles que :

$$\boldsymbol{\xi}_2 = g(\boldsymbol{\xi}_1) + \boldsymbol{\epsilon},\ \text{p.s.}$$

On peut alors réécrire le gradient de J comme :

$$\begin{aligned} \forall u \in L^2(\Xi_1, U, \mu_1),\ \nabla J(u)(\xi_1) &= \mathbb{E}\left(\nabla_u j(u(\boldsymbol{\xi}_1), g(\boldsymbol{\xi}_1) + \boldsymbol{\epsilon}) | \boldsymbol{\xi}_1 = \xi_1\right), \\ &= \mathbb{E}\left(\nabla_u j(u(\xi_1), g(\xi_1) + \boldsymbol{\epsilon})\right). \end{aligned}$$

On peut donc également appliquer l'algorithme (V.24), avec des tirages de $\boldsymbol{\epsilon}$ au lieu de $\boldsymbol{\xi}_2$;

(2) Le cas où $\boldsymbol{\xi}_1$ a un *support fini* : dans ce cas, disons que $\{\xi_1^1, \ldots, \xi_1^n\}$ constitue le support fini de $\boldsymbol{\xi}_1$. La loi de $\boldsymbol{\xi}_1$ est donc donnée par un n-uplet positif $(\pi_i)_{1 \le i \le n}$ tel que $\sum_{i=1}^n \pi_i = 1$. Le problème est alors simplement de déterminer n commandes $u(\xi_1^i) := u^i$, et l'ensemble admissible fonctionnel U^f devient un convexe en dimension finie noté U_n^f, et défini par $U_n^f = \{u \in U \ :\ \forall i = 1, \ldots, n,\ u^i \in \Gamma(\xi_i^i)\}$. On peut donc écrire l'algorithme de gradient suivant :

$$\forall i \in \{1, \ldots, n\},\ (u^i)^{k+1} := \Pi_{\Gamma(\xi_1^i)}\left((u^i)^k - \rho^k \mathbb{E}\left(\nabla_u j((u^i)^k, \boldsymbol{\xi}_2) | \boldsymbol{\xi}_1 = \xi_1^i\right)\right).$$

On est alors revenu au cadre standard de l'application de l'algorithme du gradient stochastique : on a n espérances à calculer qui suivent chacune une loi différente correspondant à la loi conditionnelle de $\boldsymbol{\xi}_2$ sachant chaque valeur possible de $\boldsymbol{\xi}_1$; on peut donc faire des tirages sur chaque composant de u, selon les lois conditionnelles associées.

C'est dans cet esprit de commandes partiellement en boucle fermée et partiellement en boucle ouverte qu'il faut comprendre la décentralisation dans cette section.

V.3.2.2. *Commandes décentralisées.* Le problème de gérer n unités affectées par n aléas se pose en toute généralité comme le problème de déterminer n commandes en boucle fermée sur les n aléas. Nous allons supposer ici que la commande de l'unité i n'est recherchée que comme une

fonction de l'aléa i propre au sous-système qu'elle dirige, c'est le cas de la gestion décentralisée. On regarde donc désormais le problème :

$$\min J(u) = \mathbb{E}\left(j(u_1(\boldsymbol{\xi}_1), \dots, u_n(\boldsymbol{\xi}_n), \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n)\right) , \tag{V.25}$$

$$\text{s.c. } u = (u_1, \dots, u_n) \in U_1^f \times \cdots \times U_n^f, \tag{V.26}$$

avec :

- pour $i = 1, \dots, n$, Ξ_i un espace métrique, et $\boldsymbol{\xi}_i$ une variable aléatoire à valeurs dans Ξ_i ; telles que pour tous $i \neq j \in \{1, \dots, n\}$, $\boldsymbol{\xi}_i$ et $\boldsymbol{\xi}_j$ soient indépendantes ;
- pour tout $i \in \{1, \dots, n\}$, U_i un espace vectoriel préhilbertien de dimension finie notée p_i, et muni d'un produit scalaire noté $\langle \cdot, \cdot \rangle_i$ et d'une norme associée ;
- pour $i = 1, \dots, n$, U_i^f un convexe fermé de $L^2(\Xi_i, U_i, \mu_i)$, qui forme un espace de Hilbert muni du produit scalaire noté $\langle u, v \rangle_{L_i^2} := \mathbb{E}\left(\langle u(\boldsymbol{\xi}_i), v(\boldsymbol{\xi}_i) \rangle_i\right)$ et de la norme associée ;
- $j : \Pi_{i=1}^n U_i \times \Pi_{i=1}^n \Xi_i \to \mathbb{R}$ une application semicontinue inférieurement.

Ce problème est donc typiquement un problème en boucle fermée, mais avec une structure décentralisée qui rappelle les remarques faites dans la sous-section précédente. Afin de tenter une décomposition de ce problème, on recourt à la théorie du principe du problème auxiliaire (PPA, cf. Appendice A). On introduit donc n fonctions auxiliaires $G_i : L^2(\Xi_i, U_i, \mu_i) \to \mathbb{R}$ pour $i = 1, \dots, n$, et on définit $G : \Pi_{i=1}^n L^2(\Xi_i, U_i, \mu_i) \to \mathbb{R}$ comme étant $G(u) = \sum_{i=1}^n G_i(u_i)$. Sous des hypothèses de forte convexité des fonctions auxiliaires, et de convexité et régularité de la fonction J, on peut alors proposer l'algorithme du PPA suivant, qui convergera vers une solution du problème initial (V.25) :

Algorithme V.5. *Étape 0 : Soit $(u_i^0, \dots, u_n^0) \in \Pi_{i=1}^n L^2(\Xi_i, U_i, \mu_i)$, $\epsilon > 0$ un seuil fixé,*
Étape k :

- *Choisir :*

$$\forall i \in \{1, \dots, n\},\ r_i^k(\cdot) = \mathbb{E}\Big(\nabla_{u_i} j(u_1^k(\boldsymbol{\xi}_1), \dots, u_{i-1}^k(\boldsymbol{\xi}_{i-1}), u_i^k(\cdot), u_{i+1}^k(\boldsymbol{\xi}_{i+1}), \dots, u_n^k(\boldsymbol{\xi}_n),$$
$$\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_{i-1}, \cdot, \boldsymbol{\xi}_{i+1}, \dots, \boldsymbol{\xi}_n)\Big) \in L^2(\Xi_i, U_i, \mu_i), \tag{V.27}$$

(V.28)

- *Pour tout $i \in \{1, \dots, n\}$, résoudre :*

$$u_i^{k+1} \in \arg\min G_i(u_i) + \langle \gamma_i^k r_i^k(\boldsymbol{\xi}_i) - \nabla G_i(u_i^k)(\boldsymbol{\xi}_i), u_i(\boldsymbol{\xi}_i) \rangle_{L_i^2}, \tag{V.29}$$
$$\text{s.c. } u_i \in U_i^f .$$

- *Si $\sum_{i=1}^n \|u_i^{k+1} - u_i^k\|_{L_i^2} \leq \epsilon$, stop, sinon aller à l'étape $k+1$.*

On a défini r_i^k par une espérance et non une espérance conditionnelle en raison de l'indépendance des $(\boldsymbol{\xi}_j)_{1 \leq j \leq n}$. Il faut comprendre la formule (V.27) comme une intégrale multiple, que l'on recomposera ensuite avec la variable aléatoire $\boldsymbol{\xi}_i$ pour obtenir le gradient.

Remarque V.6 (Noyaux de décomposition). *Les noyaux de décomposition G_i sont ici définis comme des applications d'un ensemble d'applications $(L^2(\Xi_i, U_i, \mu_i))$ dans $\mathbb{R}$. Les applications $\boldsymbol{u}_i$ sont ensuite elles-mêmes classiquement composées avec les variables aléatoires $\boldsymbol{\xi}_i$. Il s'y cache donc une opération d'espérance. En pratique, on prendra des noyaux définis comme suit :*

$$\forall u_i \in L^2(\Xi_i, U_i, \mu_i),\ G_i(u_i) = \mathbb{E}\left(g_i(u_i(\boldsymbol{\xi}_i), \boldsymbol{\xi}_i)\right),$$

avec g_i une application fortement convexe en u et vérifiant des hypothèses du type intégrande normale *transposées dans le cadre des applications.*

On voit clairement dans l'algorithme V.5 la présence dans chaque sous-problème (V.29) de termes d'aléas sur lesquels le contrôle est en boucle ouverte, et de termes d'aléas sur lesquels le contrôle est en boucle fermée. Ce type de sous-problèmes décentralisés dans lesquels demeurent de lourds calculs d'espérance en grande dimension nécessitant des approximations a déjà été remarqué dans [**42**]. Dans cet article, une approximation gaussienne de l'espérance dans les fonctions de coûts a été proposée. Nous allons ici montrer que la décentralisation des commandes

permet de s'affranchir totalement des calculs complets d'espérance.
En appliquant les remarques faites dans la sous-section précédente, on pourrait avoir envie de substituer au calcul d'espérance nécessaire à l'évaluation de chaque gradient (équation (V.27)), un simple tirage des aléas en boucle ouverte pour le sous-problème considéré. C'est ce en quoi consiste l'algorithme stochastique décrit dans la sous-section suivante.

V.3.2.3. *Algorithme et convergence.* Afin de minimiser le nombre de calculs d'espérances dans lesquels interviennent un trop grand nombre de variables aléatoires, on propose l'algorithme stochastique suivant :

ALGORITHME V.7. *Étape 0 : Soit* $(u_i^0, \dots, u_n^0) \in \Pi_{i=1}^n L^2(\Xi_i, U_i, \mu_i)$,
Étape k :
- *Soit* $(\boldsymbol{\xi}_1^{k+1}, \dots, \boldsymbol{\xi}_n^{k+1})$ *distribuée selon* μ, *indépendante des v.a. passés ;*
- *Choisir :*

$$\forall i \in \{1, \dots, n\},\ \boldsymbol{r}_i^k(\cdot) = \nabla_{u_i} j\Big(\boldsymbol{u}_1^k(\boldsymbol{\xi}_1^{k+1}), \dots, \boldsymbol{u}_{i-1}^k(\boldsymbol{\xi}_{i-1}^{k+1}), \boldsymbol{u}_i^k(\cdot), \boldsymbol{u}_{i+1}^k(\boldsymbol{\xi}_{i+1}^{k+1}), \dots, \boldsymbol{u}_n^k(\boldsymbol{\xi}_n^{k+1}),$$

$$\boldsymbol{\xi}_1^{k+1}, \dots, \boldsymbol{\xi}_{i-1}^{k+1}, \cdot, \boldsymbol{\xi}_{i+1}^{k+1}, \dots, \boldsymbol{\xi}_n^{k+1}\Big) \in L^2(\Xi_i, U_i, \mu_i), \tag{V.30}$$

(V.31)

- *Pour tout* $i \in \{1, \dots, n\}$, *résoudre :*

$$\boldsymbol{u}_i^{k+1} \in \arg\min \quad G_i(u_i) + \langle \gamma_i^k \boldsymbol{r}_i^k - \nabla G_i(\boldsymbol{u}_i^k), u_i \rangle_{L_i^2}, \tag{V.32}$$
$$\text{s.c.} \quad u_i \in U_i^f.$$

- *Si le nombre maximum d'itérations a été atteint, s'arrêter, sinon aller à l'étape* $k+1$.

REMARQUE V.8 (Algorithme de gradient perturbé). *Comme d'habitude avec le PPA, on remarque qu'en prenant des noyaux de décomposition particuliers définis comme étant pour* $i \in \{1, \dots, n\}$, $G_i(u_i) = \frac{1}{2}\|u_i\|_{L_i^2}^2$, *on obtient une formule explicite dans les sous-problèmes* (V.32) *:*

$$\forall i \in \{1, \dots, n\},\ \boldsymbol{u}_i^{k+1} = \Pi_{U_i^f}\left(\boldsymbol{u}_i^k - \gamma_i^k \boldsymbol{r}_i^k\right),$$

qui est l'analogue d'une formule de gradient projeté, à ceci près que les $\boldsymbol{r}_i^k$ *ne constitue pas un gradient, mais une approximation sans biais du gradient.*

En suivant le travail de [**32**], on montre le théorème suivant :

THÉORÈME V.9. *(i) Supposons que pour tout* $\xi \in \Pi_{i=1}^n \Xi_i$, $u \mapsto j(u, \xi)$ *soit convexe, semi-continue inférieurement, et différentiable sur* $\Pi_{i=1}^n U_i$. *Si de plus* J *est coercive sur* $\Pi_{i=1}^n U_i^f$ *qui forme un produit de convexes fermés et est de ce fait convexe fermé, alors* (V.25) *admet des solutions, et on notera* S *son ensemble de solutions.*

(ii) Supposons que pour tout $i \in \{1, \dots, n\}$, G_i *soit différentiable, et fortement convexe de module* $b_i > 0$. *Alors, pour tout* $k \in \mathbb{N}$, *les sous-problèmes* (V.32) *admettent une unique solution notée respectivement* $\boldsymbol{u}_i^{k+1} \in L^2(\Xi_i, U_i, \mu_i)$.

(iii) Supposons que pour tout $i \in \{1, \dots, n\}$, *la suite déterministe* (γ_i^k) *soit telle que :*

$$\gamma_i^k > 0, \qquad \sum_{k=1}^{\infty} \gamma_i^k = +\infty, \qquad \sum_{k=1}^{\infty} (\gamma_i^k)^2 < +\infty, \qquad \gamma_i^k = \gamma^k \tag{V.33}$$

Si de plus $j(\cdot, \xi)$ *a son gradient linéairement borné uniformément en* ξ, μ*-presque partout, alors :*

$$\lim_{k \to \infty} J(\boldsymbol{u}^k) = J(u^*) \quad \textit{presque sûrement.} \tag{V.34}$$

et la suite $(\boldsymbol{u}^k)$ *est p.s. bornée, et tout point d'accumulation de cette suite dans la topologie faible de* $\Pi_{i=1}^n L^2(\Xi_i, U_i, \mu_i)$ *est solution de* (V.25).

(iv) Si de plus $j(\cdot,\xi)$ est fortement convexe uniformément en ξ, de module $B>0$, alors S se réduit à un singleton $\{(u_1^*,\dots,u_n^*)\}$, et la suite $(\boldsymbol{u}^k)$ converge presque sûrement fortement vers $(u_1^*,\dots,u_n^*)$.

Preuve : Il suffit de suivre ligne à ligne pour tout $i\in\{1,\dots,n\}$ la preuve du théorème 2.1 de [**32**]. A l'itération k de l'algorithme, en définissant $\mathcal{F}^k$ la filtration engendrée par $(\boldsymbol{\xi}_i^1,\cdots,\boldsymbol{\xi}_i^k)_{1\leq i\leq n}$, posons $\boldsymbol{s}^k=(\boldsymbol{r}_1^k,\cdots,\boldsymbol{r}_n^k)\in\Pi_{i=1}^n L^2(\Xi_i,U_i,\mu_i)$. Ce vecteur de gradients n'est pas interprétable directement comme le gradient de la fonction de coût J. En revanche, $\mathbb{E}\left(\boldsymbol{s}^k|\mathcal{F}^k\right)=\nabla J(\boldsymbol{u}^k)$. On peut donc appliquer la même preuve que celle du théorème 2.1 de [**32**]. On donnera plus loin dans ce chapitre, à la remarque V.16, une preuve directe de ce théorème. □

V.3.2.4. *Application numérique.* Nous proposons ici une basique application numérique de l'algorithme V.7. Cet exemple n'a aucune visée pratique et n'est là que pour illustrer que l'algorithme fonctionne bien. Du reste, pour ne pas avoir à employer de techniques propres à la boucle fermée, nous nous restreignons à des contrôles en feedback linéaire, et travaillons directement sur les coefficients de ces feedbacks. On pourra pour cette raison contester l'application de la théorie qui précède à ce cadre restreint, mais qui peut le plus peut le moins. On souhaite donc résoudre le problème quadratique suivant :

$$\begin{aligned}\min_u\ &\mathbb{E}\left(u(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2)^TQu(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2)+u(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2)^Ta\right),\\ \text{s.c. } &u(\xi_1,\xi_2)=\begin{pmatrix}u_1(\xi_1)\\ u_2(\xi_2)\end{pmatrix}\in\mathbb{R}^2,\\ &u_1(\xi_1)=\alpha_1\xi_1,\ \alpha_1\in\mathbb{R},\\ &u_2(\xi_2)=\alpha_2\xi_2,\ \alpha_2\in\mathbb{R}.\end{aligned}\tag{V.35}$$

(V.36)

avec Q une matrice de taille 2×2 symétrique, définie positive, $a\in\mathbb{R}^2$, et

$$\begin{pmatrix}\boldsymbol{\xi}_1\\ \boldsymbol{\xi}_2\end{pmatrix}\sim\mathcal{N}\left(\begin{pmatrix}m_1\\ m_2\end{pmatrix},\begin{pmatrix}\sigma_1^2&0\\0&\sigma_2^2\end{pmatrix}\right).$$

On peut donc réécrire le problème (V.35) comme un problème en boucle ouverte :

$$\min_{\alpha\in\mathbb{R}^2}\alpha^T\mathcal{Q}\alpha+\alpha^T\mathcal{A},\tag{V.37}$$

avec :

$$\mathcal{Q}=\begin{pmatrix}q_{11}(m_1^2+\sigma_1^2)&q_{12}m_1m_2\\ q_{12}m_1m_2&q_{22}(m_2^2+\sigma_2^2)\end{pmatrix}\quad\text{, et, }\mathcal{A}=\begin{pmatrix}a_1m_1\\ a_2m_2\end{pmatrix}.$$

La solution optimale α^* du problème (V.37) s'obtient en égalant à 0 le gradient en α de la fonction de coût, ce qui donne :

$$\alpha^*=-\frac{1}{2}\mathcal{Q}^{-1}\mathcal{A}.$$

On souhaite maintenant utiliser l'algorithme du PPA classique (cf. algorithme V.5) et notre algorithme V.7 pour résoudre le problème (V.35). En réalité, on utilisera directement la représentation boucle ouverte du problème, i.e., (V.37). Afin de faire cela, on se donne le noyau de décomposition G défini par :

$$\forall\alpha\in\mathbb{R}^2,\ G(\alpha)=\frac{1}{2}\alpha^T\mathbb{E}\begin{pmatrix}(\boldsymbol{\xi}_1)^2&0\\0&(\boldsymbol{\xi}_2)^2\end{pmatrix}\alpha.$$

Ce noyau est bien sûr additif, et l'algorithme du PPA classique donne :

Algorithme V.10. *PPA Classique Étape 0 : Choisir $\alpha^0\in\mathbb{R}^2$,*
Étape k : Calculer :

$$\alpha^{k+1}=\alpha^k-\epsilon^k\mathbb{E}\begin{pmatrix}(\boldsymbol{\xi}_1)^2&0\\0&(\boldsymbol{\xi}_2)^2\end{pmatrix}^{-1}\left(2\mathcal{Q}\alpha^k+\mathcal{A}\right);$$

Si $\|\alpha^{k+1}-\alpha^k\|$ est suffisamment faible, stop, sinon boucler ;

La formule de mise à jour s'écrit explicitement :

$$\alpha_1^{k+1} = \alpha_1^k - \epsilon^k \left(2q_{11}\alpha_1^k + 2q_{12}\frac{m_1 m_2}{m_1^2+\sigma_1^2}\alpha_2^k + \frac{m_1}{m_1^2+\sigma_1^2}a_1 \right),$$
$$\alpha_2^{k+1} = \alpha_2^k - \epsilon^k \left(2q_{22}\alpha_2^k + 2q_{12}\frac{m_1 m_2}{m_2^2+\sigma_2^2}\alpha_1^k + \frac{m_2}{m_2^2+\sigma_2^2}a_2 \right).$$

Notre algorithme donne quant à lui :

ALGORITHME V.11 (PPA Stochastique). *Étape 0 : Choisir* $\alpha^0 \in \mathbb{R}^2$,
Étape k : Soit $\boldsymbol{\xi}_1^{k+1}, \boldsymbol{\xi}_2^{k+1}$, *i.i.d. par rapport aux v.a. passées,*
Calculer :

$$\alpha_1^{k+1} = \alpha_1^k - \epsilon^k \left(2q_{11}\alpha_1^k + 2q_{12}\frac{m_1}{m_1^2+\sigma_1^2}\alpha_2^k\boldsymbol{\xi}_2^{k+1} + \frac{m_1}{m_1^2+\sigma_1^2}a_1 \right),$$
$$\alpha_2^{k+1} = \alpha_2^k - \epsilon^k \left(2q_{22}\alpha_2^k + 2q_{12}\frac{m_2}{m_2^2+\sigma_2^2}\alpha_1^k\boldsymbol{\xi}_1^{k+1} + \frac{m_2}{m_2^2+\sigma_2^2}a_2 \right);$$

Si le nombre maximal d'itérations est atteint, s'arrêter, sinon boucler.

La figure 1 donne l'évolution de l'erreur en norme sur α au long des itérations pour chacun des deux algorithmes (à gauche pour l'algorithme classique, et à droite pour l'algorithme stochastique).

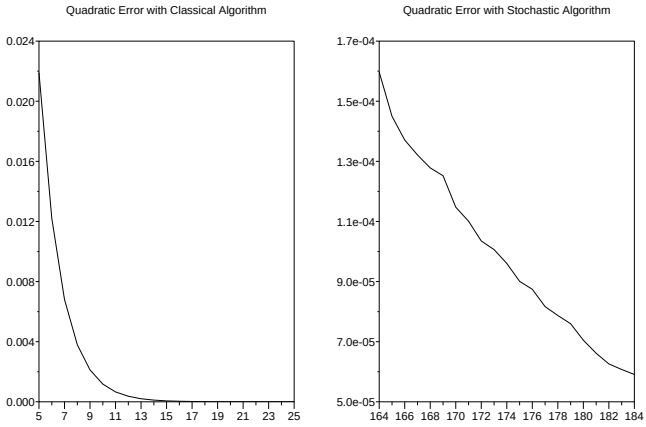


FIG. 1. Erreur quadratique sur le contrôle pour les algorithmes de décomposition au long des itérations

Il est clair que l'algorithme du PPA classique converge en moins d'itérations que l'algorithme du PPA stochastique, ce qui est dû à leur nature, mais ce dernier reste compétitif et présente l'énorme avantage de pouvoir converger dès lors que l'on sait tirer des réalisations de l'aléa, sans même en connaître au préalable la loi, alors que l'algorithme classique nécessite quant à lui de connaître cette loi afin de réaliser les calculs d'espérance dans chaque itération. Enfin, l'intérêt de l'algorithme du PPA stochastique apparaîtra surtout en plus grande dimension, lorsque les calculs exacts d'espérance s'avéreront trop coûteux en raison du trop grand nombre de variables aléatoires : le coût d'une itération de l'algorithme stochastique est indépendant du nombre de sous-problèmes, i.e. de la dimension du calcul d'espérance totale.

V.4. Principe du problème auxiliaire stochastique

V.4.1. Introduction et algorithme. À ce stade de notre réflexion, il apparaît donc que dans le cadre d'une décomposition d'un problème d'optimisation stochastique, une nouvelle source de couplage, de nature probabiliste intervient à travers la coordination : c'est pour traiter ce couplage afférent aux multiplicateurs que nous avons dû dans la section V.3 mettre en œuvre

toute cette machinerie. Afin de donner un peu de hauteur au problème, nous allons ici donner une généralisation du principe du problème auxiliaire dans un cadre hilbertien. La théorie du principe du problème auxiliaire (PPA), remontant aux travaux [**28, 29**], puis développée par exemple dans [**32**] pour une application stochastique en boucle ouverte, existe déjà dans un contexte hilbertien général. La particularité est ici d'en donner, toujours dans le cadre hilbertien, une version stochastique, c'est à dire une version dans laquelle des perturbations sont autorisées sur les outils de coordination (ici le gradient). Le but final, malheureusement non atteint ici, serait ensuite de donner une manière pratique de perturber les outils de coordination (le gradient) propre à gérer le couplage probabiliste.

Considérons le problème d'optimisation suivant :

$$\min_{u\in U^f} J(u), \tag{V.38}$$

où U^f est un sous-ensemble convexe fermé d'un espace de Hilbert U muni d'un produit scalaire noté $\langle\cdot,\cdot\rangle$ et de la norme associée. $J : U \to \mathbb{R}$ est une application convexe, coercive. Le PPA usuel (cf. Appendice A) consiste à associer au problème (V.38) la suite de problèmes auxiliaires (V.39) :

$$u^{k+1} \in \arg\min_{u\in U^f} G^k(u) + \langle \epsilon^k \nabla J(u^k) - \nabla G^k(u^k), u\rangle, \tag{V.39}$$

avec (G^k) une suite d'applications fortement convexes et différentiables, et ϵ^k une suite de réels positifs vérifiant certaines hypothèses de bornitude par rapport aux constantes de forte convexité et à la fonction J. Les avantages pratiques du PPA sont par exemple de donner une structure additive aux problèmes auxiliaires (V.39) par rapport aux composantes de u alors même que la fonction initiale J ne l'était pas forcément. Dans le cadre des problèmes stochastiques du type (I.1), le souci majeur est que l'on ne peut généralement pas évaluer précisément $\nabla J(u^k)$, pour peu que la fonction J soit trop compliquée, ou tout simplement du fait de sa nature fonctionnelle, comme cela a été abordé dans un cas particulier au chapitre IV. Nous allons donc proposer une version approchée du PPA, et montrer sa convergence.

Algorithme V.12. – *Étape 0 : choisir* $u^0 \in U^f$,
– *Étape* k *: résoudre*

$$\boldsymbol{u}^{k+1} \in \arg\min_{u\in U^f} G(u) - \langle \gamma^k \boldsymbol{s}_J^k + \nabla G(\boldsymbol{u}^k), u\rangle, \tag{V.40}$$

Si le nombre maximal d'itérations est atteint, stop, sinon, actualiser $\boldsymbol{s}_J^{k+1} \in U$ *en fonction de* $\boldsymbol{u}^{k+1}$*, et boucler.*

Dans cet algorithme, les éléments successifs $(\boldsymbol{s}_J^k)$ de l'espace U remplacent au signe près les gradients de J en les itérés courants. La question est maintenant de savoir ce que doivent satisfaire ces éléments pour que l'algorithme converge. Typiquement, des questions de sélection mesurable sur l'argmin vont se poser, etc.

V.4.2. Convergence. Nous donnons dans le théorème V.13 les hypothèses sous lesquelles l'algorithme V.12 converge.

Théorème V.13 (Principe du Problème Auxiliaire Stochastique). *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace de probabilité muni d'une filtration* $(\mathcal{F}^k)$.
(i) Supposons que $J : U \to \mathbb{R}$ *est convexe, et semicontinue inférieurement. Si de plus* J *est coercive sur* U^f *qui est un ensemble convexe fermé de* U*, alors le problème* (V.38) *admet des solutions, et l'on note* S *l'ensemble des solutions.*
(ii) Supposons de plus que $G : U \to \mathbb{R}$ *soit fortement convexe de module* $b > 0$*. Alors, les problèmes* (V.40) *ont une solution unique. Supposons que pour tout* $k \in \mathbb{N}$, $\boldsymbol{s}_J^k$ *soit une variable aléatoire* $\mathcal{F}^{k+1}$*-mesurable à valeurs dans* U*. Alors, par récurrence, pour tout* $k \in \mathbb{N}$, $\boldsymbol{u}^k$ *est une variable aléatoire dont on peut prendre une sélection* $\mathcal{F}^k$*-mesurable, à valeurs dans* U*, et l'algorithme* (V.12) *est bien défini.*

(iii) Supposons qu'il existe $A > 0$ et deux suites de réels positifs (ϵ^k, η^k) tels que :

$$\forall k \in \mathbb{N},\ \left\| \mathbb{E}\left(\boldsymbol{s}_J^k + \nabla J(\boldsymbol{u}^k) \middle| \mathcal{F}^k \right) \right\| \leq \eta^k \left(1 + \|\nabla J(\boldsymbol{u}^k)\| \right), \tag{V.41a}$$

$$\mathbb{E}\left(\|\boldsymbol{s}_J^k + \nabla J(\boldsymbol{u}^k)\|^2 \middle| \mathcal{F}^k \right) \leq A \left(1 + \frac{1}{\epsilon^k} \|\nabla J(\boldsymbol{u}^k)\|^2 \right), \tag{V.41b}$$

(iv) Supposons que les suites (ϵ^k, η^k) et (γ^k) décroissent vers 0 et soient telles que :

$$\forall k \in \mathbb{N},\ \gamma^k, \epsilon^k > 0, \quad \sum_{k\in\mathbb{N}} \gamma^k = +\infty, \quad \sum_{k\in\mathbb{N}} \frac{(\gamma^k)^2}{\epsilon^k} < +\infty, \quad \sum_{k\in\mathbb{N}} \gamma^k \eta^k < +\infty. \tag{V.42}$$

Supposons de plus que J a son gradient linéairement borné, i.e. qu'il existe deux réels $c, d > 0$, tels que pour tout $\in U$,

$$\|\nabla J(u)\| \leq c\|u\| + d, \tag{V.43}$$

alors la suite $(\boldsymbol{u}^k)$ générée par l'algorithme V.12 est telle que :

$$\lim_{k\to\infty} J(\boldsymbol{u}^k) = J(u^*),\ p.s.$$

avec $u^* \in S$. De plus, comme J est coercive, la suite $(\boldsymbol{u}^k)$ est bornée p.s. et admet des points d'accumulations dans la topologie faible, qui sont tous dans S.
(v) En outre, si J est fortement convexe de module $B > 0$, alors S se réduit à un singleton, et $(\boldsymbol{u}^k)$ converge presque sûrement fortement dans U vers l'unique solution de (V.38).

Preuve : On suit le schéma de preuve de [**32**]. L'idée de la preuve est de se donner une fonction de Lyapunov notée Λ et d'étudier sa variation entre deux itérations de l'algorithme V.12. On pose $\Lambda(u) = G(u^*) - G(u) - \langle \nabla G(u), u^* - u \rangle$. On peut dès lors remarquer que du fait de la forte convexité de G, de constante b, $\Lambda(u) \geq \frac{b}{2}\|u - u^*\|^2$. Dans la suite, on notera $\Lambda^k = \Lambda(\boldsymbol{u}^k)$. On étudie la suite (Λ^k).

$$\begin{aligned} \Lambda^{k+1} - \Lambda^k &= G(\boldsymbol{u}^k) - G(\boldsymbol{u}^{k+1}) - \langle \nabla G(\boldsymbol{u}^k), \boldsymbol{u}^k - \boldsymbol{u}^{k+1} \rangle \\ &\quad + \langle \nabla G(\boldsymbol{u}^k) - \nabla G(\boldsymbol{u}^{k+1}), u^* - \boldsymbol{u}^{k+1} \rangle, \qquad \text{(V.44)} \\ &\leq \langle \nabla G(\boldsymbol{u}^k) - \nabla G(\boldsymbol{u}^{k+1}), u^* - \boldsymbol{u}^{k+1} \rangle. \qquad \text{(V.45)} \end{aligned}$$

En effet, par convexité de G, le premier terme du deuxième membre de l'équation (V.44) est négatif. On poursuit maintenant les calculs. La solution $\boldsymbol{u}^{k+1}$ du problème (V.40) est caractérisée par l'inéquation variationnelle suivante :

$$\forall u \in U^f,\ \langle \nabla G(\boldsymbol{u}^{k+1}) - \nabla G(\boldsymbol{u}^k) - \gamma^k \boldsymbol{s}_J^k, u - \boldsymbol{u}^{k+1} \rangle \geq 0. \tag{V.46}$$

En appliquant cette caractérisation pour $u = u^*$, on obtient :

$$\begin{aligned} \langle \nabla G(\boldsymbol{u}^k) - \nabla G(\boldsymbol{u}^{k+1}), u^* - \boldsymbol{u}^{k+1} \rangle &\leq -\gamma^k \langle \boldsymbol{s}_J^k, u^* - \boldsymbol{u}^{k+1} \rangle, \\ &\leq \underbrace{-\gamma^k \langle \boldsymbol{s}_J^k, u^* - \boldsymbol{u}^k \rangle}_{B_1} + \underbrace{\gamma^k \langle -\boldsymbol{s}_J^k, \boldsymbol{u}^k - \boldsymbol{u}^{k+1} \rangle}_{B_2}. \qquad \text{(V.47)} \end{aligned}$$

La forte convexité de G donne :

$$b\|\boldsymbol{u}^{k+1} - \boldsymbol{u}^k\|^2 \leq \langle \nabla G(\boldsymbol{u}^{k+1}) - \nabla G(\boldsymbol{u}^k), \boldsymbol{u}^{k+1} - \boldsymbol{u}^k \rangle.$$

En combinant cette équation avec l'équation (V.46) appliquée à $\boldsymbol{u} = \boldsymbol{u}^k$, on obtient :

$$\begin{aligned} b\|\boldsymbol{u}^{k+1} - \boldsymbol{u}^k\|^2 &\leq -\gamma^k \langle \boldsymbol{s}_J^k, \boldsymbol{u}^k - \boldsymbol{u}^{k+1} \rangle, \\ &\leq \gamma^k \|\boldsymbol{s}_J^k\| \|\boldsymbol{u}^k - \boldsymbol{u}^{k+1}\|, \text{ par l'inégalité de Cauchy-Schwarz,} \qquad \text{(V.48)} \end{aligned}$$

ce qui finalement donne :

$$\|\boldsymbol{u}^{k+1} - \boldsymbol{u}^k\| \leq \frac{\gamma^k}{b} \|\boldsymbol{s}_J^k\|. \tag{V.49}$$

En revenant alors à l'équation (V.48), on obtient finalement la majoration de B_2 :

$$\begin{aligned} -\gamma^k \langle \boldsymbol{s}_J^k, \boldsymbol{u}^k - \boldsymbol{u}^{k+1} \rangle &\leq \frac{(\gamma^k)^2}{b} \|\boldsymbol{s}_J^k\|^2 \\ &\leq \frac{2(\gamma^k)^2}{b} \left(\|\boldsymbol{s}_J^k + \nabla J(\boldsymbol{u}^k)\|^2 + \|\nabla J(\boldsymbol{u}^k)\|^2 \right), \qquad \text{(V.50)} \end{aligned}$$

en utilisant l'inégalité usuelle $\|x+y\|^2 \leq 2(\|x\|^2+\|y\|^2)$. De plus, de par l'hypothèse de gradient linéairement borné sur J, il existe deux réels positifs c_1, c_2 tels que :

$$\|\nabla J(\boldsymbol{u}^k)\|^2 \leq c_1\|\boldsymbol{u}^k - u^*\|^2 + c_2.$$

D'où :

$$-\gamma^k\langle \boldsymbol{s}_J^k, \boldsymbol{u}^k - \boldsymbol{u}^{k+1}\rangle \leq \frac{2(\gamma^k)^2}{b}\left(\|\boldsymbol{s}_J^k + \nabla J(\boldsymbol{u}^k)\|^2 + c_1\|\boldsymbol{u}^k - u^*\|^2 + c_2\right). \tag{V.51}$$

Enfin, l'hypothèse de forte convexité sur G donne que, par définition de Λ :

$$\|\boldsymbol{u}^k - u^*\|^2 \leq \frac{2}{b}\Lambda^k.$$

ce qui, avec (V.51), amène finalement la majoration :

$$-\gamma^k\langle \boldsymbol{s}_J^k, \boldsymbol{u}^k - \boldsymbol{u}^{k+1}\rangle \leq \frac{2(\gamma^k)^2}{b}\left(\|\boldsymbol{s}_J^k + \nabla J(\boldsymbol{u}^k)\|^2 + c_1\frac{2}{b}\Lambda^k + c_2\right). \tag{V.52}$$

On s'occupe maintenant de B_1 (cf. (V.47)). Par convexité de J, il vient :

$$\begin{aligned} -\gamma^k\langle \boldsymbol{s}_J^k, u^* - \boldsymbol{u}^k\rangle &= \gamma^k\langle \nabla J(\boldsymbol{u}^k), u^* - \boldsymbol{u}^k\rangle - \gamma^k\langle \boldsymbol{s}_J^k + \nabla J(\boldsymbol{u}^k), u^* - \boldsymbol{u}^k\rangle, \\ &\leq \gamma^k\left(J(u^*) - J(\boldsymbol{u}^k)\right) - \gamma^k\langle \boldsymbol{s}_J^k + \nabla J(\boldsymbol{u}^k), u^* - \boldsymbol{u}^k\rangle, \end{aligned} \tag{V.53}$$

Finalement, en rassemblant (V.45), (V.47), (V.52), (V.53), on obtient :

$$\Lambda^{k+1} - \Lambda^k \leq \gamma^k\left(J(u^*) - J(\boldsymbol{u}^k)\right) - \gamma^k\langle \boldsymbol{s}_J^k + \nabla J(\boldsymbol{u}^k), u^* - \boldsymbol{u}^k\rangle + \frac{2(\gamma^k)^2}{b}\left(\|\boldsymbol{s}_J^k + \nabla J(\boldsymbol{u}^k)\|^2 + c_1\frac{2}{b}\Lambda^k + c_2\right). \tag{V.54}$$

On conditionne maintenant l'équation (V.54) par rapport à $\mathcal{F}^k$:

$$\begin{aligned} \mathbb{E}\left(\Lambda^{k+1} - \Lambda^k|\mathcal{F}^k\right) \leq{}& \gamma^k\left(J(u^*) - J(\boldsymbol{u}^k)\right) - \gamma^k\langle \mathbb{E}\left(\boldsymbol{s}_J^k + \nabla J(\boldsymbol{u}^k)|\mathcal{F}^k\right), u^* - \boldsymbol{u}^k\rangle \\ &+ \frac{2(\gamma^k)^2}{b}\left(\mathbb{E}\left(\|\boldsymbol{s}_J^k + \nabla J(\boldsymbol{u}^k)\|^2|\mathcal{F}^k\right) + c_1\frac{2}{b}\Lambda^k + c_2\right). \end{aligned} \tag{V.55}$$

En appliquant l'inégalité de Cauchy-Schwarz, et les hypothèses (V.41a) et (V.41b), on obtient donc :

$$\begin{aligned} \mathbb{E}\left(\Lambda^{k+1} - \Lambda^k|\mathcal{F}^k\right) \leq{}& \gamma^k\left(J(u^*) - J(\boldsymbol{u}^k)\right) + \gamma^k\eta^k(1 + \|\nabla J(\boldsymbol{u}^k)\|)\|u^* - \boldsymbol{u}^k\| \\ &+ \frac{2A(\gamma^k)^2}{b\epsilon^k}\|\nabla J(\boldsymbol{u}^k)\|^2 + \frac{2(\gamma^k)^2}{b}\left(A + c_2 + c_1\frac{2}{b}\Lambda^k\right) \end{aligned} \tag{V.56}$$

En utilisant alors les mêmes inégalités que précédemment pour travailler sur $\|\nabla J(\boldsymbol{u}^k)\|$ et son carré, il existe deux réels positifs c_3, c_4 tels que :

$$\|\nabla J(\boldsymbol{u}^k)\| \leq c_3\|\boldsymbol{u}^k - u^*\| + c_4 \leq c_3\|\boldsymbol{u}^k - u^*\|^2 + (c_4 + c_3),$$

ce qui finalement, dans (V.56) donne :

$$\begin{aligned} \mathbb{E}\left(\Lambda^{k+1} - \Lambda^k|\mathcal{F}^k\right) \leq{}& \gamma^k\left(J(u^*) - J(\boldsymbol{u}^k)\right) + \gamma^k\eta^k(1 + c_4) + \gamma^k\eta^k(1 + c_4 + c_3)\|\boldsymbol{u}^k - u^*\|^2 \\ &+ \frac{2(\gamma^k)^2}{b}\left(c_2 + A + \frac{c_2A}{\epsilon^k}\right) + \frac{2(\gamma^k)^2}{b}\left(\frac{2c_1}{b}\Lambda^k + \frac{Ac_1}{\epsilon^k}\|\boldsymbol{u}^k - u^*\|^2\right) \end{aligned} \tag{V.57}$$

L'équation (V.57) peut alors se remettre plus simplement sous la forme :

$$\mathbb{E}\left(\Lambda^{k+1} - \Lambda^k|\mathcal{F}^k\right) \leq \gamma^k\left(J(u^*) - J(\boldsymbol{u}^k)\right) + \alpha^k\Lambda^k + \beta^k, \tag{V.58}$$

avec $\alpha^k = \frac{2}{b}\left((1 + c_3 + c_4)\gamma^k\eta^k + \frac{2c_1}{b}(\gamma^k)^2 + \frac{2Ac_1}{b}\frac{(\gamma^k)^2}{\epsilon^k}\right)$, et $\beta^k = \gamma^k\eta^k(1 + c_4) + \frac{2(\gamma^k)^2}{b}\left(c_2 + A + \frac{c_2A}{\epsilon^k}\right)$, formant tous les deux les termes de séries sommables grâce aux hypothèses (V.42). Par optimalité, le premier terme du membre de droite dans (V.58) est négatif. En prenant l'espérance dans l'équation (V.58), et notant $\lambda^k = \mathbb{E}\left(\Lambda^k\right)$, on obtient :

$$\lambda^{k+1} - \lambda^k \leq \alpha^k\lambda^k + \beta^k, \tag{V.59}$$

ce qui prouve, par application du lemme A.33, que la suite (λ^k) est bornée par un réel $M > 0$.
On prouve maintenant que (Λ^k) est une quasi-martingale convergente (cf. Définition B.4) :

- (Λ^k) est par définition adaptée à $(\mathcal{F}^k)$.
- Par définition, $\Lambda^k \geq 0$ pour tout $k \in \mathbb{N}$, i.e., $\inf_{k\in\mathbb{N}} \mathbb{E}\left(\Lambda^k\right) > -\infty$.

– Posons $C_k := \{\mathbb{E}\left(\Lambda^{k+1} - \Lambda^k | \mathcal{F}^k\right) > 0\}$. Il est clair que 1_{C_k} est $\mathcal{F}^k$-mesurable. En utilisant l'inégalité (V.58), on obtient :

$$\begin{aligned}\sum_{k\in\mathbb{N}} \mathbb{E}\left(1_{C_k} \times (\Lambda^{k+1} - \Lambda^k)\right) &= \sum_{k\in\mathbb{N}} \mathbb{E}\left(1_{C_k} \times \mathbb{E}\left(\Lambda^{k+1} - \Lambda^k | \mathcal{F}^k\right)\right), \\ &\leq \sum_{k\in\mathbb{N}} \mathbb{E}\left(1_{C_k} \times (\alpha^k \Lambda^k + \beta^k)\right), \\ &\leq \sum_{k\in\mathbb{N}} (\alpha^k M + \beta^k), \\ &< +\infty,\end{aligned}$$

car les séries sont sommables.

– Il est également clair que $\sup_{k\in\mathbb{N}} \mathbb{E}\left((\Lambda^k)^-\right) < +\infty$, et par conséquent, en utilisant le théorème B.6, la suite de variables aléatoires (Λ^k) est une quasi-martingale et converge presque sûrement vers une variable aléatoire intégrable. Cette suite est donc presque sûrement bornée, et par définition, $(\boldsymbol{u}^k)$ et $(\nabla J(\boldsymbol{u}^k))$ sont donc deux suites presque sûrement bornées dans U.

On prouve maintenant que $(J(\boldsymbol{u}^k))$ converge presque sûrement vers $J(u^*)$. En revenant à l'équation (V.58) et prenant l'espérance, on obtient :

$$\gamma^k \mathbb{E}\left(J(\boldsymbol{u}^k) - J(u^*)\right) \leq \alpha^k \lambda^k + \beta^k + \lambda^k - \lambda^{k+1}.$$

On ajoute ces inégalités pour $k = 0, \ldots, n$:

$$\begin{aligned}\sum_{k=0}^{n} \gamma^k \mathbb{E}\left(J(\boldsymbol{u}^k) - J(u^*)\right) &\leq \lambda^0 - \lambda^{n+1} + \sum_{k=0}^{n} (\alpha^k M + \beta^k), \\ &\leq M + M \sum_{k=0}^{n} \alpha^k + \sum_{k=0}^{n} \beta^k. \qquad \text{(V.60)}\end{aligned}$$

En faisant alors $n \to \infty$ on trouve :

$$\sum_{k\in\mathbb{N}} \gamma^k \mathbb{E}\left(J(\boldsymbol{u}^k) - J(u^*)\right) < +\infty.$$

Par optimalité, chacun des termes de la série sous l'espérance est positif. Dès lors :

$$\sum_{k\in\mathbb{N}} \gamma^k \left(J(\boldsymbol{u}^k) - J(u^*)\right) < +\infty, \text{ p.s.} \qquad \text{(V.61)}$$

On va maintenant utiliser le lemme A.34. En effet, on a, par l'équation (V.49), et la bornitude de $\|s_J^k\|$, qu'il existe $\delta > 0$ tel que :

$$\|\boldsymbol{u}^{k+1} - \boldsymbol{u}^k\| \leq \delta \gamma^k,$$

et J est Lipschitz sur le borné contenant tous les itérés $(\boldsymbol{u}^k)$. D'où, par le lemme A.34 :

$$\lim_{k\to\infty} J(\boldsymbol{u}^k) = J(u^*) \qquad \text{(V.62)}$$

Soit $\bar{\boldsymbol{u}}$ un point d'accumulation de $(\boldsymbol{u}^k)$ dans la topologie faible. Il existe donc une sous-suite $(\boldsymbol{u}^{\phi(k)})$ qui converge faiblement vers $\bar{\boldsymbol{u}}$. Comme U^f est un sous-ensemble fermé, $\bar{\boldsymbol{u}} \in U^f$, et par semicontinuité inférieure de J, il vient :

$$J(\bar{\boldsymbol{u}}) \leq \lim_{k\to\infty} J(\boldsymbol{u}^{\phi(k)}) = J(u^*),$$

ainsi, $\bar{\boldsymbol{u}} \in S$.

Si maintenant J est fortement convexe de module $B > 0$, S est réduit alors à un singleton $\{u^*\}$. Par définition,

$$J(\boldsymbol{u}^k) - J(u^*) \geq \langle \nabla J(u^*), \boldsymbol{u}^k - u^* \rangle + \frac{B}{2} \|u^* - \boldsymbol{u}^k\|^2 \qquad \text{(V.63)}$$

Par optimalité $\langle \nabla J(u^*), \boldsymbol{u}^k - u^* \rangle \geq 0$. (V.63) donne donc la convergence forte dans U de $(\boldsymbol{u}^k)$ vers u^*, ce qui achève la preuve. □

Remarque V.14 (Noyaux changeant avec les itérations). *On pourrait s'intéresser à une version de l'algorithme V.12 dans laquelle les noyaux $G : U \to \mathbb{R}$ dépendraient des itérations, i.e. le sous-problème* (V.40) *deviendrait :*

$$\boldsymbol{u}^{k+1} \in \arg\min_{\boldsymbol{u}\in U^f} G^k(\boldsymbol{u}) - \langle \gamma^k \boldsymbol{s}_J^k + \nabla G^k(\boldsymbol{u}^k), u \rangle.$$

Le théorème de convergence V.13 resterait juste, comme un rapide coup d'œil à la preuve permet de s'en convaincre, à condition que les modules de forte convexité (b^k) *des noyaux* (G^k) *vérifient les conditions conjointes avec les suites* $\gamma^k, \epsilon^k, \eta^k$ *:*

$$\forall k \in \mathbb{N},\ \gamma^k, \epsilon^k > 0, \quad \sum_{k\in\mathbb{N}} \gamma^k = +\infty, \quad \sum_{k\in\mathbb{N}} \frac{(\gamma^k)^2}{\epsilon^k b^k} < +\infty, \quad \sum_{k\in\mathbb{N}} b_1 \frac{\gamma^k \eta^k}{b^k} < +\infty.$$

Il suffira donc de demander les hypothèses (V.42)*, et y ajouter l'hypothèse de bornitude suivante :* $\exists b > 0$ *tel que* $\forall k \in \mathbb{N},\ b_k > b$*.*

Remarque V.15 (Lien avec le chapitre IV). *Le théorème V.13 donne un autre point de vue sur les algorithmes stochastiques en dimension infinie proposés dans le chapitre IV. En effet, en prenant* $G(\cdot) = \frac{1}{2}\|\cdot\|^2$*, les sous-problèmes* (V.40) *peuvent être résolus explicitement, et l'on trouve :*

$$\boldsymbol{u}^{k+1} = \Pi_{Uf}\left(\boldsymbol{u}^k + \gamma^k \boldsymbol{s}_J^k\right),$$

ce qui correspond exactement aux algorithmes stochastiques qui y sont proposés, avec un choix adéquat de $\boldsymbol{s}_J^k$*, donné dans le cas* $U = L^2$*, et* J *différentiable par :*

$$\boldsymbol{s}_J^k(\cdot) = \nabla J(\boldsymbol{u}^k)(\boldsymbol{\xi}^{k+1}) \frac{1}{\epsilon^k} K\left(\frac{\cdot - \boldsymbol{\xi}^{k+1}}{\eta^k}\right).$$

De même, les hypothèses du théorème V.13 permettent de retrouver celles du théorème IV.5 ou du théorème IV.14. Plus précisément, les hypothèses du théorème V.13 sur la direction de descente bruitée $\boldsymbol{s}_J^k$ *s'interprètent d'une part comme une hypothèse sur l'erreur de convolution pour* (V.41a)*, et d'autre part comme une hypothèse sur la variance du noyau de convolution pour* (V.41b)*.*

Remarque V.16 (Preuve du théorème pour les commandes décentralisées). *A l'aide du théorème V.13, on peut montrer de manière immédiate le théorème V.9, en tenant le raisonnement suivant : Le point (i) est évident (cf. Appendice A). Pour tout* $k \in \{1, \dots, n\}$*, définissons* $\mathcal{F}^k = \sigma(\boldsymbol{\xi}_i^l,\ l \leq k,\ 1 \leq i \leq n)$*. Dans ces conditions, en posant*

$$\boldsymbol{s}_J^k = \begin{pmatrix} \vdots \\ -r_i^k \\ \vdots \end{pmatrix},$$

on vérifie bien que $\boldsymbol{s}_J^k$ *est* $\mathcal{F}^{k+1}$*-mesurable tandis que* u^k *est* $\mathcal{F}^k$*-mesurable. Enfin, de par la définition de l'espérance conditionnelle, et l'indépendance des tirages, il est clair que :*

$$\mathbb{E}\left(\boldsymbol{s}_J^k + \nabla J(u^k) | \mathcal{F}^k\right) = 0,$$

et donc que les hypothèses (V.41a)–(V.41b) *sont vérifiées, en prenant par exemple* $\eta^k = \epsilon^k = 1$*. On obtient donc clairement le théorème V.9 à partir du théorème V.13.*

De manière plus générale, l'intérêt du théorème V.13 et de l'algorithme V.12 est de donner une manière de faire disparaître le couplage probabiliste des outils de coordination, même si nous ne sommes pas en mesure à l'heure actuelle de donner une méthode pratique pour le faire ni d'autres applications de ces résultats.

V.5. Conclusion

En conclusion de ce chapitre, on peut pour commencer souligner une fois de plus la difficulté de donner un schéma de décomposition du type décomposition par les prix dans le cadre de l'optimisation stochastique, sans d'entrée de jeu éliminer des méthodes pourtant éprouvées de résolution des sous-problèmes, comme la programmation dynamique stochastique : ce lien entre la manière dont on décompose et la manière dont on résout les sous-problèmes (tout à fait absent dans le cas déterministe) rend le cas stochastique singulièrement difficile, et permet d'un certain côté de justifier la technique habituelle à laquelle on recourt face à un problème

stochastique de grande taille : discrétiser au plus vite l'aléa sous forme de scénarios plus ou moins arborescents pour revenir à un problème de grande taille déterministe sur lequel les algorithmes usuels s'appliqueront sans encombre.
Comme pour renforcer ce constat, la section V.3.1 de ce chapitre montre que le mariage de la décomposition par les prix avec la programmation dynamique dans le cas d'un problème linéaire quadratique se fait au prix de sous-problèmes en dimension 4, ce qui est déjà presque prohibitif du point de vue des temps de calcul. Néanmoins, les résultats de la section V.3.2 donnent une manière pratique et économique de résoudre des problèmes stochastiques lorsque l'on se restreint au cas de commandes décentralisées, faisant ainsi disparaître le couplage probabiliste.

Enfin, la section V.4 quant à elle ouvre une nouvelle voie : dans un cadre très général, on est capable de proposer un principe du problème auxiliaire stochastique, dans lequel la coordination (se réalisant d'habitude à travers le gradient de la fonction de coût à l'itéré courant) peut ne pas être exacte, et se réaliser uniquement de façon bruitée (dans certaines limites correspondant par exemple aux algorithmes développés dans le chapitre IV) sans pour autant nuire à la convergence du schéma de décomposition. Ce premier pas est encourageant pour aller vers une inversion du schéma usuel de pensée quand il s'agit de systèmes de grande taille : décomposer avant de discrétiser est envisageable grâce à ce résultat. Il serait dès lors possible de discrétiser certains sous-systèmes différemment des autres, et renvoyant ainsi une information de coordination bruitée, et malgré tout converger vers l'optimum. . .Ceci reste la principale piste d'explorations futures.

CHAPITRE VI

Fonctions non linéaires de l'espérance et boucle ouverte

VI.1. Résumé

Dans ce chapitre, une fois n'est pas coutume, nous nous intéressons aux problèmes en boucle ouverte. Néanmoins, nous ne nous intéressons pas exactement aux problèmes du type (I.1) avec un ensemble admissible ne comprenant que les fonctions constantes de l'aléa, mais aux problèmes dans lesquels le critère à optimiser est une fonction non linéaire de l'espérance. Ce type de problème apparaît en particulier, comme on le développe dans ce chapitre, dans le cas où, pour des raisons de non convexité du problème, ou pour améliorer un conditionnement, on recourt à une formulation avec Lagrangien augmenté d'un problème du type (I.1) en boucle ouverte comme celui-ci :

$$\min_{u \in U^f \subset \mathbb{R}^p} \mathbb{E}\left(j(u, \boldsymbol{\xi})\right)$$
$$\text{s.c } \Theta(u) \in C \subset \mathbb{R}^n.$$

En effet, le terme de Lagrangien augmenté s'écrit alors naturellement comme un terme non linéaire de la contrainte, et si la contrainte était une contrainte en probabilité ou en espérance, le critère est donc une fonction non linéaire de l'espérance.
Ce chapitre donne ensuite divers algorithmes pour résoudre ces problèmes. Selon la non linéarité de la fonction objectif, on argumente en faveur de telle ou telle approche. Parmi tous les algorithmes présentés, certains peuvent être étudiés en détail tandis que d'autres restent du domaine de l'heuristique : en section VI.3.1, nous présentons une méthode par estimation progressive, intuitive mais limitée sur le plan théorique, en section VI.3.2, nous proposons une méthode centrée sur la dualité de Fenchel, qui peut s'avérer intéressante lorsque le calcul de la conjuguée de la fonctionnelle non linéaire est possible, mais qui reste limitée théoriquement. En section VI.4, nous proposons sous des hypothèses de convexité et de croissance ou d'affinité une approche lagrangienne dite Arrow-Hurwicz stochastique, déclinée dans sa version Lagrangien simple et Lagrangien augmenté. En section VI.5, nous donnons un résumé des différentes méthodes et de leurs domaines d'application. Puis, pour clore le chapitre, nous passons en section VI.6 à une étude du cas de l'optimisation d'un critère en espérance sous des contraintes en espérance.
Le principal message de ce chapitre est donc de montrer une fois encore la richesse des approches du type gradient stochastique ou Arrow-Hurwicz stochastique pour traiter des problèmes d'optimisation stochastique, fussent-ils a priori peu propices à de tels traitements de par une non linéarité du critère.

VI.2. Introduction

VI.2.1. Motivation. Le but de ce chapitre est de décrire divers algorithmes stochastiques pour résoudre des problèmes du type :

$$\min_{u \in U^f \subset \mathbb{R}^p} F(u) := f\left(\mathbb{E}\left(\phi(u, \boldsymbol{\xi})\right)\right) + g(u), \tag{VI.1}$$

où $f : \mathbb{R}^n \to \mathbb{R}$ est une application non linéaire, convexe croissante $\phi : \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}^n$ et $g : \mathbb{R}^p \to \mathbb{R}$ sont des applications convexes données, U^f est un convexe fermé de $\mathbb{R}^p$, et $\boldsymbol{\xi}$ est une variable aléatoire à valeurs dans $\mathbb{R}^m$. Habituellement, le gradient stochastique est utilisé lorsque le gradient de la fonctionnelle que l'on cherche à minimiser est défini comme l'espérance d'une application. Ici, sous les hypothèses nécessaires de différentiabilité, et en notant par simplicité

$\nabla_u \phi(u,\xi)$ la jacobienne de $\phi(\cdot,\xi)$, le gradient de F, noté ∇F, s'écrit :

$$\forall u \in U^f \subset \mathbb{R}^p,\ \nabla F(u) = \mathbb{E}\left(\nabla_u \phi(u,\boldsymbol{\xi})\right)^T \nabla f\left(\mathbb{E}\left(\phi(u,\boldsymbol{\xi})\right)\right) + \nabla g(u). \tag{VI.2}$$

Ce gradient n'est donc pas une espérance, mais il en comporte deux. On ne peut pas appliquer directement un algorithme de gradient stochastique L'algorithme standard de gradient pour résoudre ce problème s'écrit :

$$u^{k+1} = \Pi_{Uf}\left(u^k - \epsilon^k \left\{\nabla g(u^k) + \mathbb{E}\left(\nabla_u \phi(u^k,\boldsymbol{\xi})\right)^T \nabla f\left(\mathbb{E}\left(\phi(u^k,\boldsymbol{\xi})\right)\right)\right\}\right). \tag{VI.3}$$

Un tel algorithme est coûteux en termes de calcul d'espérance, car il requiert à chaque itération le calcul complet de deux espérances en fonction de $\boldsymbol{\xi}$.

Nous allons dans ce chapitre regarder le cas de fonctionnelles non linéaires générales, et proposer des algorithmes de résolution de tels problèmes. Dans un premier temps, nous allons nous pencher sur le cas particulier où f est quadratique.

VI.2.2. f quadratique. Dans le cas particulier où $f(x) = \frac{1}{2}x^T A x$, avec A symétrique, l'algorithme (VI.3) se réécrit simplement :

$$\begin{aligned} u^{k+1} &= \Pi_{Uf}\left(u^k - \epsilon^k \left\{\nabla g(u^k) + \mathbb{E}\left(\nabla_u \phi(u^k,\boldsymbol{\xi})\right)^T A \mathbb{E}\left(\phi(u^k,\boldsymbol{\xi})\right)\right\}\right), \\ &= \Pi_{Uf}\left(u^k - \epsilon^k \left\{\nabla g(u^k) + \mathbb{E}\left(\nabla_u \phi(u^k,\boldsymbol{\xi}_2)^T A \phi(u^k,\boldsymbol{\xi}_1)\right)\right\}\right), \end{aligned} \tag{VI.4}$$

avec $\boldsymbol{\xi}_1$ et $\boldsymbol{\xi}_2$ indépendantes et identiquement distribuées, de même loi que $\boldsymbol{\xi}$. Le gradient s'écrivant comme une espérance, on peut donc proposer l'algorithme stochastique suivant, dont on est certain de la convergence sous des hypothèses standard de convexité ou forte convexité sur les fonctions impliquées (cf. Théorème C.5) :

$$\begin{aligned} &\text{Soient } \boldsymbol{\xi}_1^{k+1} \text{ et } \boldsymbol{\xi}_2^{k+1} \text{ i.i.d. indépendantes des v.a. passées,} \\ &\boldsymbol{u}^{k+1} = \Pi_{Uf}\left(\boldsymbol{u}^k - \epsilon^k \left\{\nabla g(\boldsymbol{u}^k) + \nabla_u \phi(\boldsymbol{u}^k,\boldsymbol{\xi}_2^{k+1})^T A \phi(\boldsymbol{u}^k,\boldsymbol{\xi}_1^{k+1})\right\}\right). \end{aligned} \tag{VI.5}$$

La présence de deux tirages dans l'algorithme (VI.5) peut provoquer une plus grande variance asymptotique et freiner la convergence de l'algorithme. À titre d'illustration, nous allons considérer le cas où ϕ est à valeurs réelles, convexe en u et strictement positive pour presque tout ξ, $g = 0$ et $U^f = \mathbb{R}^p$. On prend alors $A = 1$ et on obtient l'algorithme :

$$\begin{aligned} &\text{Soient } \boldsymbol{\xi}_1^{k+1} \text{ et } \boldsymbol{\xi}_2^{k+1} \text{ i.i.d. indépendantes des v.a. passées,} \\ &\boldsymbol{u}^{k+1} = \boldsymbol{u}^k - \epsilon^k \phi(\boldsymbol{u}^k,\boldsymbol{\xi}_1^{k+1}) \nabla_u \phi(\boldsymbol{u}^k,\boldsymbol{\xi}_2^{k+1}). \end{aligned} \tag{VI.6}$$

Cet algorithme permet donc la résolution du problème :

$$\min_{u\in\mathbb{R}^p} \frac{1}{2}\left(\mathbb{E}\left(\phi(u,\boldsymbol{\xi})\right)\right)^2, \tag{VI.7}$$

qui est équivalent sous nos hypothèses de positivité et de convexité, au problème simple :

$$\min_{u\in\mathbb{R}^p} \mathbb{E}\left(\phi(u,\boldsymbol{\xi})\right). \tag{VI.8}$$

Un algorithme de gradient stochastique standard pour le problème (VI.8) s'écrit :

$$\begin{aligned} &\text{Soit } \boldsymbol{\xi}_1^{k+1} \text{ indépendante des v.a. passées, et identiquement distribuée,} \\ &\boldsymbol{u}^{k+1} = \boldsymbol{u}^k - \epsilon^k \nabla_u \phi(\boldsymbol{u}^k,\boldsymbol{\xi}_1^{k+1}), \end{aligned} \tag{VI.9}$$

et mènera vers la même solution u^* que l'algorithme (VI.6). Disposant de deux algorithmes menant à la même solution, on souhaite désormais comparer leurs variances asymptotiques.
Le théorème central limite pour un algorithme de gradient stochastique (cf. théorème C.4) s'écrit de façon :

$$\frac{\boldsymbol{u}^k - u^*}{\sqrt{\epsilon^k}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,V),$$

avec V une matrice dépendant du problème, et (ϵ^k) suivant certaines conditions. La littérature sur les algorithmes stochastiques (par exemple [**14**]) enseigne que la forme des pas ϵ^k menant à la variance asymptotique minimale est Γ/k, avec $\Gamma \in \mathbb{R}^{p\times p}$ une matrice de taille $p \times p$. Dans ce cas, on connaît la forme de la matrice de covariance V^* du théorème central limite suivant :

$$\sqrt{k}\left(\boldsymbol{u}^k - u^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V^*),$$

Notons $H^* = \mathbb{E}\left(\nabla^2_{uu}\phi(u^*, \boldsymbol{\xi})\right)$, la hessienne du critère prise en l'optimum. On récapitule les résultats de comparaison dans le tableau 1. Ces résultats sont une application des calculs menés en C, après le théorème C.4. Ainsi, génériquement, par l'inégalité de Jensen, on sait que le

Problème (VI.8)	Problème (VI.7)
$\boldsymbol{u}^{k+1} = \boldsymbol{u}^k - \epsilon^k \nabla_u \phi(\boldsymbol{u}^k, \boldsymbol{\xi}^{k+1})$	$\boldsymbol{u}^{k+1} = \boldsymbol{u}^k - \epsilon^k \phi(\boldsymbol{u}^k, \boldsymbol{\xi}_1^{k+1}) \nabla_u \phi(\boldsymbol{u}^k, \boldsymbol{\xi}_2^{k+1})$
$V^*_{(\text{VI.8})} = (H^*)^{-1}\mathbb{E}\left(\nabla_u \phi(u^*, \boldsymbol{\xi}) \nabla_u \phi(u^*, \boldsymbol{\xi})^T\right)(H^*)^{-1}$	$V^*_{(\text{VI.7})} = \frac{\mathbb{E}\left(\phi(u^*, \boldsymbol{\xi})^2\right)}{\mathbb{E}(\phi(u^*, \boldsymbol{\xi}))^2} V^*_{(\text{VI.8})}$

Tab. 1. Théorème central limite sans et avec dédoublement de tirage

dédoublement de variable a un certain prix en terme de variance asymptotique, donné par le rapport $\mathbb{E}\left(\phi(u^*, \boldsymbol{\xi})^2\right) / \mathbb{E}\left(\phi(u^*, \boldsymbol{\xi})\right)^2 > 1$. Cependant, selon le critère considéré, ce prix est plus ou moins important, et le dédoublement peut s'avérer être une bonne solution.

Dans les sections suivantes, pour éviter de considérer les projections, nous supposerons que $U^f = \mathbb{R}^p$.

VI.3. Deux algorithmes voisins

VI.3.1. Approche par estimateur.

VI.3.1.1. *Algorithme.* Nous nous proposons ici de remplacer l'algorithme (VI.3) par l'algorithme stochastique suivant, dit par *estimateur*, évitant les calculs exhaustifs d'espérance, et fondé sur une idée simple d'estimation progressive :

Soit $\boldsymbol{\xi}^{k+1}$ indépendante des v.a. passées, et identiquement distribuée,

$$\boldsymbol{u}^{k+1} = \boldsymbol{u}^k - \epsilon^k \left(\nabla g(\boldsymbol{u}^k) + \nabla_u \phi(\boldsymbol{u}^k, \boldsymbol{\xi}^{k+1})^T \nabla f\left(\boldsymbol{z}^k\right)\right), \tag{VI.10a}$$

$$\boldsymbol{z}^{k+1} = \left(1 - \rho^k\right)\boldsymbol{z}^k + \rho^k \phi(\boldsymbol{u}^k, \boldsymbol{\xi}^{k+1}), \tag{VI.10b}$$

avec (ρ^k) et (ϵ^k) deux suites positives décroissantes. L'idée est donc très simple : elle consiste à estimer progressivement $\mathbb{E}\left(\phi(u^k, \boldsymbol{\xi})\right)$ au fil des itérations à l'aide de la formule (VI.10b), tout en espérant que la formule (VI.10a) ne fasse pas trop changer $\boldsymbol{u}^k$, selon les idées habituelles des approximations stochastiques.

VI.3.1.2. *Convergence.* Nous allons d'abord réécrire l'algorithme (VI.10) de façon plus conforme à la littérature des algorithmes stochastiques. En notant $h : \mathbb{R}^p \times \mathbb{R}^n \to \mathbb{R}^p \times \mathbb{R}$ l'application définie par :

$$\forall (u, z) \in \mathbb{R}^p \times \mathbb{R}^n,\ h(u, z) = \begin{pmatrix} \nabla g(u) + \mathbb{E}\left(\nabla_u \phi(u, \boldsymbol{\xi})\right)^T \nabla f(z) \\ z - \mathbb{E}\left(\phi(u, \boldsymbol{\xi})\right) \end{pmatrix}, \tag{VI.11}$$

l'algorithme (VI.10) se réécrit, en supposant que $\rho^k = \epsilon^k$:

(VI.12)

$$\begin{pmatrix} \boldsymbol{u}^{k+1} \\ \boldsymbol{z}^{k+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{u}^k \\ \boldsymbol{z}^k \end{pmatrix} + \rho^k \left(-h(\boldsymbol{u}^k, \boldsymbol{z}^k) + \underbrace{h(\boldsymbol{u}^k, \boldsymbol{z}^k) - \begin{pmatrix} \nabla g(\boldsymbol{u}^k) + \nabla_u \phi(\boldsymbol{u}^k, \boldsymbol{\xi}^{k+1})^T \nabla f(\boldsymbol{z}^k) \\ \boldsymbol{z}^k - \phi(\boldsymbol{u}^k, \boldsymbol{\xi}^{k+1}) \end{pmatrix}}_{\boldsymbol{w}^{k+1}} \right).$$

Dans cette écriture, $\boldsymbol{w}^{k+1}$ correspond au bruit qui est un incrément de martingale par rapport à la filtration des $\mathcal{F}^k := \sigma(\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^k)$, par indépendance des tirages successifs, tandis que $-h(\boldsymbol{u}^k, \boldsymbol{z}^k)$

correspond à la direction de descente déterministe de l'algorithme.
À supposer qu'il n'y ait pas de projection, et en utilisant la méthode de l'équation différentielle ordinaire associée (cf. [**45**]) si l'algorithme (VI.12) converge, c'est vers un point (u^*, z^*) tel que $h(u^*, z^*) = 0$, ce qui signifie :

$$\begin{aligned}\nabla g(u^*) + \mathbb{E}\left(\nabla_u \phi(u^*, \boldsymbol{\xi})\right)^T \nabla f(z^*) =& 0,\\ \mathbb{E}\left(\phi(u^*, \boldsymbol{\xi})\right) - z^* =& 0,\end{aligned}$$

ce qui implique donc que

$$\nabla g(u^*) + \mathbb{E}\left(\nabla_u \phi(u^*, \boldsymbol{\xi})\right)^T \nabla f(\mathbb{E}\left(\phi(u^*, \boldsymbol{\xi})\right)) = 0,$$

ce qui correspond à l'écriture des conditions d'optimalité sur le problème (VI.1). On peut donc être rassuré quant à la limite de l'algorithme. Reste à montrer que l'algorithme converge. Classiquement, les preuves de convergence d'algorithmes requièrent une propriété de monotonie de l'opérateur impliqué, ici l'opérateur h, même s'il est possible dans certains cas de recourir à des hypothèses plus faibles. Prenons donc $(u_1, u_2) \in \mathbb{R}^p \times \mathbb{R}^p$ et $(z_1, z_2) \in \mathbb{R}^n \times \mathbb{R}^n$. Le but est de voir si

$$H(u_1, u_2, z_1, z_2) := \langle h(u_1, z_1) - h(u_2, z_2), (u_1, z_1) - (u_2, z_2)\rangle_{\mathbb{R}^p \times \mathbb{R}^n} \geq 0.$$

En notant pour simplifier $\Phi(u) = \mathbb{E}\left(\phi(u, \boldsymbol{\xi})\right)$, et donc $\nabla\Phi(u) = \mathbb{E}\left(\nabla_u \phi(u, \boldsymbol{\xi})\right)$ pour tout $u \in \mathbb{R}^p$, on a :

$$\begin{aligned}H(u_1, u_2, z_1, z_2) =& \langle \nabla\Phi(u_1)^T \nabla f(z_1) - \nabla\Phi(u_2)^T \nabla f(z_2) + \nabla g(u_1) - \nabla g(u_2), u_1 - u_2\rangle_{\mathbb{R}^p}\\ &+ \langle z_1 - \Phi(u_1) - z_2 + \Phi(u_2), z_1 - z_2\rangle_{\mathbb{R}^n},\\ =& \langle \nabla f(z_1), \nabla\Phi(u_1)(u_1 - u_2)\rangle_{\mathbb{R}^n} - \langle \nabla f(z_2), \nabla\Phi(u_2)(u_1 - u_2)\rangle_{\mathbb{R}^n}\\ &+ \langle \nabla g(u_1) - \nabla g(u_2), u_1 - u_2\rangle_{\mathbb{R}^p} + \|z_1 - z_2\|^2_{\mathbb{R}^n} + \langle \Phi(u_2) - \Phi(u_1), z_1 - z_2\rangle_{\mathbb{R}^n}.\end{aligned}$$

Il apparaît donc difficile de montrer la monotonie de l'opérateur dans le cas général. En revanche, dans le cas où $\Phi(u) = Au + b$ et où $f(z) = \frac{1}{2}\|z\|^2_{\mathbb{R}^n}$, on observe que :

$$H(u_1, u_2, z_1, z_2) = \|z_1 - z_2\|^2_{\mathbb{R}^n} + \langle \nabla g(u_1) - \nabla g(u_2), u_1 - u_2\rangle_{\mathbb{R}^p},$$

ce qui en fait naturellement un opérateur monotone, de par la convexité de g. Mieux encore, dans ce cas, on est en mesure de montrer la convergence de l'algorithme (VI.10) vers la solution du problème :

$$\min_{u \in \mathbb{R}^p} \frac{1}{2}\|\mathbb{E}\left(A(\boldsymbol{\xi})u + b(\boldsymbol{\xi})\right)\|^2_{\mathbb{R}^n},$$

en y incluant même éventuellement un ensemble sur lequel projeter. En revanche, on est dans l'impossibilité de montrer la convergence de l'algorithme (VI.10) en toute généralité.

VI.3.2. Approche par dualité de Fenchel. On introduit ici des idées basées sur la conjugaison de Fenchel. Dans le cas où $f : \mathbb{R}^n \to \mathbb{R}$ est convexe, semicontinue inférieurement, elle est égale à sa biconjuguée de Fenchel notée f^{**} et définie par :

$$f^{**}(x) := \sup_{p \in \mathbb{R}^n} \left(p^T x - f^*(p)\right),$$

avec f^* la conjuguée de Fenchel de f, définie par $f^*(p) = \sup_{y \in \mathbb{R}^n} \left(p^T y - f(y)\right)$. Par construction, f^* est convexe (car enveloppe supérieure de fonctions affines), semicontinue inférieurement. De plus, f étant dans notre cas croissante, on a :

$$\forall x \in \mathbb{R}^n,\ \sup_{p \in \mathbb{R}^n} \left(p^T x - f^*(p)\right) = \sup_{p \in \mathbb{R}^n_+} \left(p^T x - f^*(p)\right).$$

En effet, la proposition 11.3 de [**84**] donne : $\partial f(\bar{x}) = \arg\max_{p \in \mathbb{R}^n} \left(p^T x - f^*(p)\right)$. Finalement, le problème (VI.1) se réécrit donc :

$$\min_{u \in \mathbb{R}^p} \sup_{p \in \mathbb{R}^n_+} \left(g(u) + p^T \mathbb{E}\left(\phi(u, \boldsymbol{\xi})\right) - f^*(p)\right), \tag{VI.13}$$

qui s'interprète donc comme un problème de *min-sup*, sur une fonction $g(u, p) = g(u) + p^T \mathbb{E}\left(\phi(u, \boldsymbol{\xi})\right) - f^*(p)$ concave en $p \in \mathbb{R}^n$ pour tout $u \in \mathbb{R}^p$, et convexe en $u \in \mathbb{R}^p$ pour tout

$p \in \mathbb{R}^n_+$. À supposer qu'un point-selle existe, on pourra donc appliquer un algorithme d'Arrow-Hurwicz stochastique défini par :

Soit $\boldsymbol{\xi}^{k+1}$ indépendante des v.a. passées, et identiquement distribuée,

$$\boldsymbol{u}^{k+1} = \boldsymbol{u}^k - \epsilon^k \left(\nabla g(\boldsymbol{u}^k) + \nabla_u \phi(\boldsymbol{u}^k, \boldsymbol{\xi}^{k+1})^T \boldsymbol{p}^k \right), \tag{VI.14a}$$

$$\boldsymbol{p}^{k+1} = \boldsymbol{p}^k + \rho^k \left(\phi(\boldsymbol{u}^k, \boldsymbol{\xi}^{k+1}) - \nabla f^*(\boldsymbol{p}^k) \right). \tag{VI.14b}$$

Cette approche présente un intérêt à condition que la dérivée de la conjuguée f^* soit explicitement connue. Dans notre cas, on sait par les mêmes arguments que plus haut, que pour tout $p \in \mathbb{R}^n_+$, $\partial f^*(p) = (\partial f)^{-1}(p)$.

Remarque VI.1 (f quadratique). *Dans le cas où $f(x) = \frac{x^T x}{2}$, on peut calculer $f^*(p) = \frac{p^T p}{2}$, et les algorithmes* (VI.14) *et* (VI.10) *sont exactement les mêmes.*

VI.4. Approche par dualité lagrangienne

VI.4.1. Lagrangien simple et Arrow-Hurwicz stochastique. Comme on l'a vu plus haut, le problème (VI.1) pour être bien posé, nécessite quelques hypothèses de convexité et/ou de croissance. Typiquement, nous allons demander soit

- que $f : \mathbb{R}^n \to \mathbb{R}$ soit convexe croissante, au sens où pour tous $x, y \in \mathbb{R}^n$, $x \geq y \Rightarrow f(x) \geq f(y)$, et $\phi(\cdot, \xi)$ $\mathbb{R}^n_+$-convexe pour tout ξ,
- ou que $f : \mathbb{R}^n \to \mathbb{R}$ soit convexe et $\phi(\cdot, \xi)$ affine pour tout ξ.

