

Contributions à la discretisation des contraintes de mesurabilité pour les problèmes d'optimisation stochastique

Kengy Barty

▶ **To cite this version:**

Kengy Barty. Contributions à la discretisation des contraintes de mesurabilité pour les problèmes d'optimisation stochastique. Mathématiques [math]. Ecole des Ponts ParisTech, 2004. Français. NNT : . tel-00008867

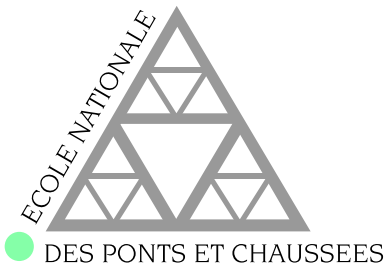
HAL Id: tel-00008867

https://pastel.hal.science/tel-00008867

Submitted on 25 Mar 2005

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 et Informatique

présentée par

Kengy BARTY

Sujet : Contributions à la discrétisation des contraintes de mesurabilité pour les problèmes d'optimisation stochastique.

Soutenue le 25 juin 2004 devant le jury composé de :

Président :	J.-M.	BONNISSEAU	Université de Paris I
Rapporteur :	J.-P.	QUADRAT	I.N.R.I.A.
Rapporteur :	J.-B.	HIRIART-URRUTY	Université Paul Sabatier (Toulouse)
Examinateur :	N.	BOULEAU	E.N.P.C.
Examinateur :	R.	AÏD	E.D.F.
Directeur de thèse :	G.	COHEN	E.N.P.C.

Remerciements

Ma plus grande reconnaissance revient à l'équipe de recherche dirigée par mon directeur de thèse Monsieur Guy Cohen et formée de Pierre Carpentier, Jean-Philippe Chancelier et Michel De Lara.

Je remercie mes rapporteurs, Jean-Baptiste Hiriart-Urruty et Jean-Pierre Quadrat, leurs remarques m'ont permis d'améliorer la qualité de ce manuscrit.

Je remercie Nicolas Bouleau, Jean-Marc Bonnisseau et René Aïd pour l'honneur qu'ils me font en acceptant de faire partie de mon jury de thèse.

Les discussions enrichissantes que j'ai eu le plaisir d'avoir avec Pierre Carpentier, ont été déterminantes dans ma formation à la recherche, elles furent une source permanente d'inspiration. Sa bonne humeur et sa gentillesse ont contribuées pour beaucoup à l'aboutissement de ce travail.

Je remercie Jean-Philippe Chancelier pour l'intérêt qu'il a bien voulu porté sur mes travaux de recherche, pour sa patiente, ses qualités humaines exceptionnelles, mais aussi pour son expertise informatique. Ce document n'aurait certainement jamais vu le jour sans son aide.

Les conseils de Michel De Lara m'ont souvent été d'un grand secours, je lui en suis très reconnaissant.

Aux thésards de l'équipe Laetitia Andrieu, Cyrille Strugarek et Anes Dallagi, j'exprime ici toute ma gratitude.

Les chercheurs du CERMICS font de ce laboratoire un environnement propice à la réflexion scientifique. Je remercie en particulier Bernard Lapeyre, Éric Cances, Jean-François Delmas, Alexandre Ern, Benjamin Jourdain, Vlad Bally, Régis Monneau, Frédéric Legoll, Yousra Gati, Tony Lelièvre, Maxime Barrault, Xavier Blanc, Claude Le Bris, Gilbert Caplain, Linda El Alaoui, Bouhari Arouna, François Lodier, Adel Ben Haj, Thérèse Guilbaud, René Lalement, Thierry Salset ainsi bien sur que Sylvie Berthe et notre brillant ingénieur système Jacques Daniel.

Sans l'amitié indéfectible de Georges Gautry ce travail n'aurait peut être même pas pu commencer, qu'il en soit ici remercié.

Je remercie ma mère, je lui dois d'être arrivé là, qu'elle trouve ici une sorte de récompense pour tous les efforts consentis pour moi.

Mes remerciement vont particulièrement à Nathalie, sans son attention je n'aurais peut être jamais abouti et cela n'aurait d'ailleurs pas eu de sens.

Enfin, j'ai une pensée émue pour mon père qui aurait souhaité voir ce moment, je lui

dédie du fond du cœur ce mémoire.

À mon père.

Résumé : Nous nous sommes penchés sur différents aspects des problèmes d'optimisation stochastique qui, à notre connaissance, ont été peu étudiés. Ainsi, nous nous sommes intéressés au problème de l'effet dual, puis à la discrétisation des contraintes de mesurabilité, à la résolution numérique de problèmes avec contraintes en information statique et enfin, nous avons étudié les conditions d'optimalité d'un problème d'optimisation stochastique, le but recherché étant de mieux comprendre comment intervient la contrainte de mesurabilité dans la caractérisation de la (ou des) solution(s) optimale(s). Notre approche numérique du problème est originale de deux points de vue :

- elle utilise les topologies sur l'espace des σ-algèbres pour mesurer la perte d'information due à la discrétisation de la contrainte de mesurabilité. L'étude de cet espace nous a permis entre autres d'apporter de nouveaux résultats qui constituent des éléments essentiels dans notre étude ;
- nous montrons que l'erreur de discrétisation provient de la contribution de deux termes d'erreur : une erreur issue de la discrétisation de la contrainte de mesurabilité et une autre erreur issue de l'approximation de l'espérance.

Nous donnons dans ce mémoire des résultats asymptotiques de convergence d'une suite de problèmes discrets vers le problème d'origine. Nous avons également, sur des problèmes particuliers, des résultats de type Lipschitz sur la fonction valeur. Par ailleurs, l'étude des conditions d'optimalité nous a permis d'obtenir deux possibilités différentes d'approche d'un problème de commande optimale stochastique.

Abstract : Our attention has been concentrated on various aspects of stochastic optimization problems which, according to our knowledge, have not been studied enough. Therefore first we shall be interested in the problem relative to the dual effect, afterwards in the discretization of the measurability constraints, in static information problem's numerical resolution and finally we shall study a stochastic optimization problem's optimality conditions with the purpose of searching for a better comprehension of the way which intervenes the measurability constraint in the optimal solution(s) characterization. Our problem's numerical approach is original of two points of view : it uses (the) topologies over the space of σ-fields in order to measure the information loss ought to the measurability constraint's discretization . Furthermore the study of this space has brought in new results which constitue essential elements of our research. We show that the discretization error results from the contribution of two other error terms : one resulting from the discretization of the measurability constraint and of another resulting from the approximation of the expectation. In this paper we give asymptotical convergence results of a series of discrete problems towards the original problem. For the same particular problem we obtain as well Lipschitz type results over the value function. Moreover by studying the optimality conditions we obtain two different possible ways of approaching a stochastic optimal control problem.

Table des matières

Glossaire **xi**

I Compte rendu de la littérature **1**
I.1 Optimisation dans l'incertain et structure d'information 1
I.1.1 Optimisation et incertain 1
I.1.2 Scénarios et structure d'information 1
I.1.3 Arbres de scénarios 2
I.2 Les catégories de problèmes 3
I.2.1 Boucle ouverte 3
I.2.2 Structure d'information statique 3
I.2.3 Structure d'information dynamique ou avec effet dual 4
I.3 Brève revue de la littérature 4
I.3.1 Effet dual 4
I.3.2 Discrétisation de filtrations 5
I.3.3 Méthodes de Monte-Carlo 5
I.3.4 Robustesse 6
I.3.5 Conditions d'optimalité 7
I.4 Originalité du mémoire 7
I.5 Plan 9

II L'effet dual en optimisation stochastique **11**
II.1 Introduction 11
II.2 Structures algébriques 13
II.2.1 Partition et relation d'équivalence 13
II.2.2 Relation de précédence 14
II.3 Effets d'un feedback sur l'information 15
II.4 Effet dual : définition et caractérisation 18
II.4.1 Absence de temps 18
II.4.2 Introduction du temps 19
II.5 Effet dual et mémoire parfaite 20
II.5.1 Effet dual sans mémoire parfaite 20
II.5.2 Effet dual et mémoire parfaite en boucle ouverte 23
II.5.3 Conclusion 25

III Variations autour des métriques sur $\mathcal{F}^*$ **27**
III.1 Introduction . . . 27
III.1.1 Notions sur les tribus . . . 28
III.2 Convergence forte de tribus . . . 29
III.2.1 Topologie de convergence forte . . . 29
III.2.2 Propriétés de continuité de l'application σ . . . 32
III.2.3 Variation autour de la topologie de la convergence forte . . . 36
III.2.4 Borne sup de tribus . . . 40
III.3 Convergence uniforme de tribus . . . 42
III.3.1 Variation autour de la convergence uniforme . . . 43
III.3.2 Convergence uniforme et convergence forte de tribus . . . 48

IV Quelques résultats asymptotiques **49**
IV.1 Introduction . . . 49
IV.1.1 Intégrandes normales . . . 49
IV.2 Pénalisation des contraintes de mesurabilité . . . 50
IV.2.1 Propriété de Lipschitz de la fonction valeur . . . 53
IV.2.2 Majoration d'erreur . . . 56
IV.3 Discrétisation d'un problème statique . . . 59
IV.3.1 Continuité "à gauche" de la valeur de l'information . . . 60
IV.3.2 Quantification de la contrainte de mesurabilité . . . 61
IV.3.3 Problème en boucle ouverte et technique de type Monte-Carlo . . . 63
IV.3.4 Approximation de la loi de ξ . . . 63
IV.3.5 Comportement asymptotique des coûts discrets . . . 66
IV.3.6 Convergence des solutions . . . 68
IV.4 Deux exemples de discrétisation . . . 69
IV.4.1 Un problème avec une contrainte de parité . . . 69
IV.4.2 Un problème L.Q.G. avec une contrainte de non-anticipativité . . . 73

V Expériences numériques **77**
V.1 Formulation du problème . . . 77
V.2 Hypothèses probabilistes . . . 78
V.3 Extensions possibles du problème . . . 79
V.4 Traitement du problème déterministe sur les chroniques . . . 79
V.5 Données numériques et fonctionnelles . . . 80
V.5.1 Demande et apports d'eau . . . 80
V.5.2 Production électrique . . . 82
V.5.3 Coût intégral . . . 82
V.5.4 Coût final . . . 82
V.5.5 Condition initiale . . . 82
V.5.6 Bornes sur l'état et la commande . . . 82
V.6 Résolution du problème déterministe . . . 83
V.6.1 Calcul formel . . . 83

V.6.2 Solutions numérique du problème déterministe 84
V.7 Quantification numérique 84
V.7.1 Cellules de Voronoï 84
V.8 Mise œuvre en pratique 86

VI Conditions de Kuhn et Tucker **97**
VI.1 Introduction 97
VI.2 Problèmes avec une contrainte de projection 98
VI.2.1 Définition du problème 98
VI.2.2 Approche par dualité 100
VI.2.3 Approche directe des conditions d'optimalité 102
VI.2.4 Application à un problème d'optimisation stochastique 103
VI.3 Conditions de Kuhn et Tucker dans un problème stochastique 104
VI.3.1 Un problème de commande optimale stochastique 104
VI.3.2 Conditions d'optimalité 105
VI.3.3 Conditions nécessaires d'optimalité, première forme 107
VI.3.4 Conditions nécessaires d'optimalité, deuxième forme 108
VI.3.5 Calcul de l'état adjoint 109

VII Conclusion et ouvertures **113**
VII.1 Résultats 113
VII.2 Ouvertures 113

A Rappels d'optimisation **115**
A.1 Optimisation convexe 115
A.2 Optimisation non convexe 117
A.2.1 Application à un problème d'optimisation non convexe 121
A.3 Fonctions marginales 121

B Rappels de probabilité **123**
B.1 Résultats sur l'espérance conditionnelle 123

Glossaire

$\stackrel{\text{def}}{=}$	Définition (le membre de gauche est défini par l'expression de droite).
$\mathbb{P}_X$	Loi image de la variable aléatoire X, $\mathbb{P}_X(A) \stackrel{\text{def}}{=} \mathbb{P}(X^{-1}(A))$.
$C^1(X)$	Ensemble des fonctions réelles continûment différentiables sur X.
$\mathbb{N},\ \mathbb{N}^*$	Ensemble des entiers naturels et ensemble des entiers naturels non nuls.
$\mathbb{R},\ \mathbb{R}^*$	Ensemble des réels et ensemble des réels privé de zéro.
$(\Omega, \mathcal{F}, \mathbb{P})$	Espace probabilisé.
$L_X^p(\Omega)$	Ensemble des applications $(\Omega, \mathcal{F})$ mesurables à valeurs dans X de puissance p^{e} intégrable.
$\langle \cdot, \cdot \rangle$	Produit scalaire ou crochet de dualité.
$\mathbb{E}$	Opérateur d'espérance.
$\mathbb{E}\left(f \mid \mathcal{B}\right)$	Espérance conditionnelle de f sachant $\mathcal{B}$.
$\mathcal{X}_A$	Fonction caractéristique de A, $\mathcal{X}_A(x) = 0$ si $x \in A$ et $+\infty$ sinon.
$\mathcal{A} \otimes \mathcal{B}$	Plus petite tribu engendrée par $\mathcal{A} \times \mathcal{B}$.
$x^\top$	Transposée du vecteur x.
f^*	Transformée de Legendre-Fenchel de l'application f.
$\left\|\left\|\cdot\right\|\right\|_X$	Norme sur l'espace vectoriel X.
$A \times B$	Produit cartésien des ensembles A et B.
$\mathbb{I}_A$	Fonction indicatrice de l'ensemble A, $\mathbb{I}_A(x) = 1$ si $x \in A$ et 0 sinon.
f'_x	Dérivée partielle de l'application f par rapport à la composante x.
$\text{co}(A)$	Enveloppe convexe de l'ensemble A.
X^*	Dual topologique de X.
v.a.	Variable aléatoire.
δ_x	Mesure de Dirac au point x.
$\mathcal{B}_\Xi$	Tribu des boréliens sur l'espace topologique Ξ.

Chapitre I

Compte rendu de la littérature

I.1 Optimisation dans l'incertain et structure d'information

I.1.1 Optimisation et incertain

Les problèmes d'*optimisation dans l'incertain* sont caractérisés par la nécessité de prendre des décisions sans savoir précisément qu'elles seront leurs conséquences. De tels problèmes apparaissent dans différents domaines d'application et soulèvent des questions aussi bien théoriques que pratiques. Commençons par définir la terminologie employée.

Optimisation : nous faisons référence à l'étude d'un problème et de sa solution, dans lequel nous devons faire un choix admissible. Les décisions admissibles sont modélisées comme les éléments d'un ensemble admissible. Le but est de trouver le "meilleur" choix (pas nécessairement unique). Les choix possibles sont comparés en fonction de la valeur que leur attribue une certaine fonction, le critère.

Incertain : le terme fait référence au fait que les données entrant en jeu dans le problème d'optimisation peuvent éventuellement être aléatoires.

I.1.2 Scénarios et structure d'information

Supposons que l'on souhaite minimiser un critère économique déterminé par l'application $J : \mathbb{R}^m \times \Xi \to \mathbb{R}$, et que ce critère dépende d'événements qui ne seront connus que progressivement. Nous dirons qu'une suite d'événements possibles est un scénario et, pour fixer les idées, ξ désignera un scénario. Par ailleurs supposons également que, même si les événements n'ont pas encore eu lieu, il faut quand même prendre une décision. Il n'est pas raisonnable de minimiser pour tous les scénarios ξ possibles la fonction $x \mapsto J(x, \xi)$. En effet, supposons que les événements possibles se résument à une succession de $+$ et de $-$. Considérons les scénarios suivants $\xi_1 = +++++$, $\xi_2 = +----$ et $\xi_3 = +-+-+$. Si à l'étape 1 du processus on observe $+$, comment déterminer s'il faut appliquer la décision associée à ξ_1, celle associée à ξ_2, ou celle associée à ξ_3. Si l'interprétation du signe $+$ est

"abondance" et celle du signe − est "pénurie", alors les décisions associées au premier événement des scénarios ξ_1 et ξ_2 peuvent être très différentes : il est légitime de penser qu'on ne ferait pas la même chose selon que l'on se prépare à une période de pénurie ou une période d'abondance. C'est la raison pour laquelle le critère d'un problème d'optimisation stochastique classique est **généralement la valeur moyenne des coûts associés à tous les scénarios possibles.** Les décisions sont supposées être prises sur la base de l'observation d'événements passés : cette contrainte se traduit mathématiquement par une contrainte de mesurabilité sur les variables de décision. Plus généralement, nous parlerons de *structure d'information* pour faire référence aux contraintes de mesurabilité auxquelles sont assujetties les variables de décision.

I.1.3 Arbres de scénarios

La résolution des problèmes d'optimisation stochastique est difficile pour deux raisons :

1. d'une part il faut calculer les espérances de fonctions aléatoires relativement à des lois continues, ou des lois dicrètes sur un grand nombre d'atomes ;
2. d'autre part il faut discrétiser les contraintes de mesurabilité du problème d'origine.

Une approche classique pour tenir compte de ces deux contraintes repose sur la technique des arbres de scénarios. L'idée est la suivante : on commence par se donner un nombre fini de scénarios, leur rôle étant de permettre l'évaluation du critère, puis on organise ces scénarios en arbre,[1] l'idée sous-jacente étant qu'un arbre est l'approximation naturelle d'une filtration[2]. Ce faisant, la discrétisation du critère et de la structure d'information est réalisée avec l'aide de l'unique structure qu'est l'arbre de scénarios. Cette technique réalise donc un compromis entre la discrétisation de la structure d'information et celle du coût. Il n'existe pas à notre connaissance de moyen permettant de distinguer la contribution de l'arbre de scénarios à la discrétisation du coût de celle allouée à la discrétisation de la structure d'information. Ce manque de clarté concernant l'apport de l'arbre de scénarios à l'une ou l'autre des discrétisations peut expliquer les difficultés rencontrées lorsque l'on souhaite juger de la qualité d'un arbre de scénarios. En fait il paraît difficile de mesurer directement la qualité de l'arbre de scénarios, étant donné qu'il cherche à réaliser un compromis entre deux problèmes très différents : un problème de nature numérique, "approcher des lois de probabilité" et un problème de nature algébrique, "approcher des σ-algèbres". Compte tenu des difficultés liées à l'évaluation de ces phénomènes, les questions liées à la qualité de l'arbre des scénarios ne trouvent pas dans la littérature de réponses directes. Le problème est contourné en considérant qu'un "bon arbre de scénarios" est une structure permettant d'obtenir une bonne approximation du coût optimal[3] [60] : ce point de vue a l'inconvénient de limiter la possibilité d'obtenir des résultats généraux sur la manière d'obtenir des arbres

[1] Un arbre est simplement une filtration discrète ou bien associée à un processus discret.

[2] Une filtration est une suite croissante de sous-tribus.

[3] Le critère de qualité aurait pu porter sur la qualité des contrôles calculés, mais cela ne donne pas un instrument de mesure "univoque" étant donné que l'unicité des solutions n'est pas garantie en général.

de scénarios, étant donné que la qualité de l'arbre est directement liée aux propriétés du problème étudié.

I.2 Les catégories de problèmes

Nous étudierons dans ce mémoire la **discrétisation des contraintes de mesurabilité** dans un problème d'optimisation stochastique, c'est la raison pour laquelle il nous parait nécessaire de déterminer précisément les différentes catégories de structure d'information, ce qui nous permettra de classer chaque problème d'optimisation dans une catégorie en fonction de sa contrainte de mesurabilité.

I.2.1 Boucle ouverte

Les problèmes d'optimisation en boucle ouverte sont les problèmes de la forme :

$$\min_{x\in\mathbb{R}^n} \mathbb{E}\left[J(x,\xi(\omega))\right]. \tag{I.1}$$

La variable de décision $x \in \mathbb{R}^n$ est prise avant l'observation d'une quelconque variable aléatoire. Ce type de problème est largement étudié [97, 98], et bien qu'en apparence très simple, il ne peut pas être abordé par des techniques classiques d'optimisation non linéaire. La difficulté principale de ce type de problème est liée à l'évaluation de $\mathbb{E}\left[J(x,\xi(\omega))\right]$ ou d'un (sous) gradient. Il existe cependant quelques cas favorables (notamment lorsque la loi de ξ est discrète) le calcul de l'espérance se résume alors à une simple somme :

$$\mathbb{E}\left[J(x,\xi(\omega))\right] = \sum_{i=1}^{N} p_i J(x,\xi_i).$$

Dans ce cas, évaluer l'espérance ou le gradient du critère consiste simplement à évaluer $J(x,\xi_i)$ ou $J_x^{'}(x,\xi_i)$ pour tout i. Cette constatation est à la base des *techniques dite de chroniques de scénarios* qui consistent à remplacer la variable aléatoire ξ par une loi discrète sur un nombre fini de scénarios dans le but d'obtenir un problème discret plus facile à résoudre. Cette approche soulève le problème du nombre de scénarios nécessaires afin de respecter une marge d'erreur fixée a priori [96, 95]. Parmi les différents angles d'approche du problème (I.1), il faut également citer celles de type gradient stochastique [76] qui sont basées sur l'idée de mener conjointement et progressivement la minimisation du critère et le calcul de l'espérance. Les problèmes en boucle ouverte pouvant être de grande taille, les techniques de décomposition [27] sont naturellement un aspect important.

I.2.2 Structure d'information statique

Les *problèmes statiques* sont ceux dont la particularité est de posséder une structure d'information indépendante de toute loi de commande. Nous dirons des problèmes statiques

qu'ils possèdent une structure d'information fixe. La décision optimale est prise après l'observation d'une variable aléatoire définie indépendamment de toute variable de décision. Un exemple de problème statique est :

$$V(\mathcal{B}) \stackrel{\text{def}}{=} \min \{\mathbb{E}\left[J(\gamma(\xi), \xi)\right] \mid \gamma \text{ est } \mathcal{B}\text{-mesurable}\}. \tag{I.2}$$

On montrera plus loin que si la tribu $\mathcal{B}$ est engendrée par une partition finie, alors $V(\mathcal{B})$ peut se ramener à un problème en boucle ouverte. Il arrive souvent en pratique que la tribu $\mathcal{B}$ soit engendrée par une variable aléatoire h, nous appellerons une telle fonction h une *fonction d'observation*. Nous insistons sur le fait que le problème d'optimisation (I.2) est seulement un exemple de problème statique ; il suffit simplement de remarquer que deux contraintes de mesurabilité disons "γ_1 est $\mathcal{A}$-mesurable et γ_2 est $\mathcal{B}$-mesurable", ne peuvent pas en général s'écrire comme une seule contrainte de mesurabilité sur le couple (γ_1, γ_2).

I.2.3 Structure d'information dynamique ou avec effet dual

Les problèmes dynamiques sont ceux qui ne font pas partie de la catégorie problèmes statiques. On trouve notamment dans cette catégorie les problèmes pour lesquels les fonctions d'observations dépendent des variables de décision.

Dans certains cas, la dépendance relativement aux variables de décision s'exprime uniquement à travers les valeurs prises par la fonction d'observation et pas à travers la σ-algèbre engendrée. Il s'agit alors de faux problèmes dynamiques : on dit alors qu'il n'y a pas d'effet dual. Un exemple de problème dynamique est :

$$\min \{\mathbb{E}\left[J(\gamma(\xi), \xi)\right] \mid \gamma \text{ est } \mathcal{B}^{\gamma}\text{-mesurable}\}. \tag{I.3}$$

Pour des développements récents au sujet de l'absence d'effet dual voir [19] et le chapitre II de ce mémoire.

I.3 Brève revue de la littérature

Nous avons simplement pour objectif de donner une idée de ce qui existe dans certains domaines de l'optimisation stochastique. Les problématiques soulevées en optimisation stochastique sont très nombreuses ; nous ne parlerons que de celles pour lesquelles nous pensons avoir apporté une contribution. Le choix des articles à résumer est très difficile étant donné le volume des publications concernant ce sujet. Néanmoins, nous avons choisi de sélectionner les articles qui nous paraissent être au fait des développements mathématiques en optimisation stochastique :

I.3.1 Effet dual [19, 11, 104, 103]

Artstein [11] s'intéresse à la résolution numérique des problèmes avec effet dual en essayant de munir l'espace des sous-tribus de $\mathcal{F}$ d'une structure d'espace vectoriel. L'intérêt d'une telle approche est d'espérer pouvoir définir une notion de calcul différentiel, mais à notre connaissance ce n'est pas encore le cas.

I.3.2 Discrétisation de filtrations [59, 29, 30, 68, 67, 24]

La discrétisation de la structure d'information est un problème qui est au cœur de l'optimisation stochastique. Cette problématique a été identifiée depuis longtemps. Rockafellar et Wets [92] évoquent la possibilité de représenter une filtration par une filtration discrète, un "arbre". La discrétisation de filtration continue a été étudiée notamment par Hoover [59] Coquet et Mémin [29, 30]. En fait un résultat de Hoover [59, théorème 7.7] permet de ramener la question de la discrétisation d'une filtration à celle de la discrétisation d'une tribu. Les propriétés topologiques de l'espace des sous-tribus d'une tribu $\mathcal{F}$ ont été étudiées par de nombreux auteurs dont je donne ici une liste qui n'est pas exhaustive : Neveu [68, 67] Boylan [24] Küdo [62] Cotter [31, 32] Alonso [6] Stinchombe [100, 99] Nghiem [69], plus récemment Piccinini [73] et Artstein [13] . Nous renvoyons le lecteur au chapitre III de ce mémoire, dans lequel nous donnons les principaux résultats relatifs à ces topologies.

I.3.3 Méthodes de Monte-Carlo [95, 97, 96, 98]

Les méthodes de type Monte-Carlo appliquées aux problèmes d'optimisation en boucle ouverte font l'objet d'intenses recherches. Ce qui motive ces recherches est en particulier le fait qu'il existe peu de résultats permettant de majorer l'erreur introduite dans la résolution numérique en remplaçant les variables aléatoires par un estimateur. Le cas particulier où une variable aléatoire est remplacée par son estimateur empirique à fait l'objet d'une attention particulière de la part notamment de Shapiro [97, 96, 98]. Soit :

$$v^* \stackrel{\text{def}}{=} \min_{x\in\mathbb{R}^n} \left\{ f(x) \stackrel{\text{def}}{=} \mathbb{E}\left[F(x,\xi)\right] \right\} \ ;$$

on construit à partir d'un N-échantillon de la variable aléatoire ξ le problème suivant :

$$\widehat{v}_N \stackrel{\text{def}}{=} \min_{x\in\mathbb{R}^n} \left\{ \widehat{f}_N(x) \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N} F(x,\xi^i) \right\}.$$

Alors pour tout $x \in \mathbb{R}^n$:

$$\mathbb{E}\left[\widehat{f}_N(x)\right] = f(x) \geq v^* = \inf_{x\in\mathbb{R}^n} \mathbb{E}\left[\widehat{f}_N(x)\right] \geq \mathbb{E}\left[\inf_{x\in\mathbb{R}^N} \widehat{f}_N(x)\right] = \mathbb{E}\left[\widehat{v}_N\right].$$

Autrement dit $\widehat{v}_N$ est un estimateur biaisé de v^*. D'autre part $\widehat{f}_N(x)$ peut être considérée comme une borne supérieure de v^*. A. Shapiro étudie dans ses articles les hypothèses sous lesquelles la suite des estimateurs $\widehat{v}_N$ converge[4] vers v^*. Les résultats sont illustrés à l'aide d'exemples notamment le problème de l'optimisation d'un critère linéaire sur un horizon fini.

[4]Différentes notions de convergences sont étudiées

I.3.4 Robustesse [33, 61, 77, 38, 10, 14]

Étudier la robustesse d'un problème d'optimisation passe généralement par établir des résultats de continuité de la fonction marginale d'un problème relativement à un paramètre. Cet aspect de l'optimisation est important pour la raison suivante : certains paramètres étant dans un espace de dimension infinie, c'est le cas par exemple des lois de probabilité, il est nécessaire de les discrétiser afin de résoudre numériquement le problème, mais il faut en plus être assuré que cette discrétisation est compatible avec l'objectif principal qui est le calcul de la solution d'un problème d'optimisation. Il est possible par exemple qu'une discrétisation de la fonction coût donne une approximation aussi précise que l'on souhaite de celle-ci, mais que quelle que soit la finesse de cette discrétisation, le coût optimal ne soit pas correctement approché (voir l'exemple du paragraphe §IV.4.1). Pour les problèmes d'optimisation stochastique ce paramètre peut être une loi de probabilité, ou bien un entier N qui contrôle la finesse d'une discrétisation. Dupačovà et Wets ont établi des théorèmes de robustesse donnant les hypothèses sous lesquelles un problème en boucle ouverte peut être approché par une suite de problèmes discrets obtenus en remplaçant la loi de la variable aléatoire ξ par une suite de lois discrètes.

Robustesse par rapport à la loi [47, 35, 72, 49, 60, 37, 36, 34]

Les propriétés de robustesse de la fonction valeur d'un problème d'optimisation en boucle ouverte par rapport à la loi des bruits est un domaine de recherche très actif. Les résultats issus de ce milieu donnent généralement des conditions sous lesquelles on peut espérer obtenir une fonction valeur lipschitzienne. Ainsi Pflug [72] montre que sous l'hypothèse que la fonction :

$$\xi \mapsto F(x, \xi) \ ;$$

soit uniformément lipschitzienne[5], la fonction valeur :

$$v(\mathbb{P}_\xi) \stackrel{\text{def}}{=} \min_{x \in \mathbb{R}^n} \left\{ f(x) \stackrel{\text{def}}{=} \mathbb{E}\left[F(x, \xi)\right] \right\} \ ;$$

est lipschitzienne, lorsque l'on munit l'espace des lois de probabilité sur $(\Omega, \mathcal{F})$ de la distance de Kantorovich-Rubinstein :

$$d(\mathbb{P}_\xi, \mathbb{P}_{\xi'}) \stackrel{\text{def}}{=} \inf_{X, X'} \left\{ \mathbb{E}\left[|X - X'|\right] \mid \mathbb{P}_X = \mathbb{P}_\xi \text{ et } \mathbb{P}_{X'} = \mathbb{P}_{\xi'} \right\}.$$

Les résultats de ce type sont, à notre connaissance, toujours obtenus pour des problèmes d'optimisation en boucle ouverte.

Robustesse par rapport à la structure d'information[10, 31, 3]

Les propriétés de robustesse par rapport à la structure d'information ont été, à notre connaissance, moins étudiées que la robustesse par rapport aux lois de probabilité. Artstein

[5] La constante de Lipschitz ne dépend pas de x.

[10] a montré que la fonction marginale :

$$V(\mathcal{B}) \stackrel{\text{def}}{=} \inf \{\mathbb{E}\left[J(\gamma(\xi),\xi)\right] \mid \gamma \text{ est } \mathcal{B}\text{-mesurable}\} \ ;$$

en continue lorsque l'on munit l'ensemble des sous-tribus de $\mathcal{F}$ la topologie de convergence forte[6]. Le même type de résultat a par ailleurs été obtenu par Cotter [31] et Allen [3].

I.3.5 Conditions d'optimalité [85, 55, 79, 89, 87, 86, 56]

Nous nous intéressons particulièrement aux conditions d'optimalité associées à des problèmes d'optimisation soumis à des contraintes de mesurabilité. Étant donné le problème suivant :

$$\mathcal{P} \left\{ \begin{array}{l} \min \mathbb{E}\left[J(\gamma(\xi),\xi)\right] \ ; \\ \gamma \text{ est } \mathcal{B}\text{-mesurable.} \end{array} \right.$$

Rockafellar [87] a montré que si $\gamma \in L^{\infty}_{\mathbb{R}^n}(\Xi)$ est solution du problème $\mathcal{P}$, il existe $\rho \in L^{1}_{\mathbb{R}^n}(\Xi)$ tel que $\mathbb{E}\left(\rho \mid \mathcal{B}\right) = 0$ et le problème $\mathcal{P}$ est équivalent (dans le cas convexe) à résoudre :

$$\min_{\gamma} \mathbb{E}\left[J(\gamma(\xi),\xi) + \langle \rho(\xi), \gamma(\xi)\rangle\right] \ ;$$

la difficulté étant surtout de montrer qu'il est possible de choisir un multiplicateur dans $L^{1}_{\mathbb{R}^n}(\Omega)$. Le même type de résultat peut être obtenu dans le cas Lipschitz [56]. Nous avons développé dans le chapitre VI une application de ces résultats à un problème de commande optimale stochastique en temps discret. L'intérêt de cette démarche réside dans l'expression des conditions d'optimalité obtenues. Nous avons en effet obtenu deux formulations équivalentes des conditions d'optimalité qui diffèrent par la nature des variables duales. Dans un cas nous avons des variables duales mesurables relativement à la filtration naturelle du bruit, dans l'autre cas nous avons des variables duales adaptées à l'observation.

I.4 Originalité du mémoire

Certains problèmes d'optimisation stochastique dont la fonction d'observation dépend des variables de décision et donc en première Analyse, faisant partie de la catégorie des problèmes avec une structure d'information dynamique, sont en fait des problèmes d'optimisation avec une structure d'information statique. Comme le souligne Varaiya et Wets [103], un enjeu majeur de l'optimisation stochastique est de pouvoir caractériser ce type de problèmes, ce qui revient quelque part à caractériser les problèmes statiques. Nous [19] avons tenté d'apporter une réponse à cette question dans le cas non trivial où la fonction d'observation dépend de la variable de décision. Notre contribution a été de déterminer des hypothèses suffisantes sous lesquelles il est possible de caractériser le plus grand ensemble de feedbacks admissibles pour lequel la tribu engendrée par la fonction d'observation est fixe même après y avoir injecté un feedback admissible.

[6] Voir la définition III.14.

Il n'est pas possible d'évaluer de combien une contrainte de mesurabilité à été violée en utilisant les normes habituelles en analyse. Si l'on munit l'intervalle réel $I = [0, 1]$ de sa tribu des boréliens, de la loi uniforme et que l'on considère les applications $h(x) = 2\varepsilon x$ et $b(x) = 0$ définies sur I. Nous avons que $||h - g||_{L^1_{\mathbb{R}}(I,\mathcal{B}_I)} = \varepsilon$, ce qui signifie que l'erreur moyenne entre g et h est d'ordre ε aussi petit que l'on veut. Malheureusement cette norme est catastrophique du point de vue de l'information, en effet $\sigma(h) = \mathcal{B}_I$ et $\sigma(g) = \{I, \emptyset\}$ autrement dit ces deux fonctions bien que très proches numériquement sont informativement très différentes. Nous avons donc étudié les différentes topologies permettant de rendre compte de la mesurabilité d'une variable aléatoire.

Nous avons par ailleurs étudié la discrétisation des problèmes d'optimisation stochastique ayant une structure d'information statique. Nous avons analysé les propriétés de continuité de la fonction valeur d'un tel problème par rapport à la structure d'information. Nous proposons une technique de discrétisation des contraintes de mesurabilité basée sur la *quantification* de la fonction d'observation . La quantification d'une variable aléatoire h définie sur Ω à valeurs dans un espace métrique $\mathcal{Y}$ est l'opération qui consiste à composer h avec une variable aléatoire discrète Q mesurable pour la tribu des boréliens sur $\mathcal{Y}$, on obtient ainsi une nouvelle variable aléatoire discrète $Q \circ h$. Cette technique de discrétisation possède a priori un avantage évident du point de vue des problèmes d'optimisation : lorsque l'on impose aux variables de décision d'être mesurables par rapport à la tribu $\sigma(Q \circ h)$ nous avons nécessairement que les décisions admissibles du problème discret le seront aussi pour le problème d'origine. Nous avons étudié les propriétés topologiques de la quantification et nous avons montré que si la suite $(Q_n)_{n\in\mathbb{N}}$ converge simplement vers l'identité alors la suite de sous-tribus $(\sigma(Q_n \circ h))_{n\in\mathbb{N}}$ converge au sens des tribus vers $\sigma(h)$. Partant de ce résultat nous proposons une approche en deux étapes pour discrétiser les problèmes avec une structure d'information statique :

1. la première étape passe par une discrétisation de la contrainte de mesurabilité par une technique de quantification ;
2. la deuxième étape est une approche numérique classique pour calculer un problème en boucle ouverte.

Nous montrerons que notre approche de la résolution numérique du problème est originale dans le sens où elle établit de manière explicite que l'erreur totale de la discrétisation que nous noterons e, est la somme de l'erreur e_1 issue de la discrétisation la structure d'information et de l'erreur e_2 issue de l'approximation du critère. À notre connaissance l'erreur e_1 commise lors de la discrétisation de la structure d'information n'a pas fait l'objet d'études mathématique très approfondies dans la littérature.

Par ailleurs notre approche est également originale dans la mesure où nous n'avons pas limité notre exposé aux contraintes de non anticipativité qui sont les contraintes de mesurabilité que l'on retrouve d'ordinaire dans la littérature. Autrement dit la méthodologie que nous proposons est suffisamment générale pour être adaptée à la plus part des problèmes statiques. Comme nous l'avons fait remarqué, la contrainte de non anticipativité est une structure d'information dont l'étude a été largement privilégiée. Par ailleurs, les problèmes sujets à cette contrainte ont la particularité suivante : la discrétisation de ces problèmes

peut se résumer simplement à la discrétisation des lois marginales du processus de bruit, c'est sans doute cette particularité qui explique pourquoi les deux sources d'erreurs que nous avons isolées sont généralement combinées au sein d'une même structure qui est l'arbre de scénarios.

I.5 Plan

Le mémoire se décompose de la manière suivante :

- Le rôle du chapitre I est de donner un bref aperçu de l'état de l'art concernant certains aspects de l'optimisation stochastique que nous allons étudier dans ce mémoire ;
- nous aborderons dans le chapitre II la question de la caractérisation de l'absence d'effet dual à partir d'outils algébriques ;
- nous allons introduire dans le chapitre III différentes notions topologiques sur l'ensemble des sous tribus d'une tribu $\mathcal{F}$;
- dans chapitre IV, nous allons montrer que la discrétisation d'un problème d'optimisation stochastique par une technique de quantification des contraintes de mesurabilité, permet d'obtenir asymptotiquement un estimateur du coût optimal du problème d'optimisation d'origine ;
- nous traiterons numériquement dans le chapitre V l'exemple d'un problème de gestion de barrage hydraulique ;
- nous allons traiter dans le chapitre VI des conditions d'optimalité d'un problème d'un commande optimale stochastique ;
- l'annexe A contient des rappels d'optimisation ;
- l'annexe B contient des rappels de la théorie des probabilités.

Chapitre II

L'effet dual en optimisation stochastique

II.1 Introduction

Dans certains cas l'observation d'un "système" peut dépendre des variables de décision agissant sur le dit système. Cette dépendance vis-à-vis des décisions peut perturber "informationnellement" le système, prenons un exemple :

Exemple II.1. Prenons l'exemple d'un réfrigérateur qui tombe en panne. On admet qu'il y a deux possibilités envisageables qui sont :

1. le propriétaire du réfrigérateur fait appel à un réparateur, cela lui coûte E_1. À nouveau il y a deux possibilités :
   - le réfrigérateur n'est pas réparable ; dans ce cas, il faut rajouter aux frais de déplacement du réparateur le coût d'un nouveau réfrigérateur soit E_2 ;
   - le réfrigérateur est réparable ; dans ce cas il faut tenir compte des frais de réparation soit E_3 ;
2. le propriétaire décide de ne rien faire, et cela ne lui coûte rien. Cependant la décision suivante est nécessairement l'achat d'un nouveau réfrigérateur dont le coût est E_2.

Nous faisons l'hypothèse naturelle que $E_3 + E_1 < E_2$, c'est-à-dire qu'il est plus avantageux de réparer son réfrigérateur (au moins dans l'immédiat !). Nous avons représenté sur la figure II.1 (qu'il faut lire de la gauche vers la droite) les différents cas possibles :

- l'évènement $\{A\}$ correspond à l'achat d'un réfrigérateur ;
- l'évènement $\{A^c\}$ correspond à la réparation du réfrigérateur.

Dans cet exemple le réfrigérateur est le "système" sur lequel le propriétaire peut agir, soit en appelant un réparateur, soit en le réparant directement. Certaines décisions prises par le propriétaire agissent sur l'information qu'il possède a priori. Ainsi, s'il décide d'appeler un réparateur cela aura une influence sur l'information qu'il possède sur l'état du réfrigérateur.

La figure II.1 est une représentation informationnelle du fonctionnement de ce système. Au départ on ne possède aucune information sur l'état du réfrigérateur $\{\emptyset, \Omega\}$. Les nœuds

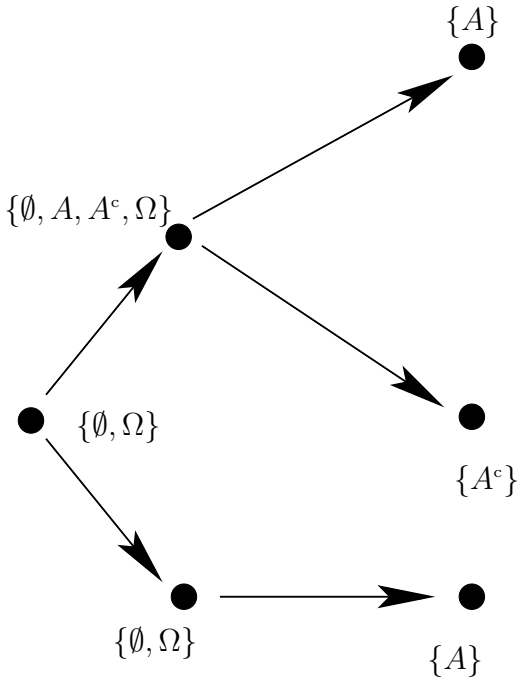


FIG. II.1 – Panne du réfrigérateur

du deuxième niveau de l'arbre de la figure II.1 ne sont pas informationnellement équivalents, c'est-à-dire que connaître $\{\Omega, \emptyset, A, A^c\}$ n'est pas équivalent à connaître $\{\Omega, \emptyset\}$. Pour savoir si le réfrigérateur en réparable $\{A^c\}$ il est nécessaire d'acheter de l'information !

Nous dirons qu'il n'y à pas d'effet dual lorsque l'observation d'un système est "informationnellement" indépendante de toute variable de décision.

On peut montrer (voir Quadrat-Viot [75, lemme 2.3]) qu'il n'y a pas d'effet dual pour les systèmes dynamiques linéaires, ($x_{t+1} = Ax_t + Bu_t + \omega_t$) dans le cas[1] ou l'observation instantanée est la variable $(x_0, \ldots, x_t)$. Les questions concernant l'effet dual, ne sont pas nouvelles. Witsenhausen [104] a montré à l'aide d'un contre-exemple linéaire quadratique gaussien devenu célèbre, que si l'observation instantanée se réduit à $x_t + v_t$, c'est-à-dire que l'on observe un état bruité, et que l'on ne garde pas en mémoire les observations passées, alors il y a effet dual, et nous ne sommes donc plus en mesure de résoudre notre problème d'optimisation par des techniques classiques. En effet dans ce dernier cas, il n'est plus possible d'estimer l'état du système grâce au filtre de Kalman puisque l'on ne doit pas utiliser les observations passées. L'absence d'effet dual implique que la structure d'observation est fixe pour le système bouclé, c'est donc un propriété numériquement intéressante. Il est donc important d'identifier les systèmes dynamiques possédant cette propriété.

Dans ce chapitre nous allons commencer par introduire des structures algébriques et nous montrerons à l'aide d'un exemple que le fait de boucler un système dynamique par une

[1]Les variables aléatoires ω_t sont supposées indépendantes gaussiennes, mais cela ne joue aucun rôle dans le résultat.

commande en feedback "perturbe" ce système. Puis, nous allons donner une caractérisation de l'absence d'effet dual à partir des contrôles constants, c'est-à-dire que nous allons montrer que s'il n'y a pas d'effet dual pour les contrôles constants alors il n'y en aura pas pour les contrôles en feedbacks sur l'observation. Nous nous intéresserons aussi aux différentes possibilités permettant de s'affranchir de l'hypothèse de mémoire parfaite.

II.2 Structures algébriques

Nous allons introduire quelques notions de structures algébriques classiques, l'objectif étant de s'affranchir de toute la technicité de la théorie de la mesure. Pour plus de détails voir G. Cohen P. Carpentier et J.-C Culioli[27].

II.2.1 Partition et relation d'équivalence

Soit Ω un ensemble de cardinal fini, $\mathrm{card}(\Omega)$ désignera le cardinal de Ω. On note 2^{Ω} la collection des sous-ensembles de Ω (incluant l'ensemble vide et lui même). Une partition σ est définie comme une famille $\{\Omega_i\}_{i\in I_\sigma}$, $-I_\sigma$ est pour l'instant un ensemble quelconque– de sous-ensembles de Ω tous non vides dont l'union recouvre Ω et disjoints deux à deux :

$$\sigma = \{\Omega_i\}_{i\in I_\sigma}\ ,\ \forall i, \Omega_i \in 2^{\Omega},\ \Omega_i \neq \emptyset,\ \cup_{i\in I_\sigma}\Omega_i = \Omega,\ i \neq j \Rightarrow \Omega_i \cap \Omega_j. = \emptyset.$$

L'ensemble des partitions de Ω est noté $\mathcal{P}(\Omega)$. Si $\mathrm{card}(I_\sigma) = n$ on dira aussi que $\mathrm{card}(\sigma) = n$. Sur l'ensemble $\mathcal{P}(\Omega)$, on peut définir la relation d'ordre suivante, soient :

$$\sigma = \{\Omega_i\}_{i\in I_\sigma} \text{ et } \sigma' = \{\Omega_j\}_{j\in I_{\sigma'}}\ ,\ \sigma' \preceq \sigma \Leftrightarrow \forall i \in I_\sigma,\ \exists j \in I_{\sigma'} : \Omega_i \in \Omega_j. \tag{II.1}$$

Muni de cette relation d'ordre, l'ensemble $\mathcal{P}(\Omega)$ est un treillis supérieur : la plus petite borne supérieure de deux partitions $\sigma = \{\Omega_i\}_{i\in I_\sigma}$ et $\sigma' = \{\Omega_i\}_{j\in I_{\sigma'}}$ notée $\sigma \vee \sigma'$, est constituée de tous les sous-ensembles obtenus en gardant toutes les intersections non vides $\Omega_i \cap \Omega_j^{'}$ pour $i \in I_\sigma$ et $j \in I_{\sigma'}$

Définition II.2. Soient σ et σ' deux partitions telles que $\sigma' \preceq \sigma$ alors, il existe une application $p : I_\sigma \to I_{\sigma'}$ (et par abus de langage $p : \sigma \to \sigma'$) dite "application père", définie de la manière suivante :

$$\forall i \in I_\sigma,\ \exists j = p(i) \in I_{\sigma'} : \Omega_i \subset \Omega_j^{'}.$$

Évidemment ce j est unique. On appellera la "multi-application fils" la multi-application $f = p^{-1}$, en rappelant que, pour une application quelconque $h : \mathcal{A} \to \mathcal{B}$, on définit la multi-application inverse h^{-1} par :

$$h^{-1} : \mathcal{B} \to \mathcal{A}, b \to h^{-1}(b) \stackrel{\text{def}}{=} \{a \in \mathcal{A} \mid h(a) = b\}.$$

Étant donnée une application h de Ω dans $\mathcal{Y}$, on peut aussi lui associer la relation d'équivalence suivante $\overset{h}{\equiv}$:

$$\forall \omega \in \Omega, \forall \omega' \in \Omega,\ \omega \overset{h}{\equiv} \omega' \Leftrightarrow h(\omega) = h(\omega') \Leftrightarrow \omega' \in h^{-1}(h(\omega)).$$

L'ensemble quotient de Ω par cette relation d'équivalence est noté Ω/h. La classe d'équivalence engendrée par $\omega \in \Omega$ (qui est un élément de Ω/h), est noté $[\omega]_h$. On le notera alors aussi σ_h. Dans la mesure où les classes d'équivalence constituent une partition de Ω, on peut considérer Ω/h comme un élément de $\mathcal{P}(\Omega)$.

II.2.2 Relation de précédence

La relation d'ordre (II.1) induit une relation de pré-ordre sur toutes les applications ayant pour domaine Ω et leur image dans des ensembles quelconques. Une relation de pré-ordre est une relation réflexive et transitive, mais non anti-symétrique. La relation de pré-ordre entre h_1 et h_2 notée $h_1 \preceq h_2$, sera définie par la relation d'ordre $\sigma_{h_1} \preceq \sigma_{h_2}$.

Définition II.3. Soient $\sigma \in \mathcal{P}(\Omega)$ et $h : \Omega \to \mathcal{Y}$: on dit que h est mesurable par rapport à σ si $\sigma_h \preceq \sigma$. Soient $h_1 : \Omega \to \mathcal{Y}_1$ et $h_2 : \Omega \to \mathcal{Y}_2$: on dit que h_1 est mesurable par rapport à h_2 si h_1 est mesurable par rapport à σ_{h_2}

Définition II.4. Soit $h : \Omega \to \mathcal{Y}$, alors :

$$\mathrm{im}h = \{y \in \mathcal{Y} \mid \exists \omega \in \Omega,\ y = h(\omega)\}.$$

Proposition II.5. *Les énoncés suivants sont équivalents :*

(i) h_1 *est mesurable par rapport à* h_2 *(notée* $h_1 \preceq h_2$*) ;*

(ii) $\sigma_{h_1} \preceq \sigma_{h_2}$ *;*

(iii) on a l'implication :

$$\forall (\omega, \omega') \in \Omega \times \Omega, \quad h_2(\omega) = h_2(\omega') \Rightarrow h_1(\omega) = h_1(\omega')\ ;$$

(iv) il existe une application $p :\ \mathrm{im}h_2 \to\ \mathrm{im}h_1$ *telle que* $h_1 = p \circ h_2$*.*

Remarque II.6. L'expression "h_1 mesurable par rapport à h_2" signifie que h_1 est moins informative que h_2.

Définition II.7. Deux applications $h_1 : \Omega \to \mathcal{Y}_1$ et $h_2 : \Omega \to \mathcal{Y}_2$ sont aussi informatives l'une que l'autre (notée $h_1 \equiv h_2$) si h_1 est mesurable par rapport à h_2 et réciproquement. Autrement dit,

$$h_1 \equiv h_2 \Leftrightarrow (h_1 \preceq h_2 \text{ et } h_2 \preceq h_1)$$

Définition II.8. Étant données deux applications $h_1 : \Omega \to \mathcal{Y}_1$ et $h_2 : \Omega \to \mathcal{Y}_2$, on désigne par $h_1 \vee h_2$ un représentant quelconque de la classe d'équivalence contenant l'élément :

$$\begin{array}{rcl} \Omega & \to & \mathcal{Y}_1 \times \mathcal{Y}_2, \\ \omega & \mapsto & (h_1(\omega), h_2(\omega)). \end{array}$$

II.3 Effets d'un feedback sur l'information

Nous allons introduire de manière assez complète les effets d'un feedback sur l'information, et montrer qu'ils ne sont pas triviaux, ils dépendent des feedbacks, de la structure d'information, et de l'ensemble admissible.

On considère $\Omega = \{1, 2, 3, 4\}$, $\mathcal{U} = \{1, 2, 3\}$, $\mathcal{Y} = \{1, 2, 3\}$ et

$$h(u, \omega) = \min(u, \omega) \ .$$

Les feedbacks $\gamma : \Omega \to \mathcal{U}$ et les observations qu'ils engendrent $\eta^\gamma \equiv h(\gamma(\cdot), \cdot) : \Omega \to \mathcal{Y}$ sont décrits par des quadruplets de nombres entiers compris entre 1 et 3. Par exemple, $\gamma = 3221$ signifie que $\gamma(1) = 3$, $\gamma(2) = 2$, $\gamma(3) = 2$, $\gamma(4) = 1$.

