



**HAL**  
open science

# Algorithmes de décomposition/coordination en optimisation stochastique

Jean-Christophe Culioli

► **To cite this version:**

Jean-Christophe Culioli. Algorithmes de décomposition/coordination en optimisation stochastique. Automatique / Robotique. École Nationale Supérieure des Mines de Paris, 1987. Français. NNT : . pastel-00834013

**HAL Id: pastel-00834013**

**<https://pastel.hal.science/pastel-00834013>**

Submitted on 13 Jun 2013

**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 à

L'ÉCOLE NATIONALE SUPÉRIEURE DES MINES DE PARIS

par

**Jean-Christophe CULIOLI**

en vue de l'obtention

DU TITRE DE DOCTEUR

EN

**MATHÉMATIQUES ET AUTOMATIQUE**

Sujet de la thèse

**ALGORITHMES DE  
DÉCOMPOSITION/COORDINATION  
EN  
OPTIMISATION STOCHASTIQUE**

soutenue le 28 Septembre 1987 devant le jury composé de

<b>MM.</b>	<b>Michel MÉTIVIER</b>	Président et rapporteur
	<b>Albert BENVENISTE</b>	Rapporteur
	<b>Guy COHEN</b>	
	<b>Patrick COLLETER</b>	Examineurs
	<b>Michel VIOT</b>	



## REMERCIEMENTS

Ma profonde gratitude va tout d'abord ici à Monsieur Guy Cohen qui a été mon initiateur en mathématiques appliquées. Il n'a cessé, depuis, de me suivre avec bienveillance, de guider fermement mes recherches, et d'enjoindre mon optimisme naïf à une rigueur salutaire.

Je dois aussi le remercier d'avoir accepté les conditions peu habituelles dans lesquelles cette thèse a été préparée, puisqu'elle a été faite en grande partie "par correspondance". La taille du courrier que j'ai reçu de Monsieur Cohen pendant ces quatre dernières années témoigne de sa disponibilité amicale et de sa ferme intention de me faire aboutir.

J'aimerais que Monsieur Michel Métivier, qui me fait l'honneur de présider le Jury, trouve ici l'expression de mon humble gratitude et du profond respect que je n'ai cessé de former à son égard, lors des nombreux cours auxquels j'ai eu le plaisir d'assister. Parmi ceux-ci, je n'oublierai jamais une certaine conférence sur "Le Hasard et les Probabilités" qui avait émerveillé la dizaine d'épistémologues en herbe que nous étions dans ma promotion.

J'exprime ici toute ma reconnaissance à Monsieur Michel Viot qui, dès le départ de ces recherches, a eu la gentillesse de lire la toute première version du théorème 2 et d'y trier le bon grain de l'ivraie. Je le remercie d'avoir accepté de juger ce travail.

J'ai suivi avec beaucoup d'intérêt les cours de Monsieur Albert Benveniste à l'Université de Jussieu sur l'"équation différentielle ordinaire", technique à laquelle il est souvent fait appel tout au long du mémoire. Je suis donc très heureux qu'il me fasse l'honneur d'en être rapporteur.

J'aimerais remercier aussi Monsieur Patrick Colleter d'avoir accepté de participer à ce Jury. J'ai travaillé pendant près de deux ans dans son équipe lors de mon passage à l'Électricité de France et je garderai toujours de cette expérience un excellent souvenir.

Enfin, j'aimerais que Messieurs John D. Ingram et Claude Baudoin, sous les responsabilités desquels j'ai eu la chance de travailler lorsque j'étais au GIERS-SCHLUMBERGER (actuellement Schlumberger Montrouge Recherches), soient ici remerciés pour la confiance amicale et l'aide technique qu'ils ont bien voulu m'accorder.



# Table des matières

<b>Introduction</b>	<b>1</b>
<b>1 Notations et résultats connus</b>	<b>7</b>
1.1 Optimisation Convexe	7
1.1.1 Optimisation Libre	8
1.1.2 Dualité	12
1.1.3 Lagrangien augmenté	15
1.1.4 Lemmes techniques	16
1.2 Modèle probabiliste	19
1.2.1 Définitions et Propriétés	19
1.2.2 Modèle issu des algorithmes stochastiques	20
1.3 Principe du Problème Auxiliaire	22
1.3.1 Quelques réflexions sur le cas $K$ quadratique	24
<b>2 Contraintes Déterministes Découplées</b>	<b>26</b>
2.1 Algorithme de décomposition stochastique	26
2.2 Convergence de l'algorithme	27
2.2.1 Inéquations variationnelles et fonction de Lyapunov	27
2.2.2 Majorations préliminaires	28
2.2.3 Convergence de la suite $\psi^k$	29
2.2.4 Limites de $J(u^k)$ et de $u^k$	30
2.3 Cas particulier : $J$ fortement convexe.	32
2.4 Application à la méthode des moindres carrés	33
2.5 Variantes de l'algorithme principal	35
2.5.1 $J$ est de la forme $J_1 + J_2$ , avec $J_2$ additive	35
2.5.2 $J$ est la composée d'une fonction croissante et d'une fonctionnelle additive	37
2.6 Comparaison avec des résultats antérieurs	38
2.6.1 Dodu, Goursat, Hertz, Quadrat, Viot (1981)	38
2.6.2 Bertran (1975)	39
2.6.3 Gladishev (1965)	40
2.6.4 Kushner & Clark (1978)	40
2.7 Tests Numériques	43
2.7.1 Optimisation Libre : $U^{ad} = U$	43
2.7.2 $\theta$ vue comme contrainte implicite définissant $U^{ad}$	47