VI.4.1.1. *f convexe croissante.* Sous les hypothèses f convexe croissante et ϕ $\mathbb{R}^n_+$-convexe, nous pouvons réécrire le problème à l'aide d'une variable annexe notée z :

$$\min_{u,z} f(z) + g(u) \tag{VI.15a}$$

$$u \in U^f,\ z \in \mathbb{R}^n,$$

$$\mathbb{E}\left(\phi(u, \boldsymbol{\xi})\right) - z \leq 0. \tag{VI.15b}$$

Le problème (VI.15), de par la $\mathbb{R}^n_+$-convexité de $\phi(u) = \mathbb{E}\left(\phi(u, \boldsymbol{\xi})\right)$ et celle de f, est convexe, et il est équivalent au problème (VI.1), grâce à la croissance de f. En composant le Lagrangien du problème (VI), défini par

$$\forall u \in \mathbb{R}^p,\ \forall z, \lambda \in \mathbb{R}^n_+,\ L(u, z, \lambda) = f(z) + g(u) + \lambda \left(\mathbb{E}\left(\phi(u, \boldsymbol{\xi})\right) - z \right),$$

on propose alors l'algorithme de Arrow-Hurwicz suivant :

Soit $\boldsymbol{\xi}^{k+1}$ indépendante des v.a. passées, et identiquement distribuée

$$\boldsymbol{u}^{k+1} = \boldsymbol{u}^k - \epsilon^k \left(\nabla g(\boldsymbol{u}^k) + \nabla_u \phi(\boldsymbol{u}^k, \boldsymbol{\xi}^{k+1})^T \boldsymbol{\lambda}^k \right), \tag{VI.16a}$$

$$\boldsymbol{z}^{k+1} = \boldsymbol{z}^k - \epsilon^k \left(\nabla f(\boldsymbol{z}^k) - \boldsymbol{\lambda}^k \right), \tag{VI.16b}$$

$$\boldsymbol{\lambda}^{k+1} = \max\left(0, \boldsymbol{\lambda}^k + \rho^k \left(\phi(\boldsymbol{u}^{k+1}, \boldsymbol{\xi}^{k+1}) - \boldsymbol{z}^{k+1} \right) \right). \tag{VI.16c}$$

VI.4.1.2. *ϕ affine en u.* Dans le cas où f est convexe, et où il existe deux applications $\phi_1 : \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^p$ et $\phi_2 : \mathbb{R}^m \to \mathbb{R}^n$ telles que pour tout $\xi \in \mathbb{R}^m$, pour tout $u \in \mathbb{R}^p$, $\phi(u, \xi) = \phi_1(\xi)u + \phi_2(\xi)$, alors le problème (VI.1) peut être réécrit à l'aide d'une variable auxiliaire z sous la forme :

$$\min_{u,z} f(z) + g(u) \tag{VI.17}$$

$$u \in U^f,\ z \in \mathbb{R}^n,$$

$$\mathbb{E}\left(\phi_1(\boldsymbol{\xi})\right) u + \mathbb{E}\left(\phi_2(\boldsymbol{\xi})\right) - z = 0. \tag{VI.18}$$

Le problème (VI.17) est convexe. À nouveau, nous formons le Lagrangien défini par :

$$\forall u \in \mathbb{R}^p,\ \forall z, \lambda \in \mathbb{R}^n,\ L(u, z, \lambda) = f(z) + g(u) + \lambda \left(\mathbb{E}\left(\phi_1(\boldsymbol{\xi})\right) u + \mathbb{E}\left(\phi_2(\boldsymbol{\xi})\right) - z \right),$$

et l'on peut considérer l'algorithme de Arrow-Hurwicz suivant :

Soit $\boldsymbol{\xi}^{k+1}$ indépendante des v.a. passées, et identiquement distribuée,

$$\boldsymbol{u}^{k+1} = \boldsymbol{u}^k - \epsilon^k \left(\nabla g(\boldsymbol{u}^k) + \phi_1(\boldsymbol{\xi}^{k+1})^T \boldsymbol{\lambda}^k \right), \tag{VI.19a}$$

$$\boldsymbol{z}^{k+1} = \boldsymbol{z}^k - \epsilon^k \left(\nabla f(\boldsymbol{z}^k) - \boldsymbol{\lambda}^k \right), \tag{VI.19b}$$

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \rho^k \left(\phi_1(\boldsymbol{\xi}^{k+1}) \boldsymbol{u}^k + \phi_2(\boldsymbol{\xi}^{k+1}) - \boldsymbol{z}^{k+1} \right). \tag{VI.19c}$$

VI.4.1.3. *Convergence.* Dans le cas de ces deux problèmes convexes (VI.15) et (VI.17), on montre la convergence des algorithmes (VI.16) et (VI.19) à l'aide du théorème général sur l'optimisation sous contraintes en espérance se trouvant dans [**33**], et rappelé en annexe (cf. Théorème C.6).

VI.4.2. Lagrangien augmenté dans le cas non-convexe. Dans le cas où f n'est pas croissante, on peut toujours considérer le problème auxiliaire suivant :

$$\min_{u,z} f(z) + g(u) \tag{VI.20}$$

$$u \in U^f,\ z \in \mathbb{R}^n,$$

$$\mathbb{E}\left(\phi(u, \boldsymbol{\xi})\right) = z. \tag{VI.21}$$

Malheureusement, ce problème n'est convexe que si f est convexe et $\phi(\cdot, \xi)$ affine pour tout ξ. En toute généralité (par exemple $\phi(\cdot, \xi)$ convexe) le problème (VI.20) n'est pas convexe à cause de la contrainte (VI.21). Il existe cependant un moyen de traiter ce problème, ou du moins de minimiser l'impact de cette non-convexité, c'est de recourir au Lagrangien augmenté[1]. Dans le cas de contraintes d'égalité comme (VI.21), on définit le Lagrangien augmenté noté $L_c : \mathbb{R}^p \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, pour tout $c > 0$, par :

$$\forall (u, z, \lambda) \in \mathbb{R}^p \times \mathbb{R}^n \times \mathbb{R}^n,\ L_c(u, z, \lambda) = f(z) + g(u) + \lambda^T (\mathbb{E}\left(\phi(u, \boldsymbol{\xi})\right) - z) + \frac{c}{2} \| \mathbb{E}\left(\phi(u, \boldsymbol{\xi})\right) - z \|^2_{\mathbb{R}^n}.$$

Remarque VI.2. *Même si $\phi(\cdot, \xi)$ est affine, l'usage du Lagrangien augmenté à la place du Lagrangien simple est intéressant pour le meilleur conditionnement du problème qu'il apporte*[2]*.*

Posons $G : \mathbb{R}^p \times \mathbb{R}^n \to \mathbb{R}$ définie par $G(u, z) = f(z) + g(u)$, et $\gamma : \mathbb{R}^p \times \mathbb{R}^n \to \mathbb{R}$ définie par $\gamma(u, z) = \mathbb{E}\left(\phi(u, \boldsymbol{\xi})\right) - z$. On peut trouver dans [**22**] la proposition suivante (Prop. 12.2) :

Proposition VI.3. *Supposons que G et γ soient deux fois dérivables en une solution (u^*, z^*) de* (VI.20)*, que ∇G et $\nabla \gamma$ soient lipschitziennes dans un voisinage de (u^*, z^*), et que λ^* soit un multiplicateur associé à (u^*, z^*) tel que :*

$$\forall h \in \mathbb{R}^n_*,\ h^T \underbrace{\begin{pmatrix} \nabla^2 g(u) + (\lambda^*)^T \mathbb{E}\left(\nabla^2_{u,u} \phi(u^*, \boldsymbol{\xi})\right) & 0 \\ 0 & \nabla^2 f(z^*) \end{pmatrix}}_{H^*} h > 0, \tag{VI.22}$$

$$\begin{aligned} (\lambda^*)^T \mathbb{E}\left(\nabla_u \phi(u^*, \boldsymbol{\xi})\right) &= 0, \\ \nabla f(z^*) - \lambda^* &= 0, \\ \mathbb{E}\left(\phi(u^*, \boldsymbol{\xi})\right) - z^* &= 0. \end{aligned} \tag{VI.23}$$

Alors il existe un voisinage V de (u^*, z^*) et $\underline{c} > 0$ tels que pour tout $c > \underline{c}$, (u^*, z^*, λ^*) soit un point selle de L_c sur $V \times \mathbb{R}$. Plus précisément, on a pour tout $(u, z, \lambda) \in (V - \{u^*, z^*\}) \times \mathbb{R}^n$,

$$L_c(u^*, z^*, \lambda) \leq L_c(u^*, z^*, \lambda^*) < L_c(u, z, \lambda^*).$$

[1] Pour une présentation claire du Lagrangien augmenté, voir [**31**], chapitre 4.

[2] cf. [**31**], chapitre 4.

En prenant donc u^* solution du problème initial (VI.1), et en posant $z^* = \mathbb{E}\left(\phi(u^*, \boldsymbol{\xi})\right)$ et $\lambda^* = \nabla f(z^*)$, on se trouve bien dans les conditions de la Proposition VI.3. En effet, les conditions (VI.23) sont bien vérifiées, tandis que la condition (VI.22) se réécrit :

$$\forall h \in \mathbb{R}^n_*,\ h^T \begin{pmatrix} \nabla^2 g(u) + \nabla^2 (f \circ \Phi)(u^*)) & 0 \\ 0 & \nabla^2 f(z^*) \end{pmatrix} h > 0,$$

si l'on suppose de la convexité localement en u^* pour $f \circ \Phi$. Ceci nous permet donc d'affirmer que :

$$g(u^*) + f(\mathbb{E}\left(\phi(u^*, \boldsymbol{\xi})\right) = \min_{u \in U^f,\ z \in \mathbb{R}^n} \max_{\lambda \in \mathbb{R}^n} L_c(u, z, \lambda) = \max_{\lambda \in \mathbb{R}^n} \min_{u \in U^f,\ z \in \mathbb{R}^n} L_c(u, z, \lambda).$$

On a dès lors l'idée d'introduire un algorithme de gradient projeté pour résoudre le problème (VI.20) :

(VI.24a)

$$u^{k+1} = \Pi_{Uf}\left(u^k - \epsilon^k \left(\nabla g(u^k) + \mathbb{E}\left(\nabla_u \phi(u^k, \boldsymbol{\xi})\right)^T \lambda^k + c\mathbb{E}\left(\nabla_u \phi(u^k, \boldsymbol{\xi})\right)^T (\mathbb{E}\left(\phi(u^k, \boldsymbol{\xi})\right) - z^k)\right)\right),$$

(VI.24b)

$$z^{k+1} = z^k - \epsilon^k \left(\nabla f(z^k) - \lambda^k - c(\mathbb{E}\left(\phi(u^k, \boldsymbol{\xi})\right) - z^k)\right),$$

(VI.24c)

$$\lambda^{k+1} = \lambda^k + c\epsilon^k \left(\mathbb{E}\left(\phi(u^k, \boldsymbol{\xi})\right) - z^k\right)$$

La convergence de cet algorithme vers la solution du problème (VI.1) peut cependant continuer à poser des problèmes[3] en toute généralité. Elle est néanmoins aisée à prouver en utilisant par exemple le théorème IV.17, lorsque $\phi(\cdot, \xi)$ est affine.

Du point de vue des algorithmes stochastiques, la formulation (VI.24) est très satisfaisante. En effet, on a maintenant un gradient qui s'écrit comme une espérance, en utilisant la sous-section VI.2.2, i.e. le dédoublement de variables. On peut donc proposer l'algorithme stochastique suivant :

Algorithme VI.4.

Soient $\boldsymbol{\xi}_1^{k+1}, \boldsymbol{\xi}_2^{k+1}$ *i.i.d. indépendantes des v.a. passées,*

(VI.25a)

$$\boldsymbol{u}^{k+1} = \Pi_{Uf}\left(\boldsymbol{u}^k - \epsilon^k \left(\nabla g(\boldsymbol{u}^k) + \nabla_u \phi(\boldsymbol{u}^k, \boldsymbol{\xi}_2^{k+1})^T \boldsymbol{\lambda}^k + c\nabla_u \phi(\boldsymbol{u}^k, \boldsymbol{\xi}_2^{k+1})^T (\phi(\boldsymbol{u}^k, \boldsymbol{\xi}_1^{k+1}) - \boldsymbol{z}^k)\right)\right),$$

(VI.25b)

$$\boldsymbol{z}^{k+1} = \boldsymbol{z}^k - \epsilon^k \left(\nabla f(\boldsymbol{z}^k) - \boldsymbol{\lambda}^k - c(\phi(\boldsymbol{u}^k, \boldsymbol{\xi}_1^{k+1}) - \boldsymbol{z}^k)\right),$$

(VI.25c)

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + c\epsilon^k \left(\phi(\boldsymbol{u}^k, \boldsymbol{\xi}_1^{k+1}) - \boldsymbol{z}^k\right)$$

A nouveau, le théorème IV.17 montre la convergence de l'algorithme VI.4 lorsque $\phi(\cdot, \xi)$ est affine.

VI.5. Résumé – Typologie

Pour résumer ce survol des techniques possibles pour traiter le problème (VI.1), nous avons donc :

[3]Sous les hypothèses de la Proposition VI.3 et de convexité conjointe en (u, z) du Lagrangien augmenté, l'algorithme (VI.24a)–(VI.24b)–(VI.24c) va converger vers le point selle de L_c, et donc vers la solution du problème (VI.1). Dire que le Lagrangien augmenté est convexe en (u, z) n'est cependant pas très éloigné de supposer que $\phi(\cdot, \xi)$ est affine...

(1) **Si f est quadratique, et $\phi(\cdot,\xi)$ quelconque, telle que $f(\mathbb{E}(\phi(\cdot,\boldsymbol{\xi})))$ soit convexe**, on peut recourir au dédoublement des tirages donné dans l'algorithme (VI.5) ;

(2) **Si f est convexe croissante, et $\phi(\cdot,\xi)$ convexe**, on peut utiliser l'algorithme d'Arrow-Hurwicz stochastique (VI.16), impliquant une variable primale et une variable duale supplémentaires ;

(3) **Si f est convexe et $\phi(\cdot,\xi)$ affine**, on peut utiliser l'algorithme d'Arrow-Hurwicz stochastique (VI.19), impliquant une variable primale et une variable duale supplémentaires ;

(4) **Si f est convexe et $\phi(\cdot,\xi)$ quelconque**, on peut appliquer l'algorithme d'Arrow-Hurwicz sur Lagrangien augmenté (VI.4), impliquant une variable primale et une variable duale supplémentaires ;

On peut fournir pour tous ces algorithmes des preuves de convergence sous quelques hypothèses supplémentaires de différentiabilité et de propriété de Lipschitz des gradients impliqués, pour peu que les suites (ρ^k,ϵ^k) introduites soient en séries divergentes. Les preuves peuvent se faire notamment en recourant au théorème IV.17 général donné dans le chapitre IV Dans tous les cas, nous avons également proposé d'autres algorithmes :

(1) L'algorithme par estimateur (VI.10), impliquant une variable primale supplémentaire ;

(2) L'algorithme par dualité de Fenchel (VI.14), impliquant une variable duale supplémentaire.

Pour ces deux propositions, qui coïncident dans le cas où f est quadratique, on ne sait donner de preuve générale de convergence, hormis dans le cas où en plus, $\phi(\cdot,\xi)$ est affine.

Remarque VI.5 (f linéaire en certaines composantes). *Lorsque f est linéaire en certaines composantes, on peut laisser ces composantes telles quelles, sans effectuer sur celles-ci de traitement particulier, et la typologie faite ci-dessus reste valable, de façon plus souple :* **les hypothèses de croissance de f ou d'affinité de $\phi(\cdot,\xi)$ ne sont nécessaires que sur les composantes non linéaires de f.**

VI.6. Contraintes en espérance convexes

VI.6.1. Dualité, introduction d'un Lagrangien augmenté. Nous allons maintenant montrer comment les différents algorithmes proposés avant peuvent trouver leur place dans une application aux problèmes sous contrainte en espérance. Tout le travail des sections précédentes était motivé par le problème général suivant :

$$\min_u J(u) := \mathbb{E}(j(u,\boldsymbol{\xi})) \tag{VI.26}$$
$$\text{s.c.} \quad u \in U^f \subset \mathbb{R}^p,$$
$$\Theta(u) := \mathbb{E}(\theta(u,\boldsymbol{\xi})) \leq 0. \tag{VI.27}$$

On suppose dans la suite que $j : \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}$, et $\theta : \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}^d$. Pour traiter de tels problèmes lorsque typiquement J et Θ sont ($\mathbb{R}^d_+$-)convexes et vérifient certaines hypothèses de régularité, on peut recourir à l'algorithme de Arrow-Hurwicz proposé dans [**33**], et pour lequel on a une preuve de convergence (cf. Théorème C.6). On sait cependant qu'une façon généralement plus favorable de considérer de tels problèmes, même si Θ est $\mathbb{R}^d_+$-convexe, et de recourir au Lagrangien augmenté. Dans le cas du problème (VI.26), le Lagrangien augmenté L_c s'écrit pour tous $u \in \mathbb{R}^p, \lambda \in \mathbb{R}^d$,

$$L_c(u,\lambda) = \mathbb{E}(j(u,\boldsymbol{\xi})) - \frac{1}{2c}\|\lambda\|^2_{\mathbb{R}^d} + \frac{1}{2c}\|\max(0,\lambda + c\mathbb{E}(\theta(u,\boldsymbol{\xi})))\|^2_{\mathbb{R}^d}. \tag{VI.28}$$

Le problème qui est ensuite considéré est un problème de min-max, ou de point selle du Lagrangien augmenté sur $U^f \times \mathbb{R}^d_+$. Posons maintenant pour tout $c \in \mathbb{R}^+$, $f_c : \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ définie par :

$$\forall (x,y,\lambda) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d,\ f_c(x,y,\lambda) = x - \frac{1}{2c}\|\lambda\|^2_{\mathbb{R}^d} + \frac{1}{2c}\|\max(0,\lambda + cy)\|^2_{\mathbb{R}^d}.$$

Posons également $\phi : \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}^{1+d}$ définie par :

$$\forall (u,\xi) \in \mathbb{R}^p \times \mathbb{R}^m,\ \phi(u,\xi) = \begin{pmatrix} j(u,\xi) \\ \theta(u,\xi) \end{pmatrix}.$$

On a donc :

$$\forall (u,\lambda) \in \mathbb{R}^p \times \mathbb{R}^d,\ f_c(\mathbb{E}\left(\phi(u,\boldsymbol{\xi})\right),\lambda) = L_c(u,\lambda),$$

et on est ainsi ramené à un cas comparable au cas des sections précédentes. En effet, par dualité, on cherche à résoudre le problème suivant, équivalent au problème (VI.26) s'il y a un point selle :

$$\max_{\lambda\in\mathbb{R}^d} \min_{u\in U^f} f_c(\mathbb{E}\left(\phi(u,\boldsymbol{\xi})\right),\lambda). \tag{VI.29}$$

Nous allons maintenant étudier la solution de (VI.29).

VI.6.2. Travail à λ fixé. Soit $\lambda \in \mathbb{R}^d$ fixé. Nous regardons le problème suivant :

$$\min_{u\in U^f} f_c(\mathbb{E}\left(\phi(u,\boldsymbol{\xi})\right),\lambda) \tag{VI.30}$$

C'est bien entendu le terme $\frac{1}{2c}\|\max(0,\lambda + c\mathbb{E}\left(\theta(u,\boldsymbol{\xi})\right))\|^2$ qui est a priori gênant. La fonction $y \mapsto \frac{1}{2c}\|\max(0,\lambda+cy)\|^2$ est convexe et croissante au sens des composantes, et $u \mapsto \mathbb{E}\left(\theta(u,\boldsymbol{\xi})\right)$ est $\mathbb{R}^d_+$-convexe. f_c est de plus linéaire en sa première composante. Par conséquent, en vertu de la Remarque VI.5, nous allons pouvoir appliquer le cas précédent *f convexe croissante et $\phi(\cdot,\xi)$ convexe*. On introduit donc une variable annexe z et on écrit le problème suivant équivalent à (VI.30) :

$$\begin{aligned} &\min_{u,z} f_c(\mathbb{E}\left(j(u,\boldsymbol{\xi})\right),z,\lambda) \\ &\text{s.c.} \quad u \in U^f,\ z \in \mathbb{R}^d, \\ &\qquad \mathbb{E}\left(\theta(u,\boldsymbol{\xi})\right) - z \le 0. \end{aligned} \tag{VI.31}$$

(VI.32)

L'algorithme de Arrow-Hurwicz utilisé ensuite, qui est le pendant de (VI.16), s'écrit, avec une nouvelle variable duale notée $\pi \in \mathbb{R}^d_+$:

Soit $\boldsymbol{\xi}^{k+1}$ indépendante des v.a. passées, et identiquement distribuée,

$$\boldsymbol{u}^{k+1} = \Pi_{U^f}\left(\boldsymbol{u}^k - \epsilon^k\left(\nabla_u j(\boldsymbol{u}^k,\boldsymbol{\xi}^{k+1}) + \nabla_u\theta(\boldsymbol{u}^k,\boldsymbol{\xi}^{k+1})^T\boldsymbol{\pi}^k\right)\right) \tag{VI.33a}$$

$$\boldsymbol{z}^{k+1} = \boldsymbol{z}^k - \epsilon^k\left(\max(0,\boldsymbol{\lambda}^k + c\boldsymbol{z}^k) - \boldsymbol{\pi}^k\right) \tag{VI.33b}$$

$$\boldsymbol{\pi}^{k+1} = \max\left(0,\boldsymbol{\pi}^k + \rho^k\left(\theta(\boldsymbol{u}^{k+1},\boldsymbol{\xi}^{k+1}) - \boldsymbol{z}^{k+1}\right)\right). \tag{VI.33c}$$

Pour les raisons évoquées avant, on sait que cet algorithme va converger sous quelques hypothèses sur les suites positives (ϵ^k,ρ^k), et sur la différentiabilité et régularité des fonctions j et θ.

VI.6.3. Mélange des itérations de décomposition et de résolution interne. Bien entendu, lorsque l'on revient au problème de point-selle (VI.29), on voit bien que derrière la résolution de (VI.30) se cachent des itérations sur le multiplicateur initial λ. Typiquement, on aurait envie d'écrire une régle de mise à jour sur λ du type :

$$\lambda^{k+1} = \lambda^k + \frac{\rho^k}{c}\left(-\lambda^k + \max(0,\lambda^k + cz^{k+1})\right). \tag{VI.33d}$$

Ceci revient à vouloir concilier le principe du problème auxiliaire (cf. Annexe A.2) avec les algorithmes stochastiques pour les contraintes en espérance. Cela est rendu possible par [**32**] qui le fait pour les critères en espérance, et [**33**] qui le fait pour les contraintes en espérance. Ainsi, en rassemblant toutes les mise-à-jour (VI.33), on obtient un algorithme stochastique pour résoudre le problème (VI.26), dont on sait prouver la convergence.

VI.6.4. Convergence. On a le théorème suivant :

THÉORÈME VI.6. *Supposons que :*

(i) $j(\cdot,\xi)$ *est strictement convexe, localement lipschitzienne sur* U^f*, et à gradients linéairement bornés,*

(ii) $\theta(\cdot,\xi)$ *est lipschitzienne, régulièrement sous-différentiable, de gradient borné uniformément en* ξ *sur* U^f*, et* $\mathbb{E}\left(\theta(\cdot,\boldsymbol{\xi})\right)$ *est* $\mathbb{R}^d_+$*-convexe,*

(iii) les suites positives (ϵ^k) *et* (ρ^k) *vérifient*

$$\sum_{k\in\mathbb{N}}\epsilon^k=+\infty,\ \sum_{k\in\mathbb{N}}\rho^k=+\infty,\ \sum_{k\in\mathbb{N}}(\epsilon^k)^2<+\infty,\ \sum_{k\in\mathbb{N}}(\rho^k)^2<+\infty,$$

et (ϵ^k/ρ^k) *est monotone au sens large.*

Alors, $(\boldsymbol{u}^k,\boldsymbol{z}^k,\boldsymbol{\pi}^k,\boldsymbol{\lambda}^k)$ *est bornée, et* $(\boldsymbol{u}^k)$ *converge presque sûrement faiblement vers* u^* *solution de* (VI.26)*. Si de plus* $\mathbb{E}\left(j(\cdot,\boldsymbol{\xi})\right)$ *est fortement convexe, alors* $(\boldsymbol{u}^k)$ *converge presque sûrement fortement vers* u^* *l'unique solution de* (VI.26)*.*

Preuve : La preuve est une application immédiate du théorème principal de [**33**] et des techniques de [**32**]. □

VI.6.5. Application numérique. À titre d'exemple, nous allons appliquer l'algorithme (VI.33) au problème suivant :

$$\begin{aligned}\min_{u\in\mathbb{R}^2}\ &\frac{1}{2}u^T\mathbb{E}\left(Q(\boldsymbol{\xi})\right)u+u^T\mathbb{E}\left(p(\boldsymbol{\xi})\right)\\ \text{s.c. }&\mathbb{E}\left(C(\boldsymbol{\xi})\right)u-\mathbb{E}\left(b(\boldsymbol{\xi})\right)\leq 0\end{aligned}\tag{VI.34}$$

avec $Q(\xi)=\begin{pmatrix}\xi^2 & \xi\\ \xi & \xi^2\end{pmatrix}$, $p(\xi)=\begin{pmatrix}\xi\\ 1\end{pmatrix}$, $C(\xi)=\begin{pmatrix}-\xi^2 & \xi\\ -\xi & \xi^2\end{pmatrix}$, et $b(\xi)=\begin{pmatrix}-\xi^2\\ -\xi\end{pmatrix}$.

On résout exactement le problème (VI.34) avec $\boldsymbol{\xi}$ suivant une loi normale centrée réduite, ce qui conduit à la solution optimale $u^*=(1\quad -1)^T$. Après 1000 itérations de notre algorithme, on trouve la solution approchée $\tilde{u}=(0,999\quad -1,08)^T$, avec les pas $\epsilon^k=\rho^k=\frac{10}{10+k}$, et la constante du Lagrangien augmenté $c=10$. La figure 1 donne l'évolution le long des itérations de l'erreur sur la solution optimale en norme euclidienne.

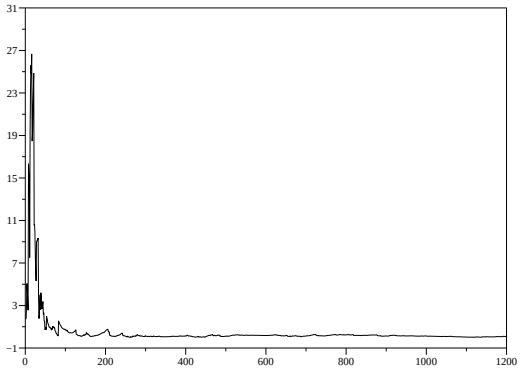


FIG. 1. Erreur sur la solution optimale

VI.7. Contraintes d'égalité en espérance non convexes

VI.7.1. Dualité et Lagrangien augmenté. On considère à nouveau le problème (VI.26), dans lequel on suppose que J est convexe, que Θ est quelconque, qu'elles sont toutes deux différentiables, mais que la contrainte est cette fois une contrainte d'égalité :

$$\min_u J(u) := \mathbb{E}\left(j(u,\boldsymbol{\xi})\right) \tag{VI.35}$$
$$\text{s.c.} \quad u \in U^f \subset \mathbb{R}^p,$$
$$\Theta(u) := \mathbb{E}\left(\theta(u,\boldsymbol{\xi})\right) = 0. \tag{VI.36}$$

On décide donc comme avant d'introduire le Lagrangien augmenté associé à ce problème pour lutter contre la non convexité de la contrainte :

$$\forall (u,\lambda) \in \mathbb{R}^p \times \mathbb{R}^d,\ L_c(u,\lambda) = \mathbb{E}\left(j(u,\boldsymbol{\xi})\right) + \lambda^T \mathbb{E}\left(\theta(u,\boldsymbol{\xi})\right) + \frac{c}{2}\|\mathbb{E}\left(\theta(u,\boldsymbol{\xi})\right)\|^2 \tag{VI.37}$$

La proposition VI.3 nous assure de l'existence d'un seuil $\underline{c} > 0$ tel que pour tout $c > \underline{c}$, le Lagrangien augmenté L_c défini par (VI.37) admet un point selle en la solution du problème initial. Par l'existence de ce point selle, on se retrouve donc à résoudre un problème de col, ce qui peut se faire par l'algorithme de Arrow-Hurwicz suivant :

Soient $\boldsymbol{\xi}_1^{k+1}, \boldsymbol{\xi}_2^{k+1}$ i.i.d. indépendantes des v.a. passées,

$$\boldsymbol{u}^{k+1} = \Pi_{U^f}\left(\boldsymbol{u}^k - \epsilon^k\left(\nabla_u j(\boldsymbol{u}^k, \boldsymbol{\xi}_1^{k+1}) + \nabla_u \theta(\boldsymbol{u}^k, \boldsymbol{\xi}_1^{k+1})^T\left(\boldsymbol{\lambda}^k + c\theta(\boldsymbol{u}^k, \boldsymbol{\xi}_2^{k+1})\right)\right)\right) \tag{VI.38a}$$
$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \rho^k \theta(\boldsymbol{u}^{k+1}, \boldsymbol{\xi}_2^{k+1}). \tag{VI.38b}$$

A nouveau, cet algorithme reste sur le plan heuristique car la non convexité du problème bloque nos outils habituels de preuve.

VI.7.2. Application numérique. Nous allons tester ce dernier algorithme sur un petit exemple non convexe :

$$\min_{u\in\mathbb{R}^2} \frac{1}{2} u^T \mathbb{E}\left(Q(\boldsymbol{\xi})\right) u + u^T \mathbb{E}\left(p(\boldsymbol{\xi})\right) \tag{VI.39}$$
$$\text{s.c.} \ \frac{1}{2} u^T \mathbb{E}\left(C(\boldsymbol{\xi})\right) u - u^T \mathbb{E}\left(b(\boldsymbol{\xi})\right) = 0$$

avec $Q(\xi) = \begin{pmatrix} \xi^2 & \xi \\ \xi & \xi^2 \end{pmatrix}$, $p(\xi) = \begin{pmatrix} \xi \\ 1 \end{pmatrix}$, $C(\xi) = \begin{pmatrix} -\xi^2 & \xi \\ -\xi & \xi^2 \end{pmatrix}$, et $b(\xi) = \begin{pmatrix} -\xi^2 \\ -\xi \end{pmatrix}$. Ce problème, en raison de la contrainte, est bien entendu non convexe. On choisit pour les tests numériques $\boldsymbol{\xi}$ selon une loi normale centrée réduite. On peut résoudre ce problème exactement en écrivant les conditions d'optimalité de Karush-Kuhn-Tucker. On compare ensuite ces résultats avec ceux calculés par l'algorithme (VI.38). On obtient, avec un bon choix du coefficient de Lagrangien augmenté, la courbe de convergence donnée par la Figure 2 (erreur quadratique sur la commande en fonction des itérations), ce qui est rassurant pour l'heuristique présentée ici.

VI.8. Contraintes d'inégalité en espérance non convexes

VI.8.1. Utilisation d'un Lagrangien augmenté. On considère à nouveau le problème (VI.26), dans lequel on suppose que J est convexe, mais que Θ n'est pas $\mathbb{R}^d_+$-convexe, et qu'elles sont toutes deux différentiables. On décide donc comme avant d'introduire le Lagrangien augmenté associé à ce problème pour lutter contre la non convexité de la contrainte. La proposition VI.3, qui existe également dans le cas de contraintes d'inégalité, nous assure de l'existence d'un seuil $\underline{c} > 0$ tel que pour tout $c > \underline{c}$, le Lagrangien augmenté L_c défini par (VI.28) admet un point selle en la solution du problème initial. Par l'existence de ce point selle, on se retrouve donc au problème (VI.29), que l'on va chercher à résoudre dans un premier temps à $\lambda \in \mathbb{R}^d$ fixé, comme en (VI.30).

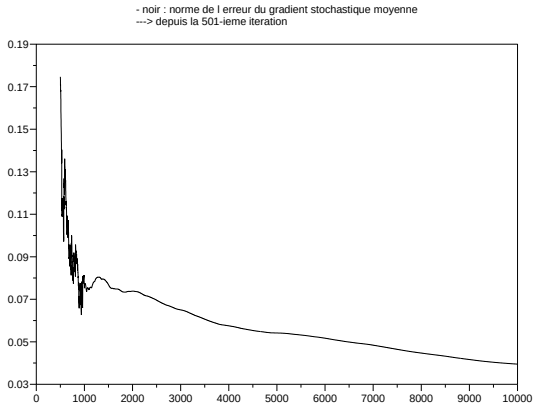


FIG. 2. Erreur quadratique sur la commande au long des itérations

A nouveau, on introduit une variable annexe z, et on considère le problème suivant, équivalent au problème (VI.30) :

$$\begin{aligned} &\min_{u,z} f_c(\mathbb{E}\left(j(u,\boldsymbol{\xi})\right), z, \lambda) \\ &\text{s.c.} \quad u \in U^f,\ z \in \mathbb{R}^d, \\ &\qquad \mathbb{E}\left(\theta(u,\boldsymbol{\xi})\right) - z = 0 \end{aligned} \tag{VI.40}$$

Le problème (VI.40) n'est pas plus convexe que le précédent, cependant, nous allons le traiter à nouveau par Lagrangien augmenté, toujours en vertu de la proposition VI.3, et selon les études préliminaires menées dans le cas non convexe. Nous introduisons donc une seconde constante de Lagrangien augmenté notée c', et nous devons résoudre, pour peu qu'il existe un point selle, le problème :

$$\begin{aligned} \max_{\lambda\in\mathbb{R}^d,\pi\in\mathbb{R}^d} \min_{u\in U^f, z\in\mathbb{R}^d} & \mathbb{E}\left(j(u,\boldsymbol{\xi})\right) - \frac{1}{2c}\|\lambda\|^2_{\mathbb{R}^d} + \frac{1}{2c}\|\max(0,\lambda+cz)\|^2_{\mathbb{R}^d} \\ & + \pi^T\left(\mathbb{E}\left(\theta(u,\boldsymbol{\xi})\right) - z\right) + \frac{c'}{2}\|\mathbb{E}\left(\theta(u,\boldsymbol{\xi})\right) - z\|^2 \end{aligned} \tag{VI.41}$$

VI.8.2. Algorithme. Dès lors, on peut penser en vertu de la section VI.4 à l'application d'un algorithme de Arrow-Hurwicz sur ces quatre variables, avec deux tirages de variables aléatoires à chaque itération pour tenir compte du terme quadratique, ce qui donne finalement les formules de mise-à-jour suivantes :

Soient $\boldsymbol{\xi}_1^{k+1}, \boldsymbol{\xi}_2^{k+1}$ i.i.d. indépendantes des v.a. passées,

(VI.42a)
$$\boldsymbol{u}^{k+1} = \Pi_{Uf}\left(\boldsymbol{u}^k - \epsilon^k\left(\nabla_u j(\boldsymbol{u}^k, \boldsymbol{\xi}_1^{k+1}) + \nabla_u\theta(\boldsymbol{u}^k, \boldsymbol{\xi}_1^{k+1})^T\left(\boldsymbol{\pi}^k + c'(\theta(\boldsymbol{u}^k, \boldsymbol{\xi}_2^{k+1}) - \boldsymbol{z}^k)\right)\right)\right)$$

(VI.42b)
$$\boldsymbol{z}^{k+1} = \boldsymbol{z}^k - \epsilon^k\left(\max(0, \boldsymbol{\lambda}^k + c\boldsymbol{z}^k) - \boldsymbol{\pi}^k - c'(\theta(\boldsymbol{u}^k, \boldsymbol{\xi}_2^{k+1}) - \boldsymbol{z}^k)\right)$$

(VI.42c)
$$\boldsymbol{\pi}^{k+1} = \boldsymbol{\pi}^k + c'\rho^k\left(\theta(\boldsymbol{u}^{k+1}, \boldsymbol{\xi}_2^{k+1}) - \boldsymbol{z}^{k+1}\right)$$

(VI.42d)
$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \frac{\rho^k}{c}\left(-\boldsymbol{\lambda}^k + \max(0, \boldsymbol{\lambda}^k + c\boldsymbol{z}^{k+1})\right).$$

Bien entendu, on ne dispose pas de preuve de convergence pour cet algorithme, faute de convexité et concavité sur le critère global du problème (VI.41). Il est donc purement heuristique et compte sur un paramètre de Lagrangien augmenté suffisamment élevé pour capter le saut de dualité.

VI.8.3. Contrainte en probabilité mollifiée. Un cas de contrainte en espérance non convexe, et même non-différentiable est le cas des contraintes en probabilité lorsque l'on souhaite les traiter par une approche stochastique plutôt qu'en considérant l'ensemble admissible défini par la contrainte comme un convexe que l'on chercherait à approcher par plans sécants. C'est ce que nous allons présenter maintenant sur un exemple :

$$\min_{u_1,u_2\in\mathbb{R}} -\mathbb{E}\left(f(x_0(1-u_1-u_2))+u_1x_0(1+b)+u_2x_0(1+\boldsymbol{\xi})\right)$$

$$\text{s.c.} \quad u_1,u_2\geq 0, \tag{VI.43a}$$

$$u_1+u_2\leq 1 \tag{VI.43b}$$

$$\mathbb{P}((1+b)u_1+(1+\boldsymbol{\xi})u_2\geq 1+l)\geq\beta \tag{VI.43c}$$

Ce problème représente un problème d'investissement : il s'agit de répartir son portefeuille entre un actif sans risque (commande u_1), un actif risqué (commande u_2), de sorte à maximiser son profit total sous une contrainte en probabilité sur le profit minimal. Ce problème a été étudié notamment dans [**2**]. Pour appliquer notre formalisme à ce problème, plusieurs remarques s'imposent :

- Les contraintes de positivité (VI.43a) seront traitées par projection, car elles définissent un convexe fermé.
- La contrainte d'inégalité (VI.43b) sera traitée par dualisation, ce qui nous donnera une variable duale supplémentaire, notée p.
- La contrainte en probabilité (VI.43c) s'écrit comme une contrainte en espérance, avec une fonction indicatrice :

$$\begin{aligned}\text{(VI.43c)} &\Leftrightarrow \beta-\mathbb{E}\left(1_{(1+b)u_1+(1+\boldsymbol{\xi})u_2-(1+l)\geq 0}\right)\leq 0,\\ &\Leftrightarrow \beta-\mathbb{E}\left(H((1+b)u_1+(1+\boldsymbol{\xi})u_2-(1+l))\right)\leq 0,\end{aligned} \tag{VI.44}$$

avec $H(y)=1_{y\geq 0}$. Nous allons approcher par convolution cette fonction indicatrice, par une fonction différentiable notée H_r, $r>0$, telle que $H_r\to H$ quand $r\to 0$. Nous aurons donc avec les notations des sections précédentes,

$$\theta_r(u,\xi)=\beta-H_r((1+b)u_1+(1+\xi)u_2-(1+l)),$$

et donc

$$\nabla_u\theta_r(u,\xi)=-H_r'((1+b)u_1+(1+\xi)u_2-(1+l))\begin{pmatrix}1+b\\1+\xi\end{pmatrix}.$$

Très intuitivement, les solutions de ce problème sont les suivantes :

- lorsque le niveau β de la contrainte (VI.43c) est proche de 1, le décideur est très averse au risque, et aura tendance à charger son portefeuille en actif sans risque (sous réserve bien sûr des paramètres choisis pour rémunérer l'actif risqué et l'actif sans risque) ;
- au contraire, lorsque le niveau β sera faible, le décideur acceptera le risque et aura tendance à charger l'actif risqué ;
- entre les deux niveaux, on assistera sans doute à un équilibrage du portefeuille défini par les paramètres des actifs.

VI.8.4. Application numérique. On considère donc le problème (VI.43), avec un paramètre de mollification donné par $r=0.01$, et la fonction H_r définie par :

$$H_r(y)=\frac{1}{1+e^{-\frac{y}{r}}},\ \text{et donc}\ H_r'(y)=\frac{e^{-\frac{y}{r}}}{r\left(1+e^{-\frac{y}{r}}\right)^2}.$$

On choisit pour la variable aléatoire $\boldsymbol{\xi}$ à valeurs réelles la fonction de répartition suivante :

$$F(\xi)=\begin{cases}0 & \text{si } \xi\leq\bar{\xi}-\sigma,\\ \frac{1}{16}\left(3\left(\frac{\xi-\bar{\xi}}{\sigma}\right)^5-10\left(\frac{\xi-\bar{\xi}}{\sigma}\right)^3+15\left(\frac{\xi-\bar{\xi}}{\sigma}\right)+8\right) & \text{si } \xi\in[\bar{\xi}-\sigma;\bar{\xi}+\sigma],\\ 1 & \text{sinon.}\end{cases} \tag{VI.45}$$

Pour simuler la variable aléatoire $\boldsymbol{\xi}$, comme cela est nécessaire pour la mise en œuvre de l'algorithme stochastique VI.42, nous utiliserons la méthode du rejet, aisée à implémenter du fait que le support de la loi de $\boldsymbol{\xi}$ est compact (il s'agit de $[\bar{\xi} - \sigma, \bar{\xi} + \sigma]$). On prendra de plus la fonction de satisfaction f suivante : $f(y) = -\frac{y^2}{2} + 2y$, et les paramètres $x_0 = 1$, $l = 0.15$, $b = 0.2$. On prendra également $\bar{\xi} = 0.4$ et $\sigma = 3$.
Pour l'application de l'algorithme VI.42, on choisit de prendre des grands pas ρ^k et ϵ^k du type $1/k^\alpha$, avec $\alpha \in (1/2, 1]$. Les figures 3 et 4 représentent l'évolution des contrôles u_1^k et u_2^k au long des itérations, dans respectivement le cas où $\beta = 0.4$ et $\beta = 0.9$, correspondant à des problèmes respectivement convexe et non convexe, comme l'indique [**2**].

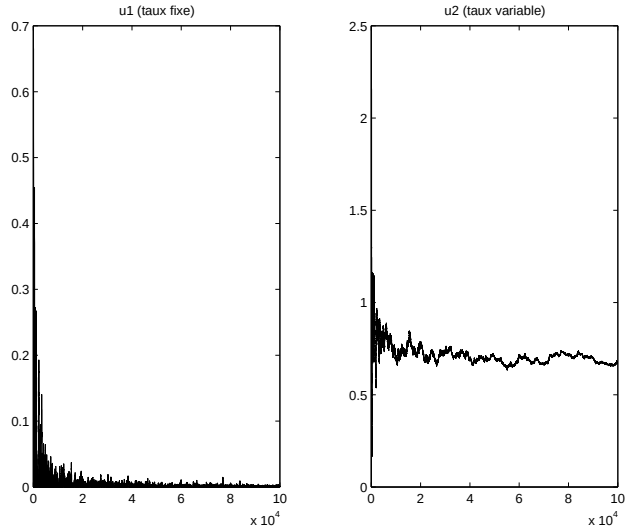


FIG. 3. Evolution des contrôles le long des itérations pour un niveau de probabilité de 0.4

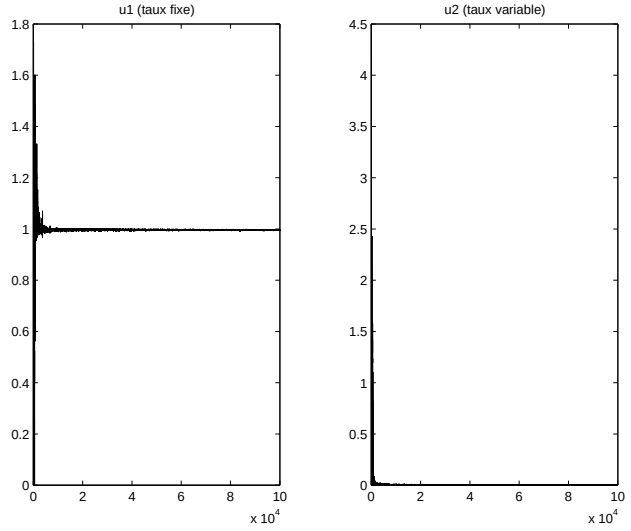


FIG. 4. Evolution des contrôles le long des itérations pour un niveau de probabilité de 0.9

Le tableau 2 donne quant à lui une synthèse des résultats obtenus dans ces deux cas. L'algorithme implémenté a donc convergé vers la solution optimale même dans un cas non-convexe, grâce notamment à l'utilisation de Lagrangiens augmentés.

Probabilité exigée β	Contrôle u_1	Contrôle u_2	Coût associé	Probabilité réalisée
0.4	0.00	0.6736	-1.54258	0.412
0.9	0.9938	0.0033	-1.20	1.

TAB. 2. Récapitulatif des résultats obtenus

VI.9. Conclusion et perspectives

On a présenté dans ce chapitre un panorama d'idées pouvant servir lors de la résolution de problèmes d'optimisation stochastique comportant un critère non linéaire en l'espérance. Cette non linéarité apparaît notamment lors du recours au Lagrangien augmenté pour traiter des problèmes contraints en espérance. Dans le cas de problèmes convexes, on a montré la convergence de divers algorithmes motivés principalement par l'algorithme de Arrow-Hurwicz stochastique que l'on peut trouver originellement notamment dans [**33**].
Des algorithmes sont également proposés pour les problèmes non-convexes, et l'on montre numériquement, sinon théoriquement, leur intérêt.

La prétention de ce chapitre est donc tout à fait humble. Avant tout, le but est de fixer sur le papier des idées diverses venant de la théorie de la dualité de Fenchel, de la théorie de la dualité de Lagrange, et du Lagrangien augmenté, et d'organiser ces idées en un faisceau de solutions pour traiter les problèmes d'optimisation stochastique en boucle ouverte sous contraintes en espérances (les contraintes en probabilité en sont un cas particulier intéressant).

Outre des expérimentations numériques toujours utiles, la perspective principale de ces travaux est selon nous de travailler aux preuves théoriques de convergence de nos algorithmes en relâchant l'hypothèse de convexité tout en préservant une certaine monotonie des opérateurs différentiels sous-jacents.

CHAPITRE VII

Conclusion

VII.1. Synthèse

VII.1.1. Résumé des contributions. Ce mémoire s'est attaché à étudier divers aspects de l'optimisation stochastique, allant des problèmes avec contraintes d'information (boucle fermée) aux problèmes à commandes constantes (boucle ouverte), d'un cadre théorique général pour la décomposition stochastique à des approches variationnelles d'un nouveau type pour la résolution pratique. Les principaux apports de cette thèse sont les suivants :

- On a prouvé que les seuls systèmes stochastiques en dimension 1 sans effet dual pour les commandes en boucle ouverte sont les systèmes linéaires, à changement de variable près.
- On a mis en évidence l'importance des contraintes de mesurabilité pour les problèmes à plusieurs pas de temps, en indiquant les limites des résultats de stabilité et des approches numériques arborescentes associées, et on a prouvé dans le cas des problèmes stochastiques *multistage* linéaires ou à critères séparés un résultat de stabilité par rapport à l'aléa mettant en lumière le rôle joué par les filtrations d'information dans ces problèmes.
- On a proposé une nouvelle approche variationnelle et les algorithmes stochastiques de descente associés pour résoudre les problèmes stochastiques convexes en boucle fermée. Cette famille d'algorithmes a d'ores et déjà donné des résultats numériques encourageants. De plus, elle ne nécessite aucune discrétisation a priori de l'aléa sous-jacent, permettant de traiter des problèmes difficiles sur la seule base de tirages indépendants des aléas.
- On a analysé les limites des approches usuelles de décomposition pour les problèmes stochastiques en boucle fermée, et expliqué les possibilités dans certains cas particuliers et impossibilités génériques de marier la décomposition avec la programmation dynamique stochastique. Puis, on a démontré un principe du problème auxiliaire stochastique, adapté au traitement des problèmes stochastiques en boucle fermée (et ainsi en dimension infinie).
- Enfin, on a donné une liste, des preuves et quelques illustrations des possibilités d'usage de lagrangiens augmentés et d'algorithmes stochastiques pour des problèmes d'optimisation stochastique en boucle ouverte non convexes comme ceux qui apparaissent lors du traitement de contraintes en probabilité.

VII.1.2. Situation. De façon plus philosophique, les contributions de cette thèse se situent à divers niveaux. La communauté de pensée dans laquelle s'inscrit cette thèse est à la fois celle de la programmation stochastique (*stochastic programming*), avec tous ses héritages liés à l'optimisation convexe et l'analyse variationnelle, et celle des algorithmes stochastiques, plus penchée vers des problématiques statistiques (problèmes d'estimation) ou de contrôle (problèmes de programmation dynamique).

Le travail présenté dans ce mémoire introduit d'une part dans la programmation stochastique une vision plus riche de l'aléa, à travers les recherches sur la stabilité et les problèmes d'information étudiés dans les chapitres II et III, mettant en lumière l'importance des structures d'information de l'aléa, et à travers l'introduction de méthodes de résolution ne nécessitant pas une discrétisation a priori de l'aléa sous forme arborescente, mais profitant de sa richesse dans une approche variationnelle d'un nouveau genre (cf chapitre IV), et permettant l'inversion du schéma habituel discrétisation-décomposition (cf. chapitre V). D'autre part, il introduit dans les problématiques d'estimation familières à l'approximation stochastique une vision plus variationnelle, dans la lignée des travaux [**59**], comme l'indique le chapitre IV. Enfin, il propose des pistes pour l'étude des problèmes de programmation dynamique stochastique de

grande taille que l'on souhaiterait décomposer, à travers des résultats de décentralisation ou d'agrégation exacte, comme cela est exposé dans le chapitre V.

Pour résumer, la contribution principale de ce mémoire est d'étudier plus profondément les problèmes d'optimisation stochastique en boucle fermée, et d'utiliser des techniques variationnelles pour résoudre des problèmes que l'on ne savait jusqu'à présent pas vraiment aborder sans les avoir d'emblée appauvris du point de vue de la représentation de l'aléa.

VII.2. Perspectives

Sur la base des travaux réalisés, de nombreux points restent à explorer. En reprenant les chapitres les uns après les autres, voici le tableau que nous pouvons dresser de possibles recherches futures :

- En ce qui concerne l'effet dual (chapitre II), le résultat que nous avons obtenu n'est valable qu'en dimension 1, et sous des hypothèses d'injectivité. Il resterait donc à déterminer quelles seraient en dimension supérieure les bonnes notions et hypothèses pour obtenir un résultat analogue, dont nous pouvons conjecturer qu'il est vrai. Les recherches de ce côté nous semblent cependant trop incertaines pour y consacrer à l'heure actuelle davantage de temps.
- Dans le chapitre III, les nouveaux résultats énoncés (cf. Théorèmes III.24 et III.27) ne sont pas valables pour des problèmes convexes généraux. Nous envisageons d'étendre nos résultats aux problèmes stochastiques *multistage* convexes dans les années qui viennent.
- Dans le chapitre IV, nous avons ouvert de nouvelles possibilités de résolution numérique de problèmes stochastique. Une voie d'utilisation intéressante de ces travaux est la résolution des problèmes de point-fixe fonctionnels apparaissant dans l'écriture des équations d'Hamilton-Jacobi-Bellman discrétisée en horizon infini. Ce genre d'équations apparaît typiquement dans la valorisation d'actifs financiers de type américain. Des études sont en cours chez EDF R&D pour développer convenablement notre méthodologie par noyaux dans ce cadre, en grande dimension (il s'agit notamment des travaux effectués en collaboration avec Pierre Girardeau, étudiant à l'Ecole Nationale Supérieure de Techniques Avancées, Jean-Sébastien Roy et Kengy Barty d'EDF R&D).
- Le chapitre V sur la décomposition stochastique peut aisément être marié avec les approches variationnelles du chapitre IV pour la résolution des sous-problèmes obtenus par usage du principe du problème auxiliaire stochastique (cf. Théorème V.13). Nous pensons dans les prochaines années étudier de près ces interactions, afin de proposer une méthode numérique efficace de décomposition. Une autre piste est également l'étude des problèmes en information décentralisée, dont la compatibilité avec la programmation dynamique stochastique est assez séduisante.
- Enfin, la motivation principale du chapitre VI était l'étude par lagrangien augmenté et algorithmes stochastiques de problèmes sous contraintes en probabilités. Cette approche qui rend le problème non-convexe ne nous semble pas à terme la mieux adaptée. Afin d'étudier de plus près les contraintes en probabilité, je travaille actuellement avec René Henrion sur ce sujet, et nous avons d'ores et déjà obtenu d'intéressants résultats de convexité. Ces résultats, pour des questions de temps, ne sont pas présentés dans cette thèse et feront l'objet d'une publication prochaine.

ANNEXE A

Optimisation

A.1. Analyse convexe

A.1.1. Ensembles convexes et projections.

DÉFINITION A.1 (Ensemble convexe). *Soit U un espace de Hilbert muni du produit scalaire noté $\langle \cdot, \cdot \rangle_U$. Soit $V \subset U$.*

(i) V est convexe *si et seulement si*

$$\forall v_1, v_2 \in V,\ \forall \alpha \in [0,1],\ \alpha v_1 + (1-\alpha) v_2 \in V.$$

(ii) V est strictement convexe *si et seulement si*

$$\forall v_1, v_2 \in V,\ \forall \alpha \in]0,1[,\ \alpha v_1 + (1-\alpha) v_2 \in \operatorname{int} V.$$

On définit maintenant la projection sur un sous-espace vectoriel fermé.

PROPOSITION A.2. *Soit U un espace de Hilbert muni du produit scalaire noté $\langle \cdot, \cdot \rangle_U$. Soit $V \subset U$ un sous-espace vectoriel fermé. Pour tout $u \in U$, le problème $\min_{v \in V} \frac{1}{2}\|u - v\|_U^2$ est bien défini, et il existe un unique minimiseur noté $\Pi_V(u)$, et appelé la projection de u sur V.*

L'opérateur Π_V défini ci-dessus possède de plus quelques propriétés intéressantes :

PROPOSITION A.3. *Soit U un espace de Hilbert muni du produit scalaire noté $\langle \cdot, \cdot \rangle_U$. Soit $V \subset U$ un sous-espace vectoriel fermé. L'opérateur de projection sur V noté $\Pi_V : U \to U$ est linéaire et autoadjoint.*

On définit maintenant la projection sur un convexe fermé.