Le treillis des partitions de l'ensemble $\Omega = \{1, 2, 3, 4\}$ comporte 15 éléments représentés sur la Figure II.2.

Il y a $3^4 = 81$ feedbacks possibles qui se répartissent sur ce treillis des partitions. Le sommet (noté $\top$ — "top") du treillis n'est pas atteignable puisqu'on ne peut avoir de fonctions injectives de Ω dans $\mathcal{U}$ étant donné que Ω est de cardinal 4 et $\mathcal{U}$ est de cardinal 3. Le Tableau II.1 donne la liste de tous les feedbacks possibles classés selon leur appartenance au treillis des partitions. Les colonnes de droite indiquent d'une part les fonctions η^γ correspondantes, d'autre part l'admissibilité du feedback (vérification de la condition $\gamma(\cdot) \preceq h(\gamma(\cdot), \cdot)$) et enfin la partition vers laquelle pointe la fonction η^γ (en nous limitant uniquement aux lignes correspondant à des feedbacks admissibles).

À partir de ce tableau exhaustif, on peut faire les observations suivantes. Observons d'abord qu'il s'agit ici d'un exemple sans principe de séparation puisque les η^γ (même en se restreignant seulement à ceux correspondant à des feedbacks admissibles) engendrent des éléments différents du treillis des partitions de Ω.

Deux feedbacks informationnellement équivalents — même admissibles — ne produisent pas des structures d'information équivalentes

Cette affirmation est illustrée par les trois feedbacks constants (premiers éléments du tableau), tous admissibles, qui engendrent des fonctions η^γ pointant vers les éléments 1, 3, 9 (situés à trois niveaux différents) du treillis des partitions.

Évidemment, dans le cas du principe de séparation, tous les feedbacks admissibles produisent la même structure d'information, donc la question ne se pose pas.

Pour deux feedbacks informationnellement équivalents, l'un peut être admissible sans que l'autre le soit

Parmi les 6 feedbacks rattachés à l'élément 2 du treillis des partitions, 3 sont admissibles et 3 ne le sont pas.

La question de savoir si un feedback informationnellement équivalent à un feedback admissible est admissible dans le cas où le principe de séparation est vérifié reste ouverte.

TAB. II.1 – Feedbacks

No.	part. γ	γ	η^γ	adm. ?	part. η^γ
1		1111	1111	O	1
2	1	2222	1222	O	3
3		3333	1233	O	9
4		1122	1122	O	2
5		1133	1133	O	2
6	2	2211	1211	N	
7		2233	1233	O	9
8		3311	1211	N	
9		3322	1222	N	
10		1222	1222	O	3
11		1333	1233	O	9
12	3	2111	1111	N	
13		2333	1233	O	9
14		3111	1111	N	
15		3222	1222	O	3
16		1212	1212	O	4
17		1313	1213	O	14
18	4	2121	1121	N	
19		2323	1223	N	
20		3131	1131	N	
21		3232	1232	O	10
22		1121	1121	O	5
23		1131	1131	O	5
24	5	2212	1212	N	
25		2232	1232	O	10
26		3313	1213	N	
27		3323	1223	N	
28		1221	1221	O	6
29		1331	1231	O	13
30	6	2112	1112	N	
31		3113	1113	N	
32		2332	1232	N	
33		3223	1223	O	11
34		1211	1211	O	7
35		1311	1211	O	7
36	7	2122	1122	N	
37		2322	1222	N	
38		3133	1133	N	
39		3233	1233	O	9
40		1112	1112	O	8
41		1113	1113	O	8
42	8	2221	1221	N	
43		2223	1223	O	11
44		3331	1231	N	
45		3332	1232	N	
46		1233	1233	O	9
47		1322	1222	N	
48	9	2133	1133	N	
49		2311	1211	N	
50		3122	1122	N	
51		3211	1211	N	
52		1232	1232	O	10
53		1323	1223	N	
54	10	2131	1131	N	
55		3121	1121	N	
56		3212	1212	N	
57		2313	1213	N	
58		1223	1223	O	11
59		1332	1232	N	
60	11	2113	1113	N	
61		3112	1112	N	
62		3221	1221	N	
63		2331	1231	N	
64		1123	1123	O	12
65		1132	1132	O	12
66	12	2213	1213	N	
67		2231	1231	N	
68		3312	1212	N	
69		3321	1221	N	
70		1231	1231	O	13
71		1321	1221	N	
72	13	2132	1132	N	
73		2312	1212	N	
74		3123	1123	N	
75		3213	1213	N	
76		1213	1213	O	14
77		1312	1212	N	
78	14	2123	1123	N	
79		2321	1221	N	
80		3132	1132	N	
81		3231	1231	N	

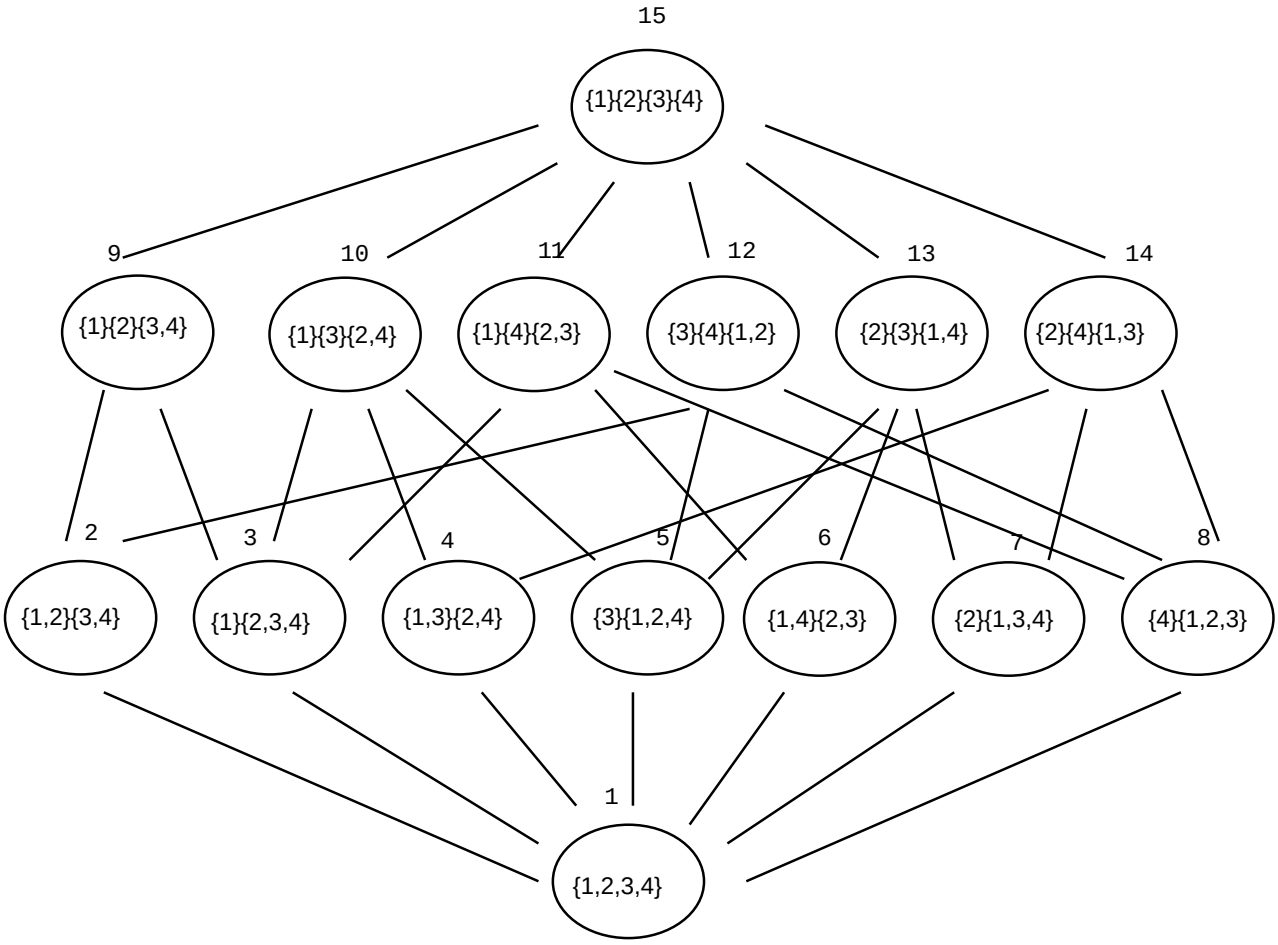


FIG. II.2 – Treillis des partitions de $\Omega = \{1, 2, 3, 4\}$

Si γ est admissible et $\gamma' \preceq \gamma$, γ' n'est pas nécessairement admissible

Par exemple, $\gamma_6 = 2211$ n'est pas admissible mais est rattaché à l'élément 2 du treillis des partitions, alors que $\gamma_{46} = 1233$ est admissible : ce feedback est rattaché à l'élément 9 du treillis des partitions et on a $\gamma_6 \preceq \gamma_{46}$.

Les informations ne varient pas de façon monotone avec les feedbacks

Par exemple, le feedback $\gamma_3 = 3333$, rattaché à l'élément 1 du treillis des partitions, engendre $\eta^{\gamma_3} = 1233$, rattaché à l'élément 9 du treillis des partitions. Par ailleurs, $\gamma_{10} = 1222$ est rattaché à l'élément 3 du treillis des partitions (on a donc $\gamma_3 \preceq \gamma_{10}$) mais $\eta^{\gamma_{10}}$ est rattaché à l'élément 3 du treillis des partitions (donc $\eta^{\gamma_{10}} \preceq \eta^{\gamma_3}$).

II.4 Effet dual : définition et caractérisation

Nous allons énumérer dans cette section les principaux résulats que nous avons obtenus dans [19] en termes de caractérisation de l'absence d'effet dual.

II.4.1 Absence de temps

Définition II.9. Soit $h : \mathcal{U} \times \Omega \to \mathcal{Y}$ une application, on appelle $\mathbb{F}_{\mathcal{U}}^{ad}(h)$ l'ensemble des feedbacks admissibles défini de la manière suivante :

$$\mathbb{F}_{\mathcal{U}}^{ad}(h) \stackrel{\text{def}}{=} \{\gamma : \Omega \to \mathcal{U} \mid \gamma(\cdot) \preceq h(\gamma(\cdot), \cdot)\}\,.$$

Définition II.10. On dit qu'il n'y a pas d'effet dual en boucle ouverte pour une fonction d'observation $h : \Omega \times \mathcal{U} \to \mathcal{Y}$ si :

$$\forall (u, , u') \in \mathcal{U} \times \mathcal{U}, \quad h(u, \cdot) \equiv h(u', \cdot)\ ;$$

dans ce cas on note ξ une application quelconque définie sur Ω telle que :

$$\forall u \in \mathcal{U} \quad h(u, \cdot) \equiv \xi.$$

Notre objectif est de déterminer une classe de fonctions plus large que $\perp$ et pour laquelle la propriété d'absence d'effet dual reste valide. C'est la raison pour laquelle nous introduisons la définition suivante :

Définition II.11. On suppose qu'il n'y a pas d'effet dual pour les feedbacks en boucle ouverte, alors le plus grand ensemble pour lequel nous n'avons pas d'effet dual est défini par :

$$\mathbb{F}_{\mathcal{U}}^{\text{nde}} = \left\{\gamma \in \mathbb{F}_{\mathcal{U}}^{\text{ad}}(h) \mid h(\gamma(\cdot), \cdot) \equiv \xi\right\}.$$

Définition II.12. Soit $h : \mathcal{U} \times \Omega \to \mathcal{Y}$ une fonction d'observation pour laquelle il n'y a pas d'effet dual pour les contrôles en boucle ouverte. On définit alors l'ensemble $\mathbb{F}_{\mathcal{U}}^{\xi}$ de la manière suivante :

$$\mathbb{F}_{\mathcal{U}}^{\xi} \stackrel{\text{def}}{=} \{\gamma : \Omega \to \mathcal{U} \mid \gamma(\cdot) \preceq \xi(\cdot)\}\,.$$

Nous sommes en mesure de donner une caractérisation de l'ensemble $\mathbb{F}_{\mathcal{U}}^{\text{nde}}(h)$ à partir des éléments de $\perp$.

Théorème II.13. *Si il n'y a pas d'effet dual en boucle ouverte pour une fonction d'observation h, alors :*

$$\mathbb{F}_{\mathcal{U}}^{\text{nde}}(h) = \mathbb{F}_{\mathcal{U}}^{\text{ad}}(h) \cap \mathbb{F}_{\mathcal{U}}^{\xi}.$$

Preuve : Voir [19, proposition 14]. □

II.4.2 Introduction du temps

Soit un ensemble $\Omega \subset \mathcal{A}^{T+1}$. L'ensemble des décisions est noté $\mathcal{U}^T \stackrel{\text{def}}{=} \underbrace{\mathcal{U} \times \ldots \times \mathcal{U}}_{T \text{ fois}}$.

Définition II.14. Pour $t = 0, \ldots, T-1$, on définit l'opérateur préfixe noté ρ_t :

$$\rho_t : \mathcal{U}^T \to \mathcal{U}^{t+1}, \quad \rho_t(u) \stackrel{\text{def}}{=} (u_0, \ldots, u_t).$$

Définition II.15. Une famille de fonctions d'observation $\{h_t\}_{t=0,\ldots,T-1}$, est une suite d'applications de la forme :

$$h_t : \mathcal{U}^T \times \Omega \to \mathcal{Y}_t.$$

On notera $h \stackrel{\text{def}}{=} \{h_t\}_{t=0,\ldots,T-1}$.

Définition II.16. Une famille de fonctions d'observation possède la propriété dite de mémoire parfaite si :

$$h_t(\cdot,\cdot) \preceq_{\mathcal{U}^T \times \Omega} h_{t+1}(\cdot,\cdot).$$

Définition II.17. Une famille $\{h_t\}_{t=0,\ldots,T-1}$ de fonctions d'observation est dite causale si elle vérifie la propriété suivante :

$$\forall t = 0, \ldots, T-1, \quad h_t(\cdot,\cdot) \preceq_{U^T \times \Omega} \rho_{t-1}(\cdot) \vee \mathrm{I}_{\mathrm{d}\Omega}(\cdot).$$

Par abus de notation on écrira indifféremment pour une fonction causale :

$$\forall u \in \mathcal{U}^T, \quad \forall i \geq t-1, \quad h_t(u,\omega) = h_t(\rho_i(u),\omega).$$

Définition II.18. On dit qu'il n'y a pas d'effet dual pour les contrôles en boucle ouverte pour une famille $\{h_t\}_{t=0,\ldots,T-1}$ de fonctions d'observation si pour tout t :

$$\forall (u,u') \in \mathcal{U}^T \times \mathcal{U}^T, \quad h_t(u,\cdot) \equiv h_t(u',\cdot),$$

dans ce cas on note $\{\xi_t\}_{t=0,\ldots,T-1}$ une famille quelconque d'applications définies sur Ω ayant la propriété suivante :

$$\forall u \in \mathcal{U}^T, \quad h_t(u,\cdot) \equiv \xi_t.$$

Définition II.19. Soient $T \in \mathbb{N}$ et $h = \{h_t\}_{t=0,\ldots,T-1}$ une famille de fonctions d'observation, on définit :

1. l'ensemble des feedbacks admissibles :

$$\mathbb{F}^{\text{ad}}(h) \stackrel{\text{def}}{=} \left\{\gamma : \Omega \to \mathcal{U}^T \mid \forall t = 0, \ldots, T-1 \quad \gamma_t \preceq h_t(\gamma(\cdot),\cdot)\right\} \ ; \tag{II.2}$$

2. dans le cas où il n'y a pas d'effet dual pour les contrôles en boucle ouverte, le plus grand ensemble sur lequel il n'y a pas d'effet dual que l'on notera $\mathbb{F}^{\text{nde}}$, de la manière suivante :

$$\mathbb{F}^{\text{nde}} \stackrel{\text{def}}{=} \left\{\gamma \in \mathbb{F}^{\text{ad}}(h) \mid \forall t = 0, \ldots, T-1 \quad h_t(\gamma(\cdot),\cdot) \equiv \xi_t\right\} \ ;$$

nous remarquons qu'il n'y à pas de symétrie entre la définition de l'ensemble $\mathbb{F}^{\text{nde}}$ dans le cas dynamique et celle du cas statique. La raison est naturelle, dans le cas séquentiel, l'information arrive de manière progressive au fur et à mesure que le temps s'écoule, il en va de même pour les décisions. Ainsi, les observations ne peuvent dépendre que des commandes passées, c'est ce que nous appelons la causalité, de même une commande à un instant donné ne peut pas dépendre d'observations ayant lieu à des instants ultérieurs. Il n'aurait pas été possible de rendre compte d'une telle structure d'information, en gardant le formalisme du cas statique.

3. soit $\left\{\widehat{\xi}_t\right\}_{t=0,\dots,T-1}$ une famille d'application définie sur Ω alors :

$$\mathbb{F}^{\widehat{\xi}} \stackrel{\text{def}}{=} \left\{\gamma : \Omega \to \mathcal{U}^T \mid \forall t = 0, \dots, T-1 \quad \gamma_t \preceq \widehat{\xi}_t\right\}. \tag{II.3}$$

Proposition II.20. *Soit $\{h_t\}_{t=0,\dots,T-1}$ une famille causale de fonctions d'observation à mémoire parfaite, ne possédant pas d'effet dual en boucle ouverte alors :*

$$\mathbb{F}^{\text{nde}} = \left\{\gamma : \Omega \to \mathcal{U}^T \mid \forall t = 0, \dots, T-1 \quad \gamma_t(\cdot) \preceq \xi_t(\cdot)\right\}.$$

Preuve : Voir [19, proposition 35]. □

II.5 Effet dual et mémoire parfaite

La caractérisation de l'absence d'effet dual obtenue dans [19], laisse ouverte la question suivante : l'hypothèse de mémoire parfaite est elle nécessaire ? Bien que cette hypothèse semble être un élément essentiel dans les démonstrations, il n'a pas été établi qu'elle était nécessaire. Nous allons tenter d'apporter quelques éléments de réponse à cette question.

II.5.1 Effet dual sans mémoire parfaite

On se pose la question de savoir ce qui est préservé lorsque l'on supprime l'hypothèse de mémoire parfaite.

Proposition II.21. *Soit $\{h_t\}_{t=0,\dots,T-1}$ une famille causale de fonctions d'observation. On suppose qu'il n'y a pas d'effet dual pour les contrôles en boucle ouverte. Alors :*

$$\left\{\gamma : \Omega \to \mathcal{U}^T \mid \forall t = 0, \dots, T-1 \quad \rho_t \circ \gamma(\cdot) \in \mathbb{F}^{\text{ad}}_{\mathcal{U}^T}(h_t) \cap \mathbb{F}^{\xi_t}_{\mathcal{U}^T}\right\} \subset \mathbb{F}^{\text{nde}}.$$

Preuve : On se donne $\gamma \in \left\{\gamma : \Omega \to \mathcal{U}^T \mid \forall t = 0, \dots, T-1 \quad \rho_t \circ \gamma(\cdot) \in \mathbb{F}^{\text{ad}}_{\mathcal{U}^T}(h_t) \cap \mathbb{F}^{\xi_t}_{\mathcal{U}^T}\right\}$, étant donné l'hypothèse d'absence d'effet dual en boucle ouverte pour la famille de fonctions $\{h_t\}_{t=0,\dots,T-1}$ nous avons d'après le théorème II.13 que :

$$\forall t = 0, \dots, T-1 \quad h_t(\rho_t \circ \gamma(\cdot), \cdot) \equiv \xi_t(\cdot).$$

Puisque la famille $\{h_t\}_{t=0,\dots,T-1}$ est causale, alors :

$$\forall t = 0, \dots, T-1 \quad h_t(\gamma(\cdot), \cdot) \equiv \xi_t(\cdot).$$

Ce qui montre que :

$$\left\{\gamma : \Omega \to \mathcal{U}^T \mid \forall t = 0, \dots, T-1 \quad \rho_t \circ \gamma(\cdot) \in \mathbb{F}^{\mathrm{ad}}_{\mathcal{U}^T}(h_t) \cap \mathbb{F}^{\xi_t}_{\mathcal{U}^T}\right\} \subset \mathbb{F}^{\mathrm{nde}}.$$

□

Définition II.22. Une famille causale $\{h_t\}_{t=0,\dots,T-1}$ de fonctions d'observation vérifie l'hypothèse (H) si pour tout $\gamma \in \mathbb{F}^{\mathrm{ad}}(h)$ nous avons :

$$\forall t = 0, \dots, T-1, \quad \rho_t \circ \gamma \preceq h_t(\gamma(\cdot), \cdot).$$

L'hypothèse (H) signifie en quelque sorte que la variable aléatoire $h_t(\gamma(\cdot), \cdot)$ permet de reconstituer l'ensemble des feedbacks antérieurs à t qui ont été injectés dans le système. Cette hypothèse restreint l'ensemble des commandes admissibles puisque :

$$\forall t = 0, \dots, T-1 \quad (\gamma_0, \dots, \gamma_t) \preceq h_t(\gamma(\cdot), \cdot) \Rightarrow \forall t = 0, \dots, T-1 \quad \gamma_t \preceq h_t(\gamma(\cdot), \cdot),$$

mais il n'y a pas d'équivalence.

Proposition II.23. *Soit $h = \{h_t\}_{t=0,\dots,T-1}$ une famille de fonctions d'observation ayant la propriété de mémoire parfaite, alors cette famille vérifie (H).*

Preuve : La propriété de mémoire parfaite est la suivante :

$$\forall t = 0, \dots, T-1 \quad h_t(\cdot, \cdot) \preceq_{\mathcal{U}^T \times \Omega} h_{t+1}(\cdot, \cdot)$$

Soit $\gamma \in \mathbb{F}^{\mathrm{ad}}(h)$ alors :

$$\forall t = 0, \dots, T-1 \quad \gamma_t(\cdot) \preceq h_t(\gamma(\cdot), \cdot).$$

Soit t fixé alors pour tout entier i inférieur à t :

$$\gamma_i \preceq h_i(\gamma(\cdot), \cdot) \preceq h_{i+1}(\gamma(\cdot), \cdot) \preceq \dots \preceq h_t(\gamma(\cdot), \cdot).$$

Donc :

$$(\gamma_0, \dots, \gamma_t) \preceq h_t(\gamma(\cdot), \cdot).$$

□

L'hypothèse (H) est strictement plus faible que l'hypothèse de mémoire parfaite, pour le voir considérons l'exemple suivant :

Exemple II.24. Considérons l'exemple suivant :

$$h_0(u,\omega) = \omega_0, \tag{II.4}$$
$$h_1(u,\omega) = u_0. \tag{II.5}$$

Nous avons clairement pour tout $\gamma : \Omega \to \mathcal{U}^T$ que $\gamma_0 \preceq \omega_0$, et $\gamma_0 \preceq \gamma_0$, donc (H) est vérifiée. Par ailleurs il est clair que $h_0(\cdot,\cdot) \not\preceq_{\mathcal{U}^T\times\Omega} h_1(\cdot,\cdot)$.

Proposition II.25. *Soit $\{h_t\}_{t=0,\dots,T-1}$ une famille causale de fonctions d'observation. On suppose que (H) est vérifié et qu'il n'y a pas d'effet dual en boucle ouverte, alors :*

$$\mathbb{F}^{\mathrm{nde}} = \left\{\gamma : \Omega \to \mathcal{U}^T \mid \forall t = 0,\dots,T-1 \quad \rho_t \circ \gamma(\cdot) \in \mathbb{F}^{\mathrm{ad}}_{\mathcal{U}^T}(h_t) \cap \mathbb{F}^{\xi_t}_{\mathcal{U}^T}\right\}.$$

Preuve : Soit $\gamma \in \mathbb{F}^{\mathrm{nde}}$ alors $\gamma \in \mathbb{F}^{\mathrm{ad}}(h)$, donc l'hypothèse (H) est vérifiée, ce qui implique que $\rho_t \circ \gamma \in \mathbb{F}^{\mathrm{ad}}_{\mathcal{U}^T}(h_t)$. Par définition de $\mathbb{F}^{\mathrm{nde}}$ nous avons que :

$$h_t(\gamma(\cdot),\cdot) \equiv \xi_t,$$

ce qui combiné avec (H) donne que $\rho_t \circ \gamma \in \mathbb{F}^{\xi_t}_{\mathcal{U}^T}$. Donc :

$$\mathbb{F}^{\mathrm{nde}} \subset \left\{\gamma : \Omega \to \mathcal{U}^T \mid \forall t = 0,\dots,T-1 \quad \rho_t \circ \gamma(\cdot) \in \mathbb{F}^{\mathrm{ad}}_{\mathcal{U}^T}(h_t) \cap \mathbb{F}^{\xi_t}_{\mathcal{U}^T}\right\}.$$

L'inclusion inverse nous est donnée par la proposition II.21. □

Le résultat suivant est très fort, la caractérisation que nous allons donner de $\mathbb{F}^{\mathrm{nde}}$ ne fera pas intervenir l'ensemble $\mathbb{F}^{\mathrm{ad}}(h)$ source de nombreuses non-linéarités.

Théorème II.26. *Soit $\{h_t\}_{t=0,\dots,T-1}$ une famille causale de fonctions d'observation. On suppose que (H) est vérifié et qu'il n'y a pas d'effet dual en boucle ouverte, alors :*

$$\mathbb{F}^{\mathrm{nde}} = \left\{\gamma : \Omega \to \mathcal{U}^T \mid \rho_t \circ \gamma(\cdot) \preceq \xi_t\right\}.$$

Preuve : Soit $\gamma : \Omega \to \mathcal{U}^T$ tel que pour tout $t \quad \rho_t \circ \gamma \preceq \xi_t$. Il est clair compte tenu de l'hypothèse de causalité que $\gamma_0 \preceq h_0$, ou bien de manière équivalente que $\rho_0 \circ \gamma \preceq h_0$. Donc, $\rho_0 \circ \gamma \in \mathbb{F}^{\mathrm{ad}}(h)$, ce qui combiné avec l'hypothèse (H) donne :

$$\forall t = 0,\dots,T-1 \quad \rho_t \circ \rho_0 \circ \gamma \preceq h_t(\rho_0 \circ \gamma(\cdot),\cdot)\ ;$$

ou plus simplement :

$$\forall t = 0,\dots,T-1 \quad \gamma_0 \preceq h_t(\rho_0 \circ \gamma(\cdot),\cdot).$$

En particulier nous avons montré que :

$$\rho_0 \circ \gamma \in \mathbb{F}^{\mathrm{ad}}_{\mathcal{U}^T}(h_1) \cap \mathbb{F}^{\xi_1}_{\mathcal{U}^T}.$$

Par une application du théorème II.13 nous avons alors que $h_1(\rho_0 \circ \gamma(\cdot),\cdot) \equiv \xi_1$, ce qui compte tenue de l'hypothèse de causalité nous donne $h_1(\gamma(\cdot),\cdot) \equiv \xi_1$. Donc :

$$h_1(\gamma(\cdot),\cdot) \equiv \xi_1 \text{ et } \gamma_1 \preceq \xi_1 \Rightarrow \gamma_1 \preceq h_1(\gamma(\cdot),\cdot).$$

Supposons que $\forall i = 0, \dots, t \quad h_i(\gamma(\cdot), \cdot) \equiv \xi_i$, posons $\tilde{\gamma} = \rho_t \circ \gamma$, il est clair que $\tilde{\gamma} \in \mathbb{F}^{\mathrm{ad}}(h)$, alors :

$$H \Rightarrow \forall k = 0, \dots, T-1 \quad \rho_k \circ \tilde{\gamma} \preceq h_k(\tilde{\gamma}(\cdot), \cdot).$$

En particulier $\tilde{\gamma} \in \mathbb{F}^{\mathrm{ad}}_{\mathcal{U}^T}(h_{t+1}) \cap \mathbb{F}^{\xi_{t+1}}_{\mathcal{U}^T}$, en effet :

$$\tilde{\gamma} = \rho_t \circ \gamma = \rho_{t+1} \circ \rho_t \circ \gamma = \rho_{t+1} \circ \tilde{\gamma} \preceq h_{t+1}(\tilde{\gamma}(\cdot), \cdot) \ ;$$

et par définition de γ :

$$\tilde{\gamma} = \rho_t \circ \gamma \preceq \rho_{t+1} \circ \gamma \preceq \xi_{t+1}.$$

Par une application du théorème II.13 nous avons alors que $h_{t+1}(\tilde{\gamma}(\cdot), \cdot) \equiv \xi_{t+1}$. Par définition de $\tilde{\gamma}$ et compte tenue de l'hypothèse de causalité :

$$h_{t+1}(\gamma(\cdot), \cdot) \equiv \xi_{t+1}.$$

Autrement dit nous avons montré que pour $t = 0, \dots, T-1 \quad h_t(\gamma(\cdot), \cdot) \equiv \xi_t$, ce combiné avec le fait que $\rho_t \circ \gamma \preceq \xi_t$ nous donne aussi que :

$$\rho_t \circ \gamma \preceq h_t(\gamma(\cdot), \cdot).$$

En résume nous avons montré que :

$$\left\{\gamma : \Omega \to \mathcal{U}^T \mid \rho_t \circ \gamma \preceq \xi_t\right\} \subset \mathbb{F}^{\mathrm{nde}}.$$

Établissons l'inclusion inverse, soit $\gamma \in \mathbb{F}^{\mathrm{nde}}$, alors $\gamma \in \mathbb{F}^{\mathrm{ad}}(h)$, donc (H) est vérifiée ce qui implique que :

$$\forall t = 0, \dots, T-1 \quad \rho_t \circ \gamma \preceq h_t(\gamma(\cdot), \cdot).$$

Mais puisque $\gamma \in \mathbb{F}^{\mathrm{nde}}$ alors $h_t(\gamma(\cdot), \cdot) \equiv \xi_t$ donc $\rho_t \circ \gamma \preceq \xi_t$. □

Reprenons l'exemple II.24 alors dans ce cas :

$$\mathbb{F}^{\mathrm{ad}}(h_1) = \left\{\gamma = (\gamma_0, \gamma_1) \mid \gamma \preceq \gamma_0(\cdot)\right\}, \quad \mathbb{F}^{\xi_1} = \left\{\gamma = (\gamma_0, \gamma_1) \mid \gamma(\cdot) \preceq \perp\right\},$$

et :

$$\mathbb{F}^{\mathrm{ad}}(h_1) \cap \mathbb{F}^{\xi_1} = \left\{\gamma = (\gamma_0, \gamma_1) \mid \gamma(\cdot) \preceq \perp\right\}.$$

D'après la proposition II.25 nous que $(\gamma_0, \gamma_1) \in \mathbb{F}^{\mathrm{ad}}(h_1) \cap \mathbb{F}^{\xi_1}$ donc nécessairement $\mathbb{F}^{\mathrm{nde}} = \perp$.

II.5.2 Effet dual et mémoire parfaite en boucle ouverte

Nous allons étudier les propriétés qui sont préservées lorsque l'on remplace l'hypothèse (H) de la définition II.22 par une hypothèse de mémoire parfaite pour les contrôles constants.

Définition II.27. Une famille de fonctions d'observation, $\{h_t\}_{t=0,\dots,T-1}$ est dite à mémoire parfaite en boucle ouverte si :

$$\forall t = 0, \dots, T-1 \quad \forall u \in \mathcal{U}^T \quad h_t(u, \cdot) \preceq h_{t+1}(u, \cdot).$$

Proposition II.28. *Soit $\{h_t\}_{t=0,\dots,T-1}$ une famille causale de fonctions d'observation, à mémoire parfaite en boucle ouverte alors :*

$$\forall \gamma \in \mathbb{F}^{\mathrm{ad}}(h), \forall t = 0, \dots, T-1 \quad (\gamma_0, \dots, \gamma_t) \preceq \bigvee_{u \in \mathcal{U}^T} h_t(u, \cdot).$$

Preuve : Soit t fixé, posons $\Lambda_t \stackrel{\text{def}}{=} \bigvee_{u \in \mathcal{U}^T} h_t(u, \cdot)$ alors :

$$\forall u \in \mathcal{U}^T \quad h_t(u, \cdot) \preceq \Lambda_t(\cdot),$$

l'hypothèse de mémoire parfaite en boucle ouverte donne :

$$\forall i = 0, \dots, t \quad \forall u \in \mathcal{U}^T \quad h_i(u, \cdot) \preceq \Lambda_t(\cdot). \tag{II.6}$$

En particulier le fait que h_0 soit causale et que $\gamma_0(\cdot) \preceq h_0(\cdot)$ a pour conséquence que :

$$\gamma_0(\cdot) \preceq \Lambda_t(\cdot). \tag{II.7}$$

De même la causalité de h_1, les inégalités (II.6) et (II.7) impliquent d'après [19, lemme 7] que :

$$h_1(\gamma(\cdot), \cdot) \preceq \Lambda_t(\cdot),$$

et par définition de $\mathbb{F}^{\mathrm{ad}}(h)$ nous avons donc $\gamma_1(\cdot) \preceq \Lambda_t(\cdot)$. En procédant à nouveau ainsi, on montre que :

$$\forall i = 0, \dots, t \quad \gamma_i(\cdot) \preceq \Lambda_t(\cdot),$$

ce qui achève la preuve. □

Proposition II.29. *Soit $\{h_t\}_{t=0,\dots,T-1}$ une famille causale de fonctions d'observation à mémoire parfaite en boucle ouverte. On suppose qu'il n'y a pas d'effet dual en boucle ouverte, alors :*

$$\mathbb{F}^{\mathrm{nde}} = \left\{ \gamma : \Omega \to \mathcal{U}^T \mid \forall t = 0, \dots, T-1 \quad \rho_t \circ \gamma(\cdot) \in \mathbb{F}^{\mathrm{ad}}_{\mathcal{U}^T}(h_t) \cap \mathbb{F}^{\xi_t}_{\mathcal{U}^T} \right\}.$$

Preuve : Soit $\gamma \in \mathbb{F}^{\mathrm{nde}}$ et t fixé alors . D'après la proposition II.28 :

$$(\gamma_0, \dots, \gamma_t) \preceq \bigvee_{u \in \mathcal{U}^T} h_t(u, \cdot),$$

ce qui combiné avec l'hypothèse d'absence d'effet dual pour les contrôles constants donne :

$$(\gamma_0, \dots, \gamma_t) \preceq \xi_t(\cdot). \tag{II.8}$$

Puisque $h_t(\gamma(\cdot), \cdot) \equiv \xi_t(\cdot)$ alors :

$$(\gamma_0, \dots, \gamma_t) \preceq h_t(\gamma(\cdot), \cdot).$$

en résumé :

$$\mathbb{F}^{\mathrm{nde}} \subset \left\{ \gamma : \Omega \to \mathcal{U}^T \mid \forall t = 0, \dots, T-1 \quad \rho_t \circ \gamma(\cdot) \in \mathbb{F}^{\mathrm{ad}}_{\mathcal{U}^T}(h_t) \cap \mathbb{F}^{\xi_t}_{\mathcal{U}^T} \right\}.$$

L'inclusion inverse nous est donnée par la proposition II.21. □

II.5.3 Conclusion

Nous avons affaibli l'hypothèse de mémoire parfaite, nous avons montré notamment que l'ensemble $\mathbb{F}^{\text{nde}}$ se réduit sous l'hypothèse (H) à l'ensemble $\mathbb{F}^{\widehat{\xi}}$. Naturellement si l'on fait cœxister l'hypothèse de mémoire parfaite pour les contrôles constants et l'hypothèse (H), on a le résultat suivant $\mathbb{F}^{\text{nde}} = \mathbb{F}^{\xi}$ sous réserve d'absence d'effet dual pour les contrôles constants. En conclusion, la mémoire parfaite pour la famille de fonctions d'observation n'est pas une hypothèse nécessaire pour caractériser le plus grand espace sur lequel il n'y a pas d'effet dual.

Chapitre III

Variations autour des métriques sur $\mathcal{F}^*$

III.1 Introduction

La structure d'information d'un problème d'optimisation stochastique, désigne la contrainte de mesurabilité portant sur les variables de décision. Il est classique en modélisation, de traduire la connaissance que l'on peut avoir d'un système en définissant dans le modèle des relations de mesurabilité par rapport à une ou plusieurs tribus. Typiquement l'observation de signaux (émis par un système) modélisés par des variables aléatoires réelles $h_1, \ldots, h_N$ se traduit dans le modèle par la contrainte que γ est $\sigma(h_1, \ldots, h_N)$-mesurable. Un résultat d'analyse nous dit que cela signifie qu'il existe une fonction mesurable f définie sur $\mathbb{R}^n$ telle que $\gamma = f(h_1, \ldots, h_N)$, autrement dit la variable de décision γ est contrainte d'être uniquement une fonction des observations. Ainsi on rencontre des tribus dans toutes les applications où il est question d'information : c'est notamment le cas dans les modèles en Mathématiques financières, ou bien plus généralement dans les modèles de décisions dans l'incertain. Malgré les nombreuses applications liées aux tribus, il existe peu d'études topologiques de l'espace (notons le $\mathcal{F}^*$) des sous-tribus d'une tribu $\mathcal{F}$ donnée. Ainsi, autant pour certains espaces (l'espace des *fonctions continues sur un compact* par exemple), on trouve pléthore de topologies et de métriques, autant en ce qui concerne l'espace $\mathcal{F}^*$, il n'existe que très peu de résultats. C'est en 1970 que Neveu [68] définit pour la première fois une topologie sur l'espace $\mathcal{F}^*$. Viennent ensuite Boylan [24] en 1971 puis Rogge [93] en 1974 qui définissent chacun une métrique sur $\mathcal{F}^*$ dont on montre facilement qu'elles sont équivalentes.

Nous étudions dans ce chapitre les topologies de convergence forte et de convergence uniforme. Nous apportons par ailleurs un résultat nouveau permettant de traduire une propriété de convergence d'une suite de variables aléatoires à valeurs vectorielles, en une propriété de convergence sur la suite des tribus associées.

III.1.1 Notions sur les tribus

Définition III.1. Deux tribus $\mathcal{G}$ et $\mathcal{G}'$ sont équivalentes si elles diffèrent seulement pour des ensembles de mesure nulle. On note cela par $\mathcal{G} \sim \mathcal{G}'$ ou bien par $\mathcal{G} \sim_{\mathbb{P}} \mathcal{G}'$ pour bien marquer la dépendance de cette relation par rapport à la probabilité $\mathbb{P}$. Si $A \in \mathcal{F}$ est de mesure nulle alors $\mathcal{G} \sim \mathcal{G} \cap A^c$.

Définition III.2. Une sous-tribu $\mathcal{B}$ de $\mathcal{F}$ est dite *complète* si elle contient toutes les parties $\mathcal{F}$-mesurables de Ω qui sont de mesure nulle.

Définition III.3. On note $\mathcal{F}^{**}$ l'ensemble des sous-tribus de $\mathcal{F}$ et $\mathcal{F}^*$ l'espace des classes d'équivalence des éléments de $\mathcal{F}^{**}$ pour la relation $\sim$. On utilisera la même notation pour désigner un élément de $\mathcal{F}^*$ ou son représentant dans $\mathcal{F}^{**}$.

Proposition III.4.

$$\forall \mathcal{B}_1, \mathcal{B}_2 \in \mathcal{F}^{**}, \quad (\mathbb{E}(f \mid \mathcal{B}_1) = \mathbb{E}(f \mid \mathcal{B}_2) \ \ \mathbb{P} - p.s. \quad \forall f \in L^1_{\mathbb{R}}(\Omega)) \Leftrightarrow \mathcal{B}_1 \sim \mathcal{B}_2.$$

Définition III.5. Soit $\mathbb{F}$ une famille de sous-ensembles de Ω . La sous-tribu engendrée par $\mathbb{F}$ est la plus petite tribu (au sens de l'inclusion) sur Ω contenant $\mathbb{F}$, elle sera notée $\sigma(\mathbb{F})$.

Définition III.6. Soit $f : (\Omega, \mathcal{F}) \to (X, \mathcal{S})$ une variable aléatoire, alors $\sigma(f) \stackrel{\text{def}}{=} f^{-1}(\mathcal{S})$.

Définition III.7. Soient A et B deux sous-ensembles de Ω ; la différence symétrique de A et B est l'ensemble noté $A\Delta B$ et défini de la façon suivante :

$$A\Delta B \stackrel{\text{def}}{=} (A \cap B^c) \cup (B \cap A^c).$$

Définition III.8. Soient $\mathbb{F}$ une famille de sous-ensembles de Ω, et F un sous-ensemble de Ω alors :

$$\mathbb{F} \cap F \stackrel{\text{def}}{=} \{B \cap F \mid B \in \mathbb{F}\}.$$

Remarque III.9. Si $\mathcal{B}$ est une sous-tribu de $\mathcal{F}$ et si F est une partie non vide de Ω alors $\mathcal{B} \cap F$ est une tribu de parties de F appelée trace de $\mathcal{B}$ sur F.

Définition III.10. Soient $\mathbb{G}$ et $\mathbb{H}$ deux familles de sous-ensembles de Ω, alors :

$$\mathbb{G} \cup \mathbb{H} \stackrel{\text{def}}{=} \{X \mid X \in \mathbb{G} \text{ ou } X \in \mathbb{H}\} \ ;$$

de même :

$$\mathbb{G} \cap \mathbb{H} \stackrel{\text{def}}{=} \{X \mid X \in \mathbb{G} \text{ et } X \in \mathbb{H}\}.$$

Lemme III.11. *Soient $\mathbb{G}$, $\mathbb{H}$ deux familles de sous-ensembles de Ω et D un sous-ensemble de Ω alors :*

$$(\mathbb{G} \cup \mathbb{H}) \cap D = (\mathbb{G} \cap D) \cup (\mathbb{H} \cap D).$$

Définition III.12. Soient $\mathbb{G}$ et $\mathbb{G}'$ deux familles de sous-ensembles de Ω. On note $\mathbb{G} \vee \mathbb{G}'$ (borne supérieure de $\mathbb{G}$ et $\mathbb{G}'$) la tribu définie de la manière suivante :

$$\mathbb{G} \vee \mathbb{G}' \stackrel{\text{def}}{=} \sigma(\mathbb{G} \cup \mathbb{G}').$$

Définition III.13. Soit $A \subset \Omega$, la fonction indicatrice de A est définie de la manière suivante :

$$\mathbb{I}_A(\omega) = \begin{cases} 1 \text{ si } \omega \in A \ ; \\ 0 \text{ sinon.} \end{cases}$$

III.2 Convergence forte de tribus

Dans *Calcul des probabilités*, J. Neveu ([67] p. 117–118) introduit en 1970 la notion de convergence forte d'espérance conditionnelle. Cotter [31] montre en 1985 que la topologie associée à la convergence forte de tribus est métrisable. Signalons les articles traitant de la discrétisation de filtration [59, 29, 30] en utilisant la notion de convergence forte de tribus.

III.2.1 Topologie de convergence forte

Nous allons définir ce qu'est la topologie de convergence forte de tribus dans $\mathcal{F}^*$.

Définition III.14. La topologie la plus grossière (contenant le moins d'ouverts) rendant continue les applications :

$$\mathcal{F}^* \to (L^1_{\mathbb{R}}(\Omega), ||\cdot||_{L^1_{\mathbb{R}}(\Omega)}), \quad \mathcal{B} \mapsto \mathbb{E}\left(f \mid \mathcal{B}\right) \ ; \tag{III.1}$$

pour tout $f \in L^1_{\mathbb{R}}(\Omega)$, sera dite topologie de convergence forte sur $\mathcal{F}^*$.

Proposition III.15 ([25, Proposition III.1]). *Soit $(\mathcal{B}_n)_{n\in\mathbb{N}}$ une suite de $\mathcal{F}^*$, et $\mathcal{B} \in \mathcal{F}^*$. Alors $\lim_n \mathcal{B}_n = \mathcal{B}$ si et seulement si pour tout $f \in L^1_{\mathbb{R}}(\Omega)$:*

$$\lim_n ||\mathbb{E}\left(f \mid \mathcal{B}_n\right) - \mathbb{E}\left(f \mid \mathcal{B}\right)||_{L^1_{\mathbb{R}}(\Omega)} = 0.$$

La topologie de convergence forte est métrisable, c'est ce que nous allons observer en introduisant la distance de Cotter.

Définition III.16 (Cotter [31]). Soit $f = (f_j)_{j\in\mathbb{N}}$ une famille dénombrable de $L^1_{\mathbb{R}}(\Omega, \mathcal{F})$, on appelle $\rho_{\mathcal{F},f}$, l'application définie sur $\mathcal{F}^{**} \times \mathcal{F}^{**}$ par :

$$\rho_{\mathcal{F},f}(\mathcal{B}, \mathcal{B}') \stackrel{\text{def}}{=} \sum_{j=1}^{\infty} \frac{1}{2^j} \min\left\{ ||\mathbb{E}\left(f_j \mid \mathcal{B}\right) - \mathbb{E}\left(f_j \mid \mathcal{B}'\right)||_{L^1_{\mathbb{R}}(\Omega)}, 1 \right\}.$$

Théorème III.17 (Cotter [31]). *Si $f = (f_j)_{j\in\mathbb{N}}$ est une famille dénombrable dense d'éléments de $L^1_{\mathbb{R}}(\Omega, \mathcal{F})$, alors $(\mathcal{F}^*, \rho_{\mathcal{F},f})$ est un espace métrique complet séparable. Deux distances utilisant des familles dénombrables denses différentes d'éléments de $L^1_{\mathbb{R}}(\Omega, \mathcal{F})$ sont uniformément équivalentes.*

Nous noterons souvent ρ ou ρ_f, quand il n'y aura pas d'ambiguïté.

Proposition III.18. *Soit $f = (f_j)_{j\in\mathbb{N}}$ une famille dense dans $L^1_{\mathbb{R}}(\Omega, \mathcal{F})$, alors $\rho_{\mathcal{F},f}$ définit la topologie de convergence forte de tribus sur $\mathcal{F}^*$.*

Preuve : Il suffit simplement de montrer qu'un ouvert de l'espace métrique $(\mathcal{F}^*, \rho_{\mathcal{F},f})$ contient un ouvert pour la topologie forte et réciproquement. Soit $\varepsilon > 0$ et considérons l'ensemble suivant :

$$V = \left\{\mathcal{B}' \mid \rho_{\mathcal{F},f}(\mathcal{B}, \mathcal{B}') < \varepsilon\right\}.$$

Nous allons montrer que V est ouvert pour la topologie forte sur $\mathcal{F}^*$. Il existe $N_\varepsilon \in \mathbb{N}$ tel que :

$$\forall \mathcal{B}, \mathcal{B}' \in \mathcal{F}^*, \quad \sum_{j=N_\varepsilon+1}^{\infty} \frac{1}{2^j} \min\left\{\left|\left|\mathbb{E}\left(f_j \mid \mathcal{B}\right) - \mathbb{E}\left(f_j \mid \mathcal{B}'\right)\right|\right|_{L^1_{\mathbb{R}}(\Omega)}, 1\right\} < \frac{\varepsilon}{2}.$$

Soit W l'ensemble suivant :

$$W = \left\{\mathcal{B}' \mid \sum_{j=1}^{N_\varepsilon} \frac{1}{2^j} \min\left\{\left|\left|\mathbb{E}\left(f_j \mid \mathcal{B}\right) - \mathbb{E}\left(f_j \mid \mathcal{B}'\right)\right|\right|_{L^1_{\mathbb{R}}(\Omega)}, 1\right\} < \frac{\varepsilon}{2}\right\}.$$

L'application

$$\mathcal{B}' \mapsto \sum_{j=1}^{N_\varepsilon} \frac{1}{2^j} \min\left\{\left|\left|\mathbb{E}\left(f_j \mid \mathcal{B}\right) - \mathbb{E}\left(f_j \mid \mathcal{B}'\right)\right|\right|_{L^1_{\mathbb{R}}(\Omega)}, 1\right\} \ ;$$

étant continue pour la topologie de convergence forte de tribus, W est donc un ouvert de cette topologie. Par ailleurs nous avons clairement que $W \subset V$. En notant T_ρ (resp. T_f) la topologie associée à ρ (resp. la topologie forte), nous avons montré que $T_\rho \subset T_f$. Nous allons montrer que pour tout f intégrable, l'application (III.1) est continue. Soient f intégrable $\varepsilon > 0$ et $\mathcal{B} \in \mathcal{F}^*$ alors :

$$\begin{aligned}\left|\left|\mathbb{E}\left(f \mid \mathcal{B}\right) - \mathbb{E}\left(f \mid \mathcal{B}'\right)\right|\right|_{L^1_{\mathbb{R}}(\Omega)} \leq \left|\left|\mathbb{E}\left(f \mid \mathcal{B}\right) - \mathbb{E}\left(f_i \mid \mathcal{B}\right)\right|\right|_{L^1_{\mathbb{R}}(\Omega)} + \left|\left|\mathbb{E}\left(f_i \mid \mathcal{B}\right) - \mathbb{E}\left(f_i \mid \mathcal{B}'\right)\right|\right|_{L^1_{\mathbb{R}}(\Omega)} \\ + \left|\left|\mathbb{E}\left(f_i \mid \mathcal{B}'\right) - \mathbb{E}\left(f \mid \mathcal{B}\right)'\right|\right|_{L^1_{\mathbb{R}}(\Omega)} \ ;\end{aligned}$$

d'où :

$$\left|\left|\mathbb{E}\left(f \mid \mathcal{B}\right) - \mathbb{E}\left(f \mid \mathcal{B}'\right)\right|\right|_{L^1_{\mathbb{R}}(\Omega)} \leq 2\left|\left|f - f_i\right|\right|_{L^1_{\mathbb{R}}(\Omega)} + 2^i \rho_{\mathcal{F},f}(\mathcal{B}, \mathcal{B}').$$

Nous choisissons f_{i_0} de sorte que $\left|\left|f - f_{i_0}\right|\right|_{L^1_{\mathbb{R}}(\Omega)} < \frac{\varepsilon}{4}$, puis nous posons $\eta = \frac{\varepsilon}{2^{i_0+1}}$. En résumé nous avons montré que :

$$\forall \varepsilon > 0 \quad \exists \eta > 0 \quad \rho_{\mathcal{F},f}(\mathcal{B}, \mathcal{B}') \leq \eta \Rightarrow \left|\left|\mathbb{E}\left(f \mid \mathcal{B}\right) - \mathbb{E}\left(f \mid \mathcal{B}'\right)\right|\right|_{L^1_{\mathbb{R}}(\Omega)} \leq \varepsilon \ ;$$

autrement dit l'application (III.1) est continue. □

Théorème III.19 (Neveu [67]). *Pour tout $p \in [1, \infty[$ et tout $X \in L^p_{\mathbb{R}}(\Omega)$, la famille :*

$$\{(\mathbb{E}(X \mid \mathcal{B}))^p \ ; \ \mathcal{B} \subset \mathcal{F}\},$$

obtenue en faisant varier $\mathcal{B}$ parmi les sous tribus de $\mathcal{F}$ est équi-intégrable . Étant donnée une suite $(\mathcal{B}_n)_{n\in\mathbb{N}}$ de sous-tribus de $\mathcal{F}$ telle que les limites $\lim_n \mathbb{E}(\mathbb{I}_A \mid \mathcal{B}_n)$ existent pour tout $A \in \mathcal{F}$ au sens de la convergence en probabilité, alors pour tout $p \in [1, \infty[$ et pour tout $X \in L^p_{\mathbb{R}}(\Omega)$ les limites $\lim_n \mathbb{E}(X \mid \mathcal{B}_n)$ existent au sens de la convergence dans $L^p_{\mathbb{R}}(\Omega)$. De plus, il existe une sous-tribu $\mathcal{B}$ de $\mathcal{F}$ telle que les limites ci-dessus coïncident p.s avec $\mathbb{E}(\mathbb{I}_A \mid \mathcal{B})$ et $\mathbb{E}(X \mid \mathcal{B})$.

Une propriété très importante de la topologie de la convergence forte est la densité des tribus engendrées par une partition finie. C'est l'objet de la proposition suivante.

Proposition III.20 (Cotter [31]). *Soit $\mathcal{B} \in \mathcal{F}^*$, alors pour tout $\epsilon > 0$ il existe une partition finie P de Ω telle que $\sigma(P) \subset \mathcal{B}$ et $\rho(\sigma(P), \mathcal{B}) < \epsilon$.*

Proposition III.21 (Cotter [31]). *La suite $(\mathcal{B}_n)_{n\in\mathbb{N}}$ converge fortement dès que l'une des conditions suffisantes suivantes est satisfaite :*

(a) pour tout $A \in \mathcal{F}$, $(\mathbb{P}(A \mid \mathcal{B}_n))_{n\in\mathbb{N}}$ converge pour la norme $L^1_{\mathbb{R}}(\Omega)$ ou en probabilité, avec $\mathbb{P}(A \mid \mathcal{B}_n) = \mathbb{E}(\mathbb{I}_A \mid \mathcal{B}_n)$,

(b) pour toute application f dans un sous ensemble dense de $L^1_{\mathbb{R}}(\Omega)$, $(\mathbb{E}(f \mid \mathcal{B}_n))_{n\in\mathbb{N}}$ converge en probabilité ou dans $L^p_{\mathbb{R}}(\Omega)$, avec $p \in [1, +\infty[$,

(c)

$$\bigvee_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \mathcal{B}_n = \bigcap_{m=1}^{\infty} \bigvee_{n=m}^{\infty} \mathcal{B}_n.$$

Définition III.22. On appelle $\mathbb{P}$-*limite supérieure* de la suite $(\mathcal{B}_n)_{n\in\mathbb{N}}$ la plus petite tribu $\mathcal{B}_0$ au sens de l'inclusion vérifiant :

$$\limsup_{n\to\infty} ||\mathbb{E}(f \mid \mathcal{B}_n)||_{L^1_{\mathbb{R}}(\Omega)} \leq ||\mathbb{E}(f \mid \mathcal{B}_0)||_{L^1_{\mathbb{R}}(\Omega)} \ ;$$

où $f \in L^1_{\mathbb{R}}(\Omega, \mathcal{F})$. On notera alors $\mathcal{B}_0 = \mathbb{P} - \limsup \mathcal{B}_n$.

Définition III.23. On appelle $\mathbb{P}$-*limite inférieure* de la suite $(\mathcal{B}_n)_{n\in\mathbb{N}}$ la plus grande tribu $\mathcal{B}_0$ au sens de l'inclusion vérifiant :

$$\liminf_{n\to\infty} ||\mathbb{E}(f \mid \mathcal{B}_n)||_{L^1_{\mathbb{R}}(\Omega)} \geq ||\mathbb{E}(f \mid \mathcal{B}_0)||_{L^1_{\mathbb{R}}(\Omega)} \ ;$$

pour toute variable aléatoire $f \in L^1_{\mathbb{R}}(\Omega, \mathcal{F})$. On notera alors $\mathcal{B}_0 = \mathbb{P} - \liminf \mathcal{B}_n$.

Théorème III.24 (Kudo [62]). *Pour toute suite $(\mathcal{B}_n)_{n\in\mathbb{N}}$ de tribus, la tribu :*

$$\mathcal{B}_0 = \left\{ A \in \mathcal{F} \mid \lim_{n\to\infty} \inf_{B\in\mathcal{B}_n} \mathbb{P}(A\Delta B) = 0 \right\} ;$$

est limite inférieure de la suite $(\mathcal{B}_n)_{n\in\mathbb{N}}$.

Théorème III.25 (Kudo [62]). *Soit $(\mathcal{B}_n)_{n\in\mathbb{N}}$ une suite de tribus alors :*

$$\mathbb{P} - \liminf \mathcal{B}_n \subset \mathbb{P} - \limsup \mathcal{B}_n ;$$

où le signe $\subset$ peut être remplacé par l'égalité si et seulement si la suite converge fortement.

Remarque III.26. Nous aurions pu définir les limites supérieure et inférieure par les ensembles $\cap_{k=1}^{\infty} \vee_{n=k}^{\infty} \mathcal{B}_n$ et $\vee_{k=1}^{\infty} \cap_{n=k}^{\infty} \mathcal{B}_n$. Si la suite $(\mathcal{B}_n)_{n\in\mathbb{N}}$ est monotone, les limites coïncident, cependant dans le cas général, on à juste :

$$\vee_{k=1}^{\infty} \cap_{n=k}^{\infty} \mathcal{B}_n \subset \mathbb{P} - \liminf \mathcal{B}_n \subset \mathbb{P} - \limsup \mathcal{B}_n \subset \cap_{k=1}^{\infty} \vee_{n=k}^{\infty} \mathcal{B}_n.$$

III.2.2 Propriétés de continuité de l'application σ

Nous étendrons dans cette section un résultat de continuité de Cotter en montrant que l'application σ qui applique à une variable aléatoire vectorielle f la tribu $f^{-1}(\mathcal{B}_{\mathbb{R}^m})$, vérifie une propriété de continuité pour la topologie de la convergence en probabilité, l'espace d'arrivée étant muni de la topologie de convergence forte de tribus.

Théorème III.27. *Soient $\mathcal{A}$ et $\mathcal{B}$ deux sous-tribus de $\mathcal{F}$ telles que $\mathcal{A} \subset \mathcal{B}$, $f = \{f_i\}_{i\in\mathbb{N}}$ une famille dense de fonctions de $L^1_{\mathbb{R}}(\Omega, \mathcal{F})$. Alors la famille $g = \{g_i\}_{i\in\mathbb{N}}$ avec $g_i = \mathbb{E}(f_i \mid \mathcal{B})$ est dense dans $L^1_{\mathbb{R}}(\Omega, \mathcal{B})$ et nous avons :*

$$\rho_{\mathcal{F},f}(\mathcal{A}, \mathcal{B}) = \rho_{\mathcal{B},g}(\mathcal{A}, \mathcal{B}).$$

Preuve : Il est clair que la famille de fonctions $\{g_i\}_{i\in\mathbb{N}}$ est une famille dénombrable d'éléments de $L^1_{\mathbb{R}}(\Omega, \mathcal{B})$, nous allons montrer qu'elle est dense dans ce même espace, soit $h \in L^1_{\mathbb{R}}(\Omega, \mathcal{B})$ alors :

$$\begin{aligned} ||g_i - h||_{L^1_{\mathbb{R}}(\Omega)} &= ||\mathbb{E}(f_i \mid \mathcal{B}) - \mathbb{E}(h \mid \mathcal{B})||_{L^1_{\mathbb{R}}(\Omega)} ; \\ &= ||\mathbb{E}(f_i - h \mid \mathcal{B})||_{L^1_{\mathbb{R}}(\Omega)} ; \\ &\leq ||f_i - h||_{L^1_{\mathbb{R}}(\Omega)} ; \end{aligned}$$

puisque la famille $\{f_i\}_{i\in\mathbb{N}}$ est dense dans $L^1_{\mathbb{R}}(\Omega, \mathcal{F})$ alors la famille $\{g_i\}_{i\in\mathbb{N}}$ est dense dans $L^1_{\mathbb{R}}(\Omega, \mathcal{B})$. Par définition de $\rho_{\mathcal{F},f}$ nous avons :

$$\rho_{\mathcal{F},f}(\mathcal{A}, \mathcal{B}) = \sum_{i=1}^{\infty} \frac{1}{2^i} \min\left\{ ||\mathbb{E}(f_i \mid \mathcal{A}) - \mathbb{E}(f_i \mid \mathcal{B})||_{L^1_{\mathbb{R}}(\Omega)}, 1 \right\}.$$

Comme $\mathcal{A} \subset \mathcal{B}$ nous avons

$$\mathbb{E}(f_i \mid \mathcal{A}) = \mathbb{E}(\mathbb{E}(f_i \mid \mathcal{B}) \mid \mathcal{A}) = \mathbb{E}(g_i \mid \mathcal{A}),$$

on obtient donc :

$$\rho_{\mathcal{F},f}(\mathcal{A},\mathcal{B}) = \sum_{i=1}^{\infty} \frac{1}{2^i} \min\left\{ ||\mathbb{E}\left(g_i \mid \mathcal{A}\right) - \mathbb{E}\left(g_i \mid \mathcal{B}\right)||_{L^1_{\mathbb{R}}(\Omega)}, 1 \right\} = \rho_{\mathcal{B},g}(\mathcal{A},\mathcal{B}).$$

□

Nous allons généraliser le résultat suivant [32, théorème 2.1] de Cotter.

Théorème III.28. *Soit $\{h_n\}_{n\in\mathbb{N}}$ une suite de variables aléatoires à valeurs dans $\mathbb{R}^m$, qui converge en probabilité vers h. Alors :*

$$\sigma(h) \subset \mathbb{P} - \liminf \sigma(h_n).$$

Preuve : Il suffit de montrer que pour tout $r > 0,\ \alpha \in \mathbb{R}^m \quad A \stackrel{\text{def}}{=} \{\omega \mid ||h(\omega) - \alpha|| < r\} \in \mathbb{P} - \liminf \sigma(h_n)$. Il existe $\delta > 0$ tel que :

$$\mathbb{P}(\{\omega \mid r - \delta < ||h(\omega) - \alpha|| < r\}) < \frac{\varepsilon}{2}.$$

Il existe N tel que si $n \geq N$ alors :

$$\mathbb{P}\left(\left\{\omega \mid ||h_n(\omega) - h(\omega)|| < \frac{\delta}{2}\right\}\right) > 1 - \frac{\varepsilon}{2}.$$

On définit :

$$W \stackrel{\text{def}}{=} \left\{\omega \mid ||h_n(\omega) - h(\omega)|| < \frac{\delta}{2}\right\} \ ;$$

et :

$$B \stackrel{\text{def}}{=} \left\{\omega \mid ||h_n(\omega) - \alpha|| \leq r - \frac{\delta}{2}\right\}.$$

Si $\omega \in W \cap B^{\text{c}}$ alors $||h(\omega) - \alpha|| > r - \delta$ et si $\omega \in W \cap B, \quad ||h(\omega) - \alpha|| < r$, ce qui implique que :

$$\mathbb{P}(\{A \Delta B\}) \leq \mathbb{P}(\{A \cap B^{\text{c}} \cap W\}) + \mathbb{P}(\{A^{\text{c}} \cap B \cap W\}) + \mathbb{P}(\{W^{\text{c}}\}) < \frac{\varepsilon}{2} + 0 + \frac{\varepsilon}{2}.$$

□

Corollaire III.29. *Soit $\{h_n\}_{n\in\mathbb{N}}$ une suite de variables aléatoires à valeurs dans $\mathbb{R}^m$ qui converge en probabilité vers h. Si $\sigma(h) = \mathcal{F}$ alors la suite $\{\sigma(h_n)\}_{n\in\mathbb{N}}$ converge fortement vers $\sigma(h)$ par rapport à $(\Omega, \mathcal{F}, \mathbb{P})$.*

Preuve : D'après le théorème III.28 :

$$\sigma(h) \subset \mathbb{P} - \liminf \sigma(h_n) \ ;$$

par ailleurs nous avons par hypothèse que $\sigma(h) = \mathcal{F}$ alors :

$$\sigma(h) \subset \mathbb{P} - \liminf \sigma(h_n) \subset \mathbb{P} - \limsup \sigma(h_n) \subset \sigma(h) \ ;$$

autrement dit d'après le théorème III.25 la suite $\{\sigma(h_n)\}_{n\in\mathbb{N}}$ converge fortement vers $\sigma(h)$. □

Le résultat que nous allons énoncer maintenant tient une place privilégiée étant donné son intérêt pratique pour les applications numériques.

Théorème III.30. *Soit $\{h_n\}_{n\in\mathbb{N}}$ une suite de variables aléatoires à valeurs dans $\mathbb{R}^m$ qui converge en probabilité vers h. Si pour tout $n \in \mathbb{N}$, $\sigma(h_n) \subset \sigma(h)$, alors la suite $\{\sigma(h_n)\}_{n\in\mathbb{N}}$ converge fortement vers $\sigma(h)$.*

Preuve : D'après le corollaire III.29, nous avons que $\sigma(h_n)$ converge fortement vers $\sigma(h)$ par rapport à $(\Omega, \sigma(h), \mathbb{P})$. D'après le théorème III.27, nous avons aussi que $\sigma(h_n)$ converge fortement vers $\sigma(h)$ par rapport à $(\Omega, \mathcal{F}, \mathbb{P})$. □

Corollaire III.31. *Soient h une variable aléatoire à valeurs dans $\mathbb{R}^m$ et $\{Q_n\}_{n\in\mathbb{N}}$ une suite d'applications mesurables de $\mathbb{R}^m$ dans $\mathbb{R}^m$. Si la suite $\{Q_n \circ h\}_{n\in\mathbb{N}}$ converge en probabilité vers h, alors la suite $\{\sigma(Q_n \circ h)\}_{n\in\mathbb{N}}$ converge fortement vers $\sigma(h)$.*

Preuve : Il suffit d'observer que nous avons nécessairement pour tout $n \in \mathbb{N}$, $\sigma(Q_n \circ h) \subset \sigma(h)$, puis d'appliquer le théorème III.30. □

Soit le problème d'optimisation suivant :

$$\mathcal{P}_n \quad \min_{Q\in D_n} ||Q(h(\xi)) - h(\xi)||^2_{L^2_{\mathbb{R}^m}(\Omega)} \ ;$$

où $D_n = \{Q : \mathcal{Y} \mapsto \mathbb{R} \mid \text{card}(\ \text{im}Q) = 2^n\}$. Alors nous avons le résultat suivant.

Proposition III.32. *Soit $h : (\mathbb{R}, \mathcal{B}_\mathbb{R}) \mapsto \mathbb{R}$ une variable aléatoire réelle positive majorée par M. Alors la suite $\{\sigma(Q_n \circ h)\}_{n\in\mathbb{N}}$ où chaque Q_n est solution de $\mathcal{P}_n$ converge fortement vers $\sigma(h)$.*

Preuve : Puisque $h \in L^\infty_\mathbb{R}(\mathbb{R})$ alors la suite $\{h_n\}_{n\in\mathbb{N}}$ définie par :

$$h_n = M \sum_{k=0}^{2^n-1} \frac{k}{2^n} \mathbb{I}_{\left\{x \mid \frac{Mk}{2^n} \le x < \frac{M(k+1)}{2^n}\right\}}(h) \ ;$$

converge vers h dans $L^\infty_\mathbb{R}(\mathbb{R})$. Par définition de Q_n nous avons alors :

$$||Q_n \circ h - h||_{L^2_\mathbb{R}(\mathbb{R})} \le ||h_n - h||_{L^2_\mathbb{R}(\mathbb{R})} \le ||h_n - h||_{L^\infty_\mathbb{R}(\mathbb{R})} \, .$$

Alors la suite $\{Q_n \circ h\}_{n\in\mathbb{N}}$ converge au sens de $L^2_{\mathbb{R}}(\mathbb{R})$ vers h, donc également en probabilité. On déduit alors grâce au théorème III.31 le résultat annoncé. □

Nous avons montré à l'aide du corollaire III.31 qu'il est possible d'obtenir la convergence de la suite $(\sigma(Q_n \circ h))_{n\in\mathbb{N}}$ dès que la suite $(Q_n \circ h)_{n\in\mathbb{N}}$ converge en probabilité vers h. Nous allons discuter maintenant de la possibilité d'obtenir la convergence de la suite $(\sigma(h \circ Q_n))_{n\in\mathbb{N}}$.

Théorème III.33. *Soit $(Q_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles à valeurs réelles et h une fonction réelle continue. On note $\mathcal{H}_n = \sigma(Q_n), \quad n \in \mathbb{N}$ et $\mathcal{H} = \sigma(Q)$. Supposons que :*

– *la suite $(Q_n)_{n\in\mathbb{N}}$ converge en probabilité vers Q ;*

– *la suite $(\mathcal{H}_n)_{n\in\mathbb{N}}$ est asymptotiquement indépendante conditionnellement à $\mathcal{H}$.*

Alors la suite $(\sigma(h \circ Q_n))_{n\in\mathbb{N}}$ converge fortement vers $\sigma(h \circ Q)$.

Preuve : Posons $\mathcal{G}_n = \sigma(h \circ Q_n)$ et $\mathcal{G} = \sigma(h \circ Q)$, nous avons clairement pour tout $n \in \mathbb{N}, \quad \mathcal{G}_n \subset \mathcal{H}_n$ et $\mathcal{G} \subset \mathcal{H}$. Alors, la suite $(\mathcal{G}_n)_{n\in\mathbb{N}}$ est asymptotiquement indépendante conditionnellement à $\mathcal{G}$. Par ailleurs l'application h étant supposée continue, nous avons aussi que la suite $(h \circ Q_n)_{n\in\mathbb{N}}$ converge en probabilité vers $h \circ Q$. Le résultat est alors un simple corollaire de [59, théorèmes 7.3 et 7.4]. □

Proposition III.34. *Soient $\mathcal{A}, \mathcal{B}$ et $\mathcal{C}$ des sous-tribus emboîtées ($\mathcal{A} \subset \mathcal{B} \subset \mathcal{C}$) de $\mathcal{F}$ alors :*

$$\rho(\mathcal{B}, \mathcal{A}) \leq \rho(\mathcal{C}, \mathcal{A}).$$

Preuve : Commençons par montrer que pour toute fonction f intégrable :

$$||\mathbb{E}(f \mid \mathcal{B})||_{L^1_{\mathbb{R}}(\Omega)} \leq ||\mathbb{E}(f \mid \mathcal{C})||_{L^1_{\mathbb{R}}(\Omega)}. \tag{III.2}$$

En effet d'après le lemme B.5 :

$$||\mathbb{E}(f \mid \mathcal{B})||_{L^1_{\mathbb{R}}(\Omega)} = ||\mathbb{E}(\mathbb{E}(f \mid \mathcal{C}) \mid \mathcal{B})||_{L^1_{\mathbb{R}}(\Omega)} ; \tag{III.3}$$

nous déduisons alors de l'inégalité de Jensen (proposition B.4) que :

$$||\mathbb{E}(\mathbb{E}(f \mid \mathcal{C}) \mid \mathcal{B})||_{L^1_{\mathbb{R}}(\Omega)} \leq ||\mathbb{E}(f \mid \mathcal{C})||_{L^1_{\mathbb{R}}(\Omega)}. \tag{III.4}$$

La combinaison des inégalités (III.3) et (III.4) permet d'obtenir l'inégalité (III.2). On remplace alors dans (III.2) f par $f - \mathbb{E}(f \mid \mathcal{A})$ et l'on obtient donc :

$$||\mathbb{E}(f \mid \mathcal{B}) - \mathbb{E}(f \mid \mathcal{A})||_{L^1_{\mathbb{R}}(\Omega)} \leq ||\mathbb{E}(f \mid \mathcal{C}) - \mathbb{E}(f \mid \mathcal{A})||_{L^1_{\mathbb{R}}(\Omega)}.$$

Cette dernière relation et la définition de ρ implique le résultat de manière évidente. □

Corollaire III.35. *Soient $\mathcal{A}, \mathcal{B}$ et $\mathcal{C}$ des sous-tribus emboîtées ($\mathcal{A} \subset \mathcal{B} \subset \mathcal{C}$) de $\mathcal{F}$ alors :*

$$\rho(\mathcal{B}, \mathcal{C}) \leq 2\rho(\mathcal{A}, \mathcal{C}).$$

Preuve : Ce résultat est une conséquence directe de l'inégalité triangulaire et de la proposition précédente. □

Corollaire III.36. *Soient $(\mathcal{G}_n)_{n\in\mathbb{N}}$ et $(\mathcal{H}_n)_{n\in\mathbb{N}}$ deux suites de sous-tribus de $\mathcal{F}$ et $\mathcal{G} \in \mathcal{F}^*$ telles que :*

$$\mathcal{G}_n \subset \mathcal{H}_n \subset \mathcal{G} \ ;$$
$$\lim_{n\to\infty} \rho(\mathcal{G}_n, \mathcal{G}) = 0 \ ;$$

alors $\lim_{n\to\infty} \rho(\mathcal{H}_n, \mathcal{G}) = 0$.

III.2.3 Variation autour de la topologie de la convergence forte

La métrique définie par Cotter présente un inconvénient pratique. Typiquement même si l'on munit Ω d'une tribu engendrée par une partition finie, la distance ρ sera quand même définie à l'aide d'une famille dénombrable. Nous allons montrer qu'il est possible d'affaiblir l'hypothèse de densité faite sur la famille de variables aléatoires entrant en jeu dans la définition de ρ .

Définition III.37. Soit E un $\mathbb{R}$-espace vectoriel normé, une famille $\{f_i\}_{i\in\mathbb{N}}$ d'éléments de E est dite totale si pour tout élément $f \in E$ et pour tout $\varepsilon > 0$ il existe N réels $\alpha_1, \ldots, \alpha_N$ tels que :

$$\left|\left| f - \sum_{i=1}^{N} \alpha_i f_i \right|\right|_E \leq \varepsilon.$$

On remarque en particulier qu'une famille dénombrable dense est nécessairement totale.

Définition III.38. Soit $f = \{f_i\}_{i\in\mathbb{N}}$ une famille totale de variables aléatoires réelles positives telle que :

$$\sum_{i=1}^{\infty} \frac{1}{2^i} \left|\left| f_i \right|\right|_{L^1_{\mathbb{R}}(\Omega)} < +\infty. \tag{III.5}$$

Alors, l'application $\tilde{\rho}_f$ est définie de la manière suivante :

$$\tilde{\rho}_f(\mathcal{B}, \mathcal{B}') = \sum_{i=1}^{\infty} \frac{1}{2^i} \left|\left| \mathbb{E}\left(f_i \mid \mathcal{B}\right) - \mathbb{E}\left(f_i \mid \mathcal{B}'\right) \right|\right|_{L^1_{\mathbb{R}}(\Omega)}.$$

Remarque III.39. Si Ω est un espace métrique et $\mathcal{F}$ la tribu des boréliens sur Ω alors, il existe une famille totale de $L^1_{\mathbb{R}}(\Omega)$ constituée de fonctions indicatrices (nous renvoyons le lecteur à [58] p. 133). Pour cette famille nous avons que :

$$\sum_{i=1}^{\infty} \frac{1}{2^i} ||f_i||_{L^1_{\mathbb{R}}(\Omega)} \leq 1.$$

Lemme III.40. *Si $\{f_i\}_{i\in\mathbb{N}}$ est une suite de variables aléatoires réelles positives vérifiant (III.5), alors la suite $\left\{\sum_{i=1}^{n} \frac{1}{2^i} f_i\right\}_{n\in\mathbb{N}}$ converge presque sûrement vers une variable aléatoire g et :*

$$\lim_{n\to\infty} \left|\left| \sum_{i=1}^{n} \frac{1}{2^i} f_i - g \right|\right|_{L^1_{\mathbb{R}}(\Omega)} = 0.$$

Preuve : La suite $\left\{\sum_{i=1}^{n} \frac{1}{2^i} f_i\right\}_{n\in\mathbb{N}}$ étant croissante, le résultat est alors une conséquence immédiate du théorème de Beppo-Levi. □

Proposition III.41. *Soit $f = \{f_i\}_{i\in\mathbb{N}}$ une famille totale d'éléments de $L^1_{\mathbb{R}}(\Omega)$ vérifiant (III.5) alors, l'application $\tilde{\rho}_f$ est une distance sur $\mathcal{F}^*$.*

Preuve : Il est clair que l'inégalité triangulaire et la symétrie sont vérifiées. Par ailleurs, il est clair également que si $\mathcal{A} = \mathcal{B}$ alors $\tilde{\rho}_f(\mathcal{A}, \mathcal{B}) = 0$. Supposons que $\tilde{\rho}_f(\mathcal{A}, \mathcal{B}) = 0$. Par définition de $\tilde{\rho}_f$ cela implique pour tout élément de la famille f_i nous avons $\mathbb{E}(f_i \mid \mathcal{A}) = \mathbb{E}(f_i \mid \mathcal{B})$. Soit f dans $L^1_{\mathbb{R}}(\Omega)$, la famille $\{f_i\}_{i\in\mathbb{N}}$ étant totale, pour tout $\varepsilon > 0$ il existe $\alpha_1, \ldots, \alpha_{N_\varepsilon}$ tels que :

$$\left|\left| f - \sum_{i=1}^{N_\varepsilon} \alpha_i f_i \right|\right|_{L^1_{\mathbb{R}}(\Omega)} \leq \varepsilon.$$

Soit $\mathcal{G}$ une sous-tribu de $\mathcal{F}$; on a alors par l'inégalité de Jensen :

$$\left|\left| \mathbb{E}(f \mid \mathcal{G}) - \mathbb{E}\left(\sum_{i=1}^{N_\varepsilon} \alpha_i f_i \mid \mathcal{G}\right) \right|\right|_{L^1_{\mathbb{R}}(\Omega)} \leq \left|\left| f - \sum_{i=1}^{N_\varepsilon} \alpha_i f_i \right|\right|_{L^1_{\mathbb{R}}(\Omega)} \leq \varepsilon. \tag{III.6}$$

D'après l'inégalité triangulaire :

$$\begin{aligned} ||\mathbb{E}(f \mid \mathcal{A}) - \mathbb{E}(f \mid \mathcal{B})||_{L^1_{\mathbb{R}}(\Omega)} \leq & \left|\left| \mathbb{E}(f \mid \mathcal{A}) - \mathbb{E}\left(\sum_{i=1}^{N_\varepsilon} \alpha_i f_i \mid \mathcal{A}\right) \right|\right|_{L^1_{\mathbb{R}}(\Omega)} \\ & + \left|\left| \mathbb{E}\left(\sum_{i=1}^{N_\varepsilon} \alpha_i f_i \mid \mathcal{A}\right) - \mathbb{E}\left(\sum_{i=1}^{N_\varepsilon} \alpha_i f_i \mid \mathcal{B}\right) \right|\right|_{L^1_{\mathbb{R}}(\Omega)} + \left|\left| \mathbb{E}\left(\sum_{i=1}^{N_\varepsilon} \alpha_i f_i \mid \mathcal{B}\right) - \mathbb{E}(f \mid \mathcal{B}) \right|\right|_{L^1_{\mathbb{R}}(\Omega)}. \end{aligned}$$

En utilisant (III.6) :

$$||\mathbb{E}(f \mid \mathcal{A}) - \mathbb{E}(f \mid \mathcal{B})||_{L^1_{\mathbb{R}}(\Omega)} \leq 2 \left|\left| f - \sum_{i=1}^{N_\varepsilon} \alpha_i f_i \right|\right|_{L^1_{\mathbb{R}}(\Omega)} ;$$

donc :

$$||\mathbb{E}(f \mid \mathcal{A}) - \mathbb{E}(f \mid \mathcal{B})||_{L^1_{\mathbb{R}}(\Omega)} \leq 2\varepsilon,$$

ε étant un nombre positif quelconque, nous avons $\mathbb{E}(f \mid \mathcal{A}) = \mathbb{E}(f \mid \mathcal{B})$ pour tout $f \in L^1_{\mathbb{R}}(\Omega)$ ce qui implique que $\mathcal{A} = \mathcal{B}$ (voir [24, théorème 2]). □

Remarque III.42. Dans le cas où Ω est un espace métrique, il existe dans $L^1_{\mathbb{R}}(\Omega)$ une famille totale de fonctions indicatrices $f = \{\mathbb{I}_{A_i}\}_{i\in\mathbb{N}}$. Il est facile de montrer que $\mathcal{F}$ peut être approché par une tribu engendrée par un nombre fini d'ensembles. Il suffit de considérer la tribu $\mathcal{B}_N$ engendrée par $\{A_1, \ldots, A_N\}$. Il est clair que nous avons alors :

$$\tilde{\rho}_f(\mathcal{B}_N, \mathcal{F}) \leq \frac{1}{2^N}.$$

Théorème III.43. *Soit $f = \{f_i\}_{i\in\mathbb{N}}$ une famille totale d'éléments de $L^1_{\mathbb{R}}(\Omega)$, vérifiant (III.5). Alors, la distance de $\tilde{\rho}_f$ définit la même topologie que la convergence forte.*

Preuve : Soit $\{\mathcal{B}_n\}_{n\in\mathbb{N}}$ une suite de tribus qui converge fortement vers $\mathcal{B}$. Alors :

$$\lim_{n\to\infty} ||\mathbb{E}(f_i \mid \mathcal{B}_n) - \mathbb{E}(f_i \mid \mathcal{B})||_{L^1_{\mathbb{R}}(\Omega)} = 0.$$

Soit $\varepsilon > 0$, il existe $N \in \mathbb{N}$ tel que :

$$\sum_{i=N}^{\infty} \frac{1}{2^i} ||\mathbb{E}(f_i \mid \mathcal{B}_n) - \mathbb{E}(f_i \mid \mathcal{B})||_{L^1_{\mathbb{R}}(\Omega)} \leq \frac{\varepsilon}{2}.$$

Il reste à choisir N_0 suffisamment grand pour que pour tout n plus grand que N_0 :

$$\sum_{i=1}^{N-1} \frac{1}{2^i} ||\mathbb{E}(f_i \mid \mathcal{B}_n) - \mathbb{E}(f_i \mid \mathcal{B})||_{L^1_{\mathbb{R}}(\Omega)} \leq \frac{\varepsilon}{2}.$$

D'où :

$$\tilde{\rho}_f(\mathcal{B}_n, \mathcal{B}) \leq \varepsilon.$$

Réciproquement nous allons supposer que :

$$\lim_{n\to\infty} \tilde{\rho}_f(\mathcal{B}_n, \mathcal{B}) = 0.$$

Alors pour tout i nous avons :

$$\lim_{n\to\infty} ||\mathbb{E}(f_i \mid \mathcal{B}_n) - \mathbb{E}(f_i \mid \mathcal{B})||_{L^1_{\mathbb{R}}(\Omega)} = 0.$$

Soient $\varepsilon > 0$ et $\mathbb{I}_A$ la fonction indicatrice de $A \in \mathcal{F}$, alors il existe $\alpha_1, \ldots, \alpha_{N_\varepsilon}$ tels que :

$$\left|\left|\mathbb{I}_A - \sum_{i=1}^{N_\varepsilon} \alpha_i f_i\right|\right|_{L^1_{\mathbb{R}}(\Omega)} \leq \frac{\varepsilon}{4}.$$

D'après l'inégalité triangulaire nous avons :

$$\begin{aligned}||\mathbb{E}(\mathbb{I}_A \mid \mathcal{B}_n) - \mathbb{E}(\mathbb{I}_A \mid \mathcal{B})||_{L^1_{\mathbb{R}}(\Omega)} \leq 2 \left|\left|\mathbb{I}_A - \sum_{i=1}^{N_\varepsilon} \alpha_i f_i\right|\right|_{L^1_{\mathbb{R}}(\Omega)} + \\ \left|\left|\mathbb{E}\left(\sum_{i=1}^{N_\varepsilon} \alpha_i f_i \mid \mathcal{B}_n\right) - \mathbb{E}\left(\sum_{i=1}^{N_\varepsilon} \alpha_i f_i \mid \mathcal{B}\right)\right|\right|_{L^1_{\mathbb{R}}(\Omega)}.\end{aligned}$$

Donc :

$$||\mathbb{E}(\mathbb{I}_A \mid \mathcal{B}_n) - \mathbb{E}(\mathbb{I}_A \mid \mathcal{B})||_{L^1_{\mathbb{R}}(\Omega)} \leq \frac{\varepsilon}{2} + \sum_{i=1}^{N_\varepsilon} |\alpha_i| \, ||\mathbb{E}(f_i \mid \mathcal{B}_n) - \mathbb{E}(f_i \mid \mathcal{B})||_{L^1_{\mathbb{R}}(\Omega)}.$$

Ce qui prouve que $\{\mathbb{E}(\mathbb{I}_A \mid \mathcal{B}_n)\}_{n\in\mathbb{N}}$ converge en norme $L^1_{\mathbb{R}}(\Omega)$ vers $\mathbb{E}(\mathbb{I}_A \mid \mathcal{B})$, donc également au sens de la convergence en probabilité. □

Théorème III.44. *Soient $f = \{f_i\}_{i\in\mathbb{N}}$ une famille totale de variables aléatoires positives de $L^1_{\mathbb{R}}(\Omega)$ vérifiant (III.5) et deux sous-tribus de $\mathcal{F}$ notées $\mathcal{B}$ et $\mathcal{B}'$. Alors pour toute application $\gamma \in L^p_{\mathbb{R}}(\Omega)$ où $\frac{1}{p} + \frac{1}{q} = 1$ nous avons :*

$$|\mathbb{E}[\langle g, \mathbb{E}(\gamma \mid \mathcal{B}) - \mathbb{E}(\gamma \mid \mathcal{B}')\rangle]| \leq ||\gamma||_{L^p_{\mathbb{R}}(\Omega)} \tilde{\rho}_f(\mathcal{B}, \mathcal{B}') \ ;$$

où $g(\omega) = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{2^i} f_i(\omega)$.

Preuve : On désigne par $\langle\cdot,\cdot\rangle$ le produit scalaire dans $\mathbb{R}$, alors :

$$\begin{aligned}\left|\mathbb{E}\left[\left\langle \sum_{i=1}^{n} \frac{1}{2^i} f_i, \mathbb{E}(\gamma \mid \mathcal{B}) - \mathbb{E}\left(\gamma \mid \mathcal{B}'\right)\right\rangle\right]\right| &= \left|\mathbb{E}\left[\sum_{i=1}^{n} \frac{1}{2^i} \left\langle \mathbb{E}(f_i \mid \mathcal{B}) - \mathbb{E}\left(f_i \mid \mathcal{B}'\right), \gamma\right\rangle\right]\right| \ ; \\ &\leq \sum_{i=1}^{n} \frac{1}{2^i} \left|\mathbb{E}\left[\left\langle \mathbb{E}(f_i \mid \mathcal{B}) - \mathbb{E}\left(f_i \mid \mathcal{B}'\right), \gamma\right\rangle\right]\right| \ ; \\ &\leq \sum_{i=1}^{n} \frac{1}{2^i} \mathbb{E}\left[\left|\left\langle \mathbb{E}(f_i \mid \mathcal{B}) - \mathbb{E}\left(f_i \mid \mathcal{B}'\right), \gamma\right\rangle\right|\right] \ ; \\ &\leq \sum_{i=1}^{n} \frac{1}{2^i} ||\gamma||_{L^p_{\mathbb{R}}(\Omega)} \left|\left|\mathbb{E}(f_i \mid \mathcal{B}) - \mathbb{E}\left(f_i \mid \mathcal{B}'\right)\right|\right|_{L^q_{\mathbb{R}}(\Omega)} \ ; \\ &\leq ||\gamma||_{L^p_{\mathbb{R}}(\Omega)} \tilde{\rho}_f(\mathcal{B}, \mathcal{B}').\end{aligned}$$

Par ailleurs :

$$\left|\left\langle \sum_{i=1}^{n} \frac{1}{2^i} f_i, \mathbb{E}\left(\gamma \mid \mathcal{B}\right) - \mathbb{E}\left(\gamma \mid \mathcal{B}'\right) \right\rangle\right| \leq \left|\left\langle g, \mathbb{E}\left(\gamma \mid \mathcal{B}\right) - \mathbb{E}\left(\gamma \mid \mathcal{B}'\right)\right\rangle\right|.$$

D'après (III.5) $g \in L^q_{\mathbb{R}}(\Omega)$ alors :

$$\left\langle g, \mathbb{E}\left(\gamma \mid \mathcal{B}\right) - \mathbb{E}\left(\gamma \mid \mathcal{B}'\right)\right\rangle \in L^1_{\mathbb{R}}(\Omega).$$

D'après le théorème de convergence dominée nous avons :

$$\lim_{n\to\infty} \mathbb{E}\left[\left\langle \sum_{i=1}^{n} \frac{1}{2^i} f_i, \mathbb{E}\left(\gamma \mid \mathcal{B}\right) - \mathbb{E}\left(\gamma \mid \mathcal{B}'\right) \right\rangle\right] = \mathbb{E}\left[\left\langle g, \mathbb{E}\left(\gamma \mid \mathcal{B}\right) - \mathbb{E}\left(\gamma \mid \mathcal{B}'\right)\right\rangle\right] \ ;$$

d'où le résultat :

$$\left|\mathbb{E}\left[\left\langle g, \mathbb{E}\left(\gamma \mid \mathcal{B}\right) - \mathbb{E}\left(\gamma \mid \mathcal{B}'\right)\right\rangle\right]\right| \leq ||\gamma||_{L^p_{\mathbb{R}}(\Omega)}\, \tilde{\rho}_f(\mathcal{B}, \mathcal{B}').$$

□

Remarque III.45. Soit une partition $P = \{\Omega_i\}_{i=1,\dots,N}$ de Ω, alors la famille $f = \{\mathbb{I}_{\Omega_i}\}_{i=1,\dots,N}$ est totale dans $L^1_{\mathbb{R}}(\Omega, \mathcal{B})$ ou $\mathcal{B} = \sigma(P)$. La distance $\tilde{\rho}_f$ se résume simplement à une somme finie :

$$\forall \mathcal{G}, \mathcal{H} \in \mathcal{F}^*, \quad \tilde{\rho}_f(\mathcal{G}, \mathcal{H}) = \sum_{i=1}^{N} \frac{1}{2^i} \left|\left|\mathbb{E}\left(\mathbb{I}_{\Omega_i} \mid \mathcal{G}\right) - \mathbb{E}\left(\mathbb{I}_{\Omega_i} \mid \mathcal{H}\right)\right|\right|_{L^1_{\mathbb{R}}(\Omega)}.$$

III.2.4 Borne sup de tribus

L'opération de borne supérieure de tribus n'est pas continue généralement. Cependant nous allons donner un cas très utile où nous pourrons affirmer la continuité de cette opération.

Proposition III.46. *Soit $\{\mathcal{A}_n\}_{n\in\mathbb{N}}$ (resp. $\{\mathcal{B}_n\}_{n\in\mathbb{N}}$) une suite de sous-tribus de $\mathcal{F}$ qui converge fortement vers $\mathcal{A}$ (resp. $\mathcal{B}$). Alors :*

$$\mathcal{A} \vee \mathcal{B} \subset \mathbb{P} - \liminf(\mathcal{A}_n \vee \mathcal{B}_n).$$

Preuve : Soit $B \in \mathcal{B}$ alors :

$$\inf_{A\in\mathcal{A}_n\vee\mathcal{B}_n} \mathbb{P}(A\Delta B) \leq \inf_{A\in\mathcal{B}_n} \mathbb{P}(A\Delta B).$$

Puisque la suite $(\mathcal{B}_n)_{n\in\mathbb{N}}$ converge fortement vers $\mathcal{B}$ alors :

$$\mathcal{B} = \mathbb{P} - \liminf_n \mathcal{B}_n \ ;$$

ce qui implique que $\lim_{n\to\infty} \inf_{A\in\mathcal{B}_n} \mathbb{P}(A\Delta B) = 0$, donc que :

$$\lim_{n\to\infty} \inf_{A\in\mathcal{A}_n\vee\mathcal{B}_n} \mathbb{P}(A\Delta B) = 0 \ ;$$

ce qui prouve que $B \in \mathbb{P} - \liminf \mathcal{A}_n \vee \mathcal{B}_n$, donc que $\mathcal{B} \subset \mathbb{P} - \liminf \mathcal{A}_n \vee \mathcal{B}_n$ et de manière symétrique $\mathcal{A} \subset \mathbb{P} - \liminf \mathcal{A}_n \vee \mathcal{B}_n$. Or $\mathbb{P} - \liminf \mathcal{A}_n \vee \mathcal{B}_n$ étant une tribu alors :

$$\mathcal{A} \vee \mathcal{B} \subset \mathbb{P} - \liminf \mathcal{A}_n \vee \mathcal{B}_n.$$

□

Remarque III.47. Nous aurions également pu remarquer que :

$$\begin{aligned}
\inf_{A \in \mathcal{B}_n} \mathbb{P}(A \Delta B) &= \inf_{A \in \mathcal{B}_n} \mathbb{E}\left[\left|\mathbb{E}\left(\mathbb{I}_B \mid \mathcal{B}_n\right) - \mathbb{I}_A\right|\right] \ ; \\
&= \mathbb{E}\left[\left|\mathbb{E}\left(\mathbb{I}_B \mid \mathcal{B}_n\right) - \mathbb{I}_{\left\{\omega \mid \mathbb{E}(\mathbb{I}_B \mid \mathcal{B}_n)(\omega) > \frac{1}{2}\right\}}\right|\right] \ ; \\
&= \mathbb{E}\left[\min\left\{\mathbb{E}\left(\mathbb{I}_B \mid \mathcal{B}_n\right), \mathbb{E}\left(\mathbb{I}_{B^c} \mid \mathcal{B}_n\right)\right\}\right] \ ; \\
&\leq \mathbb{E}\left[\left|\mathbb{E}\left(\mathbb{I}_B \mid \mathcal{B}_n\right) - \mathbb{I}_B\right|\right] \ ;
\end{aligned}$$

et suivre ensuite le même raisonnement.

Proposition III.48. *Soit $(\mathcal{A}_n)_{n\in\mathbb{N}}$ (resp. $(\mathcal{B}_n)_{n\in\mathbb{N}}$) une suite de sous-tribus de $\mathcal{F}$ qui converge fortement vers $\mathcal{A}$ (resp. $\mathcal{B}$). Si pour tout $n \in \mathbb{N}$, $\mathcal{A}_n \vee \mathcal{B}_n \subset \mathcal{A} \vee \mathcal{B}$ alors la suite $(\mathcal{A}_n \vee \mathcal{B}_n)_{n\in\mathbb{N}}$ converge fortement vers $\mathcal{A} \vee \mathcal{B}$.*

Preuve : D'après la proposition III.46 :

$$\mathcal{A} \vee \mathcal{B} \subset \mathbb{P} - \liminf \mathcal{A}_n \vee \mathcal{B}_n.$$

Alors :

$$\mathcal{A} \vee \mathcal{B} \subset \mathbb{P} - \liminf \mathcal{A}_n \vee \mathcal{B}_n \subset \mathbb{P} - \limsup \mathcal{A}_n \vee \mathcal{B}_n \subset \mathcal{A} \vee \mathcal{B}.$$

Alors, d'après le théorème III.25, la suite $(\mathcal{A}_n \vee \mathcal{B}_n)_{n\in\mathbb{N}}$ converge fortement vers $\mathcal{A} \vee \mathcal{B}$. □

Remarque III.49. La proposition III.48 est fausse pour l'opération d'intersection de deux tribus, c'est à dire pour l'application suivante :

$$\mathcal{B} \mapsto \mathcal{B} \cap \mathcal{A}.$$

En effet comme nous allons montrer grâce à l'exemple suivant il n'est pas possible de garantir la convergence de la suite $(\mathcal{B}_n \cap \mathcal{A})_{n\in\mathbb{N}}$ même dans le cas où la suite $(\mathcal{B}_n)_{n\in\mathbb{N}}$ converge fortement vers $\mathcal{F}$.

Exemple III.50. Le couple $(\Omega, \mathcal{F})$ désigne l'ensemble $[0,1] \times [0,1]$ muni de sa tribu des boréliens. Considérons l'application $g : (x,y) \mapsto -x + y$ et l'application g_n définie pour tout $n \in \mathbb{N}^*$ par :

$$g_n(x,y) = \left(\frac{2k+1}{2n}, \frac{2k'+1}{2n}\right) \quad \text{si } (x,y) \in \left]\frac{k}{n}, \frac{k+1}{n}\right] \times \left]\frac{k'}{n}, \frac{k'+1}{n}\right].$$

Il est facile de montrer que la suite $\{g_n(x,y)\}_{n\in\mathbb{N}^*}$ converge simplement vers (x,y), on déduit alors du corollaire III.29 que la suite de tribu $\{\sigma(g_n)\}_{n\in\mathbb{N}^*}$ converge fortement vers $\mathcal{F}$. Pourtant, $\sigma(g_n) \cap \sigma(g) = \{\Omega, \emptyset\}$ pour tout $n \in \mathbb{N}^*$.

III.3 Convergence uniforme de tribus

C'est en 1971 que Boylan [24] définit la topologie de convergence uniforme sur $\mathcal{F}^*$. Par la suite, plusieurs articles ont tenté de mettre en évidence les propriétés de cette distance et d'en donner des définitions topologiquement équivalentes et plus intuitives [2]. Cette section rappelle certaines propriétés relatives à la distance de Boylan et introduit une nouvelle métrique de convergence uniforme sur $\mathcal{F}^*$.

Définition III.51 (Boylan [24]). On appelle d_0 l'application définie sur $\mathcal{F}^* \times \mathcal{F}^*$ à valeurs dans $\mathbb{R}$ par :

$$d_0(\mathcal{G},\mathcal{H}) \stackrel{\text{def}}{=} \sup_{G\in\mathcal{G}} \inf_{H\in\mathcal{H}} \mathbb{P}(G\Delta H) + \sup_{H\in\mathcal{H}} \inf_{G\in\mathcal{G}} \mathbb{P}(G\Delta H) \ ;$$

Théorème III.52 (Boylan [24]). *L'application d_0 définit une distance sur $\mathcal{F}^*$.*

Théorème III.53 (Boylan [24]). *Soit $(\mathcal{G}_n)_{n\in\mathbb{N}}$ une suite de Cauchy pour d_0 d'éléments de $\mathcal{F}^*$. Il existe alors une tribu $\mathcal{G} \subset \mathcal{F}$ telle que $\lim_{n\longrightarrow+\infty} d_0(\mathcal{G}_n,\mathcal{G}) = 0$.*

Théorème III.54 (Boylan [24]). *Soit $(\mathcal{G}_n)_{n\in\mathbb{N}}$ une suite de sous-tribus de $\mathcal{F}$ et soit :*

$$\Phi = \left\{ f \in L^\infty_{\mathbb{R}}(\Omega) \mid \sup_{\omega\in\Omega} |f(\omega)| \leq 1 \quad \mathbb{P}-p.s. \right\}. \tag{III.7}$$

Alors $\lim_{n\to+\infty} ||\mathbb{E}(f \mid \mathcal{G}_n) - \mathbb{E}(f \mid \mathcal{G})||_{L^1_{\mathbb{R}}(\Omega)} = 0$ uniformément sur Φ si et seulement si :

$$\lim_{n\to\infty} d_0(\mathcal{G}_n,\mathcal{G}) = 0.$$

Nous dirons alors que la suite $(\mathcal{G}_n)_{n\in\mathbb{N}}$ converge uniformément vers $\mathcal{G}$.

Remarque III.55. Soit :

$$d_{\text{Boy}}(\mathcal{B},\mathcal{B}') \stackrel{\text{def}}{=} \sup_{f\in\Phi} ||\mathbb{E}(f \mid \mathcal{B}) - \mathbb{E}(f \mid \mathcal{B}')||_{L^p_{\mathbb{R}}(\Omega)}$$

où Φ est l'ensemble (III.7). Il est facile de montrer que d_{Boy} est une distance, de plus d'après le théorème III.54, d_{Boy} définie la topologie de la convergence uniforme de tribus.

Définition III.56 (Rogge [93]). La distance de Rogge notée δ, est définie de la manière suivante :

$$\delta(\mathcal{G},\mathcal{H}) \stackrel{\text{def}}{=} \max\left\{ \sup_{G\in\mathcal{G}} \inf_{H\in\mathcal{H}} \mathbb{P}(G\Delta H), \sup_{H\in\mathcal{H}} \inf_{G\in\mathcal{G}} \mathbb{P}(G\Delta H) \right\}.$$

La distance de Rogge est topologiquement équivalente à la distance de Boylan en effet :

$$\frac{1}{2} d_0(\mathcal{G},\mathcal{H}) \leq \delta(\mathcal{G},\mathcal{H}) \leq d_0(\mathcal{G},\mathcal{H}).$$

Le résultat suivant prouve la continuité de l'opération de borne supérieure pour la topologie induite par la distance de Boylan. Il est dû à B. Allen [3].

Théorème III.57 (Allen [3]). *Pour toutes tribus $\mathcal{G},\mathcal{G}',\mathcal{H},\mathcal{H}'$ de $\mathcal{F}^*$:*

$$d_0(\mathcal{G}\vee\mathcal{H},\mathcal{G}'\vee\mathcal{H}') \leq 4d_0(\mathcal{G},\mathcal{G}') + 4d_0(\mathcal{H},\mathcal{H}').$$

III.3.1 Variation autour de la convergence uniforme

Nous allons définir une application τ et nous allons montrer que muni de cet objet, l'ensemble $\mathcal{F}^*$ est un espace métrique. Nous montrerons également que τ est topologiquement équivalente à la distance de Boylan, nous nous intéresserons enfin aux propriétés algébriques de cette distance.

Définition III.58. Soit τ l'application définie sur $\mathcal{F}^* \times \mathcal{F}^*$ par :

$$\tau(\mathcal{G},\mathcal{H}) \stackrel{\text{def}}{=} \inf\left\{\mathbb{P}(A^c) \mid \mathcal{G}\cap A = \mathcal{H}\cap A,\ A\in\mathcal{F}\right\}.$$

Théorème III.59. *L'application τ est :*

- *symétrique :*
$$\forall(\mathcal{G},\mathcal{H})\in(\mathcal{F}^*)^2,\quad \tau(\mathcal{G},\mathcal{H}) = \tau(\mathcal{H},\mathcal{G}),$$
- *sous-additive :*
$$\forall(\mathcal{G},\mathcal{H},\mathcal{K})\in(\mathcal{F}^*)^3,\quad \tau(\mathcal{G},\mathcal{K}) \le \tau(\mathcal{G},\mathcal{H}) + \tau(\mathcal{H},\mathcal{K}),$$
- *séparante :*
$$\forall(\mathcal{G},\mathcal{H})\in(\mathcal{F}^*)^2,\quad \tau(\mathcal{G},\mathcal{H}) = 0 \Rightarrow \mathcal{G}\sim\mathcal{H}.$$

Preuve : Soient $(\mathcal{G},\mathcal{H},\mathcal{K})\in(\mathcal{F}^*)^3$, la définition de τ étant symétrique $\tau(\mathcal{G},\mathcal{H}) = \tau(\mathcal{H},\mathcal{G})$. Soient $(A,B)\in(\mathcal{F})^2$, tels que :

$$\mathcal{G}\cap A = \mathcal{H}\cap A \text{ et } \mathcal{H}\cap B = \mathcal{K}\cap B, \tag{III.8}$$

soit $C = A\cap B\in\mathcal{F}$ alors :

$$\mathcal{G}\cap C = \mathcal{H}\cap C \text{ et } \mathcal{H}\cap C = \mathcal{K}\cap C.$$

Par définition de τ : $\tau(\mathcal{G},\mathcal{K}) \le \mathbb{P}(C^c) = \mathbb{P}(A^c\cup B^c) \le \mathbb{P}(A^c) + \mathbb{P}(B^c)$, et puisque cette dernière relation est vraie pour tout $(A,B)\in(\mathcal{F})^2$ vérifiant (III.8) on a donc :

$$\tau(\mathcal{G},\mathcal{K}) \le \tau(\mathcal{G},\mathcal{H}) + \tau(\mathcal{H},\mathcal{K}).$$

Nous allons montrer que τ est séparante. Partons de l'hypothèse que $\tau(\mathcal{G},\mathcal{H}) = 0$ et montrons tout d'abord que l'on peut construire $A\in\mathcal{F}$ tel que $\mathcal{G}\cap A = \mathcal{H}\cap A$ et $\mathbb{P}(A^c) = 0$. Soit $n\in\mathbb{N}$, par définition de τ nous savons qu'il existe $A_n\in\mathcal{F}$ tel que :

$$\mathcal{G}\cap A_n = \mathcal{H}\cap A_n \quad et \quad \mathbb{P}(A_n^c) \le \frac{1}{n}.$$

Posons $A = \cup_{n\in\mathbb{N}} A_n$ alors $A\in\mathcal{F}$, $\mathbb{P}(A^c) = 0$ et :

$$\mathcal{G}\cap A = \mathcal{H}\cap A\ ;$$

en effet soit $B\in\mathcal{G}\cap A$ alors il existe $\tilde{B}\in\mathcal{G}$ tel que :

$$B = \tilde{B}\cap A\ ;$$

alors par définition de A nous avons :

$$\begin{aligned} B &= \tilde{B} \cap (\cup_{n\in\mathbb{N}} A_n) \ ; \\ &= \cup_{n\in\mathbb{N}} (\tilde{B} \cap A_n). \end{aligned}$$

Pour chaque $n \in \mathbb{N}$ il existe $\tilde{D}_n \in \mathcal{H}$ tel que $\tilde{B} \cap A_n = \tilde{D}_n \cap A_n$ donc :

$$\begin{aligned} B &= \cup_{n\in\mathbb{N}} (\tilde{D}_n \cap A_n) \ ; \\ &= (\cup_{n\in\mathbb{N}} \tilde{D}_n) \cap (\cup_{n\in\mathbb{N}} A_n) \ ; \\ &= (\cup_{n\in\mathbb{N}} \tilde{D}_n) \cap A. \end{aligned}$$

En observant que $\cup_{n\in\mathbb{N}} \tilde{D}_n \in \mathcal{H}$, nous avons montré que $B \in \mathcal{H} \cap A$ donc que $\mathcal{G} \cap A \subset \mathcal{H} \cap A$. Puisque $\mathcal{G}$ et $\mathcal{H}$ jouent un rôle symétrique :

$$\mathcal{G} \cap A = \mathcal{H} \cap A.$$

Puisque $\mathbb{P}(A^c) = 0$ alors :

$$\mathcal{G} = \mathcal{G} \cap A = \mathcal{H} \cap A = \mathcal{H}.$$

□

Proposition III.60. *Pour tout* $(\mathcal{G}, \mathcal{H}) \in (\mathcal{F}^*)^2, \quad d_0(\mathcal{G}, \mathcal{H}) \leq 2\tau(\mathcal{G}, \mathcal{H})$.

Preuve : Soit $A \in \mathcal{F}$ tel que $\mathcal{G} \cap A = \mathcal{H} \cap A$, cela signifie que, pour tout $G \in \mathcal{G}$, il existe $H \in \mathcal{H}$ tel que :

$$G \cap A = H \cap A.$$

On a donc :

$$G \cap H^c \cap A = \emptyset,$$

soit encore :

$$G \cap H^c \subset A^c.$$

Par le même raisonnement, on a aussi :

$$H \cap G^c \subset A^c,$$

et donc :

$$G \Delta H \subset A^c.$$

On a donc :

$$\forall G \in \mathcal{G} \quad \inf_{H\in\mathcal{H}} \mathbb{P}(G\Delta H) \leq \mathbb{P}(A^c),$$

qui nous assure :

$$\sup_{G\in\mathcal{G}} \inf_{H\in\mathcal{H}} \mathbb{P}(G\Delta H) \leq \mathbb{P}(A^c),$$

et nous concluons par définition de la distance de Boylan :

$$\forall A \in \mathcal{F} \quad \mathcal{G} \cap A = \mathcal{H} \cap A \Longrightarrow d_0(\mathcal{G}, \mathcal{H}) \leq 2\mathbb{P}(A^c),$$

soit par définition de τ :

$$d_0(\mathcal{G}, \mathcal{H}) \leq 2\tau(\mathcal{G}, \mathcal{H}).$$

□

Exemple III.61. Considérons les sous-tribus $\mathcal{G}$ et $\mathcal{H}$ de la tribu $\mathcal{F}$ engendrées par les partitions décrites sur la figure (III.1). On suppose ici que $\mathbb{P}$ est une probabilité uniforme. Alors

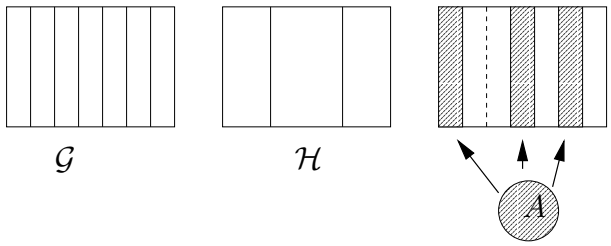


FIG. III.1 – partitions associées

$\tau(\mathcal{G}, \mathcal{H}) = \mathbb{P}(A^c)$. On remarque le choix de l'ensemble A n'est pas unique, en particulier à cause du fait que la loi est uniforme.

Proposition III.62 (Neveu [68]). *Soient $\mathcal{B}$ et $\mathcal{B}'$ deux sous tribus complètes de $\mathcal{F}$ telles que $\mathcal{B} \subset \mathcal{B}'$. Alors il existe un ensemble $A \in \mathcal{B}'$ tel que :*

$$\mathbb{P}(A^c) \leq 4d_0(\mathcal{B}, \mathcal{B}') \text{ et } \mathcal{B} \cap A = \mathcal{B}' \cap A.$$

Réciproquement si $A \in \mathcal{B}'$ est tel que $\mathcal{B} \cap A = \mathcal{B}' \cap A$ alors $d_0(\mathcal{B}, \mathcal{B}') \leq \mathbb{P}(A^c)$.

Cela signifie que si deux tribus $\mathcal{B}$ et $\mathcal{B}'$ sont emboîtées, alors il existe un ensemble A sur lequel les traces des deux tribus sont les mêmes, de plus $\mathbb{P}(A^c) \leq 4d_0(\mathcal{B}, \mathcal{B}')$.

Proposition III.63. *Soient $\mathcal{G}$ et $\mathcal{H}$ deux éléments de $\mathcal{F}^*$ alors :*

$$\tau(\mathcal{G}, \mathcal{H}) \leq 32 d_0(\mathcal{G}, \mathcal{H}).$$

Preuve : D'après le théorème III.57 on a tout d'abord :

$$d_0(\mathcal{G}, \mathcal{G} \vee \mathcal{H}) \leq 4d_0(\mathcal{G}, \mathcal{H}) \text{ et } d_0(\mathcal{H}, \mathcal{G} \vee \mathcal{H}) \leq 4d_0(\mathcal{G}, \mathcal{H}) \tag{III.9}$$

En appliquant la proposition III.62 au couple $(\mathcal{H}, \mathcal{H} \vee \mathcal{G})$ on obtient :

$$\tau(\mathcal{H}, \mathcal{H} \vee \mathcal{G}) \leq 4d_0(\mathcal{H}, \mathcal{H} \vee \mathcal{G}), \tag{III.10}$$

et de la même façon par symétrie :

$$\tau(\mathcal{G}, \mathcal{H} \vee \mathcal{G}) \leq 4d_0(\mathcal{G}, \mathcal{H} \vee \mathcal{G}), \tag{III.11}$$

ce qui associé à la sous additivité de τ nous donne :

$$\tau(\mathcal{G}, \mathcal{H}) \leq \tau(\mathcal{G}, \mathcal{H} \vee \mathcal{G}) + \tau(\mathcal{H} \vee \mathcal{G}, \mathcal{H}) \leq 4d_0(\mathcal{G}, \mathcal{H} \vee \mathcal{G}) + 4d_0(\mathcal{H}, \mathcal{H} \vee \mathcal{G}).$$

On conclut en utilisant (III.9) que :

$$\tau(\mathcal{G}, \mathcal{H}) \leq 32d_0(\mathcal{G}, \mathcal{H}).$$

□

Proposition III.64. *Soient $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ trois sous-tribus de $\mathcal{F}$ telles que : $\mathcal{A} \subset \mathcal{B} \subset \mathcal{C}$ alors $\tau(\mathcal{B}, \mathcal{C}) \leq \tau(\mathcal{A}, \mathcal{C})$.*

Preuve : Soit $A \in \mathcal{F}$ tel que $\mathcal{A} \cap A = \mathcal{C} \cap A$.

Comme $\mathcal{A} \subset \mathcal{B}$ on a $\mathcal{A} \cap A \subset \mathcal{B} \cap A$ et donc $\mathcal{C} \cap A \subset \mathcal{B} \cap A$.

De la même façon de $\mathcal{B} \subset \mathcal{C}$ on déduit que $\mathcal{B} \cap A \subset \mathcal{C} \cap A$. Donc $\mathcal{B} \cap A = \mathcal{C} \cap A$. Par définition de τ, il vient que :

$$\tau(\mathcal{B}, \mathcal{C}) \leq \mathbb{P}(A^c).$$

Cette dernière inégalité est vraie pour tout $A \in \mathcal{F}$ et donc :

$$\tau(\mathcal{B}, \mathcal{C}) \leq \inf \{\mathbb{P}(A^c) \mid A \in \mathcal{F} \text{ et } \mathcal{A} \cap A = \mathcal{C} \cap A\} = \tau(\mathcal{A}, \mathcal{C}).$$

□

Théorème III.65. *Pour toutes tribus $\mathcal{G}, \mathcal{G}', \mathcal{H}$ et $\mathcal{H}'$ de $\mathcal{F}^*$:*

$$\tau(\mathcal{G} \vee \mathcal{H}, \mathcal{G}' \vee \mathcal{H}') \leq \tau(\mathcal{G}, \mathcal{G}') + \tau(\mathcal{H}, \mathcal{H}').$$

Preuve : Soient A et B dans $\mathcal{F}$ tels que :

$$\mathcal{G} \cap A = \mathcal{G}' \cap A \text{ et } \mathcal{H} \cap B = \mathcal{H}' \cap B,$$

posons $D = A \cap B$ alors :

$$\mathcal{G} \cap D = \mathcal{G}' \cap D \text{ et } \mathcal{H} \cap D = \mathcal{H}' \cap D \Longrightarrow (\mathcal{G} \cap D) \cup (\mathcal{H} \cap D) = (\mathcal{G}' \cap D) \cup (\mathcal{H}' \cap D), \tag{III.12}$$

en appliquant le lemme III.11 à chaque terme de l'égalité (III.12) :

$$(\mathcal{G} \cup \mathcal{H}) \cap D = (\mathcal{G}' \cup \mathcal{H}') \cap D,$$

par application du lemme B.12, on obtient :

$$(\mathcal{G} \vee \mathcal{H}) \cap D = (\mathcal{G}' \vee \mathcal{H}') \cap D,$$

par définition de τ :

$$\tau(\mathcal{G} \vee \mathcal{H}, \mathcal{G}' \vee \mathcal{H}') \leq \mathbb{P}(D^c) \leq \mathbb{P}(A^c) + \mathbb{P}(B^c),$$

toujours par définition de τ :

$$\tau(\mathcal{G} \vee \mathcal{H}, \mathcal{G}' \vee \mathcal{H}') \leq \tau(\mathcal{G}, \mathcal{G}') + \tau(\mathcal{H}, \mathcal{H}').$$

□

Théorème III.66. *Soient $\mathcal{G}$ et $\mathcal{H}$ deux sous-tribus de $\mathcal{F}$ alors :*

$$\max\{\tau(\mathcal{G} \vee \mathcal{H}, \mathcal{G}), \tau(\mathcal{G} \vee \mathcal{H}, \mathcal{H})\} \leq \tau(\mathcal{G}, \mathcal{H})$$

Preuve : Soit $A \in \mathcal{F}$ tel que :

$$\mathcal{G} \cap A = \mathcal{H} \cap A, \tag{III.13}$$

nous avons par définition de $\mathcal{G} \vee \mathcal{H}$ que :

$$\mathcal{G} \vee \mathcal{H} = \sigma(\mathcal{G} \cup \mathcal{H}) \ ;$$

d'après le lemme III.11, nous avons :

$$(\mathcal{G} \cup \mathcal{H}) \cap A = (\mathcal{G} \cap A) \cup (\mathcal{H} \cap A) \ ; \tag{III.14}$$

l'équation (III.13) combinée à (III.14) donne :

$$\begin{aligned} (\mathcal{G} \cup \mathcal{H}) \cap A &= \mathcal{G} \cap A \ ; & \text{(III.15)} \\ &= \mathcal{H} \cap A \ ; & \text{(III.16)} \end{aligned}$$

on montre alors grâce au lemme B.12 que :

$$\begin{aligned} (\mathcal{G} \vee \mathcal{H}) \cap A &= \sigma((\mathcal{G} \cup \mathcal{H}) \cap A) \\ &= \mathcal{G} \cap A \\ &= \mathcal{H} \cap A \end{aligned}$$

alors pour tout $A \in \mathcal{F}$ vérifiant (III.13) nous avons :

$$\begin{aligned} \tau(\mathcal{G} \vee \mathcal{H}, \mathcal{H}) &\leq \mathbb{P}(A^c) \ ; \\ \tau(\mathcal{G} \vee \mathcal{H}, \mathcal{G}) &\leq \mathbb{P}(A^c) \ ; \end{aligned}$$

ce qui établit le résultat :

$$\max\{\tau(\mathcal{G} \vee \mathcal{H}, \mathcal{G}), \tau(\mathcal{G} \vee \mathcal{H}, \mathcal{H})\} \leq \tau(\mathcal{G}, \mathcal{H}).$$

□

Topologie sur $\mathcal{F}^*$	Convergence forte, Neveu 1970 [68]	Convergence uniforme Boylan
Distance sur $\mathcal{F}^*$	Cotter	Boylan, Rogge, Barty $\tau(\mathcal{B},\mathcal{C}) = \inf\{\mathbb{P}(A^c) \mid \mathcal{B}\cap A = \mathcal{C}\cap A\}$
$(\mathcal{A}_n)_{n\in\mathbb{N}}$ et $(\mathcal{B}_n)_{n\in\mathbb{N}}$ convergent	$(\mathcal{A}_n \vee \mathcal{B}_n)_{n\in\mathbb{N}}$ converge si $\forall n\in\mathbb{N},\ \mathcal{A}_n\vee\mathcal{B}_n \subset \mathcal{A}\vee\mathcal{B}$ Proposition III.48.	$(\mathcal{A}_n \vee \mathcal{B}_n)_{n\in\mathbb{N}}$ converge Théorème III.65
$(X_n)_{n\in\mathbb{N}}$ converge en probabilité vers X	$(\sigma(X_n))_{n\in\mathbb{N}}$ converge vers $\sigma(X)$ si $\forall n\in\mathbb{N},\ \sigma(X_n)\subset\sigma(X)$ Théorème III.30	
Densité des partitions finies dans $\mathcal{F}^*$	oui	non.
$(\mathcal{A}_n)_{n\in\mathbb{N}}$ et $(\mathcal{B}_n)_{n\in\mathbb{N}}$ convergent	$(\mathcal{A}_n \cap \mathcal{B}_n)_{n\in\mathbb{N}}$ ne converge pas Exemple III.50.	

FIG. III.2 – Résumé

III.3.2 Convergence uniforme et convergence forte de tribus

Soit $(\mathcal{G}_n)_{n\in\mathbb{N}}$ une suite de sous tribus de $\mathcal{F}$. Elle est dite monotone si elle est soit croissante ($\forall n\in\mathbb{N},\quad \mathcal{G}_n\subset\mathcal{G}_{n+1}$) soit décroissante ($\forall n\in\mathbb{N},\quad \mathcal{G}_{n+1}\subset\mathcal{G}_n$). Un résultat important de la théorie des martingales est l'existence pour tout $f\in L^1_{\mathbb{R}}(\Omega)$ d'une tribu $\mathcal{G}_\infty$ telle que

$$\lim_{n\longrightarrow\infty}\mathbb{E}(f\mid\mathcal{G}_n) = \mathbb{E}(f\mid\mathcal{G}_\infty)\quad \mathbb{P}-\text{p.s.}.$$

On notera toutefois qu'une telle suite d'éléments de $\mathcal{F}^*$ ne converge pas nécessairement pour la topologie définie par la distance de Boylan ; nous illustrons cela par l'exemple suivant :

Exemple III.67. Pour tout $n\in\mathbb{N}$ soit $\Omega_n=[0,1)$, $\mathcal{B}_n=\mathcal{B}([0,1))$, la tribu des boréliens de $[0,1)$, et $\mu_n=$ la mesure de Lebesgue sur $[0,1)$. On définit $\Omega=\Pi_{n=1}^\infty\Omega_n$, $\mathcal{F}=\Pi_{n=1}^\infty\mathcal{B}_n$ et $\mu=\Pi_{n=1}^\infty\mu_n$. Pour tout n soit $\mathcal{C}_n=\{\emptyset,\Omega_n\}$ la tribu triviale de $[0,1)$. Soit $\mathcal{G}_n=\Pi_{i=1}^n\mathcal{B}_i\times\Pi_{i=n+1}^\infty\mathcal{C}_i$. Alors $\mathcal{G}_n\uparrow\mathcal{F}$ mais $d_0(\mathcal{G}_n,\mathcal{G}_{n+1})=\frac{1}{2}$ pour tout n. En fait aucune sous suite de $(\mathcal{G}_n)$ ne converge et l'on peut montrer que $d_0(\mathcal{G}_n,\mathcal{G}_m)\geq\frac{1}{2}$, quelque soient $n\neq m$. C'est-à-dire que l'ensemble $\mathcal{F}^*$ muni de d_0 n'est pas compact.

Proposition III.68 (Cotter [31]). *La distance de Boylan est topologiquement plus forte que la distance de convergence forte. Elles sont équivalentes si et seulement si l'espace de probabilité est purement atomique.*

L'usage de la distance de convergence uniforme est limité par le fait que l'ensemble des tribus engendrées par une partitions finies de Ω n'est pas dense dans $\mathcal{F}^*$ pour la distance de convergence uniforme.

Le tableau III.2 contient le résumé des propriétés qui nous paraissent les plus importantes.

Chapitre IV

Quelques résultats asymptotiques

IV.1 Introduction

La quantité d'information dont nous disposons pour optimiser le critère d'un problème d'optimisation stochastique, peut avoir des répercussions plus ou moins importantes sur le coût de ce problème. La dépendance entre les différents paramètres de sortie d'un problème et la *structure d'information*, a été étudiée auparavant par Artstein [10] qui montre, sous des hypothèses de convexité, la continuité par rapport à la topologie de convergence forte de la fonction marginale d'un problème d'optimisation stochastique lorsque l'on perturbe la contrainte de mesurabilité. On trouve également dans les articles d'Allen [2] et de Cotter [31] le même type de résultats appliqués à un problème de maximisation d'une fonction d'utilité concave strictement monotone. Nous allons établir le même type de résultats. Nous montrerons également une propriété de Lipschitz vérifiée par la fonction valeur. Nous allons étudier la discrétisation complète d'un problème statique d'optimisation stochastique et nous allons montrer que l'erreur de discrétisation est la somme de deux erreurs dont l'une est issue de la discrétisation de la structure d'information et l'autre de l'approximation de l'espérance du critère. Nous appliquerons ensuite nos résultats asymptotiques à l'étude de deux exemples de problème d'optimisation statique.

Nous discuterons à la fin du chapitre de l'interaction entre la discrétisation et les contraintes de mesurabilité et de la manière pratique de discrétiser les contraintes de mesurabilité, en se tentant d'apporter une réponse à la question suivante "Comment traduire dans le problème discret les contraintes de mesurabilité du problème continu ?".

IV.1.1 Intégrandes normales

Nous avons rassemblé dans cette section quelques résultats concernant les intégrandes normales , le lecteur intéressé pourra consulter l'article de Berliocchi et Lasry [21] très complet sur le sujet. Dans toute la suite de ce chapitre, $(\Omega, \mathcal{F}, \mathbb{P})$ désigne un espace mesurable, muni d'une probabilité $\mathbb{P}$ et $\xi : \Omega \mapsto \Xi$ une variable aléatoire.

Définition IV.1. Si B est un borélien de $\mathbb{R}^p$, une application J de $B \times \Xi$ dans $\mathbb{R}$ est appelée intégrande normale si :
- pour presque tout $\xi \in \Xi$, l'application $J(\cdot, \xi)$ est s.c.i. sur B,
- il existe une fonction $\tilde{J} : B \times \Xi \to \mathbb{R}$ borélienne telle que $\tilde{J}(\cdot, \xi) = J(\cdot, \xi)$ pour presque tout $\xi \in \Xi$.

Une première conséquence de cette définition est que si u est une application mesurable de Ξ dans B et J une intégrande normale alors la fonction :

$$\xi \mapsto J(u(\xi), \xi),$$

est mesurable sur Ξ.

Nous énonçons maintenant quelques propriétés des intégrandes normales :
- si J est une intégrande normale, alors λJ est une intégrande normale pour tout $\lambda \in \mathbb{R}$;
- si J et J' sont des intégrandes normales, alors $(J+J')$ et $\inf(J, J')$ sont des intégrandes normales ;
- si $(J_n)_{n\in\mathbb{N}}$ est une famille d'intégrandes normales, alors $\sup_{n\in\mathbb{N}} J_n$ est une intégrande normale.

Exemple IV.2. Soit C un borélien de $B \times \Xi$; on supposera que, pour presque tout $\xi \in \Xi$, la tranche $C_\xi = \{a \in B \mid (a, \xi) \in C\}$ est fermée dans B. Alors la fonction caractéristique $\mathcal{X}_C$ de C :

$$\mathcal{X}_C(a, \xi) = \begin{cases} 0 & \text{si } (a, \xi) \in C, \\ +\infty & \text{si } (a, \xi) \notin C. \end{cases}$$

est une intégrande normale positive.

Définition IV.3. Soit B un borélien de $\mathbb{R}^n$. Une application $J : B \times \Xi \mapsto \mathbb{R}$ est dite de Carathéodory si :
- $J(\cdot, \xi)$ est continue sur B pour presque tout $\xi \in \Xi$,
- $J(u, \cdot)$ est mesurable sur Ξ pour tout $u \in B$.

Proposition IV.4. *Toute fonction de Carathéodory est une intégrande normale.*

IV.2 Pénalisation des contraintes de mesurabilité

Soient $\mathcal{B}$ une sous-tribu de $\mathcal{B}_\Xi$, $J : \mathbb{R}^n \times \Xi \mapsto \mathbb{R}$, une intégrande normale et Γ une multi-application mesurable à valeurs non vides fermées et le problème $\mathcal{P}$ suivant :

$$\mathcal{P} \begin{cases} \min \mathbb{E}\left[J(u(\xi), \xi)\right], \\ u \text{ est } \mathcal{B}\text{-mesurable}, \\ u(\xi) \in \Gamma(\xi) \quad \mathbb{P} - \text{p.s.}. \end{cases}$$

La problématique est la suivante : on cherche à déterminer des hypothèses sous lesquelles il sera possible de remplacer le problème $\mathcal{P}$ par le problèle :

$$\begin{cases} \min \mathbb{E}\left[F_{\mathcal{B}}(u(\xi),\xi)\right], \\ u \text{ est } \mathcal{B}_{\Xi}\text{-mesurable}, \\ u(\xi) \in \Gamma(\xi) \quad \mathbb{P}-\text{p.s.}. \end{cases}$$

Où $F_{\mathcal{B}}$ est une fonction qui dépend de $\mathcal{B}$. Zvi Artstein [10] à montré que dans le cas où $J(\cdot,\xi)$ est convexe et que Γ est indépendant de ξ que l'on peut prendre :

$$F_{\mathcal{B}}(\gamma) = \mathbb{E}\left[J_{\mathcal{B}}(\gamma(\xi),\xi)\right]$$

où :

$$J_{\mathcal{B}}(u,\xi) \overset{\text{def}}{=} \mathbb{E}\left(J(u,\cdot) \mid \mathcal{B}\right). \tag{IV.1}$$

Nous allons montrer que ce résultat reste valable dans d'autre cas sans supposer que $J(\cdot,\xi)$ est convexe. Nous étudierons ensuite la continuité de la fonction V. Notre approche repose sur le lemme technique suivant due à Dynkin et Evstigneev.

Lemme IV.5 (Dynkin, Evstigneev [41]). *Soient $\mathcal{B}$ une sous-tribu de $\mathcal{F}$, $J : \mathbb{R}^n \times \Xi \to \mathbb{R}$ une intégrande normale. Posons $g(u,\cdot) = \mathbb{E}\left(J(u,\cdot) \mid \mathcal{B}\right)$. On suppose qu'il existe une application $m \in L^1_{\mathbb{R}}(\mathcal{F})$ telle que $\forall u \in \mathbb{R}^n$, $|J(u,\xi)| \leq m(\xi) \quad \mathbb{P}-p.s.$, alors pour toute application $u : \Xi \to \mathbb{R}^n$ $\mathcal{B}$-mesurable on a :*

$$g(u(\xi),\xi) = \mathbb{E}\left(J(u(\xi),\xi) \mid \mathcal{B}\right).$$

Considérons le problème suivant :

$$\mathcal{P}' \begin{cases} \min \mathbb{E}\left[g(u(\xi),\xi)\right], \\ u \text{ est } \mathcal{B}_{\Xi}\text{-mesurable}, \\ u(\xi) \in \Gamma(\xi) \quad \mathbb{P}-\text{p.s.}. \end{cases}$$

Avec :

$$g(u,\xi) = \mathbb{E}\left(J(u,\xi) \mid \mathcal{B}\right).$$

Théorème IV.6. *On suppose que Γ est une multi application $\mathcal{B}$-mesurable à valeurs non vides et compactes dans $\mathbb{R}^n$ et qu'il existe une application $m \in L^1_{\mathbb{R}}(\Xi)$ telle que $\forall u \in \mathbb{R}^n$, $|J(u,\xi)| \leq m(\xi)$. Alors il existe une variable aléatoire u telle que :*

1. *u est $\mathcal{B}$-mesurable ;*
2. *u est solution du problème $\mathcal{P}'$;*
3. *u est solution du problème $\mathcal{P}$.*

Preuve : Soit la multi-application M définie par :

$$M(\xi) \overset{\text{def}}{=} \left\{ u \in \Gamma(\xi) \mid g(u,\xi) \leq \inf_{v \in \Gamma(\xi)} g(v,\xi) \right\}.$$

M est une multi-application $\mathcal{B}$-mesurable, à valeurs non vides et fermées. Par un théorème de sélection mesurable, il existe une sélection $\mathcal{B}$-mesurable de M, c'est-à-dire :

$$\exists u : \Xi \to \mathbb{R}^n \quad \mathcal{B}\text{-mesurable tel que } u(\xi) \in M(\xi).$$

Puisque $u(\xi) \in M(\xi) \ \ \mathbb{P}-$ p.s. alors $u(\xi) \in \Gamma(\xi) \ \ \mathbb{P}-$ p.s.. Par ailleurs, il est clair qu'une telle sélection mesurable réalise le minimum de $\mathcal{P}'$. Nous avons déjà que u vérifie les contraintes du problème $\mathcal{P}$, il nous reste à montrer l'optimalité de u. Soit v $\mathcal{B}$-mesurable telle que $v(\xi) \in \Gamma(\xi) \ \ \mathbb{P}-$ p.s.. Alors, par optimalité de u pour le problème $\mathcal{P}'$ on a que :

$$\mathbb{E}\left[g(u(\xi),\xi)\right] \ \leq \ \mathbb{E}\left[g(v(\xi),\xi)\right]. \tag{IV.2}$$

Par le lemme IV.5 on a :

$$g(u(\xi),\xi) = \mathbb{E}\left(J(u(\xi),\xi) \mid \mathcal{B}\right) \text{ et } g(v(\xi),\xi) = \mathbb{E}\left(J(v(\xi),\xi) \mid \mathcal{B}\right) \ ; \tag{IV.3}$$

et par une propriété de l'espérance conditionnelle :

$$\mathbb{E}\left[g(u(\xi),\xi)\right] = \mathbb{E}\left[J(u(\xi),\xi)\right] \text{ et } \mathbb{E}\left[g(v(\xi),\xi)\right] = \mathbb{E}\left[J(v(\xi),\xi)\right]$$

ce qui combiné à l'inégalité (IV.2) donne :

$$\mathbb{E}\left[J(u(\xi),\xi)\right] \ \leq \ \mathbb{E}\left[J(v(\xi),\xi)\right]; \tag{IV.4}$$

ce qui prouve que u est également solution du problème $\mathcal{P}$. □

Proposition IV.7. *On suppose que Γ est une multi application $\mathcal{B}$-mesurable à valeurs non vides et compactes et qu'il existe une application $m \in L^1_{\mathbb{R}}(\mathcal{F})$ telle que $|J(u,\xi)| \leq m(\xi)$. Soit u une solution du problème $\mathcal{P}$, alors u est solution du problème $\mathcal{P}'$.*

Preuve : Soient u une solution du problème $\mathcal{P}$ et v une solution du problème $\mathcal{P}'$. D'après le théorème IV.6, nous pouvons toujours supposer que v est $\mathcal{B}$-mesurable. L'application v vérifie donc les contraintes du problème $\mathcal{P}$ et du fait de l'optimalité de u on a :

$$\mathbb{E}\left[J(u(\xi),\xi)\right] \ \leq \ \mathbb{E}\left[J(v(\xi),\xi)\right]. \tag{IV.5}$$

Cette dernière inégalité associée à une propriété de l'espérance conditionnelle nous donne :

$$\mathbb{E}[\mathbb{E}\left(J(u(\xi),\xi) \mid \mathcal{B}\right)] \ \leq \ \mathbb{E}[\mathbb{E}\left(J(v(\xi),\xi) \mid \mathcal{B}\right)]. \tag{IV.6}$$

On associe cette fois-ci l'inégalité précédente au lemme IV.5 pour obtenir la relation suivante :

$$\mathbb{E}[g(u(\xi),\xi)] \ \leq \ \mathbb{E}[g(v(\xi),\xi)]. \tag{IV.7}$$

Étant donné que v est optimal pour $\mathcal{P}'$ alors u est également optimal pour $\mathcal{P}'$. □

IV.2.1 Propriété de Lipschitz de la fonction valeur

Nous allons nous servir des résultats du chapitre III pour établir une propriété de Lipschitz de la fonction valeur V du problème suivant :

$$\begin{cases} \min \mathbb{E}\left[J(\gamma(\xi),\xi)\right] \\ \gamma \text{ est } \mathcal{B}\text{-mesurable ;} \\ \gamma \in U_p. \end{cases} \tag{IV.8}$$

Avec $U_p \overset{\text{def}}{=} \{\gamma \in L^p_{\mathbb{R}^m}(\Xi) \mid \gamma(\xi) \in K\}$ et $\mathcal{B} \in \mathcal{B}^*_\Xi$.

Proposition IV.8. *On suppose que K est convexe alors la fonction valeur du problème IV.8 vérifie :*

$$V(\mathcal{B}) = \inf_{\gamma \in U_p} F_{\mathcal{B}}(\gamma) \ ; \tag{IV.9}$$

où :

$$F_{\mathcal{B}}(\gamma) \overset{\text{def}}{=} \mathbb{E}\left[J(\mathbb{E}\left(\gamma \mid \mathcal{B}\right)(\xi),\xi)\right].$$

Preuve : Soit $(\gamma_n)_{n\in\mathbb{N}}$ une suite minimisante associée à $V(\mathcal{B})$ alors :

$$\inf\left\{F_{\mathcal{B}}(\gamma) \mid \gamma \in U_p\right\} \leq F_{\mathcal{B}}(\gamma_n) = \mathbb{E}\left[J(\gamma_n(\xi),\xi)\right]$$

en passant à la limite on obtient :

$$\inf\left\{F_{\mathcal{B}}(\gamma) \mid \gamma \in U_p\right\} \leq V(\mathcal{B}).$$

Soit réciproquement $(\gamma_n)_{n\in\mathbb{N}}$ une suite minimisante associée à :

$$\inf\left\{F_{\mathcal{B}}(\gamma) \mid \gamma \in U_p\right\}.$$

Posons $\bar{\gamma}_n \overset{\text{def}}{=} \mathbb{E}\left(\gamma_n \mid \mathcal{B}\right)$. L'ensemble K étant convexe nous avons que :

$$\gamma_n \in U_p \Rightarrow \bar{\gamma}_n \in U_p.$$

Par ailleurs $\bar{\gamma}_n$ est $\mathcal{B}$-mesurable, il vient alors :

$$V(\mathcal{B}) \leq \mathbb{E}\left[J(\bar{\gamma}_n(\xi),\xi)\right] = F_{\mathcal{B}}(\gamma_n) \ ;$$

Il ne reste plus qu'a passer à la limite sur n, on obtient alors :

$$V(\mathcal{B}) \leq \inf\left\{F_{\mathcal{B}}(\gamma) \mid \gamma \in U_p\right\}.$$

□

À partir de ce résultat de pénalisation nous allons étudier la régularité de la fonction V. Commençons par établir quelques lemmes techniques. La norme choisie est la suivante : $||x||_{\mathbb{R}^m} = \sum_{i=1}^{m} |x_i|$.

Lemme IV.9. *On suppose que K est borné, c'est à dire qu'il existe $M \in \mathbb{R}$ tel que :*

$$\forall u \in K, \quad ||u||_{\mathbb{R}^m} \leq M \ ;$$

et qu'il existe une application $\ell \in L^q_{\mathbb{R}}(\Xi)$ avec $q = \frac{p}{p-1}$ telle que :

$$\forall u, u' \in K, \quad |J(u,\xi) - J(u',\xi)| \leq \ell(\xi)\, ||u - u'||_{\mathbb{R}^m} \,. \tag{IV.10}$$

Alors :

$$\sup_{\gamma \in U_p} |F_{\mathcal{B}}(\gamma) - F_{\mathcal{B}'}(\gamma)| \leq mM\, ||\ell||_{L^q_{\mathbb{R}}(\Xi)}\, d_{Boy}(\mathcal{B}, \mathcal{B}').$$

Preuve : Soient $\gamma \in U_p$, $\mathcal{B}$ et $\mathcal{B}'$ deux sous tribus de $\mathcal{B}_\Xi$, nous avons par hypothèse que :

$$\begin{aligned} |J(\mathbb{E}(\gamma \mid \mathcal{B})(\xi), \xi) - J(\mathbb{E}\left(\gamma \mid \mathcal{B}'\right)(\xi), \xi)| &\leq \ell(\xi) \left|\left|\mathbb{E}(\gamma \mid \mathcal{B})(\xi) - \mathbb{E}\left(\gamma \mid \mathcal{B}'\right)(\xi)\right|\right|_{\mathbb{R}^m} \\ &\leq \ell(\xi) \sum_{i=1}^{m} \left|\mathbb{E}(\gamma_i \mid \mathcal{B})(\xi) - \mathbb{E}\left(\gamma_i \mid \mathcal{B}'\right)(\xi)\right|. \end{aligned}$$

En passant à l'espérance, puis en utilisant l'inégalité de Hölder, on obtient :

$$|F_{\mathcal{B}}(\gamma) - F_{\mathcal{B}'}(\gamma)| \leq ||\ell||_{L^q_{\mathbb{R}}(\Xi)} \sum_{i=1}^{m} \left|\left|\mathbb{E}(\gamma_i \mid \mathcal{B}) - \mathbb{E}\left(\gamma_i \mid \mathcal{B}'\right)\right|\right|_{L^p_{\mathbb{R}}(\Xi)}. \tag{IV.11}$$

En remarquant que :

$$\begin{aligned} \left|\left|\mathbb{E}(\gamma_i \mid \mathcal{B}) - \mathbb{E}\left(\gamma_i \mid \mathcal{B}'\right)\right|\right|_{L^p_{\mathbb{R}}(\Xi)} &= M \left|\left|\mathbb{E}\left(\frac{\gamma_i}{M} \mid \mathcal{B}\right) - \mathbb{E}\left(\frac{\gamma_i}{M} \mid \mathcal{B}'\right)\right|\right|_{L^p_{\mathbb{R}}(\Xi)} \ ; \\ &\leq M d_{\mathrm{Boy}}(\mathcal{B}, \mathcal{B}') \ ; \end{aligned} \tag{IV.12}$$

puis combinant les inégalités (IV.11) et (IV.12), nous obtenons :

$$|F_{\mathcal{B}}(\gamma) - F_{\mathcal{B}'}(\gamma)| \leq mM\, ||\ell||_{L^q_{\mathbb{R}}(\Xi)}\, d_{\mathrm{Boy}}(\mathcal{B}, \mathcal{B}'). \tag{IV.13}$$

Le résultat s'obtient en prenant le sup sur les éléments $\gamma \in U_p$. □

Lemme IV.10. *On suppose qu'il existe une application $F_{\mathcal{B}}$ telle la relation (IV.9) soit vérifiée. Soient $\mathcal{B}, \mathcal{B}'$ deux sous tribus de $\mathcal{B}_\Xi$ et $(\gamma_n)_{n\in\mathbb{N}}$ une suite minimisante de $V(\mathcal{B}')$, c'est-à-dire telle que :*

$$\forall n \in \mathbb{N} \quad \gamma_n \in U_p, \qquad et \qquad F_{\mathcal{B}'}(\gamma_n) \leq V(\mathcal{B}') + \frac{1}{n}.$$

Alors :

$$e_n(F_{\mathcal{B}}, F_{\mathcal{B}'}) \stackrel{\text{def}}{=} F_{\mathcal{B}}(\gamma_n) - V(\mathcal{B})$$

vérifie :

$$0 \leq e_n(F_{\mathcal{B}}, F_{\mathcal{B}'}) \leq 2 \sup_{\gamma \in U_p} |F_{\mathcal{B}}(\gamma) - F_{\mathcal{B}'}(\gamma)| + \frac{2}{n} \ ;$$

et :

$$|V(\mathcal{B}) - V(\mathcal{B}')| \leq 3 \sup_{\gamma \in U_p} |F_{\mathcal{B}}(\gamma) - F_{\mathcal{B}'}(\gamma)|.$$

Remarque IV.11. Soulignons que la dépendance de l'expression $e_n(F_{\mathcal{B}}, F_{\mathcal{B}'})$ vis-à-vis de $\mathcal{B}'$ s'exprime simplement à travers la suite minimisante.

Preuve : Nous avons par hypothèse que :

$$V(\mathcal{B}) = \inf_{\gamma \in U_p} F_{\mathcal{B}}(\gamma).$$

Par définition de l'inf il existe une suite $(\mu_n)_{n\in\mathbb{N}}$ telle que :

$$F_{\mathcal{B}}(\mu_n) \leq V(\mathcal{B}) + \frac{1}{n}.$$

Posons :

$$\varepsilon = \sup_{\gamma \in U} |F_{\mathcal{B}}(\gamma) - F_{\mathcal{B}'}(\gamma)|,$$

et :

$$M_n = \left\{ \gamma \in U_p \mid F_{\mathcal{B}}(\gamma) \leq V(\mathcal{B}) + 2\varepsilon + \frac{2}{n} \right\}.$$

Supposons que $\gamma_n \notin M_n$, alors :

$$\begin{aligned} V(\mathcal{B}) + 2\varepsilon + \frac{2}{n} < F_{\mathcal{B}}(\gamma_n) \leq F_{\mathcal{B}'}(\gamma_n) + \varepsilon \leq V(\mathcal{B}') + \frac{1}{n} + \varepsilon \leq F_{\mathcal{B}'}(\mu_n) + \varepsilon + \frac{1}{n} \\ \leq F_{\mathcal{B}}(\mu_n) + \frac{1}{n} + 2\varepsilon \leq V(\mathcal{B}) + \frac{2}{n} + 2\varepsilon, \end{aligned}$$

ce qui est une contradiction. Donc $\gamma_n \in M_n$, c'est-à-dire que :

$$e_n(F_{\mathcal{B}}, F_{\mathcal{B}'}) \leq 2\varepsilon + \frac{2}{n}.$$

Le fait que $e_n(F_{\mathcal{B}}, F_{\mathcal{B}'}) \geq 0$ résulte de la définition de $V(\mathcal{B})$. Par ailleurs d'après l'inégalité triangulaire :

$$|V(\mathcal{B}) - V(\mathcal{B}')| \leq |V(\mathcal{B}) - F_{\mathcal{B}}(\gamma_n)| + |F_{\mathcal{B}}(\gamma_n) - F_{\mathcal{B}'}(\gamma_n)| + |F_{\mathcal{B}'}(\gamma_n) - V(\mathcal{B}')| \ ;$$

en utilisant

$$|V(\mathcal{B}) - V(\mathcal{B}')| \leq 3 \sup_{\gamma \in U_p} |F_{\mathcal{B}}(\gamma) - F_{\mathcal{B}'}(\gamma)| + \frac{3}{n}. \tag{IV.14}$$

□

Théorème IV.12. *Supposons K convexe alors, sous les hypothèses du lemme IV.9 :*

$$|V(\mathcal{B}) - V(\mathcal{B}')| \leq 3mM \, ||\ell||_{L^q_{\mathbb{R}}(\Xi)} \, d_{\mathrm{Boy}}(\mathcal{B}, \mathcal{B}').$$

Preuve : En utilisant la proposition IV.8 le lemme IV.10 puis le lemme IV.9 nous avons :

$$|V(\mathcal{B}) - V(\mathcal{B}')| \leq 3 \sup_{\gamma \in U_p} |F_{\mathcal{B}}(\gamma) - F_{\mathcal{B}'}(\gamma)| + \leq 3mM \, ||\ell||_{L^q_{\mathbb{R}}(\Xi)} \, d_{\text{Boy}}(\mathcal{B}, \mathcal{B}') \tag{IV.15}$$

□

IV.2.2 Majoration d'erreur

Les topologies et plus précisément les distances introduites pour comparer deux tribus, sont des outils quelquefois (sauf le cas de la remarque III.45) trop "sophistiqués", dans le sens où ils passent généralement par l'évaluation d'une infinité d'espérances conditionnelles. C'est la raison pour laquelle il est souhaitable de tenter d'évaluer l'écart $|V(\mathcal{B}) - V(\mathcal{B}')|$ sans nécessairement devoir calculer la distance entre $\mathcal{B}$ et $\mathcal{B}'$. Comme nous allons l'observer, cela est souvent possible pour une certaine classe de problèmes. Commençons par étudier un exemple : ξ désigne une variable aléatoire positive, $a \in \mathbb{R}$, $U = \{\gamma \mid \gamma(\xi) \in [-a, a]\}$, l'application J est définie de la manière suivante :

$$J(u, \xi) = \xi e^u \ ;$$

à ξ fixé il s'agit clairement d'une fonction convexe. Résoudre le problème d'optimisation associé à $V(\mathcal{B})$ pour ce J particulier est équivalent à résoudre $\tilde{V}(\mathcal{B}_\Xi)$ dans le cas où J est remplacé par $J_{\mathcal{B}}$:

$$J_{\mathcal{B}}(u, \xi) = \mathbb{E}\left(\xi \mid \mathcal{B}\right) e^u.$$

Il n'est pas nécessaire de calculer la distance entre $\mathcal{B}$ et $\mathcal{B}'$ pour obtenir une majoration de la différence entre $V(\mathcal{B})$ et $V(\mathcal{B}')$. En effet puisque nous avons :

$$\forall u, \quad J_{\mathcal{B}}(u, \xi) - J_{\mathcal{B}'}(u, \xi) = e^u(\mathbb{E}\left(\xi \mid \mathcal{B}\right) - \mathbb{E}\left(\xi \mid \mathcal{B}'\right)) \ ;$$

alors :

$$\sup_{\gamma \in U} |\mathbb{E}\left[J_{\mathcal{B}}(\gamma(\xi), \xi) - J_{\mathcal{B}'}(\gamma(\xi), \xi)\right]| \leq e^a \, ||\mathbb{E}\left(\xi \mid \mathcal{B}\right) - \mathbb{E}\left(\xi \mid \mathcal{B}'\right)||_{L^1_{\mathbb{R}}(\Xi)} \ ;$$

ce qui implique en utilisant le lemme IV.10 que :

$$|V(\mathcal{B}) - V(\mathcal{B}')| \leq 3e^a \, ||\mathbb{E}\left(\xi \mid \mathcal{B}\right) - \mathbb{E}\left(\xi \mid \mathcal{B}'\right)||_{L^1_{\mathbb{R}}(\Xi)} .$$

Cette dernière inégalité nous donne une majoration de l'erreur $|V(\mathcal{B}) - V(\mathcal{B}')|$ en fonction simplement de la différence $||\mathbb{E}\left(\xi \mid \mathcal{B}\right) - \mathbb{E}\left(\xi \mid \mathcal{B}'\right)||_{L^1_{\mathbb{R}}(\Xi)}$. Nous allons établir dans ce qui suit un résultat plus général.

Lemme IV.13. *Soient K un compact de $\mathbb{R}^m$ et J une intégrande normale de la forme :*

$$J(u, \xi) = \langle J_1(\xi), J_2(u) \rangle .$$

avec $J_1 \in L^1_{\mathbb{R}^n}(\Xi)$ *et* $J_2 : \mathbb{R}^m \mapsto \mathbb{R}^n$ *une application continue. Alors :*

$$\inf_\gamma \left\{ \mathbb{E}\left[J(\gamma(\xi), \xi) \right] \mid \gamma(\xi) \in K, \quad \gamma \textit{ est } \mathcal{B}\textit{-mesurable} \right\} = \inf_\gamma \left\{ \mathbb{E}\left[J_{\mathcal{B}}(\gamma(\xi), \xi) \right] \mid \gamma(\xi) \in K \right\} \ ;$$

où $J_{\mathcal{B}}(u, \xi) = \left\langle \mathbb{E}\left(J_1 \mid \mathcal{B} \right)(\xi), J_2(u) \right\rangle$.

Preuve : Nous allons montrer deux choses :

- premièrement que la solution du problème sans contrainte de mesurabilité est atteinte par un argmin $\mathcal{B}$-mesurable ;
- deuxièmement que pour tout γ $\mathcal{B}$-mesurable tel que $\gamma(\xi) \in K, \ \mathbb{P} - \text{p.s.}$ nous avons :

$$\mathbb{E}\left[J_{\mathcal{B}}(\gamma(\xi), \xi) \right] = \mathbb{E}\left[J(\gamma(\xi), \xi) \right].$$

Le "premièrement" revient simplement à remarquer que la multi application :

$$\Gamma(\xi) = \left\{ u \in K \mid J_{\mathcal{B}}(u, \xi) = \inf_{v \in K} J_{\mathcal{B}}(v, \xi) \right\}$$

est $\mathcal{B}$-mesurable à valeurs non vides et fermées. Par conséquent il existe une application $\mathcal{B}$-mesurable γ qui vérifie :

$$\gamma(\xi) \in \Gamma(\xi) \ ;$$

cette application est clairement solution du problème d'optimisation sans contrainte de mesurabilité, ce qui répond au premièrement. Soit maintenant γ une application $\mathcal{B}$-mesurable telle que $\gamma(\xi) \in K \ \mathbb{P} - \text{p.s.}$. Puisque J_2 est continue sur K qui est compact nous en déduisons que $J_2(\gamma) \in L^\infty_{\mathbb{R}^n}(\Xi)$, donc d'après l'inégalité de Hölder $\langle J_1, J_2(\gamma) \rangle \in L^1_{\mathbb{R}}(\Xi)$. Par ailleurs nous avons que $J_2(\gamma)$ est $\mathcal{B}$-mesurable car γ est lui même $\mathcal{B}$-mesurable, donc d'après une propriété de l'espérance conditionnelle,

$$\mathbb{E}\left[\langle J_1(\xi), J_2(\gamma(\xi)) \rangle \right] = \mathbb{E}\left[\left\langle \mathbb{E}\left(J_1 \mid \mathcal{B} \right)(\xi), J_2(\gamma(\xi)) \right\rangle \right].$$

Autrement dit :

$$\mathbb{E}\left[J_{\mathcal{B}}(\gamma(\xi), \xi) \right] = \mathbb{E}\left[J(\gamma(\xi), \xi) \right].$$

□

Théorème IV.14. *Soient $J_1 \in L^1_{\mathbb{R}^n}(\Xi)$ et $J_2 : \mathbb{R}^m \to \mathbb{R}^n$, nous notons alors :*

$$J(u,\xi) = \langle J_1(\xi), J_2(u)\rangle\,, \quad U = \{\gamma : \Xi \to \mathbb{R}^m \mid \gamma(\xi) \in K\}\ ;$$

ici $\langle\cdot,\cdot\rangle$ désigne le produit scalaire dans $\mathbb{R}^n$. On suppose que :
– *$J_1 \in L^1_{\mathbb{R}^n}(\Xi)$;*
– *J_2 est continue ;*
– *K est compact.*
Alors la fonction V définie par :

$$V(\mathcal{B}) = \inf_{\gamma\in U}\{\mathbb{E}\left[J(\gamma(\xi),\xi)\right] \mid \gamma \text{ est } \mathcal{B}\text{-mesurable}\}$$

est continue[1] *et :*

$$|V(\mathcal{B}) - V(\mathcal{B}')| \leq 3J_{\max}\,||\mathbb{E}\left(J_1 \mid \mathcal{B}\right) - \mathbb{E}\left(J_1 \mid \mathcal{B}'\right)||_{L^1_{\mathbb{R}^n}(\Xi)}$$

avec $J_{\max} = \sup_{u\in K}||J_2(u)||_{\mathbb{R}^n}$.

Preuve : Notons $J_{\mathcal{B}}(u,\xi) = \langle\mathbb{E}\left(J_1 \mid \mathcal{B}\right)(\xi), J_2(u)\rangle$ alors d'après le lemme IV.13 :

$$V(\mathcal{B}) = \inf_{\gamma\in U}\mathbb{E}\left[J_{\mathcal{B}}(\gamma(\xi),\xi)\right].$$

Pour tout $\gamma \in U$:

$$|J_{\mathcal{B}}(\gamma(\xi),\xi) - J_{\mathcal{B}'}(\gamma(\xi),\xi)| = \left|\left\langle\mathbb{E}\left(J_1 \mid \mathcal{B}\right) - \mathbb{E}\left(J_1 \mid \mathcal{B}'\right), J_2(\gamma(\xi))\right\rangle\right| \leq J_{\max}\left|\left|\mathbb{E}\left(J_1 \mid \mathcal{B}\right) - \mathbb{E}\left(J_1 \mid \mathcal{B}'\right)\right|\right|_{\mathbb{R}^n}.$$

En passant à l'espérance nous avons :

$$|\mathbb{E}\left[J_{\mathcal{B}}(\gamma(\xi),\xi)\right] - \mathbb{E}\left[J_{\mathcal{B}'}(\gamma(\xi),\xi)\right]| \leq J_{\max}\left|\left|\mathbb{E}\left(J_1 \mid \mathcal{B}\right) - \mathbb{E}\left(J_1 \mid \mathcal{B}'\right)\right|\right|_{L^1_{\mathbb{R}^n}(\Xi)}.$$

En prenant le sup sur $\gamma \in U$ nous obtenons :

$$\sup_{\gamma\in U}|\mathbb{E}\left[J_{\mathcal{B}}(\gamma(\xi),\xi)\right] - \mathbb{E}\left[J_{\mathcal{B}'}(\gamma(\xi),\xi)\right]| \leq J_{\max}\left|\left|\mathbb{E}\left(J_1 \mid \mathcal{B}\right) - \mathbb{E}\left(J_1 \mid \mathcal{B}'\right)\right|\right|_{L^1_{\mathbb{R}^n}(\Xi)}.$$

En utilisant le lemme[2] IV.10 nous obtenons alors :

$$\left|V(\mathcal{B}) - V(\mathcal{B}')\right| \leq 3J_{\max}\left|\left|\mathbb{E}\left(J_1 \mid \mathcal{B}\right) - \mathbb{E}\left(J_1 \mid \mathcal{B}'\right)\right|\right|_{L^1_{\mathbb{R}^n}(\Xi)}.$$

□

[2]La preuve du lemme IV.10 ne dépend pas du fait que $\gamma \in U_p$ ou $\gamma \in U$

Remarque IV.15. Le théorème IV.14 s'applique à un grand nombre de fonctions, pour certaines fonctions J ne satisfaisant pas les hypothèses, il est possible d'utiliser le théorème IV.14 comme un guide. C'est le cas notamment pour les applications $J(\cdot,\xi)$ qui sont deux fois différentiable, quitte à remplacer le critère par son développement de Taylor au voisinage de u_0, l'écart $|V(\mathcal{B}) - V(\mathcal{B}')|$ peut être contrôlé en maîtrisant :

$$\begin{aligned}
&||\mathbb{E}\left(J(u_0,\xi) \mid \mathcal{B}\right) - \mathbb{E}\left(J(u_0,\xi) \mid \mathcal{B}'\right)||_{L^1(\Xi)} \ ; \\
&||\mathbb{E}\left(\nabla J(u_0,\xi) \mid \mathcal{B}\right) - \mathbb{E}\left(\nabla J(u_0,\xi) \mid \mathcal{B}'\right)||_{L^1(\Xi)} \ ; \\
&||\mathbb{E}\left(HJ(u_0,\xi) \mid \mathcal{B}\right) - \mathbb{E}\left(HJ(u_0,\xi) \mid \mathcal{B}'\right)||_{L^1(\Xi)} \ ;
\end{aligned}$$

autrement dit l'erreur sur le calcul de l'espérance conditionnelle du coût, du gradient et du Hessien de $J(\cdot,\xi)$ en u_0.

IV.3 Discrétisation d'un problème statique

Soit le problème d'optimisation stochastique suivant :

$$\begin{cases} \inf F(\gamma) \stackrel{\text{def}}{=} \mathbb{E}\left[J(\gamma(\xi),\xi)\right] \ ; \\ \gamma \in \Delta_p(\mathcal{B}). \end{cases} \tag{IV.16}$$

La correspondance Δ_p est définie pour toute sous-tribu de la tribu des boréliens sur l'espace métrique Ξ.

$$\Delta_p(\mathcal{B}) \stackrel{\text{def}}{=} \left\{\gamma \in L^p_{\mathbb{R}^m}(\Xi) \mid \gamma = \mathbb{E}\left(\gamma \mid \mathcal{B}\right), \quad \gamma(\xi) \in K\right\}, \tag{IV.17}$$

où K désigne un sous ensemble de $\mathbb{R}^m$. Nous désignerons par V la fonction valeur associée au problème (IV.16), soit encore :

$$V(\mathcal{B}) \stackrel{\text{def}}{=} \inf\left\{F(\gamma) \mid \gamma \in \Delta_p(\mathcal{B})\right\}. \tag{IV.18}$$

Remarque IV.16. La fonction valeur (IV.18) est précisément celle étudiée dans [10] ; il apparaît également que cette dernière est la fonction marginale associée au couple (Δ_p, F).

Nous allons appliquer les résultats du paragraphe §A.3 afin de déduire une propriété de continuité vérifiée par la fonction valeur V. Nous montrerons que V est continue à gauche dans un sens que nous préciserons.

Hypothèses IV.17. *On suppose que :*

- *l'application J est une intégrande normale (voir définition IV.1) définie sur $\mathbb{R}^m \times \Xi$, il existe une fonction intégrable définie sur Ξ à valeurs réelle et qui minore $J(u,\xi)$ indépendemment de u. Cela afin d'éviter que l'application V prenne la valeur $-\infty$;*
- *l'ensemble K est convexe et fermé.*

Lemme IV.18. *Soient $\mathcal{A}$ et $\mathcal{B}$ deux sous-tribus de $\mathcal{B}_\Xi$ alors :*

$$\mathcal{A} \subset \mathcal{B} \Rightarrow \Delta_p(\mathcal{A}) \subset \Delta_p(\mathcal{B}) \text{ et } V(\mathcal{B}) \leq V(\mathcal{A}).$$

Preuve : Soit $\gamma \in \Delta_p(\mathcal{A})$, alors γ est $\mathcal{A}$-mesurable. Or $\mathcal{A} \subset \mathcal{B}$; γ est aussi $\mathcal{B}$-mesurable. Les autres contraintes n'étant pas liées à la tribu alors γ est aussi dans $\Delta_p(\mathcal{B})$. Par ailleurs l'inégalité $V(\mathcal{B}) \leq V(\mathcal{A})$ est simplement une conséquence de l'inclusion de l'ensemble admissible $\Delta_p(\mathcal{A})$ dans l'ensemble admissible $\Delta_p(\mathcal{B})$. □

Proposition IV.19. *On suppose que les hypothèses IV.17 sont vérifiées. La correspondance Δ_p définie par (IV.17) est semi-continue inférieurement lorsque l'on munit $\mathcal{B}^*_\Xi$ de la topologie de la convergence forte.*

Preuve : Soit $\gamma \in \Delta_p(\mathcal{B})$ et $(\mathcal{B}_k)_{k\in\mathbb{N}}$ une suite convergeant fortement vers $\mathcal{B}$. Posons $\gamma_k \overset{\text{def}}{=} \mathbb{E}(\gamma \mid \mathcal{B}_k)$, alors γ_k est $\mathcal{B}_k$ mesurable. La propriété $\gamma_k \in L^p_{\mathbb{R}^m}(\Xi)$ est déduite de l'inégalité de Jensen. L'ensemble K étant supposé convexe alors $\gamma_k(\xi) \in K$. En résumé nous avons montré que $\gamma_k \in \Delta_p(\mathcal{B}_k)$. La suite $(\mathcal{B}_k)_{k\in\mathbb{N}}$ convergeant fortement vers $\mathcal{B}$ alors :

$$\lim_k ||\mathbb{E}(\gamma \mid \mathcal{B}_k) - \mathbb{E}(\gamma \mid \mathcal{B})||_{L^p_{\mathbb{R}^m}(\Xi)} = 0.$$

Autrement dit, en remarquant que $\mathbb{E}(\gamma \mid \mathcal{B}) = \gamma$ car $\gamma \in \Delta_p(\mathcal{B})$ nous avons donc :

$$\lim_k ||\gamma_k - \gamma||_{L^p_{\mathbb{R}^m}(\Xi)} = 0.$$

□

IV.3.1 Continuité "à gauche" de la valeur de l'information

Nous allons montrer que la fonction marginale V est continue à gauche (voir définition IV.20) moyennant une hypothèse de continuité sur le critère.

Définition IV.20. Nous dirons qu'une application R définie sur $\mathcal{B}^*_\Xi$, à valeurs dans un espace métrique E, est continue à gauche en $\mathcal{B} \in \mathcal{B}^*_\Xi$, si pour toute suite $(\mathcal{B}_n)_{n\in\mathbb{N}}$ de sous-tribu de $\mathcal{B}$ qui converge fortement vers $\mathcal{B}$ la suite $(R(\mathcal{B}_n))_{n\in\mathbb{N}}$ converge vers $R(\mathcal{B})$ dans E. Si l'application R est continue à gauche pour tout $\mathcal{B} \in \mathcal{B}^*_\Xi$, alors elle sera dite continue à gauche.

Théorème IV.21. *On suppose que les hypothèses IV.17 sont vérifiées. Si F est continue sur $L^p_{\mathbb{R}^m}(\Xi)$ et $p \neq \infty$, alors la fonction marginale V est continue à gauche.*

Preuve : Soient $\mathcal{B} \in \mathcal{B}^*_\Xi$ et $(\mathcal{B}_k)_{k\in\mathbb{N}}$ une suite de sous-tribus de $\mathcal{B}$ qui converge fortement vers $\mathcal{B}$, alors d'après le lemme IV.18 :

$$\Delta_p(\mathcal{B}_k) \subset \Delta_p(\mathcal{B}) \text{ et } V(\mathcal{B}) \leq V(\mathcal{B}_k). \tag{IV.19}$$

Par ailleurs la première partie du théorème A.47 combinée à la proposition IV.19 nous donne la semi-continuité supérieure de l'application V ; autrement dit :

$$\limsup_k V(\mathcal{B}_k) \leq V(\mathcal{B}). \tag{IV.20}$$

Des équations (IV.20) et (IV.19) on tire l'inégalité suivante :

$$\limsup_k V(\mathcal{B}_k) \leq V(\mathcal{B}) \leq V(\mathcal{B}_k). \tag{IV.21}$$

Un passage à la limite inf dans l'inégalité (IV.21) donne :

$$\limsup_k V(\mathcal{B}_k) \leq V(\mathcal{B}) \leq \liminf_k V(\mathcal{B}_k).$$

L'égalité suivante est une conséquence de la définition de la limite d'une suite :

$$\lim_k V(\mathcal{B}_k) = V(\mathcal{B}).$$

□

Remarque IV.22. Étudier la continuité à gauche n'est pas très restrictif du point de vue des applications. En effet, nous proposerons dans la suite une approche numérique pour passer d'un problème d'optimisation stochastique de dimension infinie à un problème de dimension finie, qui respecte naturellement cette contrainte grace à un procédé de quantification. La quantification est l'opération qui consiste simplement à composer une variable aléatoire h et une variable aléatoire à valeurs discrètes $\mathcal{Q}$. Le fait de remplacer purement et simplement la contrainte de mesurabilité "γ est $\sigma(h)$-mesurable" par la contrainte "γ est $\sigma(Q \circ h)$-mesurable" nous fait passer de la classe des problèmes avec un nombre infini de degrés de liberté à celle avec un nombre fini. Par ailleurs, par ce procédé, nous avons automatiquement que $\sigma(Q \circ h) \subset \sigma(h)$.

IV.3.2 Quantification de la contrainte de mesurabilité

On considère ici le prototype d'un problème d'optimisation stochastique ne faisant pas intervenir de véritable structure d'information dynamique, c'est-à-dire obéissant à un principe de séparation comme ceux considérés dans le chapitre II. Soient Ξ un espace métrique, $\xi : \Omega \to \Xi$ une variable aléatoire, J une intégrande normale définie sur $\mathbb{R}^m \times \Xi$ et K un convexe fermé de $\mathbb{R}^m$. On définit le problème suivant :

$$V_1 \stackrel{\text{def}}{=} \inf \left\{ \mathbb{E}[J(\gamma(\omega), \xi(\omega))] \mid \gamma : \Omega \to K \text{ et } \gamma \text{ est } \sigma(h)\text{-mesurable} \right\}. \tag{IV.22}$$

Disons pour l'instant que h est une variable aléatoire définie sur Ω. L'espérance est prise par rapport à la variable ω qui appartient à un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$; la variable de décision est la loi de commande $\gamma : \Omega \to K$ qui est astreinte à être mesurable par rapport à h.

Quitte à renommer les variables nous pouvons toujours supposer que h est $\sigma(\xi)$-mesurable. En effet posons $z(\omega) = (\xi(\omega), h(\omega))$, $\quad \tilde{J}(x,z) = J(x,\xi)$, $\quad \tilde{h}(z_1,z_2) = z_2$ et définissons le nouveau problème :

$$V_2 \stackrel{\text{def}}{=} \min\left\{\mathbb{E}\left[\tilde{J}(\gamma(z),z)\right] \mid \gamma : \Xi \times \mathcal{Y} \to K \text{ est } \sigma(\tilde{h})\text{-mesurable}\right\}.$$

Il est clair que les problèmes V_1 et V_2 sont identiques. Mais dans le problème V_2 on constate que $\tilde{h}$ est $\sigma(z)$ mesurable. C'est la raison pour laquelle nous allons considérer sans perdre de généralité le problème suivant :

$$V \stackrel{\text{def}}{=} \min\left\{F(\gamma) \stackrel{\text{def}}{=} \mathbb{E}\left[J(\gamma(\xi),\xi)\right] \mid \gamma : \Xi \to K \text{ est } \sigma(h)\text{-mesurable}\right\} \ ; \qquad \text{(IV.23)}$$

où $h : \Xi \to \mathcal{Y}$ est une variable aléatoire $\mathcal{B}_\Xi$-mesurable.

Notre méthodologie de discrétisation est la suivante : on ramène le problème à un problème sans contrainte de mesurabilité en quantifiant l'observation, c'est-à-dire que l'on compose la variable aléatoire h avec une variable aléatoire Q à valeurs discrètes et en nombre fini. Soit :

$$V(Q_k) \stackrel{\text{def}}{=} \min\left\{\mathbb{E}\left[J(\gamma(\xi),\xi)\right] \mid \text{ sous la contrainte : } \gamma : \Xi \to K \text{ est } \sigma(Q_k \circ h)\text{-mesurable}\right\},$$

où $Q_k : \mathcal{Y} \to \mathcal{Y}$ vérifie $S_k \stackrel{\text{def}}{=} \text{card}(\ \text{im}Q_k) < \infty$, autrement dit $\text{im}Q_k = \{y_1, \ldots, y_{S_k}\}$. Posons pour tout $j = 1, \ldots, S_k$:

$$C_j^k \stackrel{\text{def}}{=} \{\xi \in \Xi \mid Q_k(h(\xi)) = y_j\} \ ;$$

alors $(C_j^k)_{j=1,\ldots,S_k}$ est une partition de Ξ. Dire que γ est $Q_k \circ h$-mesurable est équivalent à dire que γ peut se mettre sous la forme suivante :

$$\gamma(\xi) = \sum_{j=1}^{S_k} u_j \mathbb{I}_{C_j^k}(\xi).$$

On pose alors $u = (u_1, \ldots, u_{S_k}) \in K^{S_k}$ et donc :

$$V(Q_k) = \min_u \mathbb{E}\left[\sum_{j=1}^{S_k} J(u_j,\xi)\mathbb{I}_{C_j^k}(\xi)\right]. \qquad \text{(IV.24)}$$

Le problème $V(Q_k)$ obtenu par ce procédé est un problème en boucle ouverte.

IV.3.3 Problème en boucle ouverte et technique de type Monte-Carlo

Notre étude a porté jusqu'ici sur la discrétisation de la structure d'information. Cela étant, ce n'est pas la seule difficulté, même une fois ramené à un nombre fini de paramètres à optimiser, il reste quand même dans le problème à manipuler des objets de dimension infinie. Typiquement le calcul des espérances nécessite de connaître les lois de probabilité des variables aléatoires intervenant dans le critère. La question de l'approximation d'une espérance par des techniques de type Monte-Carlo fait l'objet d'une abondante littérature. En ce qui nous concerne, cette question doit être combinée avec la notion de convergence épigraphique introduite par Attouch et Wets [16] en 1981. Cette notion de convergence épigraphique est adaptée à l'optimisation, puisqu'elle implique en particulier la convergence des argmins d'un problème d'optimisation. Cette question a été abordée notamment par J. Dupačovà R. Wets [38] dans un cadre théorique très général et par Z. Artstein [14] dans le cas où l'estimateur des espérances est la moyenne empirique d'un échantillon. Certains travaux d'A. Shapiro [97, 96, 98] portent sur la détermination de vitesse de convergence des estimateurs du coût optimal en fonction à la taille de l'échantillon.

Notre but est de combiner les résultats existants concernant la discrétisation d'un problème d'optimisation en boucle ouverte aux résultats asymptotiques que nous avons obtenus à propos de la discrétisation d'un problème d'optimisation en information statique, afin d'en déduire un résultat de discrétisation des problèmes en information statique.

IV.3.4 Approximation de la loi de ξ

Nous allons appliquer les résultats obtenus par J. Dupačovà et R. Wets [38] à la discrétisation du problème en boucle ouverte (IV.24). L'ingrédient essentiel de la discrétisation consistera à remplacer la loi $\mathbb{P}_\xi$ par une suite de lois $(\mathbb{P}^n)_{n\in\mathbb{N}}$ supposée converger en loi vers $\mathbb{P}_\xi$. Nous obtenons alors un nouveau problème d'optimisation V_n, en remplaçant la loi $\mathbb{P}_\xi$ par $\mathbb{P}^n$ dans (IV.24).

$$V_n(Q_k) \stackrel{\text{def}}{=} \inf_{u_j} \int \sum_{j=1}^{S_k} J(u_j,\xi)\mathbb{I}_{C_j^k}(\xi)\mathbb{P}^n(d\xi). \tag{IV.25}$$

Hypothèses IV.23. *$J : \mathbb{R}^m \times \Xi \to]-\infty,\infty]$ est une intégrande normale, $S = \mathrm{dom}J(\cdot,\xi)$ est un sous ensemble de $\mathbb{R}^m$ indépendant de ξ, fermé et non vide. On suppose de plus que pour tout $u \in S$:*

$$\xi \mapsto J(u,\xi) \text{ est continue sur } \Xi \ ;$$

l'application

$$u \mapsto J(u,\xi) \text{ est s.c.i sur } \mathbb{R}^m,\ \mathbb{P}-p.s. \ ;$$

et est localement lipschitzienne sur S, dans le sens suivant : pour tout $u \in S$, il existe un voisinage V de u et une fonction continue et bornée $\beta : \Xi \mapsto \mathbb{R}$ telle que pour tout $u' \in V \cap S$ et $\xi \in \Xi$:

$$J(u,\xi) - J(u',\xi) \leq \beta(\xi)\,||u - u'||\,.$$

Hypothèses IV.24. *$(Z, \mathcal{Z}, \lambda)$ est un espace d'échantillons, $\{\mathcal{Z}^n\}_{n\in\mathbb{N}}$ une suite croissante de sous-tribus de $\mathcal{Z}$, $(\mathbb{P}^n)_{n\in\mathbb{N}}$ est une suite d'applications définies sur $\mathcal{B}_\Xi \times Z$ à valeurs dans $[0,1]$ telles que pour tout $\zeta \in Z$:*

$$\mathbb{P}^n(\cdot,\zeta) \text{ est une mesure de probabilité sur } (\Xi, \mathcal{B}_\Xi)\ ;$$

et pour tout $A \in \mathcal{B}_\Xi$,

$$\zeta \mapsto \mathbb{P}^n(A,\zeta) \text{ est } \mathcal{Z}^n\text{-mesurable.}$$

Pour λ-presque tout ζ, la suite $\{\mathbb{P}^n(\cdot,\zeta)\}_{n\in\mathbb{N}}$ converge en loi vers $\mathbb{P}_\xi$. On suppose également que pour tout $u \in K$:

$$\int |J(u,\xi)|\,\mathbb{P}^n(d\xi,\zeta) < \infty.$$

Cela signifie que $(\mathbb{P}^n(\cdot,\zeta))_{n\in\mathbb{N}}$ est une suite d'estimateurs d'une mesure de probabilité $\mathbb{P}$, dépendant d'un échantillon ζ. Alors,

$$V_n(Q_k) \stackrel{\text{def}}{=} \inf_{u_j} \left\{ \sum_{j=1}^{S_k} \alpha_j^k(n) \int J(u_j,\xi) \frac{\mathbb{I}_{C_j^k}(\xi)}{\alpha_j^k(n)} \mathbb{P}^n(d\xi,\zeta) \right\},$$

où $\alpha_j^k(n) \stackrel{\text{def}}{=} \int \mathbb{I}_{C_j^k}(\xi)\mathbb{P}^n(d\xi,\zeta)$. Pour (k,j) fixés, l'application :

$$\mu^n_{(j,k)}(\cdot,\zeta) : B \in \mathcal{B}_\Xi \mapsto \int \mathbb{I}_B(\xi) \frac{\mathbb{I}_{C_j^k}(\xi)}{\alpha_j^k(n)} \mathbb{P}^n(d\xi,\zeta),$$

définit une mesure de probabilité sur $(\Xi, \mathcal{B}_\Xi)$.

Lemme IV.25. *Si pour presque tout ζ, la suite $\{\mathbb{P}^n(\cdot,\zeta)\}_{n\in\mathbb{N}}$ converge en loi vers $\mathbb{P}_\xi$ et si $\mathbb{P}_\xi(\xi \in \partial C_j^k)^3 = 0$, alors, pour presque tout ζ, la suite $\left\{\mu^n_{(j,k)}(\cdot,\zeta)\right\}_{n\in\mathbb{N}}$ converge en loi vers la loi $\mu_{(j,k)}$:*

$$\mu_{(j,k)}(d\xi) = \frac{\mathbb{I}_{C_j^k}}{\mathbb{E}\left[\mathbb{I}_{C_j^k}(\xi)\right]} \mathbb{P}_\xi(d\xi).$$

Preuve : Soit $s : \Xi \to \mathbb{R}$ une fonction continue et bornée, alors :

$$\int s(\xi)\mu^n_{(j,k)}(d\xi,\zeta) \;=\; \frac{1}{\alpha_j^k(n)} \int s(\xi)\mathbb{I}_{C_j^k}(\xi)\mathbb{P}^n(d\xi,\zeta).$$

[3]Nous noterons ∂E la frontière de l'ensemble E. Par définition ∂E est l'adhérence de l'ensemble lui même privé de son intérieur.

L'ensemble des points de discontinuité de l'application $\xi \mapsto s(\xi)\mathbb{I}_{C_j^k}(\xi)$ étant contenus dans ∂C_j^k, est donc de mesure nulle pour la loi de ξ. Par conséquent :

$$\lim_{n\to\infty} \int s(\xi)\mathbb{I}_{C_j^k}(\xi)\mathbb{P}^n(d\xi,\zeta) = \int s(\xi)\mathbb{I}_{C_j^k}(\xi)\mathbb{P}_\xi(d\xi). \tag{IV.26}$$

Par ailleurs nous avons aussi que :

$$\lim_{n\to\infty} \alpha_j^k(n) = \mathbb{P}_\xi(C_j^k) \ ;$$

alors :

$$\begin{aligned} \lim_{n\to\infty} \int s(\xi)\mu_{(j,k)}^n(d\xi,\zeta) &= \frac{1}{\mathbb{E}\left[\mathbb{I}_{C_j^k}(\xi)\right]} \int s(\xi)\mathbb{I}_{C_j^k}(\xi)\mathbb{P}_\xi(d\xi) \ ; \\ &= \int s(\xi)\mu_{(j,k)}(d\xi). \end{aligned}$$

□

Théorème IV.26. *Supposons que les hypothèses IV.24, IV.23 soient vérifiées et que pour tout couple* (j,k), $\mathbb{P}_\xi(\xi \in \partial C_j^k) = 0$. *Alors :*

$$\lim_{n\to\infty} V_n(Q_k) = V(Q_k).$$

Preuve : Pour tout couple (j,k) la famille de mesures de probabilité $\left\{\mu_{(j,k)}^n(\cdot,\zeta)\right\}_{n\in\mathbb{N}}$ et la fonction J vérifiant les hypothèses de [38, théorème 3.7] alors :

$$\lim_{n\to\infty} \inf_{u_j} \int J(u_j,\xi)\mu_{(j,k)}^n(d\xi,\zeta) = \inf_{u_j} \int J(u_j,\xi)\mu_{(j,k)}(d\xi).$$

Alors :

$$\begin{aligned} \lim_{n\to\infty} V_n(Q_k) &= \sum_{j=1}^{S_k} \mathbb{P}_\xi(C_j^k) \inf_{u_j} \int J(u_j,\xi)\mu_{(j,k)}(d\xi) \ ; \\ &= \sum_{j=1}^{S_k} \mathbb{P}_\xi(C_j^k) \inf_{u_j} \int J(u_j,\xi) \frac{\mathbb{I}_{C_j^k}(\xi)}{\mathbb{P}_\xi(C_j^k)} \mathbb{P}_\xi(d\xi) \ ; \\ &= V(Q_k). \end{aligned}$$

□

Exemple IV.27. Nous allons donner un procédé de construction d'une suite de lois vérifiant les conditions du théorème. Nous disposons d'une variable aléatoire $\xi : \Omega \to \Xi$. Soit $(\Xi^{\mathbb{N}}, \mathcal{B}_{\Xi}^{\otimes\mathbb{N}}, \mathbb{P}_{\xi}^{\otimes\mathbb{N}})$ l'espace des échantillons de ξ de longueur quelconque. Soit $\zeta = \{\xi_i\}_{i\in\mathbb{N}}$ considérons alors :

$$\mathbb{P}^n(\cdot, \zeta) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} \delta_{\xi_i}.$$

Pour tout $A \in \mathcal{B}_{\Xi}$, nous avons clairement que l'application $\zeta \mapsto \mathbb{P}^n(A, \zeta)$ est mesurable par rapport à la tribu engendrée par $(\xi_1, \ldots, \xi_n)$. Pour presque tout ζ il est également clair que l'application $\mathbb{P}^n(\cdot, \zeta)$ est une mesure de probabilité sur Ξ. Soit maintenant s une variable aléatoire continue bornée sur Ξ alors :

$$\int s(\xi)\mathbb{P}^n(d\xi, \zeta) = \frac{1}{n}\sum_{i=1}^{n} s(\xi_i) \ ;$$

alors d'après la loi forte des grands nombres nous avons :

$$\lim_{n\to\infty} \int s(\xi)\mathbb{P}^n(d\xi, \zeta) = \mathbb{E}\left[s(\xi)\right], \quad \mathbb{P} - \text{p.s.}.$$

Autrement dit la suite $\{\mathbb{P}^n(\cdot, \zeta)\}_{n\in\mathbb{N}}$ converge en loi pour presque tout ζ vers la loi de ξ, c'est le théorème de Glivenko-Cantelli (voir théorème B.15). Par ailleurs, la loi empirique $\mathbb{P}^n$ conduit à une approximation naturelle du problème d'optimisation :

$$V_n(Q_k) = \inf_{u_j, j=1,\ldots,S_k} \frac{1}{n}\sum_{\ell=1}^{n}\sum_{j=1}^{S_k} J(u_j, \xi_\ell)\mathbb{I}_{C_j^k}(\xi_\ell). \tag{IV.27}$$

IV.3.5 Comportement asymptotique des coûts discrets

Nous allons maintenant combiner les résultats asymptotiques obtenus pour la discrétisation de la structure d'information et pour la discrétisation de la variable aléatoire ξ afin d'obtenir un résultat de convergence globale de la suite des valeurs des problèmes approchés (IV.25) vers la valeur du problème d'origine (IV.23). Notre démarche repose sur une simple inégalité triangulaire :

$$|V - V_n(Q_k)| \leq |V - V(Q_k)| + |V(Q_k) - V_n(Q_k)|. \tag{IV.28}$$

L'erreur totale de discrétisation $|V - V_n(Q_k)|$ peut être majorée par la somme de deux termes. Un terme qui peut s'interpréter comme l'erreur commise en discrétisant la structure d'information, un autre terme qui peut s'interpréter comme l'erreur liée à l'estimation de l'espérance du critère.

Théorème IV.28. *Soit $(Q_k)_{k\in\mathbb{N}}$ une suite de variables aléatoires à valeurs discrètes telle que $\{Q_k \circ h\}_{k\in\mathbb{N}}$ converge en probabilité vers h. On suppose que les hypothèses IV.24, IV.23 et IV.17 sont satisfaites, que l'application F est continue sur $L^1_{\mathbb{R}^m}(\Xi)$. Alors pour tout $\varepsilon > 0$ il existe $N_\varepsilon \in \mathbb{N}$ tel que :*

$$\forall k \geq N_\varepsilon, \quad \lim_{n\to\infty} |V - V_n(Q_k)| \leq \varepsilon.$$

Preuve : Soit $\varepsilon > 0$ fixé, d'après le théorème IV.21 il existe $\mathbb{N}_\varepsilon \in \mathbb{N}$ tel que pour tout $k \geq \mathbb{N}_\varepsilon$

$$|V - V(Q_k)| \leq \varepsilon.$$

Par conséquent pour tout k plus grand que N_ε :

$$|V - V_n(Q_k)| \leq |V - V(Q_k)| + |V(Q_k) - V_n(Q_k)| \leq \varepsilon + |V(Q_k) - V_n(Q_k)|$$

Le résultat se déduit simplement du théorème IV.26 par passage à la limite :

$$\lim_{n\to\infty} |V - V_n(Q_k)| \leq \varepsilon.$$

□

Interprétation des résultats

La discrétisation d'un problème d'optimisation stochastique, passe par deux étapes :
- une étape permettant la quantification de l'observation, afin de se ramener à un problème de dimension finie ;
- une autre étape permettant l'estimation des espérances (par exemple par une technique de Monte-Carlo) ;

la qualité de l'approximation du coût optimal dépend de la taille de l'échantillon, cette dernière quantité dépend de la finesse de la quantification.

La relation (IV.28) souligne un point très important de la discrétisation des problèmes d'optimisation stochastique. Elle met en évidence deux sources d'erreurs possibles attachées à la discrétisation.

1. Une première source d'erreur est liée à la quantité $|V - V(Q_k)|$: elle représente l'erreur algébrique issue de la discrétisation de la structure d'information ;
2. une deuxième source d'erreur est liée à la quantité $|V(Q_k) - V_n(Q_k)|$: elle représente l'erreur numérique issue de la discrétisation des espérances intervenant dans le problème.

On remarque que même si l'on fait tendre l'erreur numérique vers 0 en considérant des échantillons de plus en plus grands, on ne modifie en rien l'erreur algébrique !

Corollaire IV.29. *Sous les hypothèses du théorème IV.28 il existe une application croissante $k \mapsto k(n)$ telle que :*

$$\lim_{n\to\infty} V_n(Q_{k(n)}) = V.$$

Preuve : Ce résultat est une conséquence immédiate du théorème IV.28 et de la proposition B.18. □

Ce corollaire fait dépendre la quantification de la loi choisie pour remplacer la loi du problème d'origine.

IV.3.6 Convergence des solutions

Nous avons jusqu'ici traité uniquement de la question relative à la convergence du coût optimal $V_n(Q_k)$ vers le coût optimal du problème d'origine, à savoir V. Nous allons maintenant examiner le comportement de la suite de solutions optimales $(\gamma_{n,k})_{(n,k)\in\mathbb{N}\times\mathbb{N}}$ des problèmes associés à $V_n(Q_k)$.

Existence d'une solution L'existence d'une solution pour les problèmes d'allocation[4] stochastique à en particulier été étudiée dans [18, 1, 8, 10]. Les hypothèses sous lesquelles un problème d'optimisation, où les variables de décisions sont uniquement soumises à respecter des contraintes presque sûres et dont le critère se met sous la forme d'une intégrale, admet une solution, ont également été étudiées par Zolezzi [107, 21]. En dimension infinie, l'existence de solutions pour les problèmes d'optimisation passe par des arguments de convexité et de topologie faible.

Proposition IV.30. *On suppose les hypothèses du théorème IV.21 sont satisfaites. Soient $(\varepsilon_k)_{k\in\mathbb{N}}$ une suite de réels positifs qui converge vers 0 et une suite $(\gamma_k)_{k\in\mathbb{N}}$ telle que :*

$$F(\gamma_k) \leq V(Q_k) + \varepsilon_k \text{ et } \lim_{k\to\infty} ||\gamma_k - \gamma^*||_{L^p_{\mathbb{R}^m}(\Xi)} = 0.$$

Alors γ^* est solution du problème IV.23.

Preuve : D'après l'inégalité triangulaire nous avons :

$$\begin{aligned} |V - F(\gamma^*)| &\leq |V - V(Q_k)| + |V(Q_k) - F(\gamma_k)| + |F(\gamma_k) - F(\gamma^*)| \\ &\leq |V - V(Q_k)| + \varepsilon_k + \left|F(\gamma^k) - F(\gamma^*)\right|. \end{aligned}$$

En passant à la limite sur k nous avons alors $|V - F(\gamma^*)| = 0$, ce qui prouve que γ^* est solution du problème d'origine. □

[4] Un problème d'allocation stochastique est un problème d'optimisation ayant une contrainte de la forme $\mathbb{E}[X] \in C$.

Proposition IV.31. *Supposons que les hypothèses IV.24, IV.23 soient satisfaites. Si $\gamma_{n,k}$ est solution du problème d'optmisation IV.25 et que la suite $(\gamma_{n,k})_{n\in\mathbb{N}}$ converge p.s. vers γ_k, alors γ_k est solution du problème d'optimisation IV.24.*

Preuve : C'est une conséquence du théorème de Dupačovà et Wets [38, théorème 3.7] qui annonce un résultat d'épi convergence. □

IV.4 Deux exemples de discrétisation

IV.4.1 Un problème avec une contrainte de parité

Nous allons donner un exemple de problème d'optimisation stochastique afin d'illustrer certains résultats que nous avons énoncés. Le problème est le suivant : soient $\Xi = [-1, 1]$ muni de la tribu des boréliens sur Ξ et $\xi : \Omega \to \Xi$ une variable aléatoire de loi uniforme.

$$V \stackrel{\text{def}}{=} \inf \left\{ \mathbb{E}\left[(u(\xi) - \xi)^2\right] \mid u : \Xi \to \mathbb{R}, u \text{ est } \sigma(h)\text{-mesurable} \right\} \tag{IV.29}$$

où $h : z \in \Xi \to |z|$. La contrainte u est $\sigma(h)$-mesurable signifie simplement que u est une fonction mesurable paire. Nous aurons besoin du lemme technique suivant :

Lemme IV.32. *Soit $f : \Xi \to \mathbb{R}$ une variable aléatoire alors :*

$$\mathbb{E}\left(f \mid \sigma(h)\right) = \frac{f + \tilde{f}}{2},$$

où $\tilde{f} : \Xi \to \mathbb{R}, \quad z \mapsto f(-z)$.

Preuve : Il est clair que l'application :

$$z \in \Xi \mapsto \frac{f(z) + f(-z)}{2},$$

est $\sigma(h)$-mesurable. Montrons qu'il s'agit de l'espérance conditionnelle de f sachant $\sigma(h)$. Soit g une variable aléatoire réelle, puisque Ξ est symétrique alors :

$$\begin{aligned}
\mathbb{E}\left[f(\xi)g(|\xi|)\right] &= \mathbb{E}\left[f(-\xi)g(|\xi|)\right] \ ; \\
&= \frac{1}{2}\left(\mathbb{E}\left[f(\xi)g(|\xi|)\right] + \mathbb{E}\left[f(-\xi)g(|\xi|)\right]\right) \ ; \\
&= \mathbb{E}\left[\frac{f(\xi) + f(-\xi)}{2} g(|\xi|)\right].
\end{aligned}$$

□

Il est clair que l'application $J(x,z) \stackrel{\text{def}}{=} (x-z)^2$ est une intégrande normale, convexe en x, donc résoudre (IV.29) équivaut à résoudre :

$$\inf \left\{ \mathbb{E}\left[\tilde{J}(x(\xi),\xi) \right] \mid x : \Xi \to \mathbb{R} \text{ est } \mathcal{B}_\Xi\text{-mesurable} \right\} \tag{IV.30}$$

où :

$$\tilde{J}(x,z) \stackrel{\text{def}}{=} \mathbb{E}\left(J(x,\cdot) \mid \sigma(h) \right)(z).$$

D'après le lemme IV.32 nous avons alors :

$$\tilde{J}(x,z) = \frac{(x-z)^2 + (x+z)^2}{2} \; ; \tag{IV.31}$$

et donc :

$$\mathbb{E}\left[\tilde{J}(x(\xi),\xi) \right] = \mathbb{E}\left[x^2(\xi) + \xi^2 \right]. \tag{IV.32}$$

Étant donné l'expression (IV.32) il est évident que la solution du problème (IV.29) noté x^* est la variable aléatoire identiquement nulle. Le coût optimal du problème est alors :

$$V = \mathbb{E}\left[\xi^2 \right] = \frac{1}{3}.$$

Approche par le théorème IV.28 Soient $K \in \mathbb{N}$ et une application $Q_K : [0,1] \to [0,1]$. À ce stade, notre intérêt portera uniquement sur la partition associée à Q_K, que nous noterons $\mathsf{part}\, Q_K$:

$$\mathsf{part}\, Q_K \stackrel{\text{def}}{=} \left\{ I_k^K \mid k = 1, \ldots, K, \quad I_k^K = \left] \frac{k-1}{K}, \frac{k}{K} \right] \right\}.$$

Considérons alors le problème d'optimisation :

$$V_N(Q_K) \stackrel{\text{def}}{=} \min_{\{u_k\}_{k=1,\ldots,K}} \frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{N} (u_k - \xi_i)^2 \mathbb{I}_{I_k^K}(|\xi_i|).$$

On remarque que $V_N(Q_K)$ s'écrit également :

$$V_N(Q_K) = \min_{\{u_k\}_{k=1,\ldots,K}} \frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{N} (u_k^2 - 2u_k\xi_i) \mathbb{I}_{I_k^K}(|\xi_i|) + \frac{1}{N} \sum_{i=1}^{N} \xi_i^2.$$

Par ailleurs, posons :

$$\begin{aligned}
V_N^k(Q_K) &= \min_{u_k} \frac{1}{N}\sum_{i=1}^{N}(u_k^2 - 2u_k\xi_i)\mathbb{I}_{I_k^K}(|\xi_i|) \\
&= \min_{u_k} \frac{1}{N}\left(u_k^2 N_k^K - 2u_k\sum_{i=1}^{N}\xi_i\mathbb{I}_{I_k^K}(|\xi_i|)\right) \qquad \text{avec } N_k^K = \sum_{i=1}^{N}\mathbb{I}_{I_k^K}(|\xi_i|) \\
&= \min_{u_k} \frac{N_k^K}{N}\left(u_k^2 - 2u_k\frac{1}{N_k^K}\sum_{i=1}^{N}\xi_i\mathbb{I}_{I_k^K}(|\xi_i|)\right) \\
&= \min_{u_k} \frac{N_k^K}{N}\left(u_k - \frac{1}{N_k^K}\sum_{i=1}^{N}\xi_i\mathbb{I}_{I_k^K}(|\xi_i|)\right)^2 - \frac{N_k^K}{N}\left(\frac{1}{N_k^K}\sum_{i=1}^{N}\xi_i\mathbb{I}_{I_k^K}(|\xi_i|)\right)^2 \\
&= -\frac{N_k^K}{N}\left(\frac{1}{N_k^K}\sum_{i=1}^{N}\xi_i\mathbb{I}_{I_k^K}(|\xi_i|)\right)^2.
\end{aligned}$$

Ainsi nous avons donc que :

$$\begin{aligned}
V_N(Q_K) &= \sum_{k=1}^{K} V_N^k(Q_K) + \frac{1}{N}\sum_{i=1}^{N}\xi_i^2 \ ; \\
&= \sum_{k=1}^{K} -\frac{N_k^K}{N}\left(\frac{1}{N_k^K}\sum_{i=1}^{N}\xi_i\mathbb{I}_{I_k^K}(|\xi_i|)\right)^2 + \frac{1}{N}\sum_{i=1}^{N}\xi_i^2.
\end{aligned}$$

En utilisant la loi forte des grands nombres, on montre que :

$$\lim_{N\to\infty}\frac{N_k^K}{N} = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathbb{I}_{I_k^K}(|\xi_i|) = \frac{1}{K}.$$

À nouveau grâce la loi forte des grands nombres, nous avons alors :

$$\begin{aligned}
\lim_{N\to\infty}\left(\frac{1}{N_k^K}\sum_{i=1}^{N}\xi_i\mathbb{I}_{I_k^K}(|\xi_i|)\right)^2 &= \lim_{N\to\infty}\left(\frac{N}{N_k^K}\frac{1}{N}\sum_{i=1}^{N}\xi_i\mathbb{I}_{I_k^K}(|\xi_i|)\right)^2 \ ; \\
&= \left(K\mathbb{E}\left[\xi\mathbb{I}_{I_k^K}(|\xi|)\right]\right)^2.
\end{aligned}$$

Étant donné que la loi de la variable aléatoire ξ est uniforme sur $[-1,1]$, l'intégrale de ξ sur un intervalle symétrique par rapport à 0 est donc nulle. En particulier :

$$\mathbb{E}\left[\xi\mathbb{I}_{I_k^K}(|\xi|)\right] = 0.$$

Ce qui permet d'obtenir le résultat suivant :

$$\lim_{N\to\infty} V_N(Q_K) = \sum_{k=1}^{K} -K\left(\mathbb{E}\left[\xi\mathbb{I}_{I_k^K}(|\xi|)\right]\right)^2 + \mathbb{E}\left[\xi^2\right] = \mathbb{E}\left[\xi^2\right].$$

La solution u_k^N associée au k^{e} problème est calculable explicitement :

$$u_k^N = \frac{1}{N_k^K} \sum_{i=1}^{N} \xi_i \mathbb{I}_{I_k^K}(|\xi_i|).$$

Encore une fois la relation suivante est une conséquence de la loi des grands nombres :

$$\lim u_k^N = K\mathbb{E}\left[\mathbb{I}_{I_k^K}(|\xi|)\right] = 0 \quad \mathbb{P}-\text{p.s..}$$

Il est facile de voir que :

$$V(Q_K) \stackrel{\text{def}}{=} \inf_{u_k} \mathbb{E}\left[\sum_{k=1}^{K} (u_k - \xi)^2 \mathbb{I}_{I_k^K}(\xi)\right] = \mathbb{E}\left[\xi^2\right],$$

et que ce coût est atteint lorsque toutes les variables de décision sont identiquement égales à 0. En résumé, dans cet exemple l'erreur $|V - V(Q_K)|$ liée à la quantification est nulle et l'erreur liée à la discrétisation de Monte-Carlo $|V(Q_K) - V_N(Q_K)|$ converge vers 0 lorsque K est fixé et N tend vers l'infini.

Remarque IV.33. Si l'erreur liée à la quantification est nulle, c'est en partie à cause du fait que la commande optimale du problèmes lié à V est naturellement en boucle ouverte. Or, la quantification permet entre autre de se ramener à un problème en boucle ouverte.

Approche sans quantification Afin de montrer que la quantification est essentielle dans le théorème IV.28, nous allons essayer de résoudre ce problème par une technique de Monte-Carlo sans procéder à une quantification de la fonction d'observation. Soit $\xi_1, \ldots, \xi_N$ un N-échantillon de la variable aléatoire ξ. Considérons l'estimateur suivant :

$$f_N(\xi_1, \ldots, \xi_N, x) = \frac{1}{N} \sum_{i=1}^{N} J(x(\xi_i), \xi_i).$$

et le problème suivant :

$$V_N(\xi_1, \ldots, \xi_N) \stackrel{\text{def}}{=} \inf \left\{ f_N(\xi_1, \ldots, \xi_N, x) \mid \ x \text{ est } \sigma(h)\text{-mesurable} \right\}. \tag{IV.33}$$

Soit aussi l'ensemble A_N défini de la manière suivante :

$$A_N \stackrel{\text{def}}{=} \left\{ \omega \in \Omega \mid |\xi_1(\omega)|, \ldots, |\xi_N(\omega)| \text{ sont tous distincts} \right\}.$$

Nous avons clairement que A_N^{c} est négligeable pour la mesure de Lebesgue, par conséquent :

$$V_N(\xi_1, \ldots, \xi_N) = \inf_{x_i} \frac{1}{N} \sum_{i=1}^{N} (x_i - \xi_i)^2.$$

Nous observons alors aisément que $(\xi_1, \ldots, \xi_N)$, réalise le minimum du problème V_N car la valeur 0 qui est l'infimum est atteinte. Dans ce cas presque sûrement, le coût optimal V_N est toujours 0 quelle que soit la taille de l'échantillon.

IV.4.2 Un problème L.Q.G. avec une contrainte de non-anticipativité

Nous allons étudier un deuxième exemple pour lequel la contrainte de mesurabilité est simplement une contrainte de non-anticipativité.

Exemple IV.34. Soient :
- x une condition initiale aléatoire suivant une loi normale de moyenne nulle et d'écart-type σ_x ;
- u la variable de décision basée sur l'observation de x ;
- w une autre variable aléatoire suivant une loi normale de moyenne nulle et d'écart-type σ_w ; nous considérerons que x et w sont indépendantes, la matrice de covariance du couple (x, w) est :
$$\begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_w^2 \end{pmatrix}$$
- $y = x + u + w$ l'état "final" qui résulte de la réalisation du bruit et de la variable de décision ;
- l'objectif est de minimiser la fonction coût $\mathbb{E}[\varepsilon u^2 + y^2]$ —ε est un réel positif— par un choix de u comme une fonction de x.

Solution exacte

Dans le cas présent (linéaire-quadratique-gaussien et bruit de moyenne nulle), en supposant x et w indépendantes, il est bien connu que la décision optimale est donnée par un feedback linéaire sur l'observation x. Ainsi, en posant $u = cx$ dans la dynamique et la fonction coût, on obtient après des calculs élémentaires l'expression suivante de la fonction coût.
$$c^2(\varepsilon + 1)\sigma_x^2 + 2c\sigma_x^2 + \sigma_x^2 + \sigma_w^2.$$
La minimisation en c donne le cœfficient optimal du feedback :
$$c = -\frac{\sigma_x^2}{(1+\varepsilon)\sigma_x^2}$$
qui entraîne le coût optimal suivant :
$$\frac{1}{\varepsilon+1}(\varepsilon\sigma_x^2 + (\varepsilon+1)\sigma_w^2). \tag{IV.34}$$
Le feedback optimal :
$$\text{coefficient du feedback :} \quad c = -\frac{1}{\varepsilon+1} \ ;$$

Méthode de Monte-Carlo naïve

Supposons que l'on dispose de N trajectoires $(x_i, w_i),\ i = 1, \ldots, N$ issues de tirages indépendants et distribués selon la loi du couple (x, w). Pour chacune des trajectoires le coût évalué est :

$$(\varepsilon + 1)u_i^2 + 2(x_i + w_i)u_i + (x_i + w_i)^2.$$

La minimisation de cette expression en u donne la commande et le coût suivants :

$$\text{commande} : u_i = -\frac{x_i + w_i}{\varepsilon + 1}\ ; \quad \text{coût } \frac{\varepsilon(x_i + w_i)^2}{\varepsilon + 1}.$$

Nous constatons que le coût associé à une trajectoire de l'échantillon est d'ordre ε, et il en est donc de même de la moyenne des coûts associés à chacune des trajectoires de l'échantillon de taille N.

Cette approche naïve ne permet pas d'obtenir une approximation correcte du coût optimal du problème que l'on souhaite résoudre. L'orde de grandeur du coût optimal pour le problème discret est ε, alors que pour le problème d'origine il est de σ_x^2.

Approche passant par la quantification

Nous allons maintenant procéder à une discristisation respectant les hypothèses générale décrites dans ce chapitre. A savoir, une quantification des variables observées dans le but de se ramener à un problème en boucle ouverte, puis un changement de loi (on se ramène à une loi discrète) afin de pouvoir calculer les différentes expressions. Supposons que nous sommes dans le cas particulier où nous connaissons la loi (supposées gaussiennes) des variables aléatoires x et w . Nous sommes alors en mesure de construire une variables aléatoire discrète ayant les mêmes moments que les variables que l'on souhaite approcher. La figure IV.1 représente les différents scénarios associés à un couple de variables aléatoires discrètes (x, w) prenant chacune deux valeurs. Sur la figure IV.1 les deux occurrences possibles de x sont notées 1 et 2, chacune d'entre elles étant suivie des deux occurrences possibles de w notées 3 et 4 et 5 et 6. Les nœuds 1 et 2 ont une probabilité π_1 et π_2 (de somme égale à 1) et la deuxième génération de nœuds porte les probabilités $\pi_i\ i = 3, \ldots, 6$, avec $\pi_1 = \pi_3 + \pi_4$ et $\pi_2 = \pi_5 + \pi_6$.

Par ailleurs la quantification de x que nous considérons permet d'associer les décisions u_1 et u_2 aux nœuds 1 et 2 respectivement. On note $\mathbb{E}_0[y], \mathbb{E}_1[y], \mathbb{E}_2[y]$ les expressions suivantes :

$$\begin{aligned}
\mathbb{E}_0[y] &\stackrel{\text{def}}{=} \pi_1 y_1 + \pi_2 y_2\ ; \\
\mathbb{E}_1[y] &\stackrel{\text{def}}{=} \frac{\pi_3 y_3 + \pi_4 y_4}{\pi_1}\ ; \\
\mathbb{E}_2[y] &\stackrel{\text{def}}{=} \frac{\pi_5 y_5 + \pi_6 y_6}{\pi_2}.
\end{aligned}$$

Nous allons remplacer les variables aléatoires du problème d'origine, par des variables aléatoires discrètes en respectant systématiquement leurs deux premiers moments. Autrement dit on choisit π_1 et π_2 solution du système :

$$\mathbb{E}_0[x] = 0 \text{ et } \mathbb{E}_0[x^2] = \sigma_x^2\ ;$$

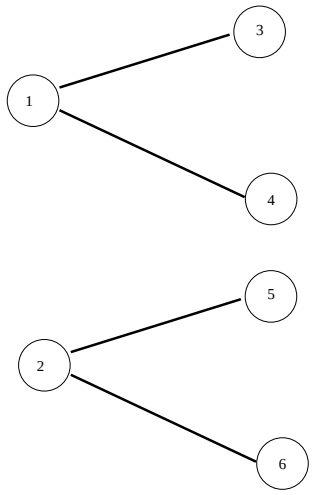


FIG. IV.1 – Scénarios

on choisit $\pi_3,\ \pi_4,\ \pi_5$ et π_6 tels que :

$$\mathbb{E}_1[w] = 0, \quad \mathbb{E}_1[w^2] = \sigma_w^2, \quad \mathbb{E}_2[w] = 0 \text{ et } \mathbb{E}_2[w^2] = \sigma_w^2.$$

Les calculs conduisent à l'expression suivante de la fonction coût :

$$\varepsilon\pi_1 u_1^2 + \pi_3(u_1 + x_1 + w_3)^2 + \pi_4(u_1 + x_1 + w_4)^2 + \varepsilon\pi_2 u_2^2 + \pi_5(u_2 + x_2 + w_5)^2 + \pi_6(u_2 + x_2 + w_6)^2,$$

qui une fois minimisée, conduit à la commande optimale :

$$\text{commandes : } u_1 = -\frac{x_1 + \mathbb{E}_1[w]}{\varepsilon + 1}, \quad u_2 = -\frac{x_2 + \mathbb{E}_2[w]}{\varepsilon + 1} \ ;$$

et au coût suivant :

$$\frac{\pi_1}{\varepsilon+1}(\varepsilon x_1^2 + (\varepsilon+1)\mathbb{E}_1[w^2] + 2\varepsilon x_1 \mathbb{E}_1[w] - (\mathbb{E}_1[w])^2) \\ + \frac{\pi_2}{\varepsilon+1}(\varepsilon x_2^2 + (\varepsilon+1)\mathbb{E}_2[w^2] + 2\varepsilon x_2 \mathbb{E}_2[w] - (\mathbb{E}_2[w])^2). \quad \text{(IV.35)}$$

La formule (IV.35) est parfaitement consistante avec la formule (IV.34).

Conclusion

La méthode de Monte-Carlo appliquée sans quantification des variables d'observation donne des résultats fantaisistes au niveau de la fonction valeur. Lorsque l'on procède au préalable à une quantification des variables d'observation, les résultats redeviennent consistants avec la solution théorique. Cet exemple montre à quel point il est risquer de s'écarter de ce schéma de discrétisation en deux parties.

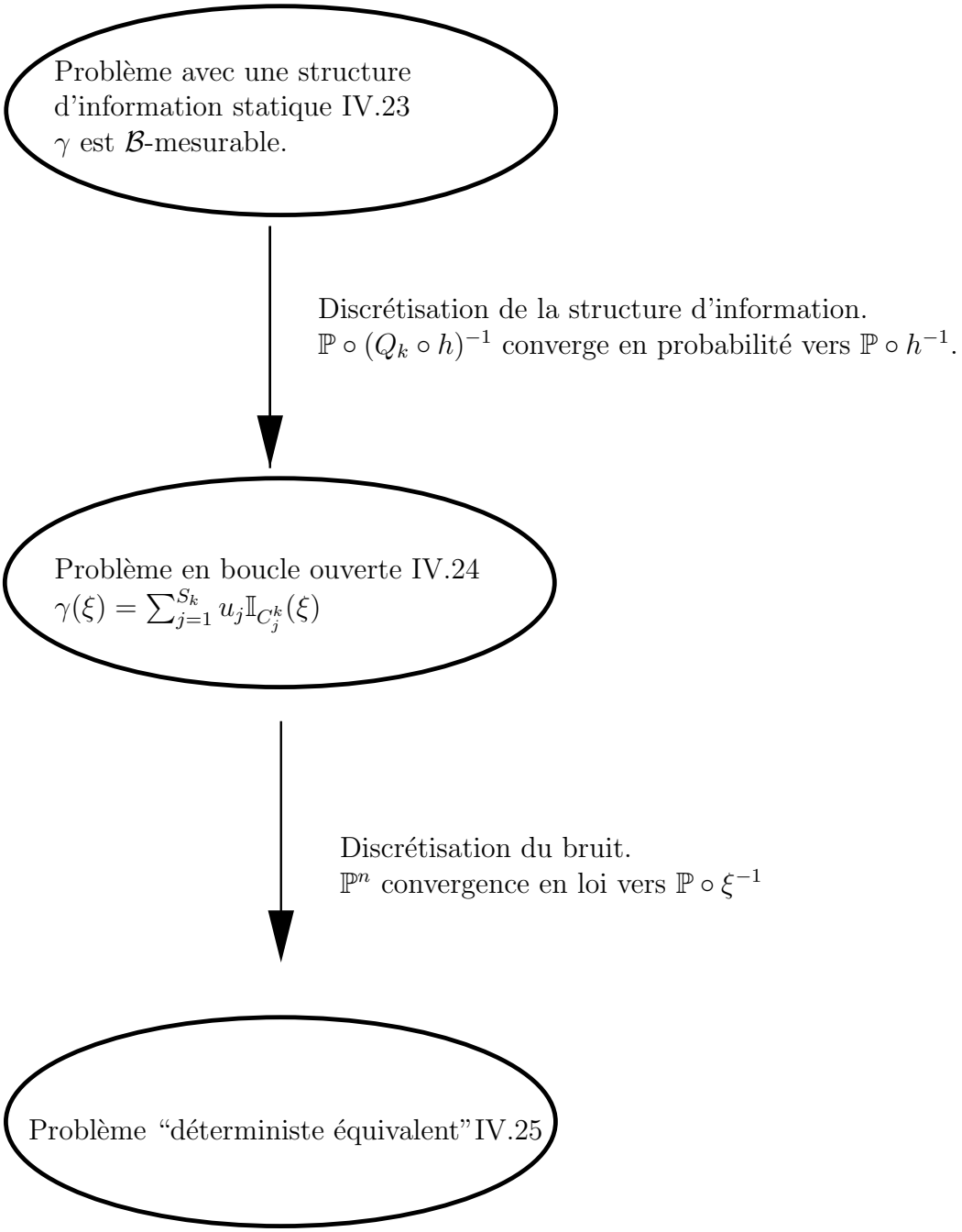


FIG. IV.2 – Résumé

Chapitre V

Expériences numériques

Dans ce chapitre, on présente un problème de commande optimale stochastique avec lequel on peut envisager de conduire des expérimentations sur la technique des chroniques arborescentes et sur la synthèse de feedback. Les caractéristiques du problème qui se prêtent à cette expérimentation et les possibilités d'extensions du problème seront discutées.

V.1 Formulation du problème

On considère un barrage hydroélectrique dont le volume d'eau en stock à l'instant t est noté $x(t)$. On a

$$\underline{x} \leq x(t) \leq \overline{x} \; . \tag{V.1}$$

On posera le problème en temps discret sur 24 heures avec des pas de temps d'une heure : alors l'indice $t = 0, 1, \ldots, T$ (avec $T = 24$) désigne les points de discrétisation, sachant que $t = 24$ correspond à la même heure que $t = 0$, mais 24 heures plus tard.

Les apports d'eau dans le barrage pendant l'heure t, $t = 1, \ldots, 24$, sont désignés par $w(t)$. Il s'agit d'un processus stochastique dont les caractéristiques seront exposées plus loin.

Le volume d'eau effectivement turbiné pendant la même heure t est désigné par $v(t-1)$. Ce décalage d'indice sera expliqué plus loin.

Ce volume d'eau turbiné résulte de la commande $u(t-1)$ qui est le volume turbiné "désiré" : si c'est possible, $v(t) = u(t)$. L'égalité est impossible si le volume d'eau résultant dans le réservoir tendrait à passer au dessous de $\underline{x}$. On a donc :

$$v(t) = \min\left(u(t), x(t) + w(t+1) - \underline{x}\right) \; . \tag{V.2}$$

Par ailleurs, lorsque le réservoir atteint son niveau maximum $\overline{x}$, il déborde (l'eau ainsi déversée ne produit pas d'électricité). On a donc :

$$x(t+1) = \min\left(x(t) - v(t) + w(t+1), \overline{x}\right) \; , \;\; t = 0, \ldots, T-1 \; , \tag{V.3}$$

Par ailleurs, la commande u est soumise à des contraintes de bornes :

$$\underline{u} \leq u(t) \leq \overline{u} \; . \tag{V.4}$$

Au volume d'eau $v(t)$ effectivement turbiné est associée une production d'électricité $p(t) = f\big(x(t), v(t)\big)$ (pouvant dépendre de la hauteur d'eau dans le barrage, et donc du volume $x(t)$). Cette production est destinée à répondre à chaque instant à une certaine demande pendant la même période qui sera notée $d(t+1)$: $d(\cdot)$ est un autre processus stochastique que l'on décrira plus loin. Si $p(t) \geq d(t+1)$, l'excès de production est revendu. Si $p(t) \leq d(t+1)$, le déficit de production doit être compensé par d'autres moyens ou bien il conduit à de la défaillance. Tout ceci se traduit par un coût $L\big(d(t+1) - p(t), t\big)$ où L est une fonction croissante par rapport à son premier argument. On précisera plus loin son expression.

Finalement, on peut résumer le problème de la façon suivante, en introduisant une rémunération du stock final $x(T)$ sous la forme $C\big(x(T)\big)$, et en supposant la condition initiale $x(0)$ donnée comme une variable déterministe ou aléatoire de loi connue :

$$\min \mathbb{E}\left[\sum_{t=0}^{T-1} L\Big(d(t+1) - f\big(x(t), v(t)\big), t\Big) + C\big(x(T)\big)\right] \tag{V.5a}$$

$$\text{sous les contraintes (V.2), (V.3), (V.4),} \tag{V.5b}$$

où $\mathbb{E}$ désigne l'espérance mathématique. Il est inutile d'imposer les contraintes de borne (V.1) qui sont satisfaites ipso facto par la paire d'équations (V.2)–(V.3).

La minimisation est faite en u en feedback causal sur x, w, d, les bruits étant supposés observables. Plus précisément, $u(t)$ peut dépendre de $\{x(s), w(s), d(s)\}_{0 \leq s \leq t}$. Ce n'est d'ailleurs qu'une fois que la loi de feedback a été spécifiée que l'espérance mathématique par rapport aux lois de probabilité sur la condition initiale et sur les bruits prend un sens, toutes les variables du problème devenant des variables aléatoires.

La structure d'information choisie correspond à une optique "décision-hasard" : la décision $u(t)$ sur le turbiné désiré est prise d'abord, l'apport d'eau $w(t+1)$ et la demande $d(t+1)$ intervenant dans le même intervalle ne sont observés qu'ensuite, ce qui résulte finalement dans le turbiné $v(t)$ et la production $p(t)$. Cette considération montre que les contraintes (V.1) ne peuvent pas être imposées à la commande dans la version stochastique (car la commande n'a pas les moyens de les respecter), mais, comme on l'a vu, la paire (V.2)–(V.3) satisfait automatiquement ces contraintes.

V.2 Hypothèses probabilistes

On suppose que $w(\cdot)$ et $d(\cdot)$ sont donnés par des bruits additifs autour d'une moyenne fonction du temps. Les trajectoires moyennes seront données. Les écarts à la moyenne seront les sorties de deux chaînes de Markov indépendantes données, centrées autour de zéro.

Par conséquent, si on voulait ramener le problème stochastique a une formulation markovienne, il faudrait trois états, à savoir x et les deux états pour les deux chaînes de Markov.

Dans l'approche par chroniques que l'on se propose de suivre, et dont on va décrire le protocole un peu plus précisément ci-après, le seul "état" à considérer est l'état physique x : on n'est pas censé connaître les caractéristiques stochastiques des bruits. Au stade de la synthèse de feedback qui est l'un des objectifs principaux de l'étude, on doit s'apercevoir qu'un feedback de $u(t)$ sur $x(t), w(t), d(t)$ est meilleur qu'un feedback sur $x(t)$ seul. Mais on peut espérer qu'un feedback sur $x(t)$ et $x(t-1)$ fournira une qualité comparable : en effet, hors des butées sur les bornes (V.1), la dynamique est quasiment linéaire et $w(t)$ s'obtient en connaissant $x(t)$ et $x(t-1)$ (et également $u(t-1)$). La situation pour $d(t)$ est différente dans la mesure où la demande n'intervient pas directement dans la dynamique. Cependant, *après feedback*, la dynamique dépend aussi de la demande via les commandes. Il est difficile de dire comment $d(t)$ est alors liée à $x(t)$ et à ses valeurs passées, mais c'est tout l'objet de l'étude de la synthèse de feedback que de déterminer la qualité de diverses politiques de bouclage envisageables (compte tenu de la structure d'information décrite précédemment).

V.3 Extensions possibles du problème

Au delà d'une première étude, pour montrer notre capacité à traiter de grands systèmes stochastiques par cette approche des chroniques arborescentes, on peut envisager d'augmenter de façon arbitraire la taille du problème en mettant en série des réservoirs dans une même vallée hydraulique : le couplage se fait par la dynamique car l'eau turbinée — ou déversée — en amont constitue — après un certain délai éventuel — un apport d'eau pour l'aval ; il se fait aussi par la fonction coût puisque l'ensemble des productions s'additionnent pour répondre à la demande commune.

On peut aussi considérer des réservoirs en parallèle dans des vallées hydrauliques distinctes, le couplage n'ayant alors lieu que par la fonction coût. Enfin, tous les montages série-parallèle sont envisageables.

Revenant au cas mono-réservoir, on peut aussi faire varier les caractéristiques des bruits (bruit blanc, chaînes de Markov à plusieurs retards, corrélation positive pour certaines chaînes de Markov, négative pour d'autres, etc.).

V.4 Traitement du problème déterministe sur les chroniques

Le but final de l'étude étant celui de la synthèse de feedback pour laquelle il s'agit de traiter des arbres ayant un maximum de nœuds, on ne cherchera pas à résoudre le problème de commande optimale déterministe sur l'arbre des chroniques avec une très grande précision. On peut souhaiter par exemple traiter le problème d'optimisation avec une méthode de gradient conjugué, ce qui suppose de préserver une forme différentiable à (V.2), (V.3) et de ne pas traiter directement les contraintes de bornes (V.4) (la projection non linéaire sur cet ensemble admissible ne préservant pas la propriété de conjugaison). On

pourra (dans le cas où les contraintes (V.4) ne sont pas traitées par projection) remplacer la commande u par

$$\min\big(\max(\underline{u},\underline{u}),\overline{u}\big) \tag{V.6}$$

qui reste automatiquement dans les bornes (V.4), la nouvelle commande $\underline{u}$ étant libre (cependant, au niveau de l'exploitation statistique des résultats, c'est avec u, et non avec $\underline{u}$ que l'on opèrera).

Les équations (V.2), (V.3) et (V.6) qui sont composées de min et max emboîtés vont être approximées par des fonctions lisses. L'approximation de base repose sur la formule

$$\min_c(x,y) = \begin{cases} y & \text{si } y \le x-c\,, \\ \dfrac{x+y}{2} - \dfrac{(x-y)^2}{4c} - \dfrac{c}{4} & \text{si } x-c \le y \le x+c\,, \\ x & \text{si } y \ge x+c\,, \end{cases} \tag{V.7}$$

fonction différentiable dépendant d'un "petit" paramètre c positif, que l'on peut obtenir par régularisation (concave) de Yosida-Moreau de la fonction $x \mapsto \min(x,y)$, régularisée que l'on translate horizontalement en x et y pour obtenir la coïncidence des branches asymptotiques (quand x ou y tend vers l'infini). La Figure **??** représente la fonction $x \mapsto \min_{0,2}(x,0)$ et sa dérivée première.

La fonction $\max_c(x,y)$, fonction "lissée" du max, est obtenue par la formule $-\min_c(-x,-y)$. Les fonctions lissées des fonctions max-min ou min-max apparaissant dans (V.6) sont obtenues par composition des fonctions précédentes, et ainsi de suite.

À partir de cette nouvelle formulation approchée différentiable, on peut envisager la résolution du problème, par exemple, par une méthode de Fletcher-Reeves (gradient conjugué).

V.5 Données numériques et fonctionnelles

V.5.1 Demande et apports d'eau

Moyenne

La composante *déterministe* (moyenne) de la demande est donnée par le vecteur

$$d = (0.9, 1, 0.7, 0.8, 0.9, 1, 0.9, 0.4, 0.3, 0.25, 0.2, 0.3, 0.4, 0.5, 0.8, 0.7, 0.7, \\ 0.8, 0.7, 0.6, 0.6, 0.8, 0.7, 0.5)\,,$$

tandis que celle des apports d'eau est donnée par le vecteur

$$w = (0.25, 0.5, 0.8, 0.9, 0.2, 0.3, 0.7, 0.5, 0.9, 0.4, 0.7, 0.8, 0.7, 1, 0.7, 0.3, \\ 0.4, 0.6, 0.8, 0.7, 0.6, 0.8, 0.9, 1)\,.$$

Ces composantes déterministes sont représentées sur la Figure V.1 avec leur différence, les valeurs cumulées sur le temps et les différences cumulées.

On voit que les apports d'eau cumulés sont en excédent par rapport à la demande cumulée en début de période mais que le bilan en fin de période est neutre (en supposant bien sûr qu'une unité d'apport d'eau est transformable approximativement en une unité de demande d'électricité). Comme, dans le début de période (correspondant à la nuit), on prendra une fonction coût qui reflète une faible valeur unitaire de la demande ("tarif de nuit"), on sera donc amené à stocker l'eau au maximum pour l'utiliser ultérieurement.

Partie stochastique

On superpose à la composante déterministe de la demande et des apports d'eau des composantes stochastiques issues de chaînes de Markov. On utilisera les deux matrices de transition ci-après d'une chaîne à 11 états, numérotés de 1 à 11 ; à chaque état sera associé une perturbation de la composante déterministe, perturbation calculée proportionnellement à l'écart entre le numéro de l'état tiré et le numéro 6 qui est l'état médian ; autrement dit, l'état médian correspond à une valeur nulle de la perturbation, l'état 1 et l'état 11 correspondant aux perturbations maximales (respectivement positive et négative). Le coefficient de proportionnalité qui a été choisi pour s'appliquer à $6-i$ (pour l'état i de la chaîne) est de 0,04. Autrement dit, à l'heure t, si l'état i de la chaîne de Markov relative à la demande est tiré, la demande est $d(t+1)+0{,}04(6-i)$.

On utilise les deux matrices de transition suivantes :

$$\begin{pmatrix}
7/12 & 1/3 & 1/12 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1/3 & 1/3 & 1/4 & 1/12 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1/12 & 1/4 & 1/3 & 1/4 & 1/12 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1/12 & 1/4 & 1/3 & 1/4 & 1/12 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1/12 & 1/4 & 1/3 & 1/4 & 1/12 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1/12 & 1/4 & 1/3 & 1/4 & 1/12 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1/12 & 1/4 & 1/3 & 1/4 & 1/12 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1/12 & 1/4 & 1/3 & 1/4 & 1/12 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1/12 & 1/4 & 1/3 & 1/4 & 1/12 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/12 & 1/4 & 1/3 & 1/3 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/12 & 1/3 & 7/12
\end{pmatrix}$$

et

$$\begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/12 & 1/3 & 7/12 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/12 & 1/4 & 1/3 & 1/3 \\
0 & 0 & 0 & 0 & 0 & 0 & 1/12 & 1/4 & 1/3 & 1/4 & 1/12 \\
0 & 0 & 0 & 0 & 0 & 1/12 & 1/4 & 1/3 & 1/4 & 1/12 & 0 \\
0 & 0 & 0 & 0 & 1/12 & 1/4 & 1/3 & 1/4 & 1/12 & 0 & 0 \\
0 & 0 & 0 & 1/12 & 1/4 & 1/3 & 1/4 & 1/12 & 0 & 0 & 0 \\
0 & 0 & 1/12 & 1/4 & 1/3 & 1/4 & 1/12 & 0 & 0 & 0 & 0 \\
0 & 1/12 & 1/4 & 1/3 & 1/4 & 1/12 & 0 & 0 & 0 & 0 & 0 \\
1/12 & 1/4 & 1/3 & 1/4 & 1/12 & 0 & 0 & 0 & 0 & 0 & 0 \\
1/3 & 1/3 & 1/4 & 1/12 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
7/12 & 1/3 & 1/12 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}$$

qui sont des matrices bi-stochastiques puisque symétriques, et admettant donc la loi uniforme comme mesure invariante. La première matrice représente une forte corrélation positive entre deux états successifs tandis que la seconde représente au contraire une forte corrélation négative. Chacune de ces deux matrices pourra être utilisée pour perturber la demande et/ou les apports d'eau.

V.5.2 Production électrique

La fonction f donne la conversion entre le turbiné v et la production électrique p lorsque le barrage est au niveau x. Pour représenter l'effet de hauteur de chute, on supposera une variation linéaire entre 0,5 et 1 par rapport à x lorsque x passe du niveau bas $\underline{x}$ au niveau haut $\overline{x}$. On adopte donc l'expression :

$$p = f(x, v) = v \times (x + \overline{x} - 2\underline{x})/2(\overline{x} - \underline{x}) \ . \tag{V.8}$$

V.5.3 Coût intégral

Le coût intégral est modulé en temps par le tarif τ représenté sur la Figure **??**. La fonction L est le produit de cette fonction par une fonction exponentielle qui vaut 0 et qui est de pente 1 au point 0. On a finalement :

$$L(y, t) = \tau(t + 1)(e^y - 1) \ . \tag{V.9}$$

Le coût marginal d'une défaillance augmente donc très vite du côté positif, tandis que le gain marginal tend vite vers zéro pour un grand excès de production par rapport à la demande (côté $y < 0$).

V.5.4 Coût final

On prend

$$C(x) = 12(x - \overline{x})^2 \ . \tag{V.10}$$

Le coût final est donc nul pour $x = \overline{x}$ et il croît si l'état final n'atteint pas la valeur maximum $\overline{x}$ (incitation à remplir le stock en fin de jeu).

V.5.5 Condition initiale

La condition initiale $x(0)$ est considérée comme une variable aléatoire de loi uniforme sur l'intervalle $[\underline{x}, \overline{x}]$.

V.5.6 Bornes sur l'état et la commande

On choisit $\underline{x} = 0$, $\underline{u} = 0$, $\overline{x} = 2$ et $\overline{u} = 1$ (en amenant le stock à son maximum à l'heure de pointe où la demande vaut 1, on peut, compte tenu de la fonction f de production, juste satisfaire la demande de pointe avec un turbiné maximum).

V.6 Résolution du problème déterministe

V.6.1 Calcul formel

Afin de valider le choix de certaines données, on a résolu le problème déterministe (demande et apports d'eau égaux à leurs moyennes données sur la Figure V.1 pour trois conditions initiales différentes ($x(0) = 0, 1$ et 2). La méthode de résolution a été un simple gradient (projeté sur l'intervalle admissible $[\underline{u}, \overline{u}]$ pour tout t — en fait, cette projection n'a presque pas joué dans nos essais). La méthode a été programmée en Scilab. Afin de permettre certaines vérifications, on donne ci-après les expressions de l'équation adjointe et du gradient par rapport à la commande. Dans ces essais numériques, la constante c intervenant dans l'approximation lisse (V.7) a été prise égale à 10^{-2}. On notera $m(x, y)$ à la place de $\min_c(x, y)$ pour raccourcir les notations. On a

$$L'_y(y,t) = \tau(t+1)e^y \qquad \text{(voir (V.9))}$$
$$f'_x(x,v) = v/2(\overline{x} - \underline{x}) \qquad \text{(voir (V.8))}$$
$$f'_v(x,v) = (x + \overline{x} - 2\underline{x})/2(\overline{x} - \underline{x}) \qquad \text{(voir (V.8))}$$
$$C'(x) = 24(x - \overline{x}) \qquad \text{(voir (V.10))}$$
$$m'_x(x,y) = \begin{cases} 0 & \text{si } y \le x - c\,, \\ 1/2 - (x-y)/2c & \text{si } x - c \le y \le x + c\,, \\ 1 & \text{si } y \ge x + c\,, \end{cases} \qquad \text{(voir (V.7))}$$
$$m'_y(x,y) = \begin{cases} 1 & \text{si } y \le x - c\,, \\ 1/2 - (y-x)/2c & \text{si } x - c \le y \le x + c\,, \\ 0 & \text{si } y \ge x + c\,, \end{cases} \qquad \text{(voir (V.7))}$$

Le Lagrangien $\mathcal{L}$ s'écrit, en introduisant l'état adjoint λ :

$$\sum_{t=0}^{T-1} \Bigg(L\Big(d(t+1) - f\big(x(t), m(u(t), x(t) + w(t+1) - \underline{x})\big), t\Big)$$
$$+ \lambda(t+1)\Big(m\big(x(t) - m(u(t), x(t) + w(t+1) - \underline{x}) + w(t+1), \overline{x}\big) - x(t+1)\Big)\Bigg) + C\big(x(T)\big)\,.$$

Alors, $\partial\mathcal{L}/\partial x(t) = 0$ donne, pour $t = T$,

$$\lambda(T) = C'\big(x(T)\big)\,,$$

et pour $t = 1, \ldots, T-1$,

$$\lambda(t) =$$
$$-L'_y\Big(d(t+1) - f\big(x(t), v(t)\big), t\Big) \times \Big(f'_x\big(x(t), v(t)\big) + f'_v\big(x(t), v(t)\big) \times m'_y\big(u(t), x(t) + w(t+1) - \underline{x}\big)\Big)$$
$$+ \lambda(t+1) \times m'_x\big(x(t) - v(t) + w(t+1), \overline{x}\big) \times \Big(1 - m'_y\big(u(t), x(t) + w(t+1) - \underline{x}\big)\Big)\,.$$

Enfin le gradient $\partial\mathcal{L}/\partial u(t)$ par rapport aux commandes s'écrit :

$$-L'_y\Big(d(t+1) - f\big(x(t), v(t)\big), t\Big) \times f'_v(x(t), v(t)\big) \\ + \lambda(t+1) \times m'_x\big(x(t) - v(t) + w(t+1), \overline{x}\big) \times m'_x\big(u(t), x(t) + w(t+1) - \underline{x}\big) \ .$$

V.6.2 Solutions numérique du problème déterministe

Ils sont résumés sur la Figure V.3.

V.7 Quantification numérique

Nous avons obtenu jusqu'ici des résultats théoriques permettant de guider la discrétisation d'un problème d'optimisation stochastique. Nous allons maintenant montrer comment nous allons nous servir de ces résultats pour calculer numériquement la solution d'un problème d'optimisation.

Nous disposerons le plus souvent dans la pratique que d'un échantillon de cette variable aléatoire. En supposant que l'échantillon dont on dispose est de taille suffisamment grande pour être représentatif de l'espace Ξ tout entier, nous choisissons de quantifier directement l'échantillon.

V.7.1 Cellules de Voronoï

Le problème est le suivant : on dispose de K couples $(v_1, \pi_1) \ldots, (v_K, \pi_K)$ avec $v_k \in \mathbb{R}^n$ et $\pi_k \geq 0$ et l'on souhaite classer les vecteurs v_k en p ($p \leq K$) classes homogènes en tenant compte de leur poids π_k. Autrement dit on cherche une partition de p classes qui soit la plus homogène que possible. Une approche pour résoudre se problème est de résoudre le problème d'optimisation suivant :

$$\min_{C_i} \min_{x_i} \sum_{i=1}^{p} \sum_{k \in C_i} \pi_k \, ||x_i - v_k||^2_{\mathbb{R}^n} \tag{V.11}$$

où $(C_i)_{i=1,\ldots,p}$ est une partition de l'ensemble d'indices $\{1, \ldots, K\}$. Dans la terminologie usuelle, les solutions $(C_i^*)_{i=1,\ldots,p}$ s'appellent les cellules de Voronoï et les éléments $(x_i^*)_{i=1,\ldots,p}$ s'appellent les centroïdes. Il est clair que connaissant C_i^*, alors x_i^* est solution du problème suivant :

$$\min_{x_i} \sum_{k \in C_i^*} \pi_k \, ||x_i - v_k||^2_{\mathbb{R}^n}$$

autrement dit :

$$x_i^* = \frac{1}{\displaystyle\sum_{k \in C_i^*} \pi_k} \sum_{k \in C_i^*} \pi_k v_k.$$

De même, supposons que les centroïdes solutions du problème (V.11) soient $\{x_i^*\}_{i=1,\dots,p}$, alors la partition optimale minimise :

$$\min_{C_i} \sum_{i=1}^{p} \sum_{k \in C_i} \pi_k \left|\left| x_i^* - v_k \right|\right|^2 . \tag{V.12}$$

Soit la partition $(C_i^*)_{i=1,\dots,p}$ définie par la relation suivante :

$$k \in C_i^* \Leftrightarrow \forall j \neq i, \quad \left|\left| x_i^* - v_k \right|\right| \leq \left|\left| x_j^* - v_k \right|\right| .$$

De même, nous définissons l'application $q^* : \{1, \dots, K\} \mapsto \{1, \dots, p\}$ définie par la relation suivante :

$$q^*(k) = i \Leftrightarrow k \in C_i^*.$$

Soit une application surjective $q : \{1, \dots, K\} \mapsto \{1, \dots, p\}$ alors :

$$\left|\left| x_{q^*(k)}^* - v_k \right|\right| \leq \left|\left| x_{q(k)}^* - v_k \right|\right| \ ;$$

donc :

$$\sum_{k=1}^{K} \pi_k \left|\left| x_{q^*(k)}^* - v_k \right|\right|^2 \leq \sum_{k=1}^{K} \pi_k \left|\left| x_{q(k)}^* - v_k \right|\right|^2 \ ;$$

ce qui signifie que la partition $\{C_i^*\}_{i=1,\dots,p}$ est solution du problème (V.12).

Construction numérique des cellules La méthode numérique que nous avons employée pour construire les classes est la suivante :

- à l'étape j, on dispose de p centroïdes $(x_i^j)_{i=1,\dots,p}$;
- on classe les vecteurs v_k de telle sorte que :

$$k \in C_i^{j+1} \Leftrightarrow \left|\left| x_i^j - v_k \right|\right|_{\mathbb{R}^n} \leq \left|\left| x_l^j - v_k \right|\right|_{\mathbb{R}^n} \quad \forall l \neq i.$$

  Il n'est pas nécessaire d'évaluer des normes pour faire ce calcul ; en effet, on peut procéder de la manière suivante :

  - supposons qu'a l'étape j nous ayons $v_k \in C_i^j$, alors pour tout $l \neq j$ on commence par déterminer l'équation de l'hyperplan médiateur H entre x_i^j et x_l^j :

$$x \in H \Leftrightarrow \quad f(x) \stackrel{\text{def}}{=} \left(x - \frac{x_i^j + x_l^j}{2} \right)^{\top} (x_i^j - x_l^j) = 0,$$

  - on évalue ensuite le signe de $f(v_k)$ et :

$$\begin{cases} f(v_k) \geq 0 \Rightarrow & v_k \in C_i^{j+1} \ ; \\ f(v_k) < 0 \Rightarrow & v_k \in C_l^{j+1}. \end{cases}$$

On recommence cette procédure $(p-1)$ fois et cela pour les K vecteurs, soit $K(p-1)$ évaluations de signe ;

– une fois que l'on a déterminé la partition $(C_i^{j+1})_{i=1,\dots,p}$, on calcule les nouvelles valeurs des centroïdes :

$$x_i^{j+1} = \frac{1}{\sum_{k \in C_i^{j+1}} \pi_k} \sum_{k \in C_i^{j+1}} \pi_k v_k \quad \forall i = 1, \dots, p \ ;$$

– on réalise le test d'arrêt suivant : si $C_i^j = C_i^{j+1} \quad \forall i = 1, \dots, p$ alors l'algorithme s'arrête, sinon on recommence la procédure avec :

$$x_i^j = x_i^{j+1} \quad \forall i = 1, \dots, p.$$

La figure V.4 montre le résultat de la classification en 20 cellules de 100000 vecteurs de $\mathbb{R}^2$ obtenus en faisant un tirage aléatoire selon une loi uniforme. On peut constater que les cellules qui se trouvent au centre ont une structure en nid d'abeille.

V.8 Mise œuvre en pratique

Au chapitre IV nous avons obtenus des résultats permettant de guider la discrétisation d'un problème d'optimisation stochastique avec une structure d'information statique. La figure IV.2 souligne les étapes importantes de cette discrétisation. Voici comment nous avons procédé pour nous mettre dans le domaine d'application du théorème IV.28.

1. Les bruits dans le problème du Barrage sont constitués de la demande en éléctricité, et des apports d'eau à chaque instant. Ces bruits sont en fait des trajectoires du bruit de dimension 25, autrement dit nous avons affaire à un bruit vectoriel $(d(t), w(t))_{t=0,\dots,24}$. Nous allons dans un premier temps simuler de manière indépendante, un nombre N de trajectoires de bruits vectoriel.
2. La première étape que nous avons décrite ne constitue pas encore une discrétisation du problème du Barrage. Les trajectoires que nous avons simulées seront celles qui serviront dans la suite pour évaluer le critère du problème en boucle ouverte que nous obtiendrons en quantifiant l'observation. La quantification permet simplement de transformer le problème d'optimisation d'origine en un problème d'optimisation en boucle ouverte. Toutefois, la qualité des commandes calculées dependra du nombre de trajectoires[1] passant par une classe de la partition de l'espace des bruits. En fait nous avons seulement besoin de partitionner les N trajectoires qui serviront à calculer les "espérances conditionnelle". La technique que nous utilisons pour créer une partition des N trajectoires vectorielles est basée sur l'emploi récursif de la technique que nous avons développée au paragraphe précédent.

[1]Dans notre cas nous allons remplacer la loi du bruit $(d(t), w(t))$ par sa loi empirique. La loi empirique est directement liée à la taille de l'échantillon, c'est la raison pour laquelle nous parlons du nombre de trajectoires

3. Une fois que nous avons déterminé une partition des N trajectoires, nous avons posé le problème. Schématiquement et pour résumer nous avons représenté sur la figure V.5 comment est formé le problème discret à partir des trajectoires simulées et de la partition que nous avons crée. Sur la figure V.5, les nœuds sur chacunes des trajectoires permettent de numéroter toutes les variables du problème d'optimisation (état, état adjoint et commandes). Les cercles qui entourent les nœuds représentent la nouvelle structure d'information discrète. Les commandes appartenant à une même classe doivent avoir la même valeur numérique. Le problème discret que nous allons résoudre est l'équivalent du problème IV.27.

Nous avons résolu numériquement le problème du Barrage en suivant la méthodologie que nous venons de décrire. Nous avons étudié la qualité des résultats obtenus en nous penchant plus particulièrement sur la qualité des argmins. En effet, une propriété importante d'une technique de résolution numérique d'un problème de commande optimale, est sa capacité à permettre de synthétiser une loi de feedback qui puissent être utilisable (et même optimale) pour le problème d'origine. Pour ce qui est du problème du Barrage, cette loi de feedback à pu être obtenue numériquement par un autre procédé : la programmation dynamique. La programmation dynamique est une technique pour résoudre les problèmes de commande optimale dans un cadre Markovien[2], malheureusement cette technique est souvent impraticable à cause de la necessité de stocker un trop grand nombre de valeurs numérique dès que la dimension de l'état[3] du système est supérieure à 3. Remarquons à cet effet que l'une des raisons du développement des techniques variationnelles du type de celle que nous avons mise en œuvre est justement de pouvoir permettre la résolution numérique de problèmes de commande optimale de grande dimension. Nous savons que dans le cas Markovien, le feedback optimal peut être une fonction de l'état du système dynamique, c'est à dire du niveau d'eau dans le Barrage. Nous avons donc calculé ce feedback par la programmation dynamique en discrétisant "finement" l'espace d'état du système, le niveau d'eau étant supposé être compris entre 0 et 1, nous avons discrétisé cet intervalle en 100 sous intervalles. Autrement dit en supposant que le niveau d'eau dans le Barrage ne peut prendre que 101 valeurs discrètes nous avons calculé le feedback optimal (approché tout de même) du problème du Barrage. Le feedback dans ce cas est simplement la quantité d'eau turbinée en fonction du niveau d'eau dans le Barrage. Par ailleurs, à partir de la commande optimale obtenue par la procédure variationnelle que nous avons décrite, nous pouvons également synthétiser une loi de feedback de la manière suivante : nous avons simplement associé à chaque valeur de l'état calculée à partir de la commande optimale, la valeur de la commande qui lui correspond. Nous avons comparé ces deux procédés (le feedback calculé par la programmation dynamique ayant le statut de feedback optimal du problème d'origine) en les représentant sur les figures suivantes.

Remarque V.1. Pour nos expériences numériques nous avons considéré $N = 1000$ trajectoires ce qui représente un nombre tout à fait insignifiant par rapport aux dimensions

[2]Cela signifie que les bruits qui existent la dynamique du problème sont indépendants

[3]Dans le cas du problème du Barrage, l'état du système est le niveau d'eau dans le Barrage, c'est donc un état de dimension 1, largement praticable par la programmation dynamique

du problème.

Analyse des résultats

- La première analyse qu'il convient de faire est que la "trace" de la quantification apparait clairement pour les deux, voir les trois premiers instants du processus. Au dela de $t = 2$, il n'y à plus de trace de l'opération de quantification. Par ailleurs les résultats semblent également être particulièrement bruité à partir de $t = 3$.
- Une autre remarque intéressante est que l'on observe que malgré une précision tout à fait médiocre, le nuage de point que l'on obtient ne rempli jamais tout l'espace, mais entoure la courbe obtenue par programmation dynamique.