<b>3</b>	<b>Décomposition par les prix</b>	<b>50</b>
3.1	Tentative d'extension directe	51
3.2	J est strictement convexe : Lagrangien ordinaire	52
3.2.1	Inéquations variationnelles et fonction de Lyapunov	53
3.2.2	Majorations préliminaires	54
3.2.3	Convergence de la suite $\psi^k$	55
3.2.4	Limites de $L(u^k, p^k)$ et de $u^k$	56
3.3	Variante lorsque J est fortement convexe	59
3.3.1	Inéquations variationnelles et fonction de Lyapunov	60
3.3.2	Majorations préliminaires	60
3.3.3	Convergence de la suite $\psi^k$	62
3.4	J non strictement convexe : Lagrangien augmenté	64
3.4.1	Inéquations variationnelles et fonction de Lyapunov	65
3.4.2	Majorations préliminaires	65
3.4.3	Convergence de la suite $\psi^k$	67
3.4.4	Limites de $L_c(u^*, p^k) - L_c(u^k, p^*)$ et de $u^k$	68
3.5	Comparaison avec des résultats antérieurs	70
3.5.1	Bertran (1975)	70
3.5.2	Kushner & Clark (1978)	71
3.6	Tests numériques	72
<b>4</b>	<b>Décomposition par les quantités</b>	<b>78</b>
4.1	Situation déterministe	83
4.1.1	Convergence de l'algorithme déterministe	83
4.2	Discussion sur une extension stochastique	87
4.2.1	L'extension directe	87
4.2.2	Une extension "freinée" en u	89
4.2.3	Une extension "freinée" en u et "normalisée" en p	91
4.3	Convergence de l'extension stochastique	91
4.3.1	Fonction de Lyapunov et Inéquations variationnelles	92
4.3.2	Majorations préliminaires	93
4.3.3	Fin de la démonstration	95
4.4	Comparaison avec des résultats antérieurs	96
4.5	Tests numériques	97
<b>5</b>	<b>Décomposition par prédiction</b>	<b>103</b>
5.1	Algorithmes à un niveau : Formulation générale	104
5.2	Résultat déterministe et discussion	105
5.3	Tests Numériques	106
<b>6</b>	<b>Conclusion et Problèmes Ouverts</b>	<b>110</b>
<b>7</b>	<b>Bibliographie</b>	<b>120</b>
7.1	Références de la synthèse	121
7.2	Références des chapitres suivants	122

# Introduction

## Avant-Propos

La plupart des problèmes de choix d'investissement ou de gestion optimale sont des problèmes de grande taille. De plus, ils font souvent intervenir un élément aléatoire provenant soit d'une connaissance probabiliste du futur (prévisions de demande, lois de pannes,...) soit de lois de la Nature par essence aléatoires (aléas climatiques, bruits, impondérables,...).

Les systèmes considérés sont souvent complexes aussi bien du point de vue de leur modélisation que de celui de leur optimisation numérique. Ils peuvent être constitués de sous-systèmes hétérogènes pour lesquels une technique globale de résolution n'est pas nécessairement **appropriée** (fonctions objectifs mixtes, contraintes locales,...) ou **possible** (engorgement mémoire prohibitif, complexité exponentielle, ...).

Même dans le cas où les sous-systèmes sont équivalents et peu nombreux, il peut arriver qu'aucune méthode de résolution globale ne soit envisageable. Il suffit de se rappeler qu'une résolution même imprécise de l'équation de Hamilton-Jacobi-Bellman **déterministe** en dimension 3 dans le cube  $[0, 1]^3$  nécessite quelques heures de temps calcul pour les ordinateurs les plus répandus dans l'Industrie.

Cartésiens pour beaucoup d'entre eux, les mathématiciens appliqués se sont, dès les années 60, intéressés à la décomposition des problèmes d'optimisation de grande taille en suites de sous-problèmes plus simples à résoudre. Les premiers travaux ont été ceux de Dantzig & Wolfe [1] (programmation linéaire) et Arrow, Hurwicz & Uzawa [1] ou Lasdon [1] (programmation convexe). Dans ces travaux précurseurs, la fonction objectif globale a une structure particulière : elle doit être **additive** par rapport à une décomposition des variables de commandes et **différentiable**. Par la suite, Takahara [1] et Mesarovic et al. [1] ont proposé des méthodes de décomposition ne nécessitant pas l'additivité mais au moins la différentiabilité.

Enfin, Cohen [2], avec le **Principe du Problème Auxiliaire**, puis Cohen & Zhu [1], ont introduit une méthode générale de décomposition, **sans hypothèse d'additivité ni même de différentiabilité**. Outre l'avantage de réduire la dimension du problème de départ, cette méthode permet de résoudre les sous-problèmes séquentiellement ou en parallèle, éventuellement avec des techniques algorithmiques hétérogènes.



C'est dans cette dernière voