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 $\xi_1 = +++++$, $\quad \xi_2 = +----$ 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 :

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

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 :

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

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 :

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

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 :

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

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 $\mathbb{E}(\rho \mid \mathcal{B}) = 0$ et le problème $\mathcal{P}$ est équivalent (dans le cas convexe) à résoudre :

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

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.