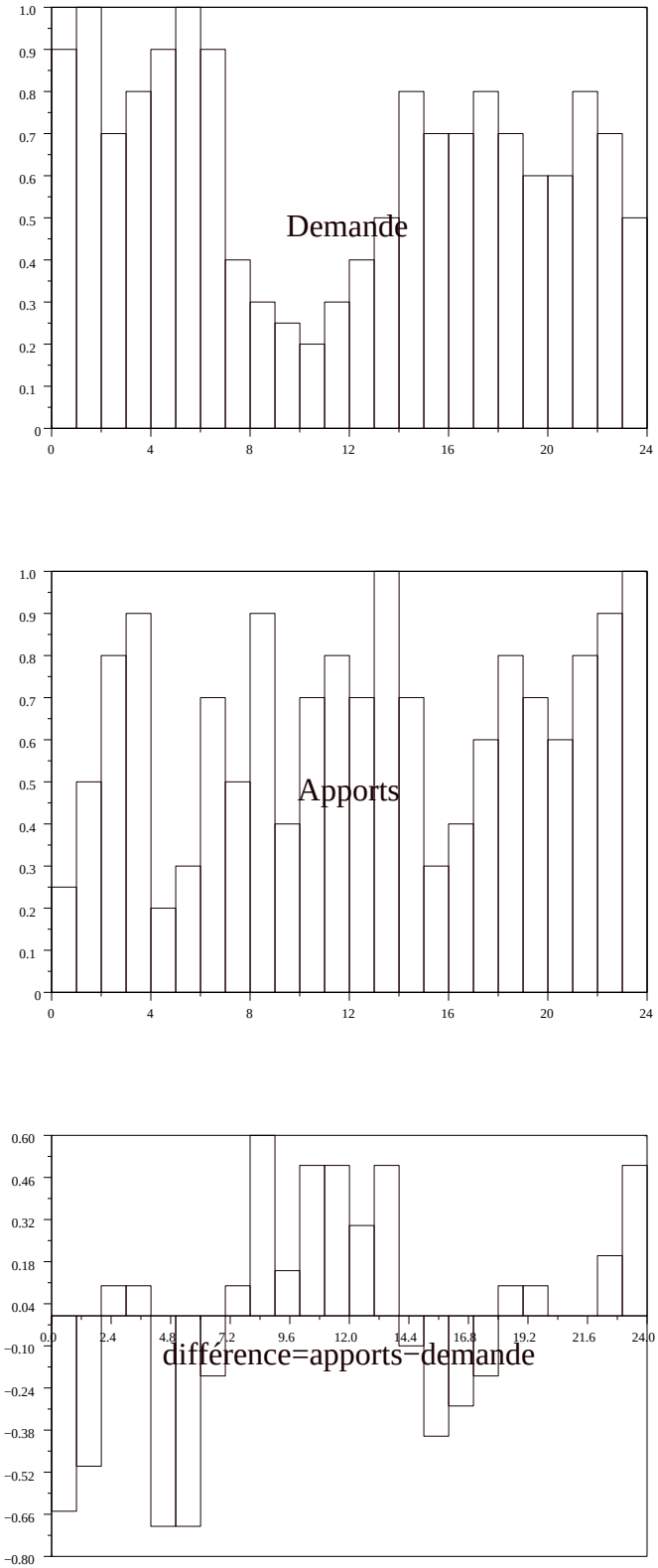


FIG. V.1 – Demande et apports d'eau

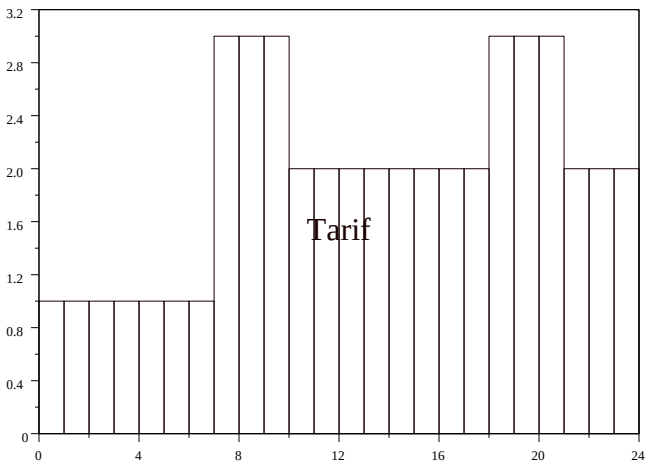


Fig. V.2 – Tarif τ en fonction du temps

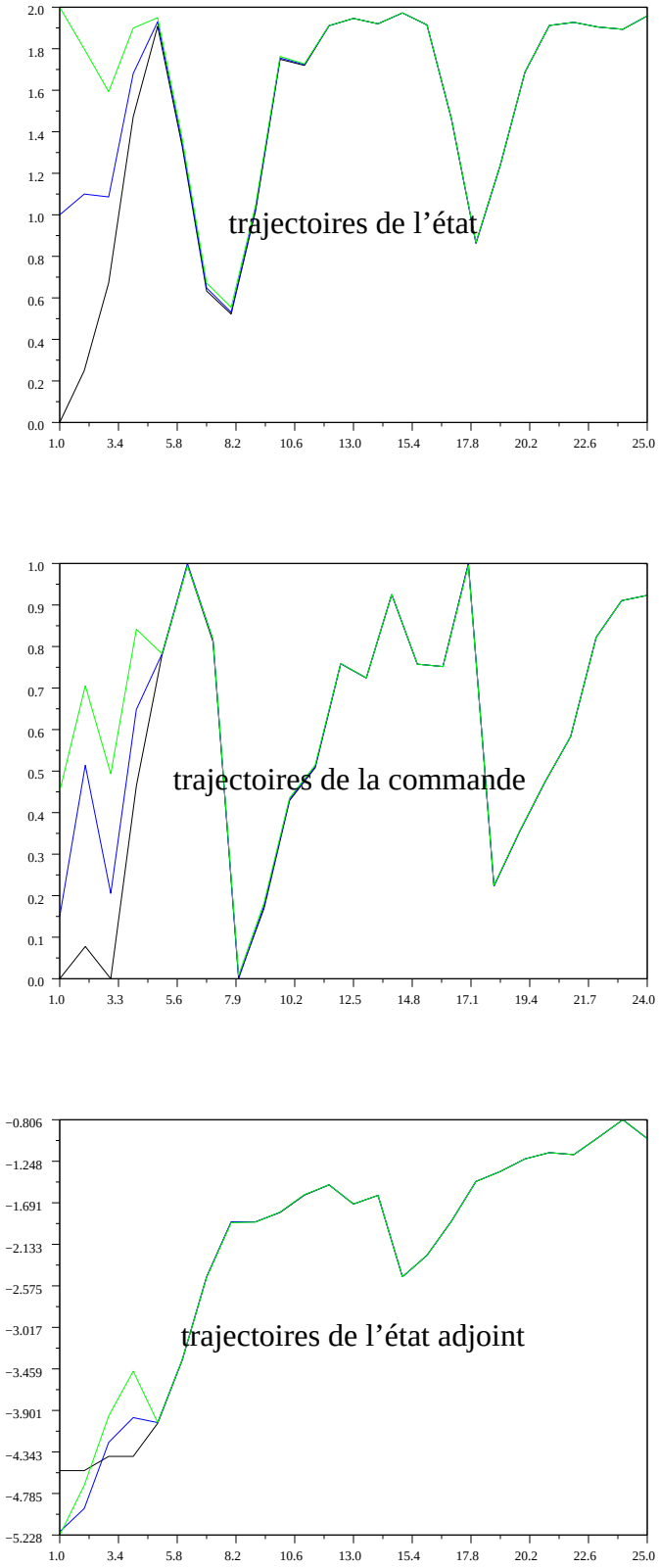


FIG. V.3 – Résultats numériques

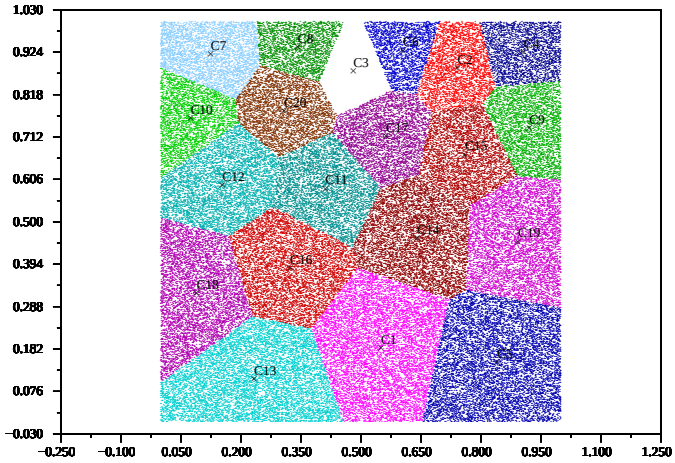


FIG. V.4 – Cellules de Voronoï d'une loi uniforme sur $[0,1] \times [0,1]$

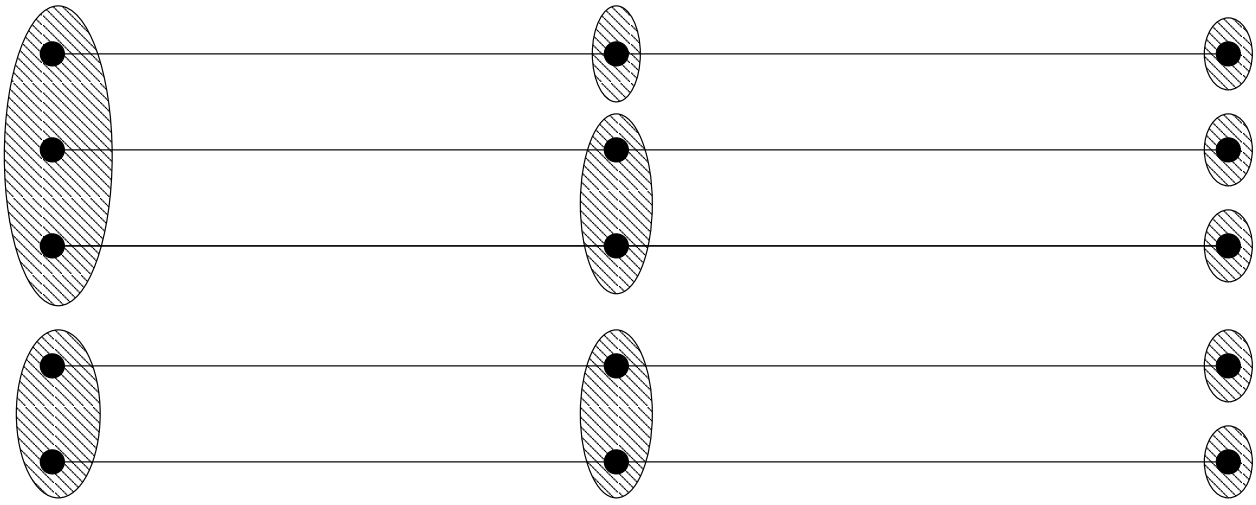


FIG. V.5 – Structure statique discrète

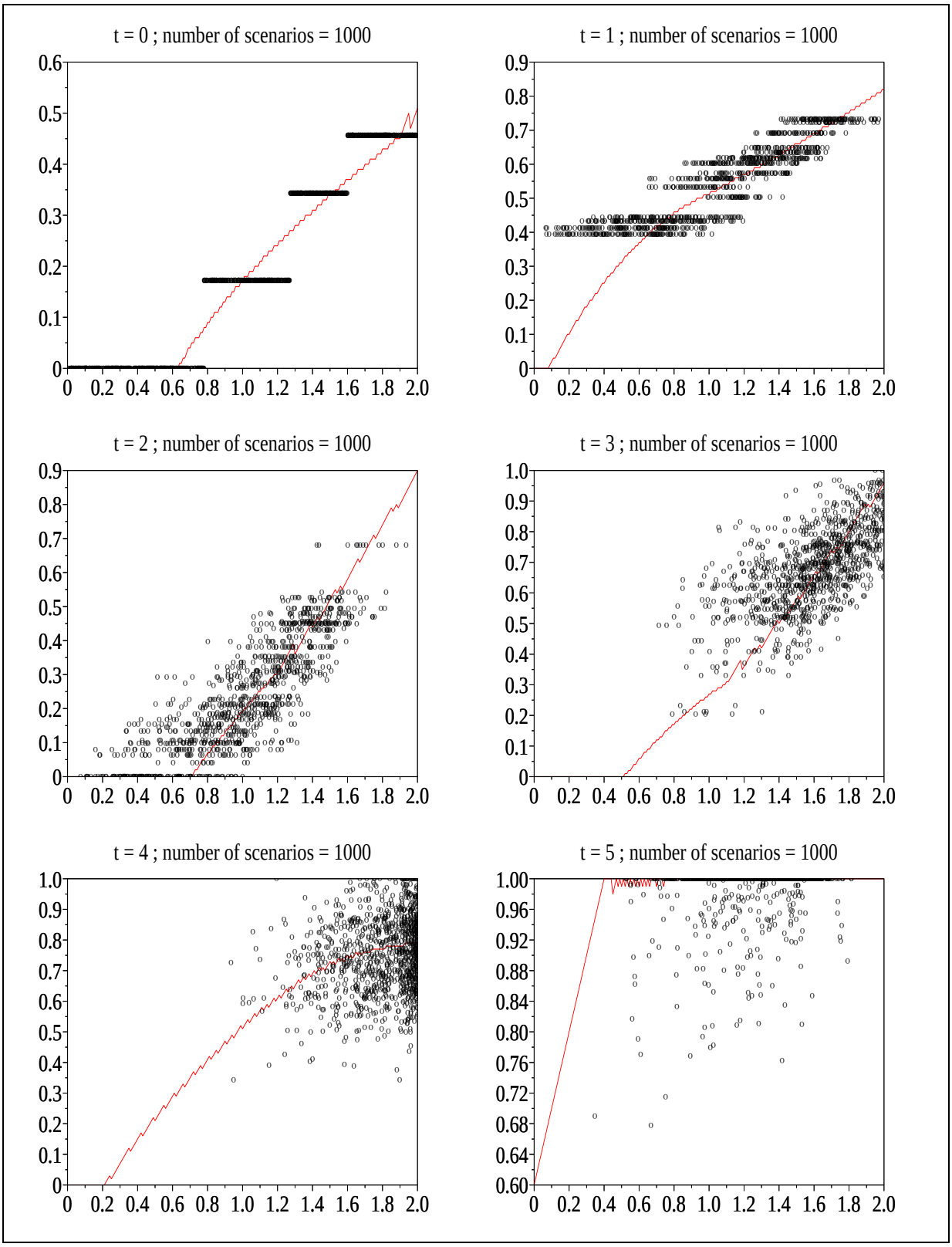


FIG. V.6 – Feeback estimé

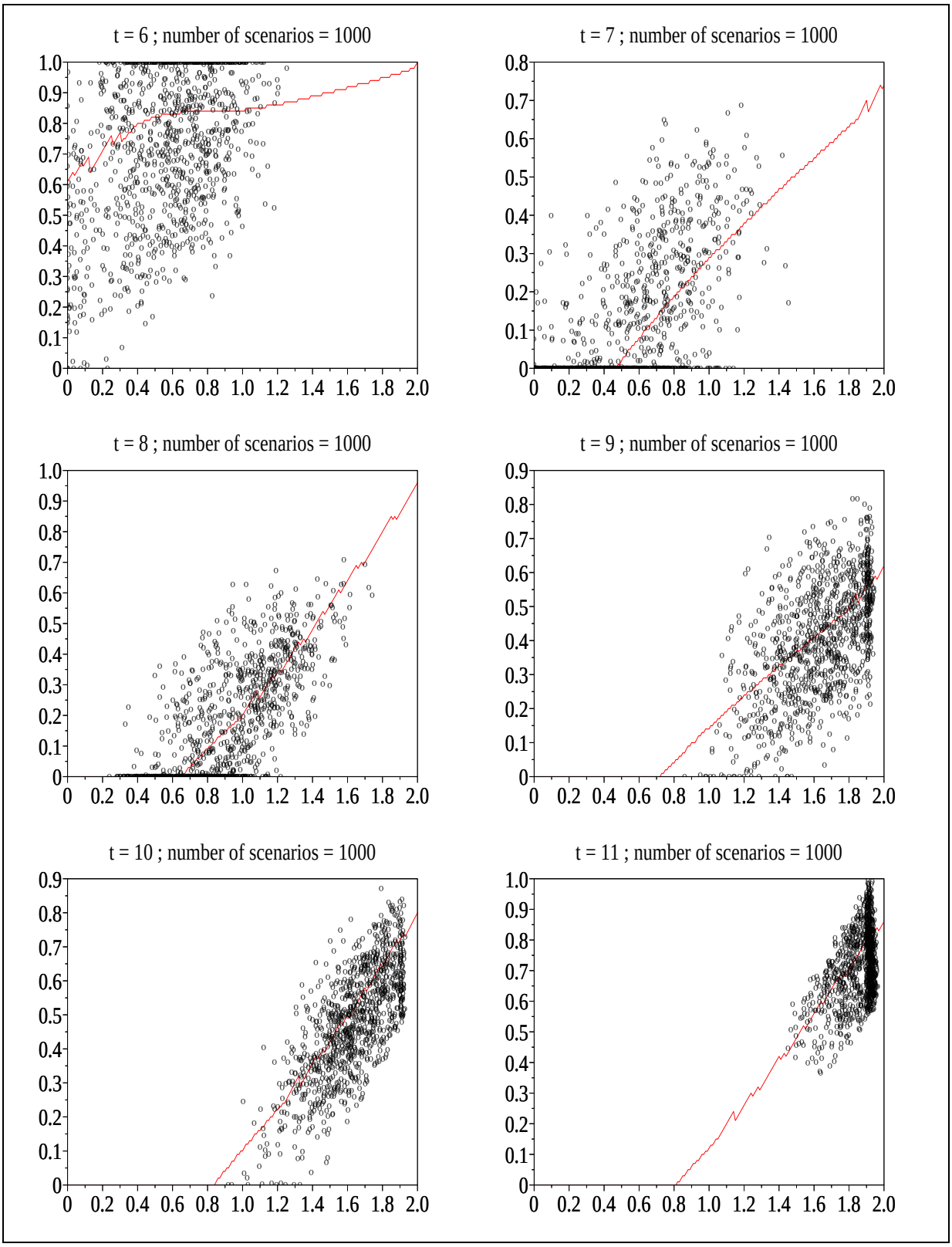


FIG. V.7 – Feeback estimé

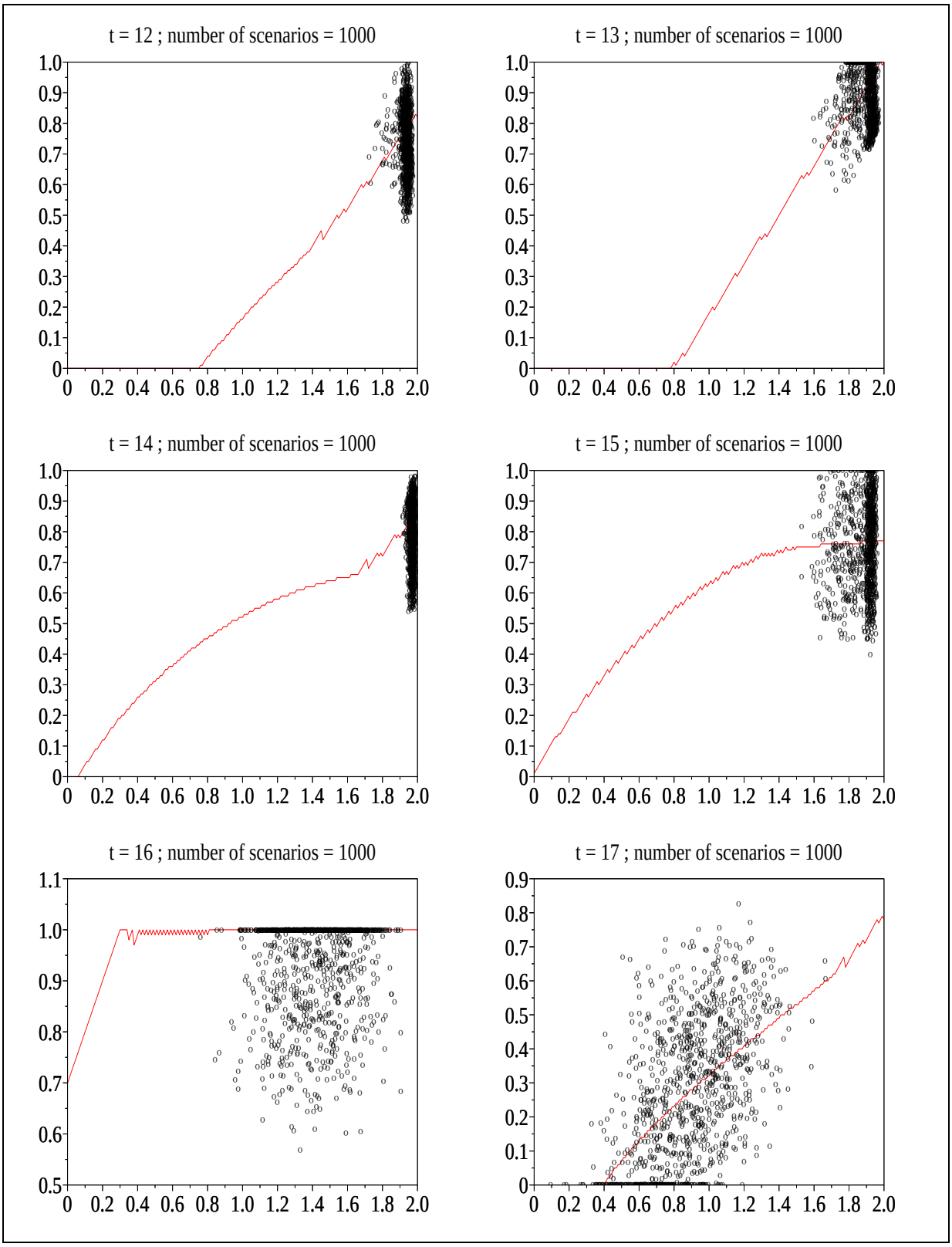


FIG. V.8 – Feeback estimé

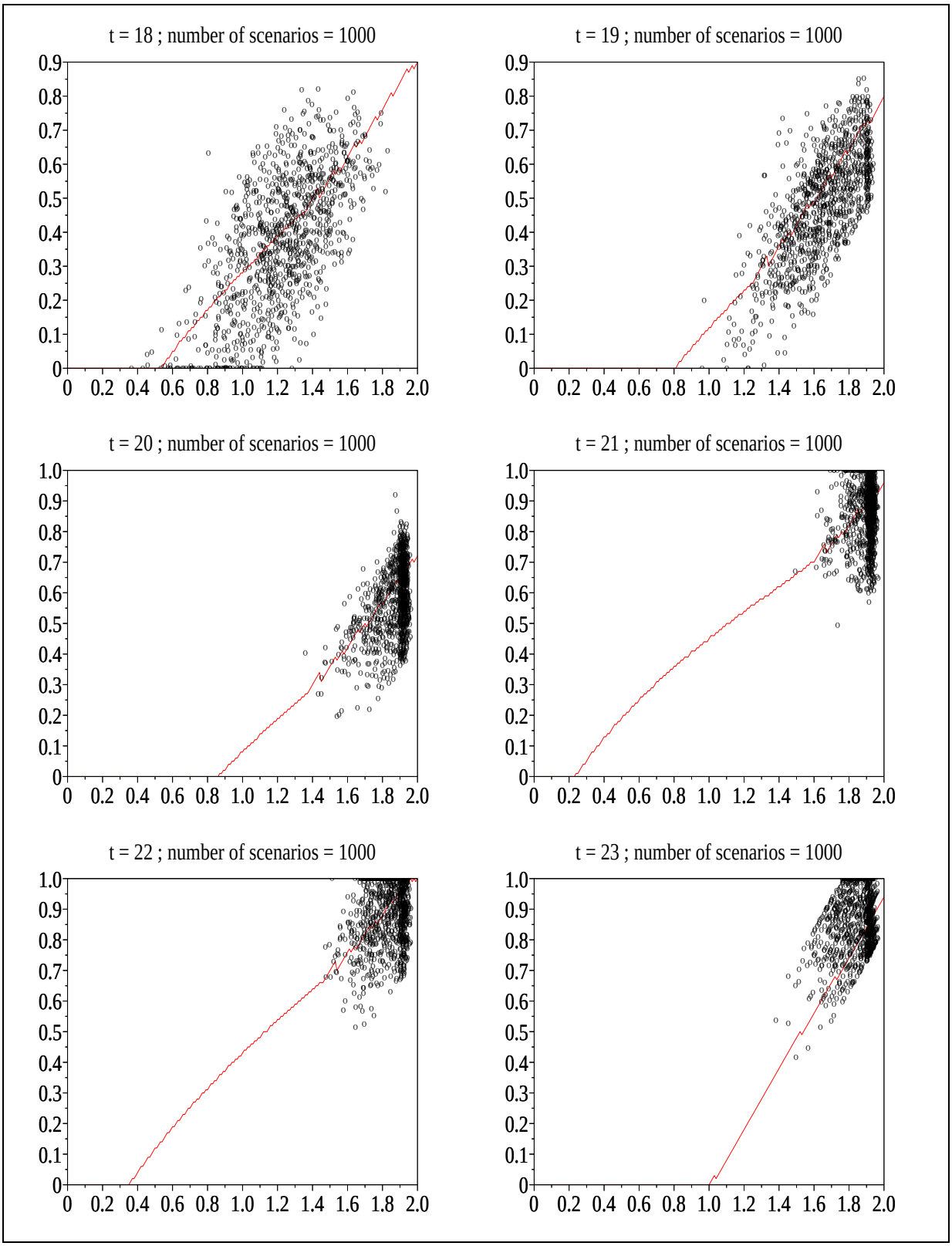


FIG. V.9 – Feeback estimé

Chapitre VI

Conditions de Kuhn et Tucker

VI.1 Introduction

La théorie du Lagrangien en optimisation permet à Arkin [7] d'écrire en 1984 les conditions d'optimalité d'un problème de commande optimale stochastique en temps discret et horizon fini dans le cas différentiable. Arkin a étudié un problème Markovien : cette restriction superflue limite les possibilités d'application de ce résultat. Par la suite, de nombreux auteurs ont abordé ce sujet pour différents problèmes d'optimisation dynamiques stochastiques. Dans le contexte où le temps est continu on peut citer Kushner [63] qui donne un résultat dans le cas du contrôle d'un processus d'Ito, Peng [71] qui traite du cas où les cœfficients du processus d'Ito dépendent du contrôle. Le lecteur intéressé pourra trouver d'autres références dans les articles cités. Pour les problèmes stochastiques dynamiques, les variables de décision sont astreintes à respecter des contraintes de mesurabilité. La théorie de la dualité en présence de contraintes de mesurabilité apparaît pour la première fois dans un article de Rockafellar [87] qui produira par la suite un article plus complet sur la question [89]. Ces résultats seront appliqués à l'étude des conditions d'optimalité d'un problème stochastique de type Bolza [85] ; Hiriart-Urruty [55] étudie un problème multipériodes dans le cadre lipschitzien. Nous avons pour objectif dans ce chapitre d'écrire les conditions d'optimalité d'un problème de commande optimale stochastique en temps discret. La particularité du problème abordé sera la présence d'une contrainte de mesurabilité portant sur les variables de décision. L'originalité de notre démarche réside dans l'interprétation des dites conditions, plus particulièrement de notre double interprétation de l'équation adjointe, c'est-à-dire du multiplicateur associé à la dynamique du système. Nous montrerons qu'il est en effet possible d'écrire deux équations adjointes pour le problème étudié, l'une faisant apparaître un processus adjoint adapté à la filtration engendrée par le processus de bruits, l'autre équation faisant intervenir un processus adjoint qui n'est pas nécessairement adapté à cette dernière filtration. Par ailleurs nous montrerons également que chacun des processus adjoints se calcule de manière récursive. Enfin, bien que nous calculions deux processus adjoints distincts, nous montrerons que les équations qui entrent en jeu lors de l'évaluation de la commande n'en sont pas modifiées pour autant.

VI.2 Problèmes avec une contrainte de projection

Nous allons étudier un cas particulier de problème d'optimisation. Il s'agit de la minimisation d'un critère, sous contrainte que la variable de décision est dans l'image d'un projecteur. Ce type de problème a déjà été étudié dans le cas particulier de contrainte de mesurabilité, typiquement la variable de décision x est assujettie à respecter une contrainte de la forme $\mathbb{E}(x \mid \mathcal{B}) = x$; nous allons étendre certains résultats de Rockafellar, Wets, Higle et Sen [87, 54] en considérant une contrainte de la forme $Ax = x$ où A est un projecteur.

VI.2.1 Définition du problème

Soient H un espace vectoriel normé, H^* son dual et $A : H \to H$ un projecteur ($AA = A$) linéaire supposé de graphe fermé. On note $A^* : H^* \to H^*$ l'adjoint de A, autrement dit l'application telle que pour tout couple $(p, x) \in H^* \times H$, $\quad \langle A^* p, x\rangle = \langle p, Ax\rangle$. Soit S le sous-ensemble de H tel que :

$$S \stackrel{\text{def}}{=} \{x \in H \mid Ax = x\},$$

et on pose :

$$M \stackrel{\text{def}}{=} S^{\perp} ;$$

c'est-à-dire que :

$$M = \left\{p \in H^* \mid \langle p, x\rangle_{H^*,H} = 0, \quad \forall x \in S\right\}.$$

Il est clair que S et M sont des sous-espaces vectoriels de H. Considérons les problèmes suivants :

$$\mathcal{P} : \left\{ \begin{array}{l} \inf J(x), \\ x \in S. \end{array} \right.$$

et :

$$\mathcal{P}' : \left\{ \begin{array}{l} \inf J(x), \\ x = z \text{ et } z \in S. \end{array} \right.$$

Ces problèmes sont clairement équivalents, cependant il est plus pratique de dualiser $\mathcal{P}'$ que $\mathcal{P}$. Le Lagrangien associé au problème $\mathcal{P}'$ est la fonction définie par :

$$L(x, z, \lambda, \mu) \stackrel{\text{def}}{=} J(x) + \langle z - x, \lambda\rangle + \langle (z - Az), \mu\rangle .$$

La fonction duale de $\mathcal{P}'$ est l'application Φ telle que :

$$\Phi(\lambda, \mu) \stackrel{\text{def}}{=} \inf_{x,z} L(x, z, \lambda, \mu).$$

On remarque que :

$$L(x, z, \lambda, \mu) = -(\langle \lambda, x\rangle - J(x)) + \langle \lambda + \mu - A^*\mu, z\rangle \ ;$$

on obtient en isolant l'expression J^* de la transformée de Legendre-Fenchel de J (voir définition A.8) :

$$\inf_x L(x, z, \lambda, \mu) = -J^*(\lambda) + \langle \lambda + \mu - A^*\mu, z \rangle .$$

Donc :

$$\Phi(\lambda, \mu) = \begin{cases} -J^*(\lambda) \text{ si } \lambda = -\mu + A^*\mu \ ; \\ -\infty \text{ sinon.} \end{cases} \tag{VI.1}$$

Commençons par établir quelques lemmes techniques.

Lemme VI.1. *Soit A un opérateur linéaire alors :*

$$\ker(I - A) \subset \ \mathrm{im}A. \tag{VI.2}$$

Preuve : Soit $x \in \ker(I - A)$ alors $Ax = x$, ce qui signifie bien que $x \in \ \mathrm{im}A$, donc que $\ker(I - A) \subset \ \mathrm{im}A$. □

Lemme VI.2. *Si A est un projecteur alors :*

$$\mathrm{im}A = \ker(I - A).$$

Preuve : Soit $y \in \ \mathrm{im}A$ alors il existe $x \in H$ tel que $y = Ax$, donc $Ay = AAx = Ax = y$, ce qui montre que $y \in \ker(I - A)$. On déduit l'égalité grâce à (VI.2). □

Remarque VI.3. Il est facile d'observer que l'expression (VI.2) est équivalente à :

$$\ker A \subset \ \mathrm{im}(I - A) \ ;$$

sous l'hypothèse que A est un projecteur cette inclusion devient une égalité d'après le lemme VI.2.

Lemme VI.4. *Si A est un projecteur, il en est de même de A^*.*

Le problème dual consiste à maximiser la fonction duale Φ, autrement dit d'après (VI.1) il s'agit alors de résoudre le problème suivant :

$$\max -J^*(\lambda) \text{ sous } \lambda \in \ \mathrm{im}(I - A^*).$$

D'après les lemmes (VI.2) et (VI.4) le problème dual est équivalent au problème suivant :

$$\mathcal{D}' : \begin{cases} \sup -J^*(p), \\ p \in \ker A^*. \end{cases}$$

Lemme VI.5. *Si A est fermé et* $\mathrm{dom}(A) = H$*, alors nous avons :*

$$(\ \mathrm{im}A)^\perp = \ker A^*. \tag{VI.3}$$

Preuve : Pour une preuve de se résultat voir [25, corollaire II.17 p. 28]. □

La conséquence du lemme (VI.5) est que $\mathcal{D}'$ est équivalent au problème suivant :

$$\mathcal{D} : \begin{cases} \sup -J^*(p), \\ p \in M. \end{cases}$$

Définition VI.6. Nous noterons v_p et v_d les valeurs optimales respectives des problèmes $\mathcal{P}$ et $\mathcal{D}$, c'est-à-dire :

$$v_p = \inf \{J(x) \mid x \in S\},$$

et :

$$v_d = \sup \{-J^*(p) \mid p \in M\}.$$

VI.2.2 Approche par dualité

Théorème VI.7. *On suppose que J est convexe et continue. Alors :*

1. $v_p \geq v_d$;
2. $\bar{x} \in S$ *est solution de* $\mathcal{P}$ *si et seulement si il existe* $\bar{p} \in M$ *tel que :*

$$v_d = -J^*(\bar{p}) = J(\bar{x}) = v_p.$$

Preuve : Par définition de J^*, quels que soient $(p, x) \in H^* \times H$:

$$J^*(p) + J(x) \geq \langle p, x \rangle_{H^*,H}.$$

Par définition de l'orthogonal d'un ensemble, on obtient alors :

$$\forall x \in S,\ \forall p \in M, \quad J(x) \geq -J^*(p),$$

et par définition de v_p et v_d,

$$v_p \geq v_d. \tag{VI.4}$$

Le problème $\mathcal{P}$ est équivalent à :

$$\inf_{x \in H} [J(x) + \mathcal{X}_S(x)].$$

D'après la proposition A.13 et la proposition A.14, $\bar{x}$ est solution de $\mathcal{P}$ si et seulement si :

$$0 \in \partial J(\bar{x}) + \partial \mathcal{X}_S(\bar{x}). \tag{VI.5}$$

Puisque S est un sous-espace vectoriel de H, alors :

$$\partial \mathcal{X}_S(\bar{x}) = S^{\perp}.$$

Étant donné que l'application J est convexe et continue, alors $\partial J(\bar{x}) \neq \emptyset$, donc l'inéquation variationnelle (VI.5) est équivalente à :

$$\exists \bar{p} \in \partial J(\bar{x}) \text{ tel que } -\bar{p} \in \partial \mathcal{X}_S(\bar{x}) = S^{\perp} \text{ c'est-à-dire } \forall x \in S,\ \langle \bar{p}, x \rangle_{H^*,H} = 0.$$

Or :

$$\bar{p} \in \partial J(\bar{x}) \Leftrightarrow J^*(\bar{p}) + J(\bar{x}) = \langle \bar{p}, \bar{x} \rangle_{H^*,H} ,$$

donc (VI.5) est équivalent à :

$$-J^*(\bar{p}) = J(\bar{x}) \text{ et } \bar{p} \in S^{\perp}.$$

De plus, $\bar{x}$ est solution de $\mathcal{P}$ alors :

$$J(\bar{x}) = v_p, \tag{VI.6}$$

et $\bar{p}$ vérifie les contraintes de $\mathcal{D}$ alors :

$$-J^*(\bar{p}) \leq v_d, \tag{VI.7}$$

donc d'après (VI.4)–(VI.6) et (VI.7) :

$$v_d = v_p.$$

□

Corollaire VI.8. *Sous les hypothèses du théorème VI.7, $\bar{x} \in S$ est solution de $\mathcal{P}$ ssi il existe un $\bar{p} \in M$ tel que :*

$$\bar{x} \in \arg\min_{x \in H}(J(x) - \langle \bar{p}, x \rangle_{H^*,H}). \tag{VI.8}$$

Preuve : D'après le théorème VI.7, $\bar{x} \in S$ est solution de $\mathcal{P}$ ssi il existe $\bar{p} \in M$ tel que :

$$J^*(\bar{p}) + J(\bar{x}) = 0. \tag{VI.9}$$

La relation (VI.9) et le fait que $\langle \bar{p}, \bar{x} \rangle_{H^*,H} = 0$ impliquent que $\bar{p} \in \partial J(\bar{x})$ or :

$$\bar{p} \in \partial J(\bar{x}) \Leftrightarrow \bar{x} \in \arg\min_{x \in H}(J(x) - \langle \bar{p}, x \rangle_{H^*,H}).$$

□

Corollaire VI.9. *On reprend les hypothèses du théorème VI.7 et on suppose de plus que A est fermée, que* $\mathrm{dom}(A) = H$ *et $AA = A$. Alors $\bar{x}$ est solution de $\mathcal{P}$ ssi il existe $p \in H^*$ tel que :*

$$\bar{x} \in \arg\min_{x \in H}(J(x) - \langle p - A^*p, x \rangle_{H^*,H}). \tag{VI.10}$$

Preuve : D'après le corollaire VI.8, nous avons que $\bar{x} \in S$ est solution de $\mathcal{P}$ si et seulement si :

$$\exists \bar{p} \in M, \quad \bar{x} \in \arg\min_{x \in H}(J(x) - \langle \bar{p}, x \rangle_{H^*,H}).$$

D'après (VI.3) $M = \ker A^*$, puisque A est un projecteur alors il en est de même de A^* et donc :

$$\ker A^* = \ \mathrm{im}(I - A^*).$$

Alors :

$$\bar{p} \in M \Leftrightarrow \exists p \in H^*, \quad \bar{p} = p - A^*p \ ;$$

le résultat est alors immédiat. □

VI.2.3 Approche directe des conditions d'optimalité

Nous avons utilisé des arguments de dualité pour déduire des conditions d'optimalité pour le problème $\mathcal{P}$. Nous allons maintenant considérer une approche plus directe des conditions d'optimalité, qui repose sur l'équivalence entre $\mathcal{P}$ et le problème suivant :

$$\mathcal{P}'' \quad \inf_{x\in H} J(Ax).$$

L'équivalence entre $\mathcal{P}$ et $\mathcal{P}''$ est une conséquence directe du lemme (VI.2). Il suffit en effet de remarquer que le problème $\mathcal{P}''$ est exactement le problème suivant :

$$\inf_{y\in H} J(y) \text{ sous } y \in \text{ im}A.$$

La proposition suivante a pour objet de préciser ce que nous entendons par "équivalents".

Proposition VI.10. *Si $\bar{x}$ est solution de $\mathcal{P}$ alors $\bar{x}$ est solution de $\mathcal{P}''$. Réciproquement si $\bar{x}$ est solution de $\mathcal{P}''$, alors $\bar{y} = A\bar{x}$ est solution de $\mathcal{P}$.*

Preuve : Soit $\bar{x}$ solution de $\mathcal{P}$ alors :

$$\forall x \in H \text{ tel que } Ax = x, \quad J(\bar{x}) \leq J(x).$$

Si $\bar{x}$ est solution de $\mathcal{P}$ alors pour tout $x \in H$, puisque A est un projecteur, $y = Ax$ vérifie $Ay = y$, donc $J(\bar{x}) \leq J(y) = J(Ax)$. Ce qui prouve que $\bar{x}$ est solution de $\mathcal{P}''$. Soit maintenant $\bar{x}$ solution de $\mathcal{P}''$, alors pour tout $x \in H$ tel que $Ax = x$ nous avons $J(\bar{x}) \leq J(Ax) = J(x)$. Ce qui prouve que $\bar{x}$ est également solution de $\mathcal{P}$. □

Remarque VI.11. Si nous avions introduit dans le problème la contrainte $x \in K$, il aurait fallu supposer que $AK \subset K$ pour obtenir l'équivalence entre $\mathcal{P}$ et $\mathcal{P}''$.

Nous allons faire référence dans le théorème suivant au sous-differentiel généralisé de Clarke, voir définition A.23 dans l'annexe A.

Théorème VI.12. *Supposons que J soit localement lipschitzienne et que A soit un opérateur linéaire continu. Si $\bar{x}$ est solution de $\mathcal{P}''$ il existe $\bar{p} \in H^*$ tel que $A^*\bar{p} = 0$ et $\bar{p} \in \partial J(\bar{x})$.*

Preuve : L'application $J(A\cdot)$ étant localement lipschitzienne, les conditions d'optimalité sont :

$$0 \in \partial J(A\cdot)(\bar{x}).$$

D'après [28, Théorème 2.3.10], nous avons :

$$\partial J(A\cdot)(\bar{x}) \subset A^*\partial J(A\bar{x}).$$

Puisque $A\bar{x} = \bar{x}$, alors :

$$0 \in A^*\partial J(\bar{x}) \ ;$$

autrement dit, il existe $\bar{p} \in \partial J(\bar{x})$ tel que $A^*\bar{p} = 0$. □

Corollaire VI.13. *Soit A un projecteur continu, si $\bar{x}$ est solution de $\mathcal{P}$ alors il existe $\bar{p} \in H^*$ tel que $A^*\bar{p} = 0$ et $\bar{p} \in \partial J(\bar{x})$.*

Preuve : D'après la proposition VI.10 nous avons que $\bar{x}$ est aussi solution de $\mathcal{P}''$. Par conséquent, $\bar{x}$ vérifie les conditions d'optimalité du problème $\mathcal{P}''$ données par le théorème VI.12. □

VI.2.4 Application à un problème d'optimisation stochastique

Le cadre mathématique du problème primal $\mathcal{P}$ est adapté à la situation des problèmes d'optimisation stochastique. Les variables de décision dans les problèmes d'optimisation stochastique sont toujours assujetties à respecter une contrainte de mesurabilité. Celle-ci peut s'exprimer de différentes façons. Soit $(\Omega, \mathcal{B}_\Omega, \mathbb{P})$ un espace probabilisé, $\{\mathcal{B}_1, \ldots, \mathcal{B}_T\}$ une famille de sous-tribus de $\mathcal{B}_\Omega$, et $X = (X_1, \ldots, X_T)$, puis posons :

$$\Pi_{L^p_{\mathbb{R}^n}(\Omega)}(X) \stackrel{\text{def}}{=} (\mathbb{E}\,(X_1 \mid \mathcal{B}_1), \ldots, \mathbb{E}\,(X_T \mid \mathcal{B}_T)).$$

Nous avons clairement que $\Pi_{L^p_{\mathbb{R}^n}(\Omega)}$ est un opérateur linéaire continu, et que c'est une projection. L'espérance conditionnelle dans la définition de $\Pi_{L^p_{\mathbb{R}^n}(\Omega)}$ existe puisqu'elle est définie sur $L^p_{\mathbb{R}^n}(\Omega)$, $\quad p \geq 1$. La continuité de $\Pi_{L^p_{\mathbb{R}^n}(\Omega)}$ est une conséquence de l'inégalité de Jensen :

$$\left|\left|\left(\Pi_{L^p_{\mathbb{R}^n}(\Omega)}(X)\right)_k\right|\right|^p_{L^p_{\mathbb{R}^n}(\Omega)} = ||\mathbb{E}\,(X_k \mid \mathcal{B}_k)||_{L^p_{\mathbb{R}^n}(\Omega)} \leq ||X_k||_{L^p_{\mathbb{R}^n}(\Omega)}\,.$$

La proposition suivante donne l'expression précise de l'adjoint de $\Pi_{L^p_{\mathbb{R}^n}(\Omega)}$.

Proposition VI.14 (Eisner et Olsen [42]). *Soit $p \in [1, \infty[$ alors $\Pi^*_{L^p_{\mathbb{R}^n}(\Omega)} = \Pi_{L^q_{\mathbb{R}^n}(\Omega)}$ avec $\frac{1}{p} + \frac{1}{q} = 1$.*

Considérons les notations suivantes :
– $H \stackrel{\text{def}}{=} L^1_{\mathbb{R}^m}(\Omega)$, donc naturellement $H^* = L^\infty_{\mathbb{R}^m}(\Omega)$;
– pour tout $x \in H$, $\quad A(x) \stackrel{\text{def}}{=} \mathbb{E}\,(x \mid \mathcal{B})$;
et considérons le problème d'optimisation suivant :

$$\left\{ \begin{array}{l} \min \mathbb{E}[J(x(w), w)] \\ \text{sous les contraintes} \left\{ \begin{array}{l} x = A(x), \\ x \in H. \end{array} \right. \end{array} \right. \tag{VI.11}$$

Il consiste en la minimisation d'un critère en espérance, sous des contraintes de mesurabilité. Nous ferons l'hypothèse que l'application :

$$\begin{array}{rcl} L^1_{\mathbb{R}^m}(\Omega) & \to & \mathbb{R}\ ; \\ x & \mapsto & \mathbb{E}\,[J(x(w), w)]\ ; \end{array}$$

est convexe. Il est connu que A est un projecteur fermé et que son adjoint est l'espérance conditionnelle sur H^*. D'après le corollaire VI.8 il existe $p \in (\ \mathrm{im}A)^\perp$ tel que si $\bar{x}$ est solution du (VI.11) ; alors, $\bar{x}$ est également solution de :

$$\min_{x \in H} \mathbb{E}\left[J(x(w), w) + \langle p(w), x(w)\rangle\right].$$

Puisque $(\ \mathrm{im}A)^\perp = \ker A^*$ alors $\mathbb{E}\left(p \mid \mathcal{B}\right) = 0$. D'après le corollaire VI.9 il existe $p \in L^\infty_{\mathbb{R}^m}(\Omega)$ tel que si $\bar{x}$ est solution du (VI.11) alors $\bar{x}$ est aussi solution de :

$$\min_{x \in H} \mathbb{E}\left[J(x(w), w) + \langle p(w) - \mathbb{E}\left(p \mid \mathcal{B}\right)(w), x(w)\rangle\right].$$

En supposant $J(\cdot, \xi)$ convexe, alors une solution $\bar{x}$ du problème (VI.11) vérifie aussi :

$$\bar{x}(w) \in \arg\min_x (J(x, w) + \langle p(w) - \mathbb{E}\left(p \mid \mathcal{B}\right)(w), x(w)\rangle).$$

VI.3 Conditions de Kuhn et Tucker dans un problème stochastique

L'espace de probabilité associé aux éléments aléatoires du problème est $(\Omega, \mathcal{F}, \mathbb{P})$, où Ω est un sous-ensemble borélien de $(\mathbb{R}^n)^{T+1}$, $\mathcal{F}$ désigne la tribu des boréliens sur Ω et $\mathbb{P}$ est une mesure de probabilité sur $(\Omega, \mathcal{F})$. Un élément ω de Ω est de la forme :

$$(\omega_0, \ldots, \omega_T) \in (\mathbb{R}^n)^{T+1}.$$

VI.3.1 Un problème de commande optimale stochastique

Nous allons considérer un problème dans lequel les décisions doivent être prises en réponse à l'observation progressive d'une certaine variable aléatoire et de manière à minimiser l'espérance d'une fonction coût, tout en respectant des contraintes sur les variables de décision.

Les variables $w = (w_0, \ldots, w_T)$ et $u = (u_0, \ldots, u_T) \in (\mathbb{R}^m)^T$ représentent respectivement des éléments aléatoires et des décisions. Elles permettent d'intégrer une dynamique discrète :

$$\begin{aligned} x_{t+1} &= F(x_t, u_t, w_{t+1}, t) \quad t = 0, \ldots, T-1 \ ; \\ x_0 &= \eta(w_0) \ ; \end{aligned} \tag{VI.12}$$

et déterminent un "coût trajectoriel" noté $J_0(x, u, w)$:

$$J_0(x, u, w) \stackrel{\text{def}}{=} K(x_T) + \sum_{t=0}^{T-1} L(x_t, u_t, w_{t+1}, t).$$

Soit une famille croissante de sous-tribus complètes de $\mathcal{F}$ noté $\mathcal{G}$:

$$\mathcal{G} = \{\mathcal{G}_0, \ldots, \mathcal{G}_T\} \quad \text{avec} \quad \mathcal{G}_0 \subset \mathcal{G}_1 \subset \ldots \subset \mathcal{G}_T.$$

Une stratégie admissible u devra être un élément de $\mathcal{N}_p$:

$$\mathcal{N}_p \stackrel{\text{def}}{=} \left\{ u = (u_0, \ldots, u_{T-1}) \in L^p_{(\mathbb{R}^m)^T}(\Omega) \mid u_t \text{ est } \mathcal{G}_t\text{-mesurable} \quad \forall t = 0, \ldots, T-1 \right\} ;$$

Le problème de commande optimale considéré s'exprime alors de la manière suivante :

$$\inf \left\{ f(u) \stackrel{\text{def}}{=} \mathbb{E}\left[J(u(w), w)\right] \mid \text{ sous la contrainte } u \in \mathcal{N}_p \right\}, \tag{VI.13}$$

avec $J(u, \omega) \stackrel{\text{def}}{=} J_0(x(u, \omega), u, \omega)$, $x(u, \omega)$ étant obtenu en intégrant la dynamique discrète (VI.12) à l'aide de la stratégie u.

Hypothèses (H)

1. l'application $L(\cdot, \cdot, w_{t+1}, t)$ est $C^1(\mathbb{R}^\ell \times \mathbb{R}^m)$ $\quad \mathbb{P} - \text{p.s.}$;
2. l'application $F(\cdot, \cdot, w_{t+1}, t)$ est $C^1(\mathbb{R}^\ell \times \mathbb{R}^m)$ $\quad \mathbb{P} - \text{p.s.}$;
3. l'application $K(\cdot)$ est $C^1(\mathbb{R}^\ell)$;

Hypothèse (A) Il existe $c > 0$ tel que :

$$\left|\left| J'(u, w) \right|\right| \leq c(1 + ||u||^{p-1}).$$

L'hypothèse (A) permet [28, Théorème 2.7.5] simplement d'avoir l'inclusion suivante :

$$\partial f(u) \subset \{p \mid p(w) \in \partial J(\cdot, w)(u(w))\}.$$

Elle signifie également que f est lipschitzienne sur tout sous ensemble borné de $L^p_{(\mathbb{R}^m)^T}(\Omega)$.

VI.3.2 Conditions d'optimalité

Il est classique pour les problèmes de commande optimale d'exprimer le critère à minimiser en éliminant les variables d'état. Le calcul du gradient du nouveau critère passe alors par un procédé qui consiste à introduire une nouvelle variable qui simplifiera le calcul du gradient.

Proposition VI.15. *Si (H) est vérifiée, alors l'application* $J(\cdot, w) \in C^1((\mathbb{R}^m)^T)$ *et :*

$$\forall u \in (\mathbb{R}^m)^T, \quad (J'_{u_t})^\top(u, w) = (L'_u)^\top(x_t, u_t, w_{t+1}, t) + (F'_u)^\top(x_t, u_t, w_{t+1}, t)\lambda_{t+1} \ ;$$

où :

$$\left\{ \begin{array}{rcl} \lambda_T & = & (K'_x)^\top(x_T) \ ; \\ \lambda_t & = & (L'_x)^\top(x_t, u_t, w_{t+1}, t) + (F'_x)^\top(x_t, u_t, w_{t+1}, t)\lambda_{t+1}. \end{array} \right.$$

Preuve : Commençons par écrire $J(u, w)$ au moyen d'un état adjoint qui est ici un ensemble de $T+1$ vecteurs $\lambda = (\lambda_t)_{t=0,\ldots,T}$:

$$J(u, w) = K(x_T) + \sum_{t=0}^{T-1} L(x_t, u_t, w_{t+1}, t) + \lambda_{t+1}^\top (F(x_t, u_t, w_{t+1}, t) - x_{t+1}) \ ;$$

Cette identité a lieu quelque soit λ pourvu que x vérifie la dynamique (VI.12). On choisira λ ultérieurement de manière à éliminer la dérivée de la fonction $u \mapsto x(u)$. Nous pouvons à présent calculer la dérivée de $J(\cdot, w)$ en u :

$$\begin{aligned} J'(u, w).v = K'(x_T)x'_T(u).v + \sum_{t=0}^{T-1} & L'_x(x_t, u_t, w_{t+1}, t)x'_t(u).v + L'_u(x_t, u_t, w_{t+1}, t).v_t \\ & + \lambda_{t+1}^\top (F'_x(x_t, u_t, w_{t+1}, t)x'_t(u).v - x'_{t+1}(u).v + F'_u(x_t, u_t, w_{t+1}, t).v_t). \end{aligned}$$

En procédant à un regroupement des termes :

$$\begin{aligned} J'(u, w).v = K'(x_T)x'_T(u).v + \sum_{t=0}^{T-1} & (L'_x(x_t, u_t, w_{t+1}, t)x'_t(u) + \lambda_{t+1}^\top F'_x(x_t, u_t, w_{t+1}, t)x'_t(u) \\ & - \lambda_{t+1}^\top x'_{t+1}(u)).v + \sum_{t=0}^{T-1} (L'_u(x_t, u_t, w_{t+1}, t).v_t + \lambda_{t+1}^\top F'_u(x_t, u_t, w_{t+1}, t).v_t). \end{aligned}$$

En choisissant λ_t comme indiqué nous obtenons alors la formule suivante :

$$\begin{aligned} J'(u, w).v = \lambda_T^\top x'_T(u).v. + \sum_{t=0}^{T-1} & (\lambda_t^\top x'_t(u) - \lambda_{t+1}^\top x'_{t+1}(u)).v + \\ & \sum_{t=0}^{T-1} (L'_u(x_t, u_t, w_{t+1}, t).v_t + \lambda_{t+1}^\top F'_u(x_t, u_t, w_{t+1}, t).v_t). \end{aligned}$$

Cette dernière expression se simplifie en remarquant que :

$$\lambda_T^\top x'_T(u).v + \sum_{t=0}^{T-1} (\lambda_t^\top x'_t(u) - \lambda_{t+1}^\top x'_{t+1}(u)).v = \lambda_0^\top x'_0(u).v = 0 \ ;$$

donc :

$$J'(u, w).v = \sum_{t=0}^{T-1} L'_u(x_t, u_t, w_{t+1}, t).v_t + \lambda_{t+1}^\top F'_u(x_t, u_t, w_{t+1}, t).v_t.$$

□

Théorème VI.16. *Supposons que (A) soit vérifiée. Si $\bar{u}$ est une solution du problème (VI.13), alors il existe une application mesurable $\bar{p} = (\bar{p}_0, \ldots, \bar{p}_{T-1}) \in L^q_{(\mathbb{R}^m)^T}(\Omega)$ telle que :*

$$\bar{p}(w) \in \partial J(\cdot, w)(\bar{u}(w)) \text{ et } \mathbb{E}\left(\bar{p}_t \mid \mathcal{G}_t\right) = 0, \quad \forall t = 0, \ldots, T-1.$$

Preuve : L'opérateur A défini sur $L^p_{(\mathbb{R}^m)^T}(\Omega)$ par :

$$A(x_0, \ldots, x_{T-1}) = \left(\mathbb{E}\left(x_0 \mid \mathcal{G}_0\right), \ldots, \mathbb{E}\left(x_{T-1} \mid \mathcal{G}_{T-1}\right)\right)$$

est un projecteur, c'est-à-dire que $AA = A$. L'ensemble $\mathcal{N}_p$ est simplement l'image de ce projecteur :

$$\mathcal{N}_p = \left\{ u \in L^p_{(\mathbb{R}^m)^T}(\Omega) \mid Au = u \right\} ;$$

le problème d'optimisation (VI.13) s'écrit plus simplement :

$$\inf_u f(Au) \text{ sous } Au = u.$$

D'après le théorème VI.12 :

$$\exists \bar{p} = (\bar{p}_0, \ldots, \bar{p}_{T-1}) \in L^q_{(\mathbb{R}^m)^T}(\Omega), \quad A^* \bar{p} = 0$$

et :

$$\bar{p} \in \partial f(\bar{u}).$$

D'après l'hypothèse (A) nous avons par [28, Théorème 2.7.5] :

$$\partial f(\bar{u}) \subset \left\{ p \in L^q_{(\mathbb{R}^m)^T}(\Omega) \mid p(w) \in \partial J(\cdot, w)(\bar{u}(w)) \right\}.$$

En résumé, il existe $\bar{p}_t \in L^q_{\mathbb{R}^m}(\Omega)$ tel que $\mathbb{E}\left(\bar{p}_t \mid \mathcal{G}_t\right) = 0$ et :

$$\bar{p} = (\bar{p}_0, \ldots, \bar{p}_{T-1}) \in \partial J(\cdot, w)(\bar{u}(w)).$$

□

VI.3.3 Conditions nécessaires d'optimalité, première forme

Corollaire VI.17. *Supposons que (A) et (H) soient satisfaites. Si $\bar{u}$ est solution du problème $\mathcal{P}$ alors $\bar{u}$ vérifie :*

$$\mathbb{E}\left((L'_u)^\top(x_t, \bar{u}_t, w_{t+1}, t) + (F'_u)^\top(x_t, \bar{u}_t, w_{t+1}, t)\lambda_{t+1} \mid \mathcal{G}_t \right) = 0, \tag{VI.14}$$

où :

$$\begin{aligned} \lambda_T &= (K'_x)^\top(x_T) ; \\ \lambda_t &= (L'_x)^\top(x_t, \bar{u}_t, w_{t+1}, t) + (F'_x)^\top(x_t, \bar{u}_t, w_{t+1}, t)\lambda_{t+1}. \end{aligned} \tag{VI.15}$$

Preuve : Ce résultat est une conséquence directe de la proposition VI.15 et du théorème VI.16. □

Remarque VI.18. L'intérêt du corollaire VI.17 est de donner une caractérisation de la solution sans supposer l'indépendance des variables aléatoires $(w_t)_{t=0,\dots,T}$: ceci est un avantage sur la technique de programmation dynamique.

VI.3.4 Conditions nécessaires d'optimalité, deuxième forme

La filtration naturelle $(\mathcal{F}_t)_{t=0,\dots,T-1}$ associée au processus $(w_t)_{t=0,\dots,T-1}$ est définie de la manière suivante :

$$\mathcal{F}_t \stackrel{\text{def}}{=} \sigma(w_0, \dots, w_t), \quad \forall t = 0, \dots, T-1.$$

Corollaire VI.19. *Supposons que (A) et (H) soient satisfaites et que $\mathcal{G}_t \subset \mathcal{F}_{t+1}$. Si $\bar{u}$ est solution du problème (VI.13), alors :*

$$\mathbb{E}\left((L_u^{'})^\top(x_t, \bar{u}_t, w_{t+1}, t) + (F_u^{'})^\top(x_t, \bar{u}_t, w_{t+1}, t)\Lambda_{t+1} \mid \mathcal{G}_t\right) = 0, \tag{VI.16}$$

où :

$$\Lambda_T = (K_x^{'})^\top(x_T), \tag{VI.17}$$

$$\Lambda_t = \mathbb{E}\left((L_x^{'})^\top(x_t, \bar{u}_t, w_{t+1}, t) + (F_x^{'})^\top(x_t, \bar{u}_t, w_{t+1}, t)\Lambda_{t+1} \mid \mathcal{F}_t\right). \tag{VI.18}$$

Preuve : On applique le corollaire VI.17 et l'on pose $\Lambda_t = \mathbb{E}(\lambda_t \mid \mathcal{F}_t)$. Alors en prenant l'espérance conditionnelle dans l'équation (VI.15) on obtient :

$$\Lambda_t = \mathbb{E}\left((L_x^{'})^\top(x_t, u_t, w_{t+1}, t) + (F_x^{'})^\top(x_t, u_t, w_{t+1}, t)\lambda_{t+1} \mid \mathcal{F}_t\right).$$

Étant donné que $\mathcal{F}_t \subset \mathcal{F}_{t+1}$, que $x_t, \bar{u}_t$ et $F(x, u, w_{t+1}, t)$ sont $\mathcal{F}_{t+1}$ mesurables nous déduisons du lemme B.5 :

$$\Lambda_t = \mathbb{E}\left((L_x^{'})^\top(x_t, \bar{u}_t, w_{t+1}, t) + (F_x^{'})^\top(x_t, \bar{u}_t, w_{t+1}, t)\Lambda_{t+1} \mid \mathcal{F}_t\right).$$

Puisque $\mathcal{G}_t \subset \mathcal{F}_{t+1}$ nous déduisons à nouveau du lemme B.5 et de la formule VI.14 :

$$\mathbb{E}\left((L_u^{'})^\top(x_t, \bar{u}_t, w_{t+1}, t) + (F_u^{'})^\top(x_t, \bar{u}_t, w_{t+1}, t)\Lambda_{t+1} \mid \mathcal{G}_t\right) = 0.$$

□

VI.3.5 Calcul de l'état adjoint

Cette section présente une autre manière de calculer l'état adjoint, intervenant dans le principe du minimum. Nous donnons un théorème qui est une version temps-discret de l'équation adjointe introduite par Haussman [53].

Définition VI.20. On définit pour tout $t \in \mathbb{N}$ la dynamique discrète suivante :

$$\left\{ \begin{array}{lcl} A_{t,t} & = & I \ ; \\ A_{t,s+1} & = & A_{t,s}(F_x^{'})^{\top}(x_s, u_s, w_{s+1}, s) \text{ si } s \geq t \quad s \in \mathbb{N}. \end{array} \right.$$

Lemme VI.21. *Les matrices $A_{t,s}$ vérifient les relations suivantes :*

$$A_{t,s} = (F_x^{'})^{\top}(x_t, u_t, w_{s+1}, t)A_{t+1,s}.$$

Preuve : Pour alléger les notations posons $\gamma(t) \stackrel{\text{def}}{=} (F_x^{'})^{\top}(x_t, u_t, w_{t+1}, t)$, alors :

$$\begin{array}{rcl} A_{t,s+1} & = & A_{t,s}\gamma(s) \ ; \\ & = & A_{t,s-1}\gamma(s-1)\gamma(s) \ ; \\ & = & A_{t,t}\gamma(t)\ldots\gamma(s-1)\gamma(s) \ ; \\ & = & \gamma(t)\ldots\gamma(s-1)\gamma(s) \ ; \end{array}$$

d'où :

$$A_{t,s} = \gamma(t)\underbrace{\gamma(t+1)\ldots\gamma(s-1)}_{A_{t+1,s}} \ ;$$

et :

$$A_{t,s} = \gamma(t)A_{t+1,s}.$$

□

Théorème VI.22. *L'état adjoint Λ du théorème VI.17 se calcule selon la formule suivante :*

$$\Lambda_T = (K_x^{'})^{\top}(x_T),$$

$$\Lambda_t = \mathbb{E}\left(\sum_{s=t}^{T-1} A_{t,s}(L_x^{'})^{\top}(x_s, u_s, w_{s+1}, s) + A_{t,T}(K_x^{'})^{\top}(x_T) \mid \mathcal{F}_t \right).$$

Preuve : Il est clair que :

$$\Lambda_T = \mathbb{E}\left(A_{T,T}(K_x^{'})^{\top}(x_T) \mid \mathcal{F}_T \right).$$

On suppose que la propriété est vraie pour Λ_{t+1}, alors :

$$\begin{aligned}
\Lambda_t &= \mathbb{E}\left((L_x^{'})^{\top}(x_t,u_t,w_{t+1},t)+(F_x^{'})^{\top}(x_t,u_t,w_{t+1},t)\Lambda_{t+1}\mid\mathcal{F}_t\right),\\
\Lambda_t &= \mathbb{E}\Bigg((F_x^{'})^{\top}(x_t,u_t,w_{t+1},t)\mathbb{E}\left(\sum_{s=t+1}^{T-1}A_{t+1,s}(L_x^{'})^{\top}(x_s,u_s,w_{s+1},s)+A_{t+1,T}(K_x^{'})^{\top}(x_T)\mid\mathcal{F}_{t+1}\right)\\
&\quad +(L_x^{'})^{\top}(x_t,u_t,w_{t+1},t)\Big|\mathcal{F}_t\Bigg),\\
\Lambda_t &= \mathbb{E}\Bigg((L_x^{'})^{\top}(x_t,u_t,w_{t+1},t)+\sum_{s=t+1}^{T-1}(F_x^{'})^{\top}(x_t,u_t,w_{t+1},t)A_{t+1,s}(L_x^{'})^{\top}(x_s,u_s,w_{s+1},s)\\
&\quad +(F_x^{'})^{\top}(x_t,u_t,w_{t+1},t)A_{t+1,T}(K_x^{'})^{\top}(x_T)\Big|\mathcal{F}_t\Bigg),\\
\Lambda_t &= \mathbb{E}\left((L_x^{'})^{\top}(x_t,u_t,w_{t+1},t)+\sum_{s=t+1}^{T-1}A_{t,s}(L_x^{'})^{\top}(x_s,u_s,w_{s+1},s)+A_{t,T}(K_x^{'})^{\top}(x_T)\mid\mathcal{F}_t\right),\\
\Lambda_t &= \mathbb{E}\left(\sum_{s=t}^{T-1}A_{t,s}(L_x^{'})^{\top}(x_s,u_s,w_{s+1},s)+A_{t,T}(K_x^{'})^{\top}(x_T)\mid\mathcal{F}_t\right).
\end{aligned}$$

□

Proposition VI.23. *L'état adjoint λ_t du théorème VI.19 se calcule de la manière suivante :*

$$\lambda_t=\sum_{s=t}^{T-1}A_{t,s}(L_x^{'})^{\top}(x_s,u_s,w_{s+1},s)+A_{t,T}(K_x^{'})^{\top}(x_T).$$

Preuve : La démonstration est identique à la précédente. □

Théorème VI.24. *Supposons que les fonctions $L_u^{'},F_u^{'},L_x^{'},F_x^{'},K_x^{'}$ sont bornées, que $\mathcal{G}_t=\mathcal{F}_t$ et que $w_t(\omega)=\omega_t$. Alors le système d'équations (VI.16), (VI.17) et (VI.18), peut être vérifié par des applications $\bar{u}_t$ et Λ_t mesurables par rapport à la variable aléatoire x_t.*

Preuve : La preuve se fait par une récurrence rétrograde sur l'indice t. Commençons par poser :

$$\Phi_t(x,u)=\mathbb{E}\left[(L_u^{'})^{\top}(x,u,w_{t+1},t)+(F_u^{'})^{\top}(x,u,w_{t+1},t)\Lambda_{t+1}\right],$$

et :

$$M_t(x)\overset{\text{def}}{=}\{u\mid\Phi_t(x,u)=0\}.$$

Il est clair que Λ_T est x_T mesurable, grâce à la dynamique nous pouvons exprimer x_T en fonction de x_{T-1} et de u_{T-1}. M_T étant une multi application mesurable à valeurs non vides et fermées, il existe une application γ_T telle que :

$$\gamma_T(x)\in M_T(x),$$

par conséquent $\bar{u}_T = \gamma_T(x_T)$ convient. Suppose qu'il existe une application mesurable Λ^x_{t+1} telle que $\Lambda_{t+1} = \Lambda^x_{t+1}(x_{t+1})$ alors :

$$\Lambda_t = \mathbb{E}\left((L'_x)^\top(x_t, \gamma_t(x_t), w_{t+1}, t) + (F'_x)^\top(x_t, \gamma_t(x_t), w_{t+1}, t)\Lambda^x_{t+1}(F(x_t, \gamma_t(x_t), w_{t+1}, t)) \mid \mathcal{F}_t\right).$$

Puisque w_{t+1} est indépendante de $\mathcal{F}_t$, alors Λ_t peut s'exprimer en fonction de x_t et u_t. À nouveau grâce à un argument de sélection mesurable nous avons l'existence d'une variable aléatoire mesurable γ_t vérifiant :

$$\gamma_t(x) \in M_t(x) \ ;$$

et $\bar{u}_t = \gamma_t(x_t)$ convient. □

Chapitre VII

Conclusion et ouvertures

VII.1 Résultats

Nous avons montré que la discrétisation d'un problème d'optimisation stochastique avec une structure d'information statique produit nécessairement deux sources d'erreurs. L'une provient de la discrétisation de la structure d'information, l'autre de celle de l'espérance du critère. Nous avons donné un procédé permettant de contrôler ces deux erreurs et nous avons montré que ce procédé permet asymptotiquement d'approcher la solution du problème d'origine. Par ailleurs le procédé que nous avons développé au chapitre IV possède un intérêt pratique, au moins en ce qui concerne la discrétisation de la structure d'information. Nous avons en effet que la solution du problème discret est aussi une solution admissible (du point de vue de la contrainte de mesurabilité) pour le problème d'origine.