PROPOSITION A.4. *Soit U un espace de Hilbert muni du produit scalaire noté $\langle \cdot, \cdot \rangle_U$. Soit $V \subset U$ un sous-ensemble convexe fermé. Pour tout $u \in U$, le problème $\min_{v \in V} \frac{1}{2}\|u - v\|_U^2$ est bien défini, et il existe un unique minimiseur noté $\Pi_V(u)$, et appelé la projection de u sur V.*

Preuve : cf. [**60**], Section III.3. □

On a alors les propriétés suivantes :

PROPOSITION A.5. *Soit U un espace de Hilbert muni du produit scalaire noté $\langle \cdot, \cdot \rangle_U$. Soit $V \subset U$ un sous-ensemble convexe fermé.*

(i) Soit $u \in U$. $v_u \in V$ est la projection $\Pi_V(u)$ si et seulement si

$$\forall v \in V,\ \langle u - v_u, v - v_u \rangle_U \leq 0.$$

(ii) Pour tous $u_1, u_2 \in U$, on a :

$$\|\Pi_V(u_1) - \Pi_V(u_2)\|^2 \leq \langle \Pi_V(u_1) - \Pi_V(u_2), u_1 - u_2 \rangle_U,$$

ce qui implique que Π_V est monotone et lipschitzienne.

Enfin, on a des propriétés particulières dans le cas de la projection sur un cône, le cône étant, comme le fait remarquer [**60**], un intermédiaire entre le convexe fermé et le sous-espace vectoriel.

DÉFINITION A.6 (Cône polaire). *Soit $K \subset U$ un cône convexe. Le cône polaire de K noté K^o est défini par :*

$$K^o = \{u \in U\ :\ \forall v \in K,\ \langle u, v \rangle_U \leq 0\}.$$

PROPOSITION A.7. *Soit $K \subset U$ un cône convexe fermé. Soit $u \in U$. Alors, $u_K = \Pi_K(u)$ si et seulement si :*

$$u_K \in K, \quad u - u_K \in K^o, \quad \langle u - u_K, u_K \rangle_U = 0.$$

Preuve : cf. [**60**], Section III.3. □

LEMME A.8 (Sous-espace vectoriel). *Soit V_1 un sous-espace vectoriel fermé de U et V_2 un convexe fermé de U. On note alors $V = V_1 \cap V_2$ qui est un convexe fermé. Si de plus $\Pi_{V_2}(V_1) \subset V_1$, alors les opérateurs de projection vérifient la règle de composition suivante :*

$$\Pi_V = \Pi_{V_2} \circ \Pi_{V_1}.$$

Preuve : On utilise pour ce faire la caractérisation (i) de la projection dans la proposition A.5. Soit $u \in U$, et $v \in V$. On a déjà que $\Pi_{V_2}(\Pi_{V_1}(u)) \in V_1 \cap V_2$. De plus :

$$\begin{aligned} \langle u - \Pi_{V_2}(\Pi_{V_1}(u)), v - \Pi_{V_2}(\Pi_{V_1}(u)) \rangle_U &= \langle u - \Pi_{V_1}(u), v - \Pi_{V_2}(\Pi_{V_1}(u)) \rangle_U \\ &\quad + \langle \Pi_{V_1}(u) - \Pi_{V_2}(\Pi_{V_1}(u)), v - \Pi_{V_2}(\Pi_{V_1}(u)) \rangle_U, \\ &\leq \langle u - \Pi_{V_1}(u), v - \Pi_{V_2}(\Pi_{V_1}(u)) \rangle_U, \end{aligned} \tag{A.1}$$

car par la caractérisation (i) de la proposition A.5 appliquée à la projection de $\Pi_{V_1}(u)$ sur V_2, le second terme de la somme du membre de droite est négatif, du fait que $v \in V_2$. En revenant à l'équation (A.1), on obtient, de par la linéarité et le caractère auto-adjoint de Π_{V_1} et le fait que $v \in V_1$, et $\Pi_{V_2}(\Pi_{V_1}(u)) \in V_1$:

$$\begin{aligned} \langle u - \Pi_{V_1}(u), v - \Pi_{V_2}(\Pi_{V_1}(u)) \rangle_U &\leq \langle u - \Pi_{V_1}(u), \Pi_{V_1}\left(v - \Pi_{V_2}(\Pi_{V_1}(u))\right) \rangle_U, \\ &= \langle \Pi_{V_1}\left(u - \Pi_{V_1}(u)\right), v - \Pi_{V_2}(\Pi_{V_1}(u)), \text{ car } \Pi_{V_1} \text{ est autoadjoint}, \\ &= 0, \end{aligned}$$

ce qui achève de caractériser la projection sur V. □

REMARQUE A.9 (Sans condition d'inclusion). *On peut remarquer que la condition d'inclusion $\Pi_{V_2}(V_1) \subset V_1$ dans le lemme A.8 est très importante. En effet, considérons dans $\mathbb{R}^2$, le sous-espace vectoriel $V_1 = \left\{(x, y) \in \mathbb{R}^2 : y = 0\right\}$ et le convexe fermé $V_2 = \left\{(x, y) \in \mathbb{R}^2 : x^2 + (y - 1/2)^2 \leq 1\right\}$ (i.e. V_2 est la boule centrée en $(0, 1/2)$, de rayon 1). Bien entendu, ici, $\Pi_{V_2}(V_1) \nsubseteq V_1$. En prenant alors le point $v = (-3, 1/2)$, on constate en effet que $\Pi_{V_1 \cap V_2}(v) \neq \Pi_{V_2}\left(\Pi_{V_1}(v)\right)$, comme l'indique notamment la figure 1, sur laquelle v_1 désigne $\Pi_{V_2}(\Pi_{V_1}(v))$ et v^* désigne le projeté sur l'intersection.*

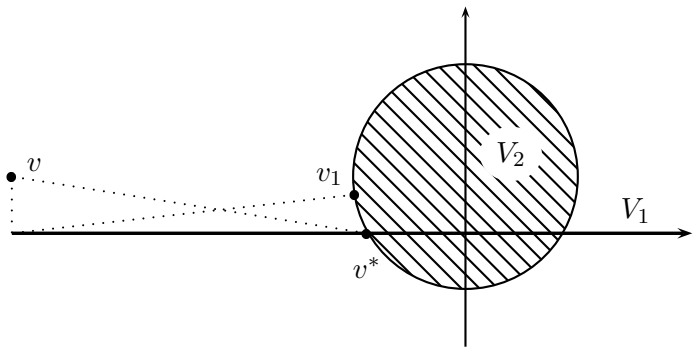


FIG. 1. Cas où la condition d'inclusion sur les projecteurs n'est pas vérifiée

LEMME A.10 (Cône convexe). *Soit V_1 un cône convexe fermé de U et V_2 un convexe fermé de U. On note alors $V = V_1 \cap V_2$ qui est un convexe fermé. Si de plus $(Id - \Pi_{V_2})(V_1) \subset V_1$, et si $\Pi_{V_2}(V_1) \subset V_1$, alors les opérateurs de projection vérifient la règle de composition suivante :*

$$\Pi_V = \Pi_{V_2} \circ \Pi_{V_1}.$$

Preuve : On utilise pour ce faire la caractérisation (i) de la projection dans la proposition A.5. Soit $u \in U$, et $v \in V$. On a déjà que $\Pi_{V_2}(\Pi_{V_1}(u)) \in V_1 \cap V_2$. De plus :

$$\begin{aligned}\langle u - \Pi_{V_2}(\Pi_{V_1}(u)), v - \Pi_{V_2}(\Pi_{V_1}(u))\rangle_U &= \langle u - \Pi_{V_1}(u), v - \Pi_{V_2}(\Pi_{V_1}(u))\rangle_U \\ &\quad + \langle \Pi_{V_1}(u) - \Pi_{V_2}(\Pi_{V_1}(u)), v - \Pi_{V_2}(\Pi_{V_1}(u))\rangle_U, \\ &\leq \langle u - \Pi_{V_1}(u), v - \Pi_{V_2}(\Pi_{V_1}(u))\rangle_U, \qquad \text{(A.2)}\end{aligned}$$

car par la caractérisation (i) de la proposition A.5 appliquée à la projection de $\Pi_{V_1}(u)$ sur V_2, le second terme de la somme du membre de droite est négatif, du fait que $v \in V_2$. On s'intéresse maintenant à l'autre terme :

$$\begin{aligned}\langle u - \Pi_{V_1}(u), v - \Pi_{V_2}(\Pi_{V_1}(u))\rangle_U &= \langle u - \Pi_{V_1}(u), v - \Pi_{V_1}(u)\rangle_U + \langle u - \Pi_{V_1}(u), \Pi_{V_1}(u) - \Pi_{V_2}(\Pi_{V_1}(u))\rangle_U \\ &\leq \langle \underbrace{u - \Pi_{V_1}(u)}_{\in V_1^o}, \underbrace{(Id - \Pi_{V_2})(\Pi_{v_1}(u))}_{\in V_1}\rangle_U, \qquad \text{(A.3)} \\ &\leq 0, \text{ comme } V_1 \text{ est un cône.} \qquad \text{(A.4)}\end{aligned}$$

La première inégalité s'obtient par caractérisation de la projection sur le convexe fermé V_1, et (A.3) achève la preuve. □

Remarque A.11 (Sans la deuxième condition d'inclusion). *La condition d'inclusion $(Id - \Pi_{V_2})(V_1) \subset V_1$ dans le lemme A.10 apparaît même comme nécessaire. Voici en effet un exemple. Considérons dans $\mathbb{R}^2$ le cas où $V_1 = \{(x, y) \in \mathbb{R}^2 \ : \ x \geq 0,\ y \geq 0\}$, i.e. V_1 est l'orthant positif, tandis que $V_2 = \{(x, y) \in [0, 2]^2 \ : \ y = 2 - x\}$. On a bien que $\Pi_{V_2}(V_1) \subset V_1$, en revanche, on n'a pas que $(Id - \Pi_{V_2})(V_1) \subset V_1$: en prenant par exemple le point $(1/2, 0)$, on vérifie que l'inclusion est fausse. Considérons maintenant le point $v = (-1, 1)$. On note $v_1 = \Pi_{V_2}(\Pi_{V_1}(v))$, et $v^* = \Pi_{V_1 \cap V_2}(v)$. Il est clair que $v_1 \neq v^*$, comme le montre la figure 2, ce qui achève de mettre en lumière le caractère suffisant de l'hypothèse d'inclusion sur $Id - \Pi_{V_2}$.*

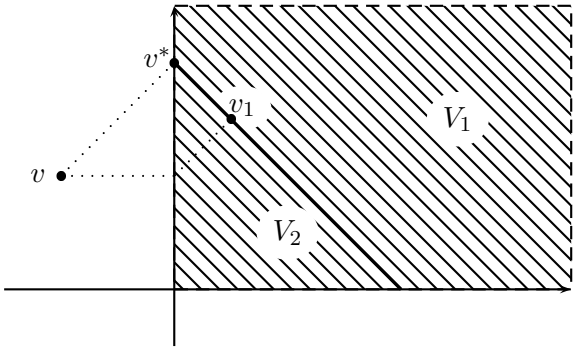


Fig. 2. Cas où la deuxième condition d'inclusion sur les projecteurs n'est pas vérifiée

Remarque A.12 (Sans cône). *Enfin, même lorsque les deux conditions d'inclusion sont vérifiées, mais que V_2 et V_1 sont des convexes fermés quelconques, la propriété de composition n'est pas vérifiée. En effet, prenons dans $\mathbb{R}^2$, $V_1 = [0, 1]^2$, et $V_2 = \{(x, y) \in [0, 1]^2 \ : \ y \leq 1 - x\}$. Alors, on vérifie bien les deux conditions d'inclusion, mais en revanche, si l'on considère le point $v = (2, 1/2)$, on observe que $v^* = \Pi_{V_1 \cap V_2}(v) \neq \Pi_{V_2}(\Pi_{V_1}(v)) = v_1$, comme l'indique la figure 3.*

A.1.2. Fonctions convexes.

Définition A.13 (Epigraphes et domaines). *Soit U un espace de Hilbert muni du produit scalaire noté $\langle \cdot, \cdot \rangle_U$. Soit $f : U \to \mathbb{R}$ une application.*

*(i) L'*épigraphe *de f noté* $\mathrm{epi}(f)$ *est défini par* $\mathrm{epi}(f) = \{(u, y) \in U \times \mathbb{R} \ : \ f(u) \leq y\} \subset U \times \mathbb{R}$ *;*

(ii) Le domaine *de f noté* $\mathrm{dom}(f)$ *est la projection de* $\mathrm{epi}(f)$ *sur U, i.e.,* $\mathrm{dom}(f) = \{u \in U \ : \ \exists y \in \mathbb{R},\ f(u) \leq y\} = \{u \in U,\ f(u) < +\infty\}$.

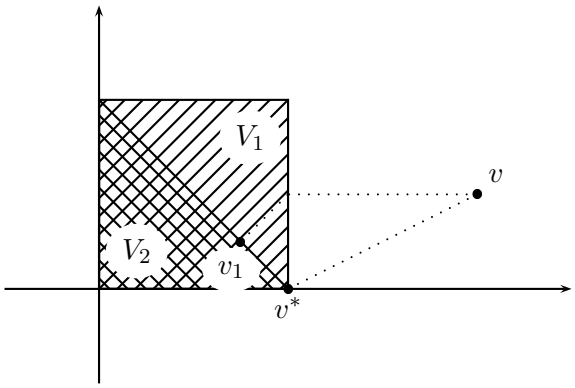


FIG. 3. Cas où les ensembles sont des convexes quelconques

DÉFINITION A.14 (coercivité). *Soit U un espace de Hilbert muni du produit scalaire noté $\langle\cdot,\cdot\rangle_U$. Une application $f : U \to \mathbb{R}$ est dite* coercive *sur U si et seulement si* $\lim_{\|u\|\to\infty} f(u) = +\infty$.

DÉFINITION A.15 (Fonctions convexes). *Soit U un espace de Hilbert muni du produit scalaire noté $\langle\cdot,\cdot,\rangle_U$. Soit $f : U \to \mathbb{R}$ une application.*

(i) f est dite convexe *si et seulement si son épigraphe* $\mathrm{epi}(f)$ *est convexe ; Si de plus l'ensemble convexe* $\mathrm{dom}(f)$ *est non-vide et si la restriction de f à* $\mathrm{dom}(f)$ *est finie, f est dite propre.*

(ii) f est dite strictement convexe *si et seulement si son épigraphe est strictement convexe ;*

(iii) f est dite fortement convexe *de module $b > 0$ si et seulement si $f(\cdot) - \frac{b}{2}\|\cdot\|_U^2$ est convexe.*

On a également les caractérisations suivantes de la convexité :

PROPOSITION A.16. *Soit U un espace de Hilbert muni du produit scalaire noté $\langle\cdot,\cdot\rangle_U$. Soit $f : U \to \mathbb{R}$ une application.*

(i) f est convexe si et seulement si

$$\forall u, v \in U,\ \forall \alpha \in [0,1],\ f(\alpha u + (1-\alpha)v) \leq \alpha f(u) + (1-\alpha) f(v); \tag{A.5}$$

(ii) f est strictement convexe si et seulement si l'équation (A.5) *est valable avec l'inégalité stricte ;*

(iii) f est fortement convexe de module $b > 0$ si et seulement si

$$\forall u, v \in U,\ \forall \alpha \in [0,1],\ f(\alpha u + (1-\alpha)v) \leq \alpha f(u) + (1-\alpha) f(v) - \frac{b\alpha(1-\alpha)}{2}\|u - v\|_U^2. \tag{A.6}$$

Preuve : Immédiat. cf. [**60**], IV, Proposition 1.1.2. pour le point (iii). □

Partant de ces caractérisations de la convexité, nous allons donner quelques propriétés de différentiabilité.

DÉFINITION A.17 (Dérivées directionnelles et différentiabilité). *Soit U un espace de Hilbert muni du produit scalaire noté $\langle\cdot,\cdot\rangle_U$. Soit $f : U \to \mathbb{R}$ une application.*

(i) Soit $u \in U$. f est dite directionnellement différentiable *dans la direction $v \in U$ si et seulement si en notant*

$$f'(u; v) = \lim_{\lambda \searrow 0} \frac{f(u + \lambda v) - f(u)}{\lambda},$$

alors $f'(u; v)$ existe et est finie (on l'appelle la dérivée directionnelle de f en u dans la direction v). f est dite directionnellement différentiable si elle l'est en toute les directions.

(ii) Soit $u \in U$. f est dite différentiable *en u si et seulement si il existe $g_u \in U$ tel que :*

$$\forall v \in U,\ f'(u; v) = \langle g_u, v\rangle_U.$$

Dans ce cas, on note $\nabla f(u) \in U$ l'élément g_u, appelé gradient de f en u.

Définition A.18 (Sous-différentiabilité des fonctions convexes). *Soit U un espace de Hilbert muni du produit scalaire noté $\langle\cdot,\cdot\rangle_U$. Soit $f : U \to \mathbb{R}$ une application convexe propre. Pour tout $u \in U$, on définit le* sous-différentiel *de f en u comme étant l'ensemble*

$$\partial f(u) = \{g \in U \ : \ \forall v \in U,\ f(v) \geq f(u) + \langle g, v-u\rangle_U\}\,. \tag{A.7}$$

On appelle sous-gradients de f en u les éléments de $\partial f(u)$.
f est dite sous-différentiable *en u si et seulement si $\partial f(u) \neq \emptyset$.*

Théorème A.19. *Soit U un espace de Hilbert muni du produit scalaire noté $\langle\cdot,\cdot\rangle_U$. Soit $f : U \to \mathbb{R}$ une application convexe. Soit $u \in U$ tel que $f(u) < \infty$. $g \in U$ est un sous-gradient de f en u si et seulement si*

$$\forall v \in U,\ f'(u;v) \geq \langle g, v\rangle_U.$$

Preuve : cf. [**80**], Theorem 23.2. □

Théorème A.20. *Soit U un espace de Hilbert muni du produit scalaire noté $\langle\cdot,\cdot\rangle_U$. Soit $f : U \to \mathbb{R}$ une application convexe. Soit $u \in U$ tel que $f(u) < \infty$. Si f est différentiable en u, alors $\nabla f(u)$ est l'unique sous-gradient de f en u, et en particulier :*

$$\forall v \in U,\ f(v) \geq f(u) + \langle \nabla f(u), v-u\rangle_U.$$

Réciproquement, si f a un unique sous-gradient en u, alors f est différentiable en u.

Preuve : cf. [**80**], Theorem 25.1. □

Dans le cas de fonctions différentiables, on peut caractériser autrement la forte convexité et la convexité :

Théorème A.21. *Soit U un espace de Hilbert muni du produit scalaire noté $\langle\cdot,\cdot\rangle_U$. Soit $f : U \to \mathbb{R}$ une application différentiable sur U^f un convexe fermé de U.*

(i) f est convexe sur U^f si et seulement si :

$$\forall u, v \in U^f,\ f(v) \geq f(u) + \langle \nabla f(u), v-u\rangle_U; \tag{A.8}$$

(ii) f est strictement convexe sur U^f si et seulement si l'inégalité (A.8) *est stricte dès que $u \neq v$;*

(iii) f est fortement convexe sur U^f de module $b > 0$ si et seulement si :

$$\forall u, v \in U^f,\ f(v) \geq f(u) + \langle \nabla f(u), v-u\rangle_U + \frac{b}{2}\|u-v\|_U^2. \tag{A.9}$$

Preuve : cf. [**60**], IV, Theorem 4.1. □

Enfin, on donne une condition d'optimalité :

Théorème A.22. *Soit U un espace de Hilbert muni du produit scalaire noté $\langle\cdot,\cdot\rangle_U$. Soit $f : U \to \mathbb{R}$ une application différentiable sur U^f un convexe fermé de U. Les trois propositions suivantes sont équivalentes :*

(i) $u^* \in U^f$ minimise f sur U^f ;

(ii) Pour tout $v \in U^f$, $\langle \nabla f(u^*), v-u^*\rangle_U \geq 0$;

(iii) $0 \in \partial f(u^*) + N_{U^f}(u^*)$, avec $N_{U^f}(u^*) = \{g \in U \ : \ \forall v \in U^f, \langle g, v-u^*\rangle_U \leq 0\}$

Preuve : cf. [**60**], VII, Theorem 1.1.1. □

On donne maintenant la définition d'intégrande normale. Cette notion, particulièrement développée au chapitre 14 de [**84**], est adaptée au cas de l'optimisation stochastique, car elle permet par exemple de conserver les propriétés de mesurabilité du contrôle lorsqu'il est composé avec une intégrande normale. Le lecteur intéressé pourra se reporter à [**84**] pour un exposé complet de cette notion et des propriétés qui y sont rattachées.

Définition A.23 (Intégrande normale). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité, $\boldsymbol{\xi}$ une variable aléatoire à valeurs dans Ξ, un espace métrique muni de la tribu $\mathcal{B}_\Xi$ qui est égale à $\mathcal{F}$ transportée par $\boldsymbol{\xi}$, U un espace de Hilbert muni de sa tribu borélienne $\mathcal{B}_U$, et $f : U \times \Xi \to \mathbb{R}$ une application. On dit que f est une* intégrande normale *si et seulement si :*

(i) f est $\mathcal{B}_U \otimes \mathcal{B}_\Xi$-mesurable ;

(ii) Pour $\mathbb{P}$-presque tout $\omega \in \Omega$, $f(\cdot, \boldsymbol{\xi}(\omega)) : U \to \mathbb{R}$ est semicontinue inférieurement.

Définition A.24 (Fonction de Carathéodory). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité, $\boldsymbol{\xi}$ une variable aléatoire à valeurs dans Ξ, un espace métrique muni de la tribu $\mathcal{B}_\Xi$ qui est égale à $\mathcal{F}$ transportée par $\boldsymbol{\xi}$, U un espace de Hilbert muni de sa tribu borélienne $\mathcal{B}_U$, et $f : U \times \Xi \to \mathbb{R}$ une application. On dit que f est une fonction de Carathéodory si et seulement si :*

(i) Pour tout $u \in U$, $f(u, \cdot) : \Xi \to \mathbb{R}$ est $\mathcal{B}_\Xi$-mesurable ;

(ii) Pour $\mathbb{P}$-presque tout $\omega \in \Omega$, $f(\cdot, \boldsymbol{\xi}(\omega)) : U \to \mathbb{R}$ est continue.

La proposition suivante découle des deux précédentes définitions :

Proposition A.25. *Toute fonction de Carathéodory est une intégrande normale.*

A.2. Principe du problème auxiliaire (PPA)

Dans cette section, nous allons brièvement rappeler le principe du problème auxiliaire (PPA). Ce principe, dont les origines remontent à [**28**],[**29**], est assez exhaustivement présenté dans [**31**]. L'exposé rapide que nous allons donner ici est là uniquement pour fournir au lecteur les idées et intuitions de base de ce principe.

L'idée originelle du PPA est de fournir un cadre général aux méthodes de décomposition-coordination pour les grands systèmes. En particulier, le PPA permet de montrer que l'ensemble des algorithmes de décomposition à deux niveaux ressort à un même principe fondamental que les algorithmes usuels de descente en optimisation convexe. Pour comprendre mieux cette idée, écrivons le problème considéré :

$$\min_{u \in U^f} J(u) = J^D(u) + J^S(u), \tag{A.10}$$

avec U un espace de Banach, U' son dual, et l'on notera $\langle \cdot, \cdot \rangle$ le crochet de dualité entre U et U'. U^f est un sous-ensemble convexe fermé de U, et $J, J^D, J^S : U \to \mathbb{R}$ des applications telles que J^D est convexe, semicontinue inférieurement et différentiable, et J^S seulement convexe et semicontinue inférieurement. Le problème (A.10) est appelé le problème maître.

Nous allons maintenant introduire le problème auxilaire, dépendant d'un $v \in U$, $\epsilon > 0$ et d'une application $G : U \to \mathbb{R}$ convexe semicontinue inférieurement et différentiable :

$$\min_{u \in U^f} A^v(u) = G(u) + \langle \epsilon (J^D)'(v) - G'(v), u \rangle + \epsilon J^S(u), \tag{A.11}$$

où l'accent représente la dérivée. On a alors le lemme suivant :

Lemme A.26. *Si $v \in U$ est tel que $v \in \arg\min_{u \in U^f} A^v(u)$, alors v est solution du problème* (A.10).

Preuve : Le résultat est une application de l'inéquation variationnelle caractérisant le minimum $u^* \in U^f$ d'une application $J^S + J^D$:

$$\forall u \in U^f,\ \langle (J^D)'(u^*), u - u^* \rangle + J^S(u) - J^S(u^*) \geq 0.$$

□

Ce lemme donne alors l'idée d'introduire un algorithme de point fixe pour résoudre le problème (A.10). C'est l'algorithme du problème auxiliaire :

Algorithme A.27. – *Étape 0 : choisir $u^0 \in U^f$, k=0, un seuil $\epsilon > 0$;*
– *Étape k : Résoudre le problème :*

$$\min_{u \in U^f} G^k(u) + \langle \epsilon^k (J^D)'(u^k) - (G^k)'(u^k), u \rangle + \epsilon^k J^S(u) \tag{A.12}$$

on note alors u^{k+1} la solution.
Si $\|u^k - u^{k+1}\| < \epsilon$, on s'arrête, sinon k devient $k+1$ et on boucle.

On donne ensuite un théorème de convergence pour cet algorithme :

THÉORÈME A.28 (Cohen, 1980). *(i) Supposons que J^S soit convexe, semicontinue inférieurement, et que J^D soit convexe, semicontinue inférieurement, 0-coercive sur U^f qui est un sous-ensemble convexe fermé de U, et que la dérivée de J^D notée $(J^D)'$ soit lipschitzienne de constante A. Alors le problème* (A.10) *admet des solutions, et l'on note U^* l'ensemble de ses solutions.*

(ii) Supposons que les fonctions G^k soient convexes, semicontinues inférieurement, et de dérivées uniformément lipschitizennes de constantes $B^k \leq B$ et uniformément fortement monotones de constantes $b^k \geq b$. Alors, le problème auxiliaire (A.12) *admet une unique solution notée u^{k+1}.*

(iii) Si de plus il existe $\alpha > 0, \beta > 0$ tels que pour tout $k \in \mathbb{N}$,

$$\alpha \leq \epsilon^k \leq \frac{2b^k}{A+\beta},$$

Alors $(J(u^k))$ décroît strictement (sauf si l'algorithme a convergé en un nombre fini d'itérations), et $(J^D)'(u^k)$ converge fortement vers $(J^D)'(u^*)$ (unique, même si U^* n'est pas réduit à un singleton), et la suite (u^k) est bornée, donc faiblement compacte, et tout point d'accumulation dans la topologie faible est dans U^*.