Le rôle de la méthode numérique que nous avons utilisée pour discrétiser la structure d'information peut s'expliquer clairement avec l'aide du résultat du paragraphe IV.2. La quantification de la structure d'information nous permet simplement de supprimer les contraintes de mesurabilité en transformant le critère du problème à optimiser, le nouveau critère étant l'espérance conditionnelle du critère d'origine par rapport à une tribu discrète. C'est également ce qui se produit lorsque l'on utilise les techniques dites de chroniques de scénarios, la contrainte de mesurabilité n'est plus supportée par les variables de décisions, mais directement par le critère.

VII.2 Ouvertures

Nous avons choisi une approche directe de la discrétisation, une autre approche aurait pu consister à étudier la discrétisation des conditions d'optimalité d'un problème d'optimisation. Même si les conditions d'optimalité sont théoriquement équivalentes, elles n'en demeurent pas moins intéressantes pour la raison suivante : l'équivalence des conditions d'optimalité repose sur une propriété intrinsèque de l'espérance conditionnelle, celle d'être un projecteur. L'idée sous-jacente de l'approche par la quantification est simplement de remplacer une espérance conditionnelle, par une espérance conditionnelle discrète, le nouvel

opérateur discret étant lui même un projecteur (puisque c'est une espérance conditionnelle) devrait conduire numériquement aux mêmes résultats que l'on passe par l'une ou l'autre des expressions des conditions d'optimalité. Mis à part ce cas, les opérateurs discrètes pouvant remplacer une espérance conditionnelle ne sont pas des projections en général. Autrement dit la discrétisation de conditions d'optimalités qui sont théoriquement équivalentes devraient a priori conduire à des résultats numériques différents.

Annexe A

Rappels d'optimisation

A.1 Optimisation convexe

X désigne par défaut un espace vectoriel normé dont le dual topologique est noté X^*.

Définition A.1. On dit que $f : X \to \mathbb{R} \cup \{+\infty\}$ est convexe si et seulement si (ssi) :

$$\forall x_1, x_2 \in \mathrm{dom}(f), \quad \forall \lambda \in [0,1], \quad f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2).$$

On dit que f est strictement convexe si :

$$\forall x_1, x_2 \in \mathrm{dom}(f),\ x_1 \neq x_2, \quad \forall \lambda \in]0,1[, \quad f(\lambda x_1 + (1-\lambda)x_2) < \lambda f(x_1) + (1-\lambda) f(x_2).$$

Théorème A.2. *Soit X un espace vectoriel normé, et soit $f : X \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe majorée au voisinage de $a \in \mathrm{dom}(f)$, alors f est continue en a.*

Théorème A.3. *Si K est un ouvert convexe non vide d'un espace de Banach X et si f est une fonction convexe semi-continue inférieurement sur K, alors f est continue sur K.*

Proposition A.4. *Soit f une fonction convexe, $a \in \mathrm{dom}(f)$ et $u \in X$, alors la fonction φ :*

$$\varphi(t) = \frac{f(a+tu) - f(a)}{t}$$

est croissante sur $]0, +\infty[$.

Définition A.5. On pose pour f convexe :

$$f'(a,u) = \inf_{t>0} \frac{f(a+tu) - f(a)}{t}.$$

On dit que $f'(a,\cdot)$ est la dérivée directionnelle de f en a. S'il existe $\ell \in X^*$ telle que $f'(a,\cdot) = \ell(\cdot)$ on dit que f est Gâteaux différentiable en a.

Remarque A.6. Si f est Gâteaux différentiable alors, $f'(a, \lambda u) = \lambda f'(a, u)$ pour tout $u \in X$ et $\lambda > 0$, on dit que la dérivée directionnelle $f'(a, \cdot)$ est positivement homogène de degré 1. On remarque qu'alors, $f'(a, \cdot)$ est convexe.

Théorème A.7. *Soit $f : X \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe. Alors les propriétés suivantes sont équivalentes :*

1. $f(u) = \inf_{v \in X} f(v)$,
2. $f'(u, v - u) \geq 0$ *pour tout* $v \in \mathrm{dom}(f)$.

Définition A.8. Soit $f : X \to \mathbb{R} \cup \{+\infty\}$ une fonction. La transformée de Legendre Fenchel ou conjuguée de f est la fonction $f^* : X^* \to \mathbb{R} \cup \{+\infty\}$ définie par :

$$f^*(x^*) \stackrel{\text{def}}{=} \sup_{x \in X}(\langle x^*, x\rangle - f(x)), \quad \forall x^* \in X^*.$$

Définition A.9. Soient $f : X \to \mathbb{R} \cup \{+\infty\}$ et $x_0 \in \mathrm{dom}(f)$. Le sous- différentiel $\partial f(x_0)$ de f en x_0 est l'ensemble (éventuellement vide) des $x^* \in X^*$ tels que :

$$f(x) \geq f(x_0) + \langle x^*, x - x_0\rangle \quad \forall x \in X$$

On dit que x^* est un sous-gradient de f en x_0.

Définition A.10. Soit $A \subset X$ un ensemble et $a \in A$. Le cône normal de A en $a \in A$ est l'ensemble :

$$N_A(a) \stackrel{\text{def}}{=} \{x^* \in X^* \mid \langle x^*, x - a\rangle \leq 0, \quad \forall x \in A\}.$$

Proposition A.11. *Soient $f : X \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe et $x_0 \in \mathrm{dom}(f)$. Alors les propriétés suivantes sont équivalentes :*

1. $x^* \in \partial f(x_0)$,
2. $f(x_0) + f^*(x^*) \leq \langle x^*, x_0\rangle$,
3. $f(x_0) + f^*(x^*) = \langle x^*, x_0\rangle$,
4. $\langle x^*, \cdot - x_0\rangle + f(x_0)$ *est une minorante affine de* f.

Proposition A.12. *Soit $f : X \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe continue en $x_0 \in \mathrm{dom}(f)$. Alors $\partial f(x_0) \neq \emptyset$ et $\partial f(x_0)$ est borné et $\sigma(X^*, X)$[1] compact.*

Proposition A.13. *Soient f et $g : X \to \mathbb{R} \cup \{+\infty\}$ des fonctions convexes telles qu'il existe $x_0 \in \mathrm{dom}(g)$ avec f continue en x_0. Alors pour tout $x \in X$ on a :*

$$\partial(f + g)(x) = \partial f(x) + \partial g(x).$$

Proposition A.14. *Soit $f : X \to \mathbb{R} \cup \{+\infty\}$, $f \not\equiv +\infty$. Alors $\bar{x}$ est un minimum global si et seulement si $0 \in \partial f(x_0)$.*

[1] faiblement compact

A.2 Optimisation non convexe

Les détails des preuves des résultats d'optimisation énoncés dans cette section se trouvent dans [83, 84].

Théorème A.15 (Rademacher). *Soit Ω un ouvert de $\mathbb{R}^n$ et $f : \Omega \to \mathbb{R}$ une fonction localement lipschitzienne sur Ω. Alors f est différentiable presque partout sur Ω.*

Définition A.16. Soit A un sous ensemble de E, le polaire négatif de A est l'ensemble noté A° tel que :

$$A^\circ = \{x \in E' \mid \langle x, y\rangle \leq 0 \ \forall y \in A\}$$

Lemme A.17 (Farkas). *Soient $\mathcal{E}$ et $\mathcal{I}$ des ensembles finis ou vides, et $(a_i)_{i\in\mathcal{E}\cup\mathcal{I}}$ une famille d'éléments de $\mathbb{R}^p$. Posons :*

$$A = \{v \mid \langle a_i, v\rangle = 0 \ \forall i \in \mathcal{E}, \quad \langle a_i, v\rangle \leq 0 \ \forall i \in \mathcal{I}\},$$

$$B = \left\{ \sum_{i\in\mathcal{E}\cup\mathcal{I}} \lambda_i a_i \mid \lambda_i \in \mathbb{R} \ \forall i \in \mathcal{E}, \quad \lambda_i \geq 0 \ \forall i \in \mathcal{I} \right\}.$$

Alors $A^\circ = B$ et $A = B^\circ$.

Définition A.18. La dérivée directionnelle d'une application $f : \Omega \to \mathbb{R}$ en x_0 dans la direction v est l'élément noté $Df(x_0)(v)$ défini par :

$$Df(x_0)(v) = \lim_{t\to 0^+} \frac{f(x_0 + tv) - f(x_0)}{t}$$

Définition A.19. Soient Ω un ouvert d'un espace vectoriel normé E, $f : \Omega \to \mathbb{R}$ une application localement lipschitzienne, $x \in \Omega$ alors :

$$D_c f(x)(v) \stackrel{\text{def}}{=} \limsup_{\substack{y\to x \\ t\to 0^+}} \frac{f(y + tv) - f(y)}{t}$$

c'est la dérivée directionnelle généralisée au sens de Clarke.

Définition A.20. Soit Ω un ouvert d'un espace de Banach E. Une application $f : \Omega \to \mathbb{R}$ est strictement différentiable en $x_0 \in \Omega$ si il existe $\varphi \in E'$ tel que :

$$f(z) - f(y) = \varphi(z - y) + ||z - y||\, \epsilon(z, y) \quad \epsilon(z, y) \to 0 \text{ quand } z \to x_0, \ y \to x_0$$

Remarque A.21. f strictement différentiable implique que f est différentiable.

Proposition A.22. *Soit $f : \Omega \to \mathbb{R}$, $x_0 \in \Omega$ et f différentiable en tout point de la boule $B(x_0, r) \subset \Omega$ de centre x_0 de rayon $r > 0$. Si l'application $x \mapsto f'(x)$ définie sur $B(x_0, r)$ est continue en x_0 alors f est strictement différentiable en x_0.*

Définition A.23. On définit le gradient généralisé au sens de Clarke $\partial f(x)$ de f au point x par :

$$\partial f(x) \stackrel{\text{def}}{=} \{\varphi \in E' \mid \varphi(v) \leq D_c f(x)(v) \quad \forall v \in E\}$$

Proposition A.24. *Si f est strictement différentiable en $x \in \Omega$ alors :*

$$\partial f(x) = \{f'(x)\}$$

Définition A.25. On dira que f lipschitzienne au voisinage de x est régulière si :
- $Df(x)(v)$ existe pour tout $v \in E$;
- $Df(x)(v) = D_c f(x)(v)$ pour tout $v \in E$.

Exemple A.26. Les fonctions convexes continues sont régulières, de même pour les fonctions strictement différentiables en x.

Proposition A.27. *Soit $f : \Omega \to \mathbb{R}$ localement lipschitzienne et $x \in \Omega$ alors :*
- *$\partial f(x) \neq \emptyset$,*
- *$\partial f(x)$ est convexe et faiblement étoile compact dans E',*
- *$D_c f(x)(v) = \max_{\varphi \in \partial f(x)} \varphi(v)$.*

Proposition A.28. *Soient f et g deux fonctions localement lipschitziennes alors :*

$$\forall x \in \Omega \quad \partial(f+g)(x) \subset \partial f(x) + \partial g(x)$$

Proposition A.29. *Soient f et g deux fonctions régulières en x, alors $f+g$ est régulière en x et :*

$$\partial(f+g)(x) = \partial f(x) + \partial g(x)$$

Définition A.30. Soit K une partie non vide de E ; on notera d_K l'application :

$$x \in E \mapsto d_K(x) \stackrel{\text{def}}{=} \inf\{||x-y|| \mid y \in K\}$$

Proposition A.31. *L'application d_K est 1 lipschitzienne sur E.*

Définition A.32. Soit $x \in K$, on dira que $v \in E$ est un vecteur tangent à K en x si $D_c d_K(x)(v) = 0$. On notera :

$$T_K(x) \stackrel{\text{def}}{=} \{v \in E \mid D_c d_K(x)(v) = 0\},$$

le cône tangent à K en x.

Définition A.33. Soit $x \in K$, on définit le cône normal au sens de Clarke $N_K(x)$ à K en x comme le polaire négatif de $T_K(x)$.

$$N_K(x) \stackrel{\text{def}}{=} \{\varphi \in E' \mid \varphi(v) \leq 0, \quad \forall v \in T_K(x)\}$$

$N_K(x)$ est un cône convexe de sommet $0_{E'}$ et w^* fermé.

Proposition A.34. *Soit $x \in K$ alors :*

$$N_K(x) = \overline{\bigcup_{\lambda \geq 0} \lambda \partial d_K(x)}^{w^*}$$

Proposition A.35. *Soit f une application k lipschitzienne sur un ouvert $S \subset E$. Considérons $x \in K \subset S$ tel que f réalise son minimum sur K au point x. Alors pour tout $\widehat{k} \geq k$ la fonction :*

$$g(y) \stackrel{\text{def}}{=} f(y) + \widehat{k} d_K(y)$$

réalise en x son minimum sur S. De plus si $\widehat{k} > k$ et si K est fermé alors tout autre point qui minimise g sur S est dans K.

Corollaire A.36. *Soit f une application lipschitzienne sur un voisinage de $x \in K$ tel que f atteigne son minimum sur K en x, alors $0_{E'} \in \partial f(x) + N_K(x)$.*

On utilisera la notation suivante :

$$(x', \alpha') \downarrow_f x \Leftrightarrow (x', \alpha') \to (x, f(x)) \text{ (en restant) dans } epi f$$

Définition A.37. Une application f est dite directionnellement lipschitzienne en x si il existe au moins un élément $y \in E$ tel que :

$$\limsup_{\substack{(x',\alpha') \downarrow_f x \\ y' \to y,\ t \downarrow 0}} \frac{f(x' + ty') - \alpha'}{t} < \infty$$

Théorème A.38. *Les assertions suivantes impliquent que f est directionnellement lipschitzienne en un point x (où $f(x)$ est fini).*

1. *f est lipschitzienne dans un voisinage de x.*
2. *f est convexe et majorée supérieurement dans le voisinage d'un point (pas nécessairement le point x).*
3. *$E = \mathbb{R}^n$ f est s.c.i dans un voisinage de x, et $\partial f(x)$ est non vide et ne contient pas entièrement de droite.*

Théorème A.39. *Soit Ω un ouvert de E, $f : \Omega \to \mathbb{R}$ une application localement lipschitzienne et $K \stackrel{\text{def}}{=} \{x \in \Omega \mid f(x) \leq 0\}$. Soit $\bar{x}$ tel que $f(\bar{x}) = 0$ et vérifiant $0_{E'} \notin \partial f(\bar{x})$, alors K est épi-lipschitzien en x et :*

$$\{y \in E \mid D_c f(\bar{x})(y) \leq 0\} \subset T_K(\bar{x})$$

$$N_K(\bar{x}) \subset \{\lambda z \mid \lambda \geq 0,\ z \in \partial f(x)\},$$

si de plus f est régulière au point $\bar{x}$ alors :

$$B_K(\bar{x}) = T_K(\bar{x}) = \{v \in E \mid D_c f(\bar{x})(v) \leq 0\}$$

Corollaire A.40. *Soit* $K = \{x' \mid f_i(x') \leq 0, \quad i = 1, \ldots, m\}$. *Pour* $x \in K$ *soit* $I(x) = \{i \mid f_i(x) = 0\}$. *On suppose que les fonctions* f_i *sont lipschitziennes dans un voisinage de* x *et que :*

$$0 \notin co\,\{\partial f_i(x) \mid i \in I(x)\}$$

alors K *est épi-lipschitzien en* x *et :*

$$T_K(x) \supset \{y \mid \langle y, z_i\rangle \leq 0 \ \forall z_i \in \partial f_i(x),\ i \in I(x)\}$$

$$\mathrm{int}T_K(x) \supset \{y \mid \langle y, z_i\rangle < 0 \ \forall z_i \in \partial f_i(x), i \in I(x)\}$$

$$N_K(x) \subset \left\{ \sum_{i \in I(x)} \lambda_i z_i \mid \lambda_i \geq 0,\ z_i \in \partial f_i(x) \right\}$$

Théorème A.41. *Soient* C_1 *et* C_2 *deux sous-ensembles de* E, *et soit* $x \in C_1 \cap C_2$. *Supposons que :*

$$T_{C_1}(x) \cap \mathrm{int}T_{C_2}(x) \neq \emptyset,$$

et que C_2 *est épi-lipschitzien en* x. *Alors :*

$$T_{C_1 \cap C_2}(x) \supset T_{C_1}(x) \cap T_{C_2}(x),$$

$$N_{C_1 \cap C_2}(x) \subset N_{C_1}(x) + N_{C_2}(x).$$

Théorème A.42. *Soit* $S = \bigcap_{j=1}^{m} S_j$ *où* $S_j = \{x \in \mathbb{R}^p \mid h_j(x) = 0\}$. *Si les fonctions* h_j *sont différentiables en* $x_0 \in S$ *alors :*

$$T_S(x_0) \subset \bigcap_{j=1}^{m} \{d \in \mathbb{R}^p \mid \langle \nabla h_j(x_0), d\rangle = 0\}. \tag{A.1}$$

Si les fonctions $h_j : \Omega \to \mathbb{R}$ *sont continûment différentiables en* x_0 *et que :*

$$\nabla h_1(x_0), \ldots, \nabla h_m(x_0)$$

sont linéairement indépendants alors on a l'égalité dans la formule (A.1) et :

$$N_S(x_0) = \left\{ \sum_{j=1}^{m} \lambda_j \nabla h_j(x_0) \mid \lambda_j \in \mathbb{R},\ \forall j = 1, \ldots, m \right\}$$

A.2.1 Application à un problème d'optimisation non convexe

Considérons le problème suivant :

$$\begin{cases} \min f(x) \\ x \in \mathbb{R}^p \\ x \in K = \{x \in \Omega \mid g_i(x) \leq 0,\ i = 1, \ldots, m\} \\ x \in L = \{x \in \Omega \mid h_j(x) = 0,\ j = 1, \ldots, n\} \end{cases}$$

Théorème A.43. *Soit $x \in K \cap L$ une solution du problème précédent, alors sous les hypothèses suivantes :*

1. *l'application $f : \Omega \to \mathbb{R}$ est localement lipschitzienne ;*
2. *les applications $(g_i)_{i=1,\ldots,m}$ et $(h_j)_{j=1,\ldots,n}$ sont localement lipschitziennes et continûment différentiables ;*
3. *$0 \notin co\{\nabla g_i(x) \mid i \in I(x)\}$ avec $I(x) = \{i \mid g_i(x) = 0\}$;*
4. *$(\nabla h_1(x), \ldots, \nabla h_n(x))$ sont linéairement indépendants ;*
5. *il existe $d \in \mathbb{R}^p$ tel que :*

$$\langle \nabla g_i(x), d\rangle < 0,\ i \in I(x);$$
$$\langle \nabla h_j(x), d\rangle = 0,\ j = 1, \ldots, n;$$

il existe $(\lambda_i)_{1\leq i\leq m}$ et $(\mu_j)_{1\leq j\leq n}$ tels que :

$$\forall i = 1, \ldots, m\ \lambda_i \geq 0,\ \lambda_i g_i(x) = 0 \quad \forall j = 1, \ldots, n\ \mu_j \in \mathbb{R}.$$

$$0_{E'} \in \partial f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) + \sum_{j=1}^{n} \mu_j \nabla h_j(x).$$

A.3 Fonctions marginales

Dans toute cette section les ensembles E et F désignent par défaut des espaces métriques.

Définition A.44. Soient E et F deux espaces topologiques. Une multiapplication $C : E \rightrightarrows F$ est dite *semi-continue supérieurement* en $x \in E$, si pour tout voisinage $\mathcal{U}$ de $C(x)$, il existe un voisinage $\mathcal{V}(x)$ de x tel que :

$$\forall x' \in \mathcal{V}(x),\ C(x') \subset \mathcal{U}.$$

Elle est dite *semi-continue supérieurement* si elle est semi-continue supérieurement en tout point de E.

Définition A.45. Une multiapplication $C : E \rightrightarrows F$ est dite *semi-continue inférieurement* en $x \in \mathrm{dom}(C)$[2] si pour tout $y \in C(x)$ et pour toute suite d'éléments $(x_n)_{n\in\mathbb{N}}$ convergeant vers x, il existe une suite d'éléments $y_n \in C(x_n)$ convergeant vers y.

[2] $\mathrm{dom}(C) \stackrel{\text{def}}{=} \{x \in E \mid C(x) \neq \emptyset\}$

Définition A.46. Soit C une multiapplication $C : E \rightrightarrows F$ et f une fonction numérique $f : \text{Graph(C)}^3 \to \mathbb{R}$. La fonction $g : E \to \mathbb{R} \cup \{+\infty\}$ définie par :

$$g(x) \stackrel{\text{def}}{=} \inf_{y \in C(x)} f(x, y), \tag{A.2}$$

est dite fonction marginale de (C, f).

Théorème A.47. *Soit deux espaces métriques E et F, une multiapplication $C : E \rightrightarrows F$ ainsi qu'une fonction $f : \text{Graph(C)} \to \mathbb{R}$.*

- *Si f est semi-continue supérieurement et C est semi-continue inférieurement, alors la fonction marginale (A.2) de (C, f) est semi-continue supérieurement ;*
- *si f est semi-continue inférieurement et C est semi-continue supérieurement et si les valeurs de C sont compactes, la fonction marginale (A.2) de (C, f) est semi-continue inférieurement.*

Preuve : On peut trouver une démonstration de ce théorème dans [17, p. 284]. □

[3] $\text{Graph(C)} \stackrel{\text{def}}{=} \{(x, y) \in E \times F \mid y \in C(x)\}$

Annexe B

Rappels de probabilité

B.1 Résultats sur l'espérance conditionnelle

Définition B.1. Pour Y une v.a. intégrable sur $(\Omega, \mathcal{F}, \mathbb{P})$ et $\mathcal{O}$ une sous-tribu de $\mathcal{F}$, on appelle espérance conditionnelle par rapport à $\mathcal{O}$, la v.a. $\mathcal{O}$ mesurable et intégrable unique pour l'égalité p.s. notée $\mathbb{E}(Y \mid \mathcal{O})$, telle que pour tout $A \in \mathcal{O}$:

$$\mathbb{E}[Y\mathbb{I}_A] = \mathbb{E}[\mathbb{E}(Y \mid \mathcal{O})\,\mathbb{I}_A].$$

Définition B.2. Une famille $X_1, \ldots, X_n$ de n v.a. définies sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans un espace mesurable $(E_1, \mathcal{E}_1)$. sont indépendantes lorsque :

$$\mathbb{P}_{(X_1,\ldots,X_n)} = \mathbb{P}_{X_1} \otimes \ldots \otimes \mathbb{P}_{X_n}.$$

Définition B.3. On dit que n v.a $X_1, \ldots, X_n$ sont conditionnellement indépendantes par rapport à une tribu $\mathcal{C}$ si pour toute suite $f_1, \ldots, f_n$ de v.a bornées définies sur $\mathbb{R}$, on a :

$$\mathbb{E}(f_1(X_1)\ldots f_n(X_n) \mid \mathcal{C}) = \mathbb{E}(f_1(X_1) \mid \mathcal{C})\ldots\mathbb{E}(f_n(X_n) \mid \mathcal{C}).$$

Proposition B.4 (Inégalité de Jensen). *Soit C un convexe de $\mathbb{R}^n$ et Φ une fonction convexe de C dans $\mathbb{R}$. Soit X un vecteur aléatoire défini sur $(\Omega, \mathcal{F})$, à valeurs dans C. Pour X et $\Phi(X)$ intégrables :*

1. *$\mathbb{E}(X \mid \mathcal{O})$ est à valeurs dans C ;*
2. *$\mathbb{E}(\Phi(X) \mid \mathcal{O}) \geq \Phi(\mathbb{E}(X \mid \mathcal{O}))$;*

si Φ est strictement convexe, on a égalité sur $\Gamma \in \mathcal{O}$ si $\mathbb{I}_\Gamma(X)$ est p.s. $\mathcal{O}$ mesurable. La relation $\mathbb{E}[\Phi(X)] = \mathbb{E}[\Phi(\mathbb{E}(X \mid \mathcal{O}))]$ signifie donc que X est p.s. $\mathcal{O}$ mesurable.

Lemme B.5. *Soit $z \in L^1_{\mathbb{R}^n}(\Omega)$, $\mathcal{A}$ et $\mathcal{B}$ deux sous-tribus de $\mathcal{F}$ telles que $\mathcal{A} \subset \mathcal{B}$ alors :*

$$\mathbb{E}(z \mid \mathcal{A}) = \mathbb{E}(\mathbb{E}(z \mid \mathcal{B}) \mid \mathcal{A}).$$

Définition B.6. Une famille $\{X_i\}_{i\in I}$ de v.a. réelles intégrables est dite *équi-intégrable* si on a :

$$\lim_{a\to\infty} \sup_I \int_{|X_i|>a} |X_i| = 0.$$

Proposition B.7. *Une famille $\{X_i\}_{i\in I}$ de v.a. réelles intégrables est équi-intégrable si et seulement si elle vérifie les deux conditions :*

- *(équi-continuité) pour tout $\varepsilon > 0$, il existe un $\eta_\varepsilon > 0$ tel que :*

$$\sup_I \int_A |X_i| \leq \varepsilon \text{ dès que } \mathbb{P}(A) \leq \eta_\varepsilon \quad (A \in \mathcal{F}) \ ;$$

-

$$\sup_I \int_A |X_i| < \infty.$$

Définition B.8 (Convergence en probabilité). Une suite $\{X_n\}_{n\in\mathbb{N}}$ *converge en probabilité* vers X si pour tout $\varepsilon > 0$:

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

Définition B.9 (Convergence en loi). Une suite de mesures finies $\{\mu^n\}_{n\in\mathbb{N}}$ *converge en loi* vers μ si pour toute fonction bornée telle que l'ensemble :

$$A = \{\xi \in \Xi \mid f \text{ est discontinue en } \xi\},$$

soit de mesure nulle, on a :

$$\lim_{n\to\infty} \int f(\xi)\mu^n(d\xi) = \int f(\xi)\mu(d\xi).$$

Proposition B.10. *Soit $p \in [1, \infty[$. Pour toute suite de v.a. $(X_n)_{n\in\mathbb{N}}$ dans $L^p_{\mathbb{R}}(\Omega)$ et pour toute v.a. X les deux conditions suivantes sont équivalentes :*

- *$\{|X_i|^p \mid i \in \mathbb{N}\}$ est équi-intégrable et $(X_n)_{n\in\mathbb{N}}$ converge vers X en probabilité ;*
- *$X \in L^p_{\mathbb{R}}(\Omega)$ et $\lim_{n\to\infty} ||X_n - X||_{L^p_{\mathbb{R}}(\Omega)} = 0$.*

Théorème B.11. *Soit $(\mathcal{B}_n)_{n\in\mathbb{N}}$ une suite croissante de sous-tribus de $\mathcal{F}$ et soit E un espace de Banach séparable. Alors pour toute variable aléatoire intégrable $f : \Omega \mapsto E$, la convergence :*

$$\mathbb{E}(f \mid \mathcal{B}_n) \to \mathbb{E}(f \mid \mathcal{B}^\infty)$$

a lieu dans E pour presque tout ω, si $\mathcal{B}^\infty$ désigne la tribu engendrée par les $\mathcal{B}_n$ ($n \in \mathbb{N}$). En outre, la suite $(\mathbb{E}(f \mid \mathcal{B}_n), \quad n \in \mathbb{N})$ converge vers $\mathbb{E}(f \mid \mathcal{B}^\infty)$ dans $L^1_E(\Omega)$.

Lemme B.12 (Neveu [67] p. 18). *Si A est une partie non vide de Ω et si F est une classe de parties de Ω engendrant la tribu $\mathcal{F}$, alors $F \cap A$ engendre $\mathcal{F} \cap A$.*

Remarque B.13. Si $\mathcal{B}$ est une sous-tribu de $\mathcal{F}$ qui n'est pas complète dans $(\Omega, \mathcal{F}, \mathbb{P})$, désignons par $\tilde{\mathcal{B}}$ la tribu complète qu'elle engendre. Notons alors qu'une fonction réelle $\mathcal{F}$-mesurable est $\tilde{\mathcal{B}}$ mesurable si et seulement si il existe une fonction $\mathcal{B}$-mesurable à laquelle elle est égale $\mathbb{P}-$p.s..

Définition B.14. La loi empirique F_n de n observations $X_1, \ldots, X_n$ à valeurs dans $(E, \mathcal{E})$ est définie pour chaque $\omega \in \Omega$ par :

$$F_n(\omega, \cdot) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{X_i}(\omega)$$

Théorème B.15 (Glivenko-Cantelli).
- *Pour une suite (X_n) de vecteurs aléatoires à valeurs dans $\mathbb{R}^k$, indépendantes, de loi F, les répartitions empiriques $(F_n(\omega, \cdot))$ convergent étroitement vers F pour presque tout ω.*
- *Pour $k = 1$, la convergence des fonctions de répartition est p.s. uniforme, pour presque tout ω :*

$$\lim_{n\to\infty} \sup_x |F_n(\omega, x) - F(x)| = 0.$$

Définition B.16. Soit $\Gamma : \Xi \mapsto \mathbb{R}^n$ une multi application. Nous dirons que Γ est mesurable si pour tout sous ensemble fermé $F \subset \mathbb{R}^n$:

$$\Gamma^{-1}(F) = \{\xi \in \Xi \mid \Gamma(\xi) \cap F \neq \emptyset\} \in \mathcal{F}.$$

Proposition B.17 (Inégalité de Hölder). *Pour $p \in [0, \infty]$ et $\frac{1}{p} + \frac{1}{q} = 1$: Si $X \in L^p(\mathcal{F})$ et $Y \in L^q(\mathcal{F})$ alors $XY \in L^1(\mathcal{F})$ et*

$$||XY||_{L^1_{\mathbb{R}}(\Omega)} \leq ||X||_{L^p_{\mathbb{R}}(\Omega)} ||Y||_{L^q_{\mathbb{R}}(\Omega)}.$$

Proposition B.18 (Attouch [15, Corollaire 1.16]). *Soient (X, d) un espace métrique et $\{x_{\nu,\mu} \mid \nu \in \mathbb{N}, \quad \mu \in \mathbb{N}\}$ une suite d'éléments de X telle que :*

$$\lim_{\nu\to\infty} d(x_{\nu,\mu}, x_\mu) = 0 \ ;$$

et :

$$\lim_{\mu\to\infty} d(x_\mu, x) = 0.$$

Alors, il existe une application croissante $\nu \mapsto \mu(\nu)$ telle que :

$$\lim_{\nu\to\infty} d(x_{\nu,\mu(\nu)}, x) = 0.$$

Proposition B.19. *On suppose que IV.17 est vérifiée. Soient $\mathcal{B}$ une tribu, K une multi-application $\mathcal{B}$-mesurable à valeurs convexes fermées et non vides, X une variable aléatoire telle que $||X||^2_{\mathbb{R}^n}$ soit intégrable. Si $X(\xi) \in K(\xi)$ $\mathbb{P}-$p.s. alors $\mathbb{E}(X \mid \mathcal{B})(\xi) \in K(\xi)$ $\mathbb{P}-$p.s..*

Preuve : Soit $P_{K(\xi)}$ la projection sur l'ensemble convexe fermé $K(\xi)$, on défini alors :

$$p = \mathbb{E}\left(X \mid \mathcal{B}\right) - P_K(\mathbb{E}\left(X \mid \mathcal{B}\right)).$$

Il est clair que p est une variable aléatoire $\mathcal{B}$-mesurable. Par ailleurs nous avons aussi :

$$\langle p(\xi), X(\xi)\rangle \leq \langle p(\xi), \mathbb{E}\left(X \mid \mathcal{B}\right)(\xi)\rangle - ||p(\xi)||^2_{\mathbb{R}^n} \ ;$$

par conséquent :

$$\mathbb{E}\left[\langle p(\xi), X(\xi)\rangle\right] \leq \mathbb{E}\left[\langle p(\xi), \mathbb{E}\left(X \mid \mathcal{B}\right)(\xi)\rangle\right] - \mathbb{E}\left[||p(\xi)||^2_{\mathbb{R}^n}\right] \ ;$$

il faut donc nécessairement que :

$$\mathbb{E}\left[||p(\xi)||^2_{\mathbb{R}^n}\right] = 0 \ ;$$

c'est-à-dire que $p(\xi) = 0$ $\mathbb{P}-$ p.s. autrement dit que $\mathbb{E}\left(X \mid \mathcal{B}\right)(\xi) \in K(\xi)$ $\mathbb{P}-$ p.s.. □

Bibliographie

[1] J. Abrham. An approximate method for solving a continuous time allocation problem. *Econometrica*, 38(3) :473–481, 1970.

[2] B. Allen. Neighboring information and distributions of agents characteristics under uncertainty. *Journal of mathematical economics*, 12 :63–101, 1983.

[3] B. Allen. Uniform continuity of information combination. *Journal of mathematical economics*, 21 :453–459, 1992.

[4] Robert M Gray Allen Gersho. *Vector Quantization and signal compression.* Kluwer Academic Publishers, 1992.

[5] A. Alonso. A counterexample on the continuity of conditional expectation. *Journal of mathematical analysis and applications*, 129 :1–5, 1988.

[6] A. Alonso. L^p continuity of conditional expectation. *Journal of mathematical analysis and applications*, 221 :161–176, 1998.

[7] V.I. Arkin. Abstracts of the international conference on stochastic optimization. *Kiev V.I Glushkov Institute, Ukrenian Academy of Sciences*, 1984.

[8] Z. Artstein. On a variational problem. *Journal of mathematical analysis and applications*, 45 :404–415, 1974.

[9] Z. Artstein. A note on Fatou's lemma in several dimensions. *Journal of mathematical economics*, pages 277–282, 1979.

[10] Z. Artstein. Sensitivity to σ-fields of information in stochastic allocation. *Stochastics and Stochastics Report*, pages 41–63, 1991.

[11] Z. Artstein. Sensors and information in optimization under stochastic uncertainty. *Math of Operation Research*, 18(3), 1993.

[12] Z. Artstein. Gains and costs of information in stochastic programming. *Annals of operations research*, 85 :129–152, 1999.

[13] Z. Artstein. Compact convergence of σ-fields and relaxed conditional expectation. *Probability theory and related fields*, 120 :369–394, 2001.

[14] Z. Artstein and R. Wets. Consistency of minimizers and SLLN for stochastic programs. *Journal of convex analysis*, pages 1–20, 1995.

[15] H. Attouch. *Variational convergence for functions and operators.* Pitman, 1984.

[16] H. Attouch and R. Wets. Approximation and convergence in nonlinear optimization. *Nonlinear programming*, 4 :367–394, 1981.

[17] J.-P. Aubin. *Initiation à l'analyse appliquée*. Masson, 1994.

[18] R.J. Aumann and M. Perles. A variational problem arising in economics. *J Math Anal Appl*, 11 :488–503, 1965.

[19] K. Barty, P. Carpentier, J-P. Chancelier, G. Cohen, M. Cohen-De-Lara, and T. Guilbaud. Dual effect free stochastic controls. Stochastic Programming E-print Series `http://www.speps.info/`, 18, 2003 à paraître dans Annals of Operation Research, 2004.

[20] H. Berliocchi and J.-M Lasry. Nouvelles applications des mesures paramétrées. *C. R Acad sc Paris*, 274 :1623–1626, 1972.

[21] H. Berliocchi and J.-M Lasry. Intégrandes normales et mesures paramétrées en calcul des variations. *Bull soc math France*, 101 :129–184, 1973.

[22] N. Bouleau. *Probabilités de l'ingénieur*. Herman, 1986.

[23] N. Bouleau. On effective computation of expectations in large or infinite dimension. *J of computational and app. Math*, 33 :23–24, 1990.

[24] Boylan. Equi-convergence of martingales. *Annals of mathematical statistics*, 42 :552–559, 1971.

[25] H. Brezis. *Analyse fonctionnelle, théorie et applications*. Masson, 1983.

[26] L.D. Brown and R. Purves. Measurable selections of extrema. *The annals of statistics*, 1(5) :902–912, 1973.

[27] P. Carpentier, G. Cohen, and J.C. Culioli. Stochastic optimal control and decomposition-coordination methods–part I : Theory. In Roland Durier and Christian, editors, *Recent Developements in Optimization*. Michelot (Eds), LNEMS Springer-Verlag, Berlin, 1995.

[28] F.H. Clarke. *Optimization and nonsmooth analysis*. Wiley, 1983.

[29] F. Coquet, V. Mackevicius, J. Mémin, and Shige Peng. Stability in $\mathbb{D}$ of martingales and backward equations under discretization of filtration. *Stochastic Processes and their applications*, 75 :235–248, 1998.

[30] F. Coquet, V. Mackevicius, J. Mémin, and Shige Peng. Corrigendum of : Stability in $\mathbb{D}$ of martingales and backward equations under discretization of filtration. *Stochastic Processes and their applications*, 82 :335–338, 1999.

[31] K.D Cotter. Similarity of information and behavior with a pointwise convergence topology. *Journal of mathematical economics*, 15 :25–38, 1986.

[32] K.D Cotter. Convergence of information random variables and noise. *Journal of mathematical economics*, 16 :39–51, 1987.

[33] J. Dupačovà. Stability and sensitivity analysis for stochastic programming. *Annals of operations research*, 27 :115–142, 1990.

[34] J. Dupačovà. Postoptimality for multistage stochastic linear programs. *Annals of operations research*, 56 :65–78, 1995.

[35] J. Dupačovà. Stochastic programming : Approximation via scenarios : Proceedings of 3rd Caribbean Conference on Approximation and Optimization, 1995. `http://citeseer.nj.nec.com/dupacova96stochastic.html`.

[36] J. Dupačovà. Scenario-based stochastic programs : Resistance with respect to sample. *Annals of operations research*, 64 :21–38, 1996.

[37] J. Dupačovà, N. Gröwe-Kuska, and W. Römisch. Scenario reduction in stochastic programming : an approach using probability metrics. *Mathematical Programming*, A95 :493–511, 2003.

[38] J. Dupačovà and R. Wets. Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *The annals of statistics*, 16 :1517–1549, 1988.

[39] J. Dupačovà G. Consigli and S-W. Wallace. Scenarios for multistage stochastic programs. *Annals of operations research*, 100 :25–53, 2000.

[40] D. Dacunha-Castelle and M. Duflo. *Probabilité et statistique, Problèmes à temps fixe*. Masson, 1994.

[41] E.B. Dynkin and I.V Evstigneev. Regular conditional expectation of correspondances. *Theory of probability and its applications*, XXI(2), 1976.

[42] M.J. Eisner and P. Olsen. Duality for stochastic programming interpreted as L. P. in L_p-space. *SIAM J Appli Math*, 28(4) :779–792, 1975.

[43] I. Ekeland. *Convexity Methods in Hamiltonian Mechanics*. Springer-Verlags, 1989.

[44] I. Ekeland and R. Temam. *Analyse convexe et problèmes variationnels*. Dunod, 1974.

[45] H. Fetter. On the continuity of conditional expectation. *Journal of mathematical analysis and applications*, 61 :227–231, 1977.

[46] W.H. Fleming and R-W. Rishel. *Deterministic and Stochastic Optimal Control*. Springer-Verlag, 1975.

[47] K. Frauendorfer. Barycentric scenario trees in convex multistage stochastic programming. *Mathematical Programming*, 75 :277–293, 1996.

[48] A.L. Gibbs and F.E. Su. On choosing and bounding probability metrics. *International Statistical Review*, 70 :419–435, 2001.

[49] N. Gröwe-Kuska, H. Heitsch, and W. Römisch. Scenario reduction and scenario tree construction for power management problems. *IEEE Bologna Power Tech Proceedings*, 2003.

[50] N. Gröwe-Kuska, Kiwiel Nowak Römisch, and Wegner. *Power management in hydrothermal system under uncertainty by lagrangian relaxation in : Decision Making under Uncertainty : Energy and Power*, volume 128. IMA Volumes in mathematics and its applications, 2002.

[51] N. Gröwe-Kuska and W. Römish. Stochastic unit commitment in hydro-thermal power production planning. *SIAM Series in Optimization*, 2004 (to appear).

[52] D.B. Hausch and W.T. Ziemba. Bounds on the value of information in uncertain decision problem. *Stochastics*, 10 :181–217, 1983.

[53] U.G. Haussmann. On the adjoint process for optimal control of diffusion processes. *SIAM J. Control and Optimization*, 19 :221–244, 1981.

[54] J. Higle and S. Sen. Multistage stochastic convex programs duality and its implications. To appear in Annals of Operation Research.

[55] J.-B. Hiriart-Urruty. Conditions nécessaires d'optimalité pour un programme stochastique avec recours. *SIAM J.Control and optimization*, 16(2) :317–329, 1978.

[56] J.-B. Hiriart-Urruty. Extension of lipschitz integrands and minimization of nonconvex integral functionals : Applications to the optimal recourse problem in dicrete time. *Probability and mathematical statistics*, 3 :19–36, 1982.

[57] J.-B. Hiriart-Urruty and C. Lemarechal. *Convex analysis and minimization algorithms*. Springer-Verlag, 1993.

[58] F. Hirsch. *Éléments d'analyse fonctionnelle*. Masson, 1997.

[59] D.N. Hoover. Convergence in distribution and Skorokhod convergence for the general theory of processes. *Probability theory and related fields*, 89 :239–259, 1991.

[60] M. Kaut and S.W. Wallace. Evaluation of scenario-generation methods for stochastic programming, 2003. `http://citeseer.nj.nec.com/597868.html`.

[61] A.J. King and R. Wets. Epi-consistency of convex stochastic programs. *Stochastics and stochastics reports*, 34 :83–92, 1991.

[62] H. Kūdo. A note of the strong convergence of σ-algebras. *Annals of probability*, 2 :76–83, 1974.

[63] H. J. Kushner. Necessary conditions for continuous parameter stochastic optimization problems. *SIAM J. Control*, 10(3) :550–565, 1972.

[64] H.J. Kushner. On the stochastic maximum principle with average constraints. *Journal of mathematical analysis and applications*, 12 :13–26, 1965.

[65] D. Landers and L. Rogge. An inequality for the Hausdorff-metric of σ-fields. *Annals of probability*, 14 :724–730, 1986.

[66] E. Michael. Continuous selections I. *Annals of mathematics*, 63(2) :361–381, 1956.

[67] J. Neveu. *Bases mathématiques du calcul des probabilités*. Masson, 1970.

[68] J. Neveu. Note on the tightness of the metric on the set of complete sub σ-algebras of probability space. *The annals of mathematical statistics*, 43(4) :1369–1371, 1972.

[69] D.N. Nghiem. Convergence forte des espérances conditionnelles et des projecteurs d'un espace de Hilbert. *Annales de L'Institut Henri Poincaré*, 4(1) :9–13, 1970.

[70] P. Olsen. Discretizations of multistage stochastic programming problems. *Mathematical programming study*, 6 :111–124, 1976.

[71] S. Peng. A general stochastic maximum principle for optimal control problem. *SIAM J. Control and Optimization*, 28(4) :966–979, 1990.

[72] G. Pflug and A. Swietanowski. Optimal scenario tree generation for multiperiod financial optimization. *Math. Programming*, 89, 2001. `http://citeseer.nj.nec.com/pflug98optimal.html`.

[73] L. Piccinini. Convergence of sub-σ-fields and convergence of associated subspaces $L^p(\mathcal{B}_n)(p \in [1, +\infty[)$. *J. Math. Anal. Appl.*, 225(1) :73–90, 1998.

[74] M. L. Puterman. *Markov decision processes*. Wiley, 1994.

[75] J.-P. Quadrat and M. Viot. Introduction à la commande stochastique, 2002. `http://www-rocq.inria.fr/metalau/quadrat/`.

[76] J.P. Quadrat, M. Gousat, A. Hertz, and M. Viot. Méthodes de gradient stochastique pour l'optimisation des investissements dans un réseau électrique. *EDF, Bulletin de la Direction des Etudes et Recherches*, pages 1933–1967, 1981.

[77] T. Rachev and W. Römisch. Quantitative stability in stochastic programming : the method of probability metrics. *Mathematics of operation research*, 27 :792–818, 2002.

[78] M.M. Rao. *Measure theory and integration*. Wiley-Interscience, 1987.

[79] R.T. Rockafellar. Integrals which are convex functionals. *Pacific J Math*, 24 :525–540, 1968.

[80] R.T. Rockafellar. Measurable dependence of convex sets and functions on parameters. *Journal of mathematical analysis and applications*, 28 :4–25, 1969.

[81] R.T. Rockafellar. Conjugate duality and optimization. *Conference Board of Math Sciences Series, SIAM Publications*, 16 :79, 1974.

[82] R.T. Rockafellar. Integral functionals, normal integrands and measurable selections. *Lecture Notes in Mathematics*, 543, 1976.

[83] R.T. Rockafellar. Directionally lipschitzian functions and subdifferential calculus. *Proc London Math Sco*, pages 331–335, 1978.

[84] R.T. Rockafellar. Generalized directional derivatives and subgradients of nonconvex functions. *Canad J Math*, 32 :257–280, 1980.

[85] R.T. Rockafellar. Deterministic and stochastic optimization problems of Bolza type in discrete time. *Stochastics*, 10 :273–312, 1981.

[86] R.T. Rockafellar and R. Wets. Stochastic convex programming : Kuhn Tucker conditions. *J. Mathematical Economics*, 2 :349–370, 1975.

[87] R.T. Rockafellar and R. Wets. Nonanticipativity and $\mathcal{L}^1$-martingales in stochastic optimization problems. *Mathematical programming study*, 6 :170–187, 1976.

[88] R.T. Rockafellar and R. Wets. Measures as Lagrange multipliers in multistage stochastic programming. *Journal of mathematical analysis and applications*, 60 :301–313, 1977.

[89] R.T. Rockafellar and R. Wets. The optimal recourse problem in discrete time : $\mathcal{L}^1$-multipliers for inequality constraints. *SIAM Journal Control and Optimization*, 16(18) :16–36, 1978.

[90] R.T. Rockafellar and R. Wets. On the interchange of subdifferentiation and conditional expectation for convex functionals. *Stochastics*, 7 :173–182, 1982.

[91] R.T. Rockafellar and R. Wets. The aggregation principle in scenario analysis and stochastic optimization. *Algorithms and model formulations in mathematical programming*, 51 :92–113, 1989.

[92] R.T. Rockafellar and R. Wets. Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of operations research*, 16 :119–147, 1991.

[93] L. Rogge. Uniform inequalities for conditional expectations. *Annals of probability*, 2 :486–489, 1974.

[94] T. Rzezuchowski. Strong convergence of selections implied by weak. *Bull Austral Math Soc*, 39 :201–214, 1989.

[95] A. Shapiro. Statistical inference of multistage stochastic programming problems. In *Probabilistic constrained optimisation : Methodology and application*, pages 282–304. Uryasev, S. (Ed.), Kluwer Academic Publishers, 2000.

[96] A. Shapiro, Homem de Mello, and T. J. Kim. Conditioning of convex piecewise linear stochastic programs. *Mathematical Programming*, 94 :1–19, 2002.

[97] A. Shapiro and T. Homem-De-Mello. On a rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs. *SIAM Journal on Optimization*, 11 :70–86, 2000.

[98] A. Shapiro and A.J. Kleywegt. The sample average approximation mathod for stochastic discrete optimization. *SIAM J Optimization*, 12 :479–502, 2001.

[99] M.B. Stinchcombe. Bayesian information topologies. *Journal of Mathematical economics*, 19 :233–253, 1993.

[100] M.B. Stinchcombe. A further note on bayesian information topologies. *Journal of Mathematical economics*, 22 :189–193, 1993.

[101] C. Striebel. Sufficient statistics in the optimum control of stochastic systems. *Math anal appl*, 12 :576–592, 1965.

[102] M. Valadier. Intégration de convexes fermés notamment d'épigraphes Inf-convolution continue. *R.I.R.O*, pages 57–73, 1970.

[103] P. Varaiya and R. Wets. Stochastic dynamic optimization approches and computation, 1989. `http://www.citeseer.nj.nec.com/70487.html`.

[104] H.S. Witsenhausen. A counterexample in stochastic optimal control. *SIAM J Control*, 6(1) :131–147, 1968.

[105] H.S. Witsenhausen. Separation of estimation and control for discrete time systems. *Proceedings of the IEEE*, 69(11) :1557–1566, 1971.

[106] L.C. Young. Generalized surfaces in the calculus of variations. *Annals of mathematics*, 42 :84–103, 1942.

[107] T. Zolezzi. Well-posed optimization problems for integral functionals. *Journal of optimization theory and applications*, pages 417–430, 1980.

Index

A
Arbres de scénarios 2
Asymptotiquement indépendante 35

B
Borne sup de tribus 40
Boucle ouverte 3

C
Carathéodory 50
Chroniques de scénarios 3
Commande optimale stochastique ... 104
Continuité à gauche 60
Contrainte de parité 69
Convergence
  en loi 124
  uniforme de tribus 42
  en probabilité 124
Correspondance
  s.c.i. 121
  s.c.s. 121

D
Delta, Δ
  différence symétrique 28
Distance 29
  de Boylan 42
  de Cotter 29
  de Rogge 42

E
Effet dual 4, 18
Equation adjointe 108
Equi-intégrable 31, 124
Etat adjoint 106

F
Famille
  causale 19
  totale 36
Filtration 5
Fonction indicatrice 29
Fonction d'observation 4

I
Inégalité
  de Hölder 125
  de Jensen 123
Intégrandes normales 49

K
Kantorovich-Rubinstein 6
Kuhn Tucker 104

L
Limite
  inférieure de tribus 31
  supérieure de tribus 31

M
Mémoire parfaite 19, 20, 23
Multi-application mesurable 125

Q
Quantification 8, 74

S
Sigma, σ 28, 32
Structure d'information 1
  dynamique 4
  statique 3

T
Tau, τ 43
Tribu complète 28

V

Valeur de l'information 60
Voronoï
    Cellule 84