(iv) Si de plus $(J^D)'$ est fortement monotone de constante a sur un borné contenant la suite (u^k), alors la suite (u^k) converge fortement vers l'unique solution u^* de (A.10)*. De plus, on a la majoration a posteriori de l'erreur :*

$$\|u^{k+1} - u^*\| \leq \frac{\frac{B^k}{\epsilon^k} + A}{a} \|u^{k+1} - u^k\|.$$

Preuve : cf. [**31**], Théorème 3.6. □

On ne s'étendra pas ici plus longtemps sur les applications de ce principe à la décomposition-coordination des grands systèmes, qui est en partie traitée pour le cas stochastique dans le présent mémoire (chapitre V), et que l'on trouvera abondamment traitée dans [**31**]. L'intérêt majeur à retenir du PPA est que la difficulté (ou facilité) des problèmes auxiliaires (A.12) est presque entièrement reportée dans le choix des noyaux de décomposition G^k, puisque le gradient de J^D n'apparaît que linéairement dans cette formulation. Le PPA permet donc a priori de remplacer un problème difficile par une succession de problèmes d'autant plus simples que le choix des noyaux G^k a été effectué intelligemment.

A.3. Analyse fonctionnelle

DÉFINITION A.29 (μ-équivalence). *Soient $(E, \mathcal{B}_E)$ et $(F, \mathcal{B}_F)$ deux espaces de Banach mesurables, dont on notera les normes respectives $\|\cdot\|_E$ et $\|\cdot\|_F$. Soit μ une mesure sur $(E, \mathcal{B}_E)$. Soit u, v deux applications mesurables de E dans F. On dira que u et v sont μ-équivalentes si et seulement si $u = v$ sauf éventuellement sur un ensemble $N \subset E$ tel que $\mu(N) = 0$.*

DÉFINITION A.30 (Espaces $\mathcal{L}^p$ et L^p). *Soient $(E, \mathcal{B}_E)$ et $(F, \mathcal{B}_F)$ deux espaces de Banach mesurables, dont on notera les normes respectives $\|\cdot\|_E$ et $\|\cdot\|_F$. Soit μ une mesure sur $(E, \mathcal{B}_E)$. Pour tout $p \in [1, +\infty)$,*

(1) on notera $\mathcal{L}^p(E, F, \mu)$ l'ensemble des applications mesurables $v : E \to F$ telles que $\int_E \|v(x)\|_F^p d\mu(x) < +\infty$;

(2) on notera $L^p(E, F, \mu)$ l'espace quotient de $\mathcal{L}^p(E, F, \mu)$ par la relation de μ-équivalence.

PROPOSITION A.31 (Complétude des espaces L^p). *Soient $(E, \mathcal{B}_E)$ et $(F, \mathcal{B}_F)$ deux espaces de Banach mesurables, dont on notera les normes respectives $\|\cdot\|_E$ et $\|\cdot\|_F$. Soit μ une mesure sur $(E, \mathcal{B}_E)$. Pour tout $p \in [1, +\infty]$, l'espace $L^p(E, F, \mu)$ muni de la norme*

$$\forall v \in L^p(E, F, \mu)\ \|v\|_{L^p(E,F,\mu)} = \left(\int_E \|v(x)\|_F^p d\mu(x) \right)^{\frac{1}{p}},$$

est un espace de Banach. Si de plus F est un espace de Hilbert et que la norme $\|\cdot\|_F$ dérive du produit scalaire, alors $L^2(E,F,\mu)$, muni du produit scalaire

$$\forall u,v \in L^2(E,F,\mu),\ \langle u,v\rangle_{L^2(E,F,\mu)} = \int_E \langle u(x),v(x)\rangle_F d\mu(x),$$

est un espace de Hilbert.

Preuve : cf. [**44**], Théorème 5.2.1. □

PROPOSITION A.32 (Dualité des espaces L^p). *Soient $(E,\mathcal{B}_E)$ et $(F,\mathcal{B}_F)$ deux espaces de Hilbert mesurables, munis respectivement des produits scalaires notés $\langle\cdot,\cdot\rangle_E$ et $\langle\cdot,\cdot\rangle_F$. Soit μ une mesure σ-finie sur E. Pour tout $p\in[1,+\infty)$, le dual de $L^p(E,F,\mu)$ est $L^q(E,F,\mu)$, avec $\frac{1}{p}+\frac{1}{q}=1$.*

Preuve : cf. [**44**], Théorème 6.4.1. □

A.4. Lemmes techniques

On donne ici trois lemmes techniques utiles dans les preuves de convergence des algorithmes (éventuellement stochastiques) dérivés du principe du problème auxiliaire. On en donne ici une preuve pour plus d'exhaustivité. Les deux premiers lemmes remontent aux travaux [**30**]. Le dernier est propre à cette thèse, mais est un dérivé du deuxième.

LEMME A.33. *Soit $(x_k)_{k\in\mathbb{N}}$ une suite de réels positifs. Soit $(\alpha_k)_{k\in\mathbb{N}}$ et $(\beta_k)_{k\in\mathbb{N}}$ deux suites de réels positifs telles que $\sum_{k\in\mathbb{N}}\alpha_k<+\infty$ and $\sum_{k\in\mathbb{N}}\beta_k<+\infty$. Si de plus on a :*

$$\forall k\in\mathbb{N},\ x_{k+1}-x_k\le\alpha_k x_k+\beta_k,$$

alors la suite $(x_k)_{k\in\mathbb{N}}$ est bornée.

Preuve : En sommant l'hypothèse entre 0 et n, on obtient :

$$x_{n+1}\le x_0+\sum_{k=0}^{n}\beta_k+\sum_{k=0}^{n}\alpha_k x_k.$$

Dès lors, en définissant $m_k=\max_{0\le l\le k}x_l$, on a par positivité la même inégalité avec la suite m :

$$m_{n+1}\le x_0+\sum_{k=0}^{n}\beta_k+\sum_{k=0}^{n}\alpha_k m_k.$$

Nous allons montrer que (m_n) est bornée. Soit $1>\epsilon>0$. Par l'hypothèse de sommabilité sur les suites α,β, il existe $n_0\in\mathbb{N}$ tel que pour tout $n\ge n_0$, $\sum_{k=n_0}^{n}\alpha_k<\epsilon$. On réécrit donc l'inégalité précédente en coupant les sommes, pour tout $n\ge n_0$:

$$\begin{aligned}
m_{n+1} &\le x_0+\sum_{k=0}^{n}\beta_k+\sum_{k=0}^{n_0-1}\alpha_k m_k+\sum_{k=n_0}^{n}\alpha_k m_k,\\
&\le x_0+\sum_{k=0}^{n}\beta_k+\sum_{k=0}^{n_0-1}\alpha_k m_k+m_{n+1}\sum_{k=n_0}^{n}\alpha_k,\\
&\le \frac{x_0+\sum_{k=0}^{n}\beta_k+\sum_{k=0}^{n_0-1}\alpha_k m_k}{1-\epsilon},\\
&\le \frac{x_0+\sum_{k=0}^{\infty}\beta_k+\sum_{k=0}^{n_0-1}\alpha_k m_k}{1-\epsilon},
\end{aligned}$$

du fait de la positivité de la suite (β_k), ce qui montre que la suite (m_n) est bornée, et achève la preuve. □

LEMME A.34. *Soit J une application d'un espace de Hilbert H dans $\mathbb{R}$, lipschitzienne de module L. Soit $(u_k)_{k\in\mathbb{N}}$ une suite d'éléments de H et $(\epsilon_k)_{k\in\mathbb{N}}$ une suite de réels positifs tels que :*

(i) $\sum_{k\in\mathbb{N}}\epsilon_k=+\infty$,

(ii) $\exists\mu\in\mathbb{R},\ \sum_{k\in\mathbb{N}}\epsilon_k|J(u_k)-\mu|<+\infty$,

(iii) $\exists\delta>0,\ \forall k\in\mathbb{N},\ \|u_{k+1}-u_k\|\le\delta\epsilon_k$.

Alors $(J(u_k))_{k\in\mathbb{N}}$ converge vers μ.

Preuve : Pour tout $\alpha \in \mathbb{R}$, définissons le sous-ensemble N_α de $\mathbb{N}$ tel que :

$$N_\alpha := \{k \in \mathbb{N} \ : \ |J(u_k) - \mu| \leq \alpha\}.$$

On notera alors N_α^c le complémentaire de N_α dans $\mathbb{N}$. Les hypothèses $(i-ii)$ impliquent que N_α n'est pas fini.
En appliquant l'hypothèse (ii), on a :

$$+\infty > \sum_{k\in\mathbb{N}} \epsilon_k |J(u_k) - \mu| \geq \sum_{k\in N_\alpha^c} \epsilon_k |J(u_k) - \mu| \geq \alpha \sum_{k\in N_\alpha^c} \epsilon_k.$$

Cela prouve que pour tout $\beta > 0$, il existe un $n_\beta \in \mathbb{N}$ tel que $\sum_{k\in N_\alpha^c,\ k\geq n_\beta} \epsilon_l \leq \beta$.
Soit alors $\epsilon > 0$. Prenons $\alpha = \epsilon/2$ et $\beta = \epsilon/(2L\delta)$, avec L le module de Lipschitz de J. Pour tout $k \geq n_\beta$, on a deux possibilités :

- Si $k \in N_\alpha$, alors $|J(u_k) - \mu| \leq \alpha < \epsilon$.
- Si $k \in N_\alpha^c$, soit m le plus petit élément de N_α tel que $m \geq k$ (on sait qu'il existe car N_α n'est pas fini). On peut alors écrire :

$$\begin{aligned} |J(u_k) - \mu| &\leq |J(u_k) - J(u_m)| + |J(u_m) - \mu| \leq L\|u_k - u_m\| + \alpha, \\ &\leq L\delta \left(\sum_{l=k}^{m-1} \epsilon_l \right) + \alpha \leq L\delta \left(\sum_{l\in N_\alpha^c,\ l\geq n_\beta} \epsilon_l \right) + \alpha \leq \epsilon, \end{aligned}$$

ce qui achève la preuve. □

LEMME A.35. *Soit J une application d'un espace de Hilbert H dans $\mathbb{R}$, et soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité muni d'une filtration $(\mathcal{F}^k)$. Soit (u_k) une suite de variables aléatoires à, valeurs dans H telle que pour tout $k \in \mathbb{N}$, u_k est $\mathcal{F}_k$-mesurable, et soit (γ_k) une suite de réels positifs telles que :*

(i) $\sum_{k\in\mathbb{N}} \gamma_k = +\infty$,
(ii) $\exists \mu \in \mathbb{R}$, $\sum_{k\in\mathbb{N}} \gamma_k \left(J(u_k) - \mu\right) < +\infty$, *et* $\forall k \in \mathbb{N}$, $J(u_k) - \mu \geq 0$, *p.s.*
(iii) $\exists \delta > 0$, $\forall k \in \mathbb{N}$, $J(u_k) - \mathbb{E}\left(J(u_{k+1})|\mathcal{F}_k\right) \leq \delta\gamma_k$, *p.s.*

Then $(J(u_k))$ converge presque sûrement vers μ.

Preuve : Pour tout $\alpha \in \mathbb{R}$, définissons le sous-ensemble N_α de $\mathbb{N}$ tel que :

$$N_\alpha := \{k \in \mathbb{N} \ : \ J(u_k) - \mu \leq \alpha, \text{ p.s.}\}.$$

On notera alors N_α^c le complémentaire de N_α dans $\mathbb{N}$. Les Hypothèses $(i-ii)$ impliquent que N_α n'est pas fini.
Selon l'hypothèse (ii), il vient :

$$+\infty > \sum_{k\in\mathbb{N}} \gamma_k \left(J(u_k) - \mu\right) \geq \sum_{k\in N_\alpha^c} \gamma_k \left(J(u_k) - \mu\right) \geq \alpha \sum_{k\in N_\alpha^c} \gamma_k.$$

Cela prouve que pour tout $\beta > 0$, il existe un $n_\beta \in \mathbb{N}$ tel que $\sum_{k\in N_\alpha^c,\ k\geq n_\beta} \gamma_l \leq \beta$.
Soit $\epsilon > 0$. Prenons $\alpha = \epsilon/2$ et $\beta = \epsilon/(2\delta)$. Pour tout $k \geq n_\beta$, on a deux possibilités :

- Si $k \in N_\alpha$, alors $J(u_k) - \mu \leq \alpha < \epsilon$.
- Si $k \in N_\alpha^c$, soit m le plus petit éléments de N_α tel que $m \geq k$ (on sait qu'il existe car N_α n'est pas fini). On peut dès lors écrire :

$$\begin{aligned} J(u_k) - \mu &= J(u_k) - \mathbb{E}\left(J(u_m)|\mathcal{F}_k\right) + \mathbb{E}\left(J(u_m)|\mathcal{F}_k\right) - \mu \\ &= \mathbb{E}\left(\sum_{l=k}^{m-1} J(u_l) - \mathbb{E}\left(J(u_{l+1})|\mathcal{F}_l\right) |\mathcal{F}_k \right) + \mathbb{E}\left(J(u_m)|\mathcal{F}_k\right) - \mu, \\ &\leq \delta \left(\sum_{l=k}^{m-1} \gamma_l \right) + \alpha \leq \delta \left(\sum_{l\in N_\alpha^c,\ l\geq n_\beta} \gamma_l \right) + \alpha \leq \epsilon, \end{aligned}$$

ce qui achève la preuve.

□

ANNEXE B

Quasimartingales

On se donne dans tout ce chapitre un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ muni d'une filtration $(\mathcal{F}_k)$. La présentation que nous donnons ici des quasimartingales est issue de [**66**]. Soit $\boldsymbol{x} = (\boldsymbol{x}_k)_{k\in\mathbb{N}}$ un processus stochastique à valeurs réelles adapté à $(\mathcal{F}_k)$. On suppose dans toute cette section qu'il est intégrable, i.e. pour tout $k \in \mathbb{N}$, $\mathbb{E}(|X_k|) < +\infty$.

Définition B.1 (contenu). *On appelle contenu du processus stochastique $\boldsymbol{x}$ l'application $\lambda_{\boldsymbol{x}}(k, F)$ définie pour tout $k \in \mathbb{N}$ et tout $F \in \mathcal{F}_k$ par*

$$\lambda_{\boldsymbol{x}}(k, F) = \mathbb{E}(1_F \times (\boldsymbol{x}_{k+1} - \boldsymbol{x}_k)).$$

Définition B.2 (sur- et sous-martingale). *Le processus $\boldsymbol{x}$ est une sur- (resp. sous-) martingale si et seulement si*

$$\forall k \in \mathbb{N},\ \forall F \in \mathcal{F}_k,\ \lambda_{\boldsymbol{x}}(k, F) \leq 0 \quad (\text{resp. } \geq 0).$$

Si $\boldsymbol{x}$ est à la fois une sur- et sous-martingale, alors c'est une martingale.

Définition B.3 (variation du contenu). *La variation du contenu de $\boldsymbol{x}$ en $k \in \mathbb{N}$ notée $|\lambda_{\boldsymbol{x}}|(k)$ est définie par :*

$$|\lambda_{\boldsymbol{x}}|(k) = \mathbb{E}(|\mathbb{E}(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k|\mathcal{F}_k)|).$$

La variation totale du contenu, notée $|\lambda_{\boldsymbol{x}}|$ est alors définie par la somme possiblement divergente suivante

$$|\lambda_{\boldsymbol{x}}| = \sum_{k\in\mathbb{N}} |\lambda_{\boldsymbol{x}}|(k).$$

Définition B.4 (quasimartingale). *Le processus $\boldsymbol{x}$ est une quasimartingale si et seulement si sa variation totale du contenu est finie, i.e.*

$$\sum_{k\in\mathbb{N}} \mathbb{E}(|\mathbb{E}(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k|\mathcal{F}_k)|) < +\infty.$$

Il est donc clair qu'une martingale est une quasimartingale.

On peut alors démontrer la proposition suivante :

Proposition B.5 (Métivier). *Posons pour tout $k \in \mathbb{N}$, $G_k = \{\omega \in \Omega : \mathbb{E}(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k|\mathcal{F}_k)(\omega) > 0\}$. Si $\boldsymbol{x}$ vérifie les deux conditions :*

$$\sum_{k\in\mathbb{N}} \mathbb{E}(1_{G_k} \times (\boldsymbol{x}_{k+1} - \boldsymbol{x}_k)) < +\infty,$$

$$\inf_{k\in\mathbb{N}} \mathbb{E}(\boldsymbol{x}_k) > -\infty,$$

alors c'est une quasimartingale.

Preuve : cf. [**66**], Proposition 9.5. □

Enfin, on dispose du théorème de convergence suivant pour les quasimartingales :

Théorème B.6 (Métivier). *Supposons que $\boldsymbol{x}$ soit une quasimartingale et vérifie la condition suivante :*

$$\sup_{k\in\mathbb{N}} \mathbb{E}(\boldsymbol{x}_k^-) < +\infty, \quad \textit{(avec } x^- = -\min(0, x)\textit{)},$$

alors $(\boldsymbol{x}_k)$ converge presque sûrement vers une variable aléatoire intégrable $\boldsymbol{x}_\infty$, et l'on a :

$$\mathbb{E}(|\boldsymbol{x}_\infty|) \leq \inf_{k\in\mathbb{N}} \mathbb{E}(|\boldsymbol{x}_k|).$$

Preuve : cf. [**66**], Theorem 9.4. □

ANNEXE C

Algorithmes stochastiques en dimension finie

On s'intéresse ici au problème de l'approximation stochastique. Ce domaine dont les origines remontent aux articles fondateurs [**77, 63**], a connu depuis un développement très important, décrit par exemple dans [**65**]. Les méthodes d'approximations stochastiques ont trouvé de nombreux et naturels débouchés en optimisation stochastique, à travers la méthode du gradient stochastique, développée notamment en Europe de l'Est et par l'école dite de Kiev, par exemple [**72, 67**]. On pourra également consulter [**50**] pour un recueil de techniques. Un exemple du mariage entre l'approximation stochastique et l'optimisation est donné par [**32**] ou [**33**].

Divers chapitres de ce mémoire utilisent des idées de l'approximation stochastique. Il nous a donc semblé utile dans cet appendice de donner les idées principales et quelques résultats existant sur ces techniques. Tous les résultats qui seront donnés ici le seront en dimension finie. La plupart des résultats que nous allons donner dans ce chapitre se trouvent dans [**46, 14**]. Certes, les travaux [**32**] se placent dans le cadre hilbertien, et donc a priori sans limitation de dimension, mais n'incluent pas les problèmes en boucle fermée, i.e. fonctionnels.

Des travaux ont été entrepris en dimension infinie pour le contexte fonctionnel, parmi lesquels on peut citer par exemple [**74, 75, 97, 27**]. On discute plus en détails de ces résultats en dimension infinie dans le chapitre IV du mémoire.

C.1. Schéma de Robbins-Monro

L'algorithme de Robbins-Monro [**77**] en toute généralité permet de rechercher les zéros d'une application $H : X \to X$, avec X un espace de Hilbert de dimension finie, définie par

$$\forall x \in X,\ H(x) := \mathbb{E}\left(h(x, \boldsymbol{\xi})\right),$$

avec $\boldsymbol{\xi}$ une variable aléatoire sur l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans un espace Ξ donné, et de loi μ, et $h : X \times \Xi \to X$ une application mesurable en son deuxième argument. L'algorithme de Robbins-Monro s'écrit alors comme suit :

ALGORITHME C.1 (Robbins-Monro). *Étape -1 : choisir $x^0 \in X$,*
Étape $k \geq 0$:
- *Tirer $\boldsymbol{\xi}^{k+1}$ selon la loi μ indépendamment des tirages passés,*
- *Mettre à jour :*

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \rho^k h(\boldsymbol{x}^k, \boldsymbol{\xi}^{k+1}).$$

La convergence de cet algorithme a ensuite été abondamment étudiée. Nous allons donner ici les principaux résultats de convergence. Ils sont de deux types. Tout d'abord, on dispose de résultats de *convergence presque sûre* de l'algorithme C.1, le terme presque sûr étant compris par rapport aux échantillons comme nous l'expliquerons après. Ensuite, on peut trouver des résultats de normalité asymptotique de l'algorithme C.1, qu'on interprète naturellement comme des résultats de vitesse.

THÉORÈME C.2 (Convergence presque sûre). *Notons pour tout $k \in \mathbb{N}$, $\mathcal{F}^k = \sigma(\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^k)$. Supposons que :*
(i) Il existe $x^* \in X$ tel que $H(x^*) = 0$, et $\forall x \in X$, tel que $x \neq x^*$, $\langle x^* - x, H(x)\rangle_X > 0$;
(ii) Il existe $K \geq 0$ tel que l'algorithme C.1 vérifie

$$\forall k \in \mathbb{N},\ \mathbb{E}\left(\|h(\boldsymbol{x}^k, \boldsymbol{\xi}^{k+1})\|_X^2 | \mathcal{F}^k\right) \geq K(1 + \|\boldsymbol{x}^k\|_X^2).$$

(iii) La suite (ρ^k) *vérifie*

$$\sum_{k\in\mathbb{N}} \rho^k = +\infty,\ \sum_{k\in\mathbb{N}} (\rho^k)^2 < +\infty.$$

Alors, la suite $(\boldsymbol{x}^k)$ *engendrée par l'algorithme C.1 converge presque sûrement vers* x^**.*

On peut relâcher l'hypothèse (ii) du théorème en utilisant une caractérisation à l'aide de fonctions de Lyapunov et en faisant une hypothèse de bornitude uniforme sur tout compact de la fonction h.

Nous donnons maintenant le lemme dit de Robbins-Siegmund, introduit dans [**78**], qui constitue la brique élémentaire pour la démonstration du théorème C.2 et d'autres du même type.

Lemme C.3 (Robbins-Siegmund). *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace de probabilité muni d'une filtration* $(\mathcal{F}^k)$*. Soient* $(\boldsymbol{V}^k, \boldsymbol{\beta}^k, \boldsymbol{\chi}^k, \boldsymbol{\eta}^k)$ *quatre suites positives de variables aléatoires adaptées à* $(\mathcal{F}^k)$*. On suppose de plus que*

$$\sum_{k\in\mathbb{N}} \boldsymbol{\beta}^k < +\infty, \quad \sum_{k\in\mathbb{N}} \boldsymbol{\chi}^k < +\infty, \quad \text{p.s.}$$

Si on a

$$\forall k \in \mathbb{N},\ \mathbb{E}\left(\boldsymbol{V}^{k+1} \middle| \mathcal{F}^k\right) \leq (1+\boldsymbol{\beta}^k)\boldsymbol{V}^k + \boldsymbol{\chi}^k - \boldsymbol{\eta}^k.$$

Alors, $(\boldsymbol{V}^k)$ *converge presque sûrement vers une variable aléatoire finie* $\boldsymbol{V}^\infty$*, et* $\sum_{k\in\mathbb{N}} \boldsymbol{\eta}^k < +\infty$*.*

Preuve : cf. par exemple [**45**], Théorème 2.III.3. □

Pour donner un résultat de normalité asymptotique, on se restreint à $\rho^k = \frac{\rho}{(1+k)^\alpha}$ pour tout $k \in \mathbb{N}$.

Théorème C.4 (Normalité asymptotique). *Posons* $\Sigma = \mathbb{E}\left(h(x^*, \boldsymbol{\xi})h(x^*, \boldsymbol{\xi})^T\right)$*. Supposons que les hypothèses du théorème C.2 sont vérifiées. Supposons de plus que :*

(i) H *est* C^1 *au voisinage* $V(x^*)$ *de* x^**, et il existe* A *une matrice symétrique définie positive telle que :*

$$\forall x \in V(x^*),\ H(x) = A(x^* - x) + o(x - x^*).$$

(ii) Il existe $\beta > 0$ *tel que* $\sup_{k\in\mathbb{N}} \mathbb{E}\left(\|h(\boldsymbol{x}^k, \boldsymbol{\xi}^{k+1}) - H(\boldsymbol{x}^k)\|_X^{2+\beta}\right) < +\infty$*.*

(iii) Il existe $\epsilon > 0$ *tel que la famille* $\{h(x, \boldsymbol{\xi}) \ :\ \|x - x^*\|_x \leq \epsilon\}$ *est uniformément intégrable.*

Alors la suite $\left(\frac{\boldsymbol{x}^k - x^*}{\sqrt{\rho^k}}\right)$ *converge en loi vers une loi normale centrée de variance* V *définie par :*

(a) Si $\alpha = 1$ *et* $\rho A - \frac{I}{2} > 0$ *alors*

$$V = \rho \int_0^\infty \exp\left(\left(\frac{I}{2} - \rho A\right)t\right) \Sigma \exp\left(\left(\frac{I}{2} - \rho A\right)t\right) dt.$$

(b) Si $\alpha < 1$*, alors*

$$V = \rho \int_0^\infty \exp\left(-\rho A t\right) \Sigma \exp\left(-\rho A t\right) dt.$$

On trouve souvent dans la littérature le résultat précédent sous la forme suivante. Sous les mêmes hypothèses que dans le théorème C.4, la covariance optimale V est solution de l'équation de Lyapunov suivante, pour $\alpha = 1$:

$$\Sigma + \left(\nabla H(x^*) + \frac{I}{2}\right) V + V \left(\nabla H(x^*) + \frac{I}{2}\right)^T = 0. \tag{C.1}$$

En prenant alors des pas matriciels pour l'algorithme C.1, du type $\rho^k = \frac{\Gamma}{(1+k)^\alpha}$, l'équation de Lyapunov devient

$$\Sigma + \left(\Gamma\nabla H(x^*) + \frac{I}{2}\right) V + V \left(\Gamma\nabla H(x^*) + \frac{I}{2}\right)^T = 0.$$

et la variance optimale est donnée par $V^* = \nabla H(x^*)^{-1}\Sigma\nabla H(x^*)^{-1}$, c'est à dire, dans le cas où $H(x) = -\mathbb{E}\left(\nabla_u j(u, \boldsymbol{\xi})\right)$, avec j convexe en u, par :

$$V^* = \mathbb{E}\left(\left[\nabla^2_{uu} j(u^*, \boldsymbol{\xi})\right]^{-1}\right)\mathbb{E}\left(\left[\nabla_u j(u^*, \boldsymbol{\xi})\right]^T\left[\nabla_u j(u^*, \boldsymbol{\xi})\right]\right)\mathbb{E}\left(\left[\nabla^2_{uu} j(u^*, \boldsymbol{\xi})\right]^{-1}\right).$$

C.2. Application au gradient stochastique

L'algorithme C.1 peut être utilisé dans le cadre de l'optimisation stochastique. En effet, regardons le problème suivant :

$$\min_{x\in\mathbb{R}^p} \mathbb{E}\left(f(x, \boldsymbol{\xi})\right), \tag{C.2}$$

où $\boldsymbol{\xi}$ est une variable aléatoire à valeurs dans Ξ et $f : \mathbb{R}^p \times \Xi \to \mathbb{R}$ est une application convexe en x et mesurable en ξ. En notant X^* l'ensemble des solutions du problème (C.2), les conditions d'optimalité du problème s'écrivent :

$$\forall x^* \in X^*,\ \mathbb{E}\left(\nabla_x f(x^*, \boldsymbol{\xi})\right) = 0.$$

Résoudre les conditions d'optimalité revient donc à trouver les zéros de la fonction $H : \mathbb{R}^p \to \mathbb{R}^p$ définie par $H(x) = -\mathbb{E}\left(\nabla_x f(x, \boldsymbol{\xi})\right)$; autrement dit, on est dans le cadre de l'algorithme C.1, qui s'écrit ici pour sa phase d'actualisation :

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - \rho^k \nabla_x f(\boldsymbol{x}^k, \boldsymbol{\xi}^{k+1}). \tag{C.3}$$

Naturellement, des hypothèses de convexité sur f et de Lipschitz sur son gradient permettent de vérifier les conditions $(i-ii)$ du théorème C.2, ce qui montre que l'algorithme (C.3) converge presque sûrement vers la solution du problème (C.2). On peut illustrer l'algorithme (C.3) de manière plus directe. Un algorithme de gradient standard sur le problème (C.2) s'écrirait :

$$x^{k+1} = x^k - \rho^k \mathbb{E}\left(\nabla_x f(x^k, \boldsymbol{\xi})\right), \tag{C.4}$$

avec (ρ^k) possiblement constante, ce qui ne nuirait pas à la convergence.
Ainsi, l'algorithme de gradient stochastique peut-il s'interpréter comme un algorithme de descente dans lequel la véritable direction de descente est estimée itérativement, à l'aide de tirages successifs, en utilisant une suite (ρ^k) tendant vers 0 comme $(1/k)$, permettant à la fois d'assurer que l'itéré courant ne *bouge* pas trop, et que la véritable direction de descente, autrement dit une espérance, soit reconstituée par un argument du type Monte Carlo. Cette seconde interprétation du gradient stochastique (C.3) va nous permettre d'introduire le cas de problèmes contraints.

En effet, de façon plus générale, les problèmes d'optimisation du type (C.2) sont souvent contraints. Typiquement, on demande que $x \in X^f$, avec X^f un convexe fermé de $\mathbb{R}^p$. On peut alors avoir recours, en suivant l'optimisation classique, à des algorithmes de gradient projeté, du type :

$$x^{k+1} = \Pi_{X^f}\left(x^k - \rho^k \mathbb{E}\left(\nabla_x f(x^k, \boldsymbol{\xi})\right)\right),$$

ce qui, de par l'interprétation donnée plus haut des algorithmes de gradient stochastique, donne envie de considérer l'algorithme stochastique projeté suivant :

$$\boldsymbol{x}^{k+1} = \Pi_{X^f}\left(\boldsymbol{x}^k - \rho^k \nabla_x f(\boldsymbol{x}^k, \boldsymbol{\xi}^{k+1})\right). \tag{C.5}$$

La convergence de cet algorithme est donnée par le théorème C.5, dont on peut trouver une démonstration dans le contexte général de la décomposition et du principe du problème auxiliaire dans [**32**].

THÉORÈME C.5 (Cohen-Culioli, 90). *(i) Supposons $x \mapsto f(x, \xi)$ convexe, semi continue inférieurement, et différentiable sur X^f, pour tout $\xi \in \Xi$, et $\xi \mapsto f(x, \xi)$ mesurable pour tout $x \in X^f$. Si de plus X^f est un convexe fermé sur lequel $\mathbb{E}\left(f(\cdot, \boldsymbol{\xi})\right) : \mathbb{R}^p \to \mathbb{R}$ est coercive, alors* (C.2) *sous la contrainte $x \in X^f$ admet des solutions, qu'on notera X^*.*

(ii) Supposons que la suite (ρ^k) soit telle que :

$$\sum_{k\in\mathbb{N}} \rho^k = +\infty,\ \sum_{k\in\mathbb{N}} (\rho^k)^2 < +\infty, \tag{C.6}$$

et qu'il existe deux constantes $\alpha, \beta \geq 0$ telle que :

$$\forall x \in X^f,\ \forall \xi \in \Xi,\ \|\nabla_x f(x,\xi)\| \leq \alpha\|x - x^*\| + \beta. \tag{C.7}$$

Alors, presque sûrement, $\lim_{k\to\infty} \mathbb{E}\left(f(x^k, \boldsymbol{\xi})|\mathcal{F}^k\right) = \mathbb{E}\left(f(x^*, \boldsymbol{\xi})\right)$ pour $x^* \in X^*$, et la suite $(\boldsymbol{x}^k)$ est bornée et tout point d'accumulation est dans X^*.

(iii) Si de plus $\mathbb{E}\left(f(\cdot, \boldsymbol{\xi})\right)$ est fortement convexe, alors $X^* = \{x^*\}$ et $(\boldsymbol{x}^k)$ converge presque sûrement vers x^*.

Ce théorème est valable même lorsque x appartient à un espace de dimension infinie. Des arguments de preuve similaires se trouvent dans [**59**], pour la démonstration de la convergence d'algorithmes plus généraux encore que (C.5). On ne dispose en revanche pas à ce jour de théorème central limite pour les algorithmes stochastiques projetés.

Enfin, on trouve, dans [**33**] ou [**59**], des algorithmes du type (C.5) avec une contrepartie dans le dual, dans le cas de problèmes de point-selle ou avec contraintes explicites, permettant de panacher à loisir les actualisations de type stochastique avec les actualisations de type déterministe. Plus précisèment, si l'on considère le problème :

$$\begin{aligned} &\min_{x\in X^f \subset \mathbb{R}^p} \mathbb{E}\left(f(x, \boldsymbol{\xi})\right) \\ &\qquad \text{s.c.} \mathbb{E}\left(\theta(x, \boldsymbol{\xi})\right) \in -C \subset \mathbb{R}^m, \end{aligned} \tag{C.8}$$

on peut écrire l'algorithme de Arrow-Hurwicz stochastique suivant :

$$\boldsymbol{x}^{k+1} = \Pi_{X^f}\left(x^k - \rho_x^k\left(\nabla_x f(x^k, \boldsymbol{\xi}^{k+1}) + (\nabla_x \theta(\boldsymbol{x}^k, \boldsymbol{\xi}^{k+1}))^T \boldsymbol{p}^k\right)\right) \tag{C.9a}$$

$$\boldsymbol{p}^{k+1} = \Pi_{C'}\left(\boldsymbol{p}^k + \rho_p^k \theta(\boldsymbol{x}^{k+1}, \boldsymbol{\xi}^{k+1})\right) \tag{C.9b}$$

On peut alors montrer le théorème suivant :

THÉORÈME C.6 (Cohen-Culioli, 94). *Supposons que f et θ sont bien définies et mesurables en leur deuxième argument, que X^f est un convexe fermé de $\mathbb{R}^p$, et C un convexe fermé de $\mathbb{R}^m$. Supposons également que le lagrangien associé au problème* (C.8) *admet un point-selle noté (x^*, p^*), et que :*

(i) $\mathbb{E}\left(j(\cdot, \boldsymbol{\xi})\right) : \mathbb{R}^p \to \mathbb{R}$ est strictement convexe sur X^f, et pour tout $\xi \in \Xi$, $j(\cdot, \xi)$ est localement lipschitzienne sur X^f ;

(ii) $\theta(\cdot, \xi) : \mathbb{R}^p \to \mathbb{R}^m$ est lipschitzienne, différentiable, et de gradient borné, le tout uniformément en $\xi \in \Xi$. De plus, $\mathbb{E}\left(\theta(\cdot, \boldsymbol{\xi})\right) : \mathbb{R}^p \to \mathbb{R}^m$ est C-convexe.

(iii) Il existe des constantes $\alpha, \beta \geq 0$ telles que :

$$\forall \xi \in \Xi,\ \forall x \in X^f,\ \|\nabla_x f(x,\xi)\| \leq \alpha\|x - x^*\| + \beta; \tag{C.10}$$

(iv) Les suites (ρ_x^k, ρ_p^k) sont telles que :

$$\sum_{k\in\mathbb{N}} \rho_x^k = +\infty,\ \sum_{k\in\mathbb{N}} \rho_p^k = +\infty,\ \sum_{k\in\mathbb{N}} (\rho_x^k)^2 < +\infty,\ \sum_{k\in\mathbb{N}} (\rho_p^k)^2 < +\infty, \tag{C.11}$$

et la suite (ρ_x^k/ρ_p^k) est monotone au sens large.

(v) Il existe des constantes $\gamma, \delta > 0$ telles que :

$$\forall x \in X^f,\ \mathbb{E}\left(\left(\theta(x, \boldsymbol{\xi}) - \mathbb{E}\left(\theta(x, \boldsymbol{\xi})\right)\right)^2\right) < \gamma\|x - x^*\|^2 + \delta.$$

Alors, presque sûrement, la suite $(\boldsymbol{x}^k, \boldsymbol{p}^k)$ *engendrée par* (C.9a)–(C.9b) *est bornée, et* $(\boldsymbol{x}^k)$ *converge faiblement vers* x^*. *Si de plus,* $\mathbb{E}\,(j(\cdot, \boldsymbol{\xi}))$ *est fortement convexe, la convergence est forte.*

Ce théorème présente en soit un intérêt, mais sa preuve est également extrêmement importante. En effet, c'est sur le schéma qui y a été proposé par Cohen et Culioli que nous avons construit une part importante des démonstrations du chapitre V, et du chapitre IV.

Bibliographie

1. B. Allen, *Neighboring information and distributions of agents' characteristics under uncertainty*, J. Math. Econ. (1983), no. 12, 63–101.
2. L. Andrieu, *Optimisation sous contrainte en probabilité*, Thèse de doctorat, École Nationale des Ponts et Chaussées, 2004.
3. Z. Artstein, *Sensitivity to σ-fields of information in stochastic allocation*, Stochastic Stochastics Rep. (1991), no. 36, 41–63.
4. ———, *Probing for information in two-stage stochastic programming and the associated consistency*, Asymptotic Satistics (P. Mandl and M. Huskova, eds.), Springer Verlag, Berlin, 1994, pp. 21–33.
5. ———, *Gains and costs of information in stochastic programming*, Ann. Oper. Res. (1999), no. 85, 129–152.
6. J.-P. Aubin and H. Frankowska, *Set-valued analysis*, Birkhäuser, Boston, 1990.
7. V. Barbu and Th. Precupanu, *Convexity and optimization in banach spaces*, Kluwer, Dordrecht, 1986.
8. K. Barty, *Contributions à la discrétisation des contraintes de mesurabilité pour les problèmes d'optimisation stochastique*, Thèse de doctorat, École Nationale des Ponts et Chaussées, 2004.
9. K. Barty, P. Carpentier, J.-P. Chancelier, G. Cohen, M. de Lara, and T. Guilbaud, *Dual effect free stochastic controls*, Ann. Oper. Res. (2004).
10. K. Barty, J.-S. Roy, and C. Strugarek, *A perturbed gradient algorithm in Hilbert spaces*, Optimization Online (2005), http ://www.optimization-online.org/DB_HTML/2005/03/1095.html.
11. ———, *A stochastic gradient type algorithm for closed loop problems*, submitted, available on-line at http ://dochost.rz.hu-berlin.de/speps/ (2005).
12. ———, *Temporal difference learning with kernels for pricing american-style options*, submitted (2005), http ://www.optimization-online.org/DB_HTML/2005/05/1133.html.
13. R. Bellman and S.E. Dreyfus, *Functional approximations and dynamic programming*, Math tables and other aides to computation **13** (1959), 247–251.
14. A. Benvéniste, M. Métivier, and P. Priouret, *Adaptive algorithms and stochastic approximation*, Springer Verlag, New York, 1990.
15. H. Berliocchi and J.-M. Lasry, *Nouvelles applications des mesures paramétrées*, C. R. Acad. Sci., Paris **274** (1972), 1623–1626.
16. P. Bernhard and G. Cohen, *Commande optimale, décentralisation et jeux dynamiques*, Dunod, Paris, 1976.
17. D.P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, 1999.
18. D.P. Bertsekas and S.E. Shreve, *Stochastic optimal control : the discrete-time case*, Athena Scientific, Belmont, 1996.
19. D.P. Bertsekas and J.N. Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, Belmont, 1996.
20. ———, *Gradient convergence in gradient methods*, SIAM J. Optim. **10** (2000), no. 3, 627–642.
21. S. Bochner, *Harmonic analysis and the theory of probability*, University of California Press, Berkeley, 1955.
22. J.F. Bonnans, J.-Ch. Gilbert, C. Lemaréchal, and C. Sagastizábal, *Optimisation numérique*, Springer Verlag, 1997.
23. J.F. Bonnans and A. Shapiro, *Perturbation analysis of optimization problems*, Springer Verlag, Berlin Heidelberg, 2000.
24. E.S. Boylan, *Equiconvergence of martingales*, The Annals of Mathematical Statistics **42** (1971), no. 2, 552–559.
25. P. Carpentier, G. Cohen, and J.-C. Culioli, *Stochastic optimal control and decomposition-coordination methods - Part I : Theory*, Recent Developments in Optimization. Lecture Notes in Economics and Mathematical Systems (R. Durier and C. Michelot, eds.), no. 429, Springer Verlag, Berlin, 1995, pp. 72–87.
26. ———, *Stochastic optimal control and decomposition-coordination methods - Part II : Applications*, Recent Developments in Optimization. Lecture Notes in Economics and Mathematical Systems (R. Durier and C. Michelot, eds.), no. 429, Springer Verlag, Berlin, 1995, pp. 88–103.
27. X. Chen and H. White, *Asymptotic properties of some projection-based Robbins-Monro procedures in a Hilbert space*, Stud. Nonlinear Dyn. Econom. **6** (2002), 1–53.

28. G. Cohen, *Optimization by decomposition and coordination : a unified approach*, IEEE Trans. Autom. Control **23** (1978), no. 2, 222–232.

29. ———, *Auxiliary Problem Principle and decomposition of optimization problems*, J. Optimization Theory Appl. (1980), no. 32, 277–305.

30. ———, *Décomposition et coordination en optimisation déterministe différentiable et non-différentiable*, Thèse de doctorat d'État, Université de Paris IX Dauphine, 1984.

31. ———, *Optimisation des grands systèmes*, Cours du DEA MMME, Université de Paris I, 2003.

32. G. Cohen and J.-C. Culioli, *Decomposition Coordination Algorithms for Stochastic Optimization*, SIAM J. Control Optimization **28** (1990), no. 6, 1372–1403.

33. ———, *Optimisation stochastique sous contraintes en espérance*, Rapport interne Centre Automatique et Systèmes, Ecole des Mines de Paris (1994), no. A–288.

34. K.D. Cotter, *Similarity of information and behavior with a pointwise convergence topology*, J. Math. Econ. (1986), no. 15, 25–38.

35. ———, *Convergence of information, random variables and noise*, J. Math. Econ. (1987), no. 16, 39–51.

36. J.-C. Culioli, *Algorithmes de décomposition/coordination en optimisation stochastique*, Thèse de doctorat, École des Mines de Paris, 1987.

37. D. Dacunha-Castelle and M. Duflo, *Probabilités et statistiques, problèmes à temps fixe*, Masson, 1994.

38. A. Dallagi, *a definir*, Thèse de doctorat, École Nationale des Ponts et Chaussées, 2006.

39. D.P. de Farias and B. Van Roy, *The Linear Programming Approach to Approximate Dynamic Programming*, Oper. Res. **51** (2003), no. 6, 850–856.

40. P. Deheuvels, *Sur l'estimation séquentielle de la densité*, C. R. Acad. Sci. Paris Ser. A-B **276** (1973), 1119–1121.

41. ———, *Estimation séquentielle de la densité*, Contribuciones en Probabilidad y Estadistica Matematica Enseñanza de la Matematica y Analysis, Grindley, Granada, Espagne, 1979, pp. 156–168.

42. F. Delebecque and J.-P. Quadrat, *Contribution of stochastic control singular perturbation averaging and team theories to an example of large-scale systems : Management of hydropower production*, IEEE Transactions on Automatic Control **23** (1978), no. 2, 209–222.

43. L. Devroye, *A course in density estimation*, Birkhäuser, Boston, 1987.

44. R.M. Dudley, *Real analysis and probability*, Cambridge Uuniversity Press, Cambridge, UK, 2002.

45. M. Duflo, *Algorithmes stochastiques*, Springer Verlag, Berlin, 1996.

46. ———, *Random iterative models*, Springer Verlag, Berlin, 1997.

47. J. Dupačová, N. Gröwe-Kuska, and W. Römisch, *Scenario reduction in stochastic programming. An approach using probability metrics*, Math. Program. **95** (2003), no. 3, 493–511.

48. J. Dupačová and R.J.-B. Wets, *Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems*, Ann. Stat. **16** (1988), no. 4, 1517–1549.

49. Y. Ermoliev, V. Norkin, and R.J.-B. Wets, *The minimization of semicontinuous functions : Mollifier subgradients*, SIAM J. Control Optimization **33** (1995), no. 1, 149–167.

50. Y. Ermoliev and R.J.-B. Wets (eds.), *Numerical techniques for stochastic optimization problems*, Springer, Berlin, 1988.

51. O. Fiedler and W. Römisch, *Stability in multistage stochastic programming*, Ann. Oper. Res. (1995), no. 56, 79–93.

52. R. Fortet and E. Mourier, *Convergence de la répartition empirique vers la répartition théorique*, Annales scientifiques de l'Ecole Normale Supérieure **70** (1953), no. 3, 267–285.

53. P. Glasserman, *Monte-Carlo methods in financial engineering*, Springer Verlag, Berlin, 2003.

54. L. Greengard and J. Strain, *The fast gauss transform*, SIAM Journal on Scientific and Statistical Computing **12** (1991), no. 1, 79–94.

55. H. Heitsch and W. Römisch, *Scenario reduction algorithms in stochastic programming*, Comput. Optim. Appl. (2003), no. 24, 187–206.

56. ———, *Scenario tree modelling for multistage stochastic programs*, Preprint 05-19, Institut für Mathematik, Humboldt-Universität zu Berlin, and submitted (2005).

57. H. Heitsch, W. Römisch, and C. Strugarek, *Stability of multistage stochastic programs*, Preprint, DFG Research Center Matheon "Mathematics for key technologies", (to appear in SIAM J. Optim.) (2005), no. 255.

58. J.L. Higle and S. Sen, *Stochastic decomposition*, Kluwer, Dordrecht, 1996.

59. J.-B. Hiriart-Urruty, *Algorithmes de résolution d'équations et d'inéquations variationnelles*, Z. Wahrscheinlichkeitstheorie verw. Gebiete **33** (1975), 167–186.

60. J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, Springer Verlag, Berlin, 1996.

61. C.C. Holt, F. Modigliani, and H.A. Simon, *A linear decision rule for production and employment scheduling*, Manage. Sci. **2** (1955), no. 1.

62. D.N. Hoover, *Convergence in distribution and Skorokhod convergence for the general theory of processes*, Probab. Theory Relat. Fields (1991), no. 89, 239–259.

63. J. Kiefer and J. Wolfowitz, *Stochastic estimation of the maximum of a regression function*, Annals of Mathematical Statistics **23** (1952), 462–466.

64. H. Kudo, *A note on the strong convergence of σ-algebras*, Ann. Probab. **2** (1974), 76–83.

65. T.L. Lai, *Stochastic Approximation*, Ann. Stat. **31** (2003), no. 2, 391–406.

66. M. Métivier, *Semimartingales*, De Gruyter, Berlin, 1982.

67. M.B. Nevel'son and R.Z. Has'minskii, *Stochastic approximation and recursive estimation*, American Mathematical Society, Providence, RI, 1973.

68. J. Neveu, *Note on the tightness of the metric on the set of complete sub σ algebras of a probability space*, The Annals of Mathematical Statistics (1972), no. 4, 1369–1371.

69. E. Parzen, *On estimating of a probability density and mode*, Annals of Mathematical Statistics **35** (1962), 1065–1076.

70. T. Pennanen, *Epi-convergent discretizations of multistage stochastic programs*, Mathematics of Operations Research **30** (2005), 245–256.

71. G.Ch. Pflug, *Scenario tree generation for multiperiod financial optimization by optimal discretization*, Math. Program. (2001), no. 89, 251–271.

72. B.T. Polyak and Y.Z. Tsypkin, *Pseudogradient adaptation and training algorithms*, Autom. Remote Control **12** (1973), 83–94.

73. S.T. Rachev and W. Römisch, *Quantitative stability in stochastic programming : the method of probability metrics*, Math. Oper. Res. **27** (2002), 792–818.

74. P. Révész, *Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes, I.*, Studia Sci. Math. Hungar. **8** (1973), 391–398.

75. _____, *Robbins-Monro procedure in a Hilbert space, II.*, Studia Sci. Math. Hungar. **8** (1973), 469–472.

76. _____, *How to apply the method of stochastic approximation in the non-parametric estimation of a regression function*, Math. Operationsforsch. Statist. Ser. Statistics **8** (1977), no. 1, 119–126.

77. H. Robbins and S. Monro, *A stochastic approximation method*, Annals of Mathematical Statistics **22** (1951), 400–407.

78. H. Robbins and D. Siegmund, *A convergence theorem for nonnegative almost supermartingales and some applications*, Optimizing Methods in Statistics (J.S. Rustagi, ed.), Academic Press, New York, 1971, pp. 233–257.

79. S.M. Robinson, *Analysis of sample-path optimization*, Math. Oper. Res. **21** (1996), no. 3.

80. R.T. Rockafellar, *Convex analysis*, Princeton University Press, Princeton, 1970.

81. R.T. Rockafellar and R.J.-B. Wets, *Stochastic convex programming : basic duality*, Pacific J. Math. **62** (1976), no. 1, 173–195.

82. _____, *Stochastic convex programming : singular multipliers and extended duality, singular multipliers and duality*, Pacific J. Math. **62** (1976), no. 2, 507–522.

83. _____, *Scenarios and policy aggregation in optimization under uncertainty*, Math. Oper. Res. **16** (1991), no. 1, 119–147.

84. _____, *Variational analysis*, Springer Verlag, Berlin Heidelberg, 1998.

85. L. Rogge, *Uniform inequalities for conditional expectations*, Ann. Probab. **2** (1974), 486–489.

86. W. Römisch, *Stability of stochastic programs*, Handbooks in Operations Research and Management Science (A. Ruszczynski and A. Shapiro, eds.), vol. 10, Elsevier, Amsterdam, 2003, pp. 483–554.

87. M. Rosenblatt, *Remarks on some nonparametric estimates of a density function*, Annals of Mathematical Statistics **27** (1956), 832–837.

88. D. Salinger, *A splitting algorithm for multistage stochastic programming with application to hydropower scheduling*, PhD Thesis, University of Washington, 1997.

89. L. Schwartz, *Analyse*, Hermann, 1993.

90. A. Shapiro and T. Homem-de Mello, *On the rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs*, SIAM J. Optim. **11** (2000), no. 1, 70–86.

91. A. Shapiro and A. Ruszczynski (eds.), *Stochastic Programming*, Elsevier, Amsterdam, 2003.

92. S. Smale and Y. Yao, *Online learning algorithms*, Preprint, www.tti-c.org/smale_papers/ (2004).

93. J. Strain, *The fast gauss transform with variable scales*, SIAM Journal on Scientific and Statistical Computing **12** (1991), 1131–1139.

94. T.J. Wagner and C.T. Wolverton, *Recursive estimates of probability densities*, IEEE Trans. Syst. Man. Cybern. **5** (1969), 307.

95. H.S. Witsenhausen, *A counterexample in stochastic optimal control*, SIAM Journal of Control **2** (1968), no. 6, 149–160.

96. C. Yang, R. Duraiswami, N. Gumerov, and L. Davis, *Improved fast gauss transform and efficient kernel density estimation*, IEEE International Conference on Computer Vision (2003), 464–471.

97. G. Yin and Y.M. Zhu, *On H-valued Robbins-Monro processes*, J. Multivariate Anal. **34** (1990), 116–140.

Index

Algorithme
  de décomposition par les prix, 92
  de gradient perturbé, 64–73
  de gradient projeté, 46
  de gradient stochastique, 45, 58, 140
  de Robbins-Monro, 138
  du problème auxiliaire, 131
Arbres d'aléas, 89

Contraintes
  de mesurabilité, 23, 57
  en probabilité, 120

Distance
  de Boylan, 33
  de Cotter, 34
  de Fortet-Mourier, 25
  de Rogge, 33
Double effet, voir Effet dual
Dual effect, voir Effet dual

Effet dual, 5, 8
Espaces de noyaux reproduisants, 79

Grandes déviations, 53

Information, voir Structure d'information

Lagrangien augmenté, 113
Lemme de Robbins-Siegmund, 69, 139

Mollification, 47

Principe du problème auxiliaire, 89, 101, 131
  stochastique, 101
Problème
  auxiliaire, voir Principe du problème auxiliaire
  en boucle fermée, 3, 88
  en boucle ouverte, 3, 87
  en boucle fermée, 46
  multistage, 5, 30, 36–40, 87
Programmation dynamique stochastique, 90–100
Projection
  mesurable, 57
  sur une intersection, 57, 127–128

Quantification, 29
Quasimartingale, 52, 67, 73, 136

Règles de décision linéaires, 76–79
RKHS, voir Espaces de noyaux reproduisants
Robbins-Monro, voir Algorithme de
Robbins-Siegmund, voir Lemme de

Structure d'information
  décentralisée, 95
  dynamique, 4
  en mémoire parfaite, 5
  statique, 4

Topologies sur les tribus, 31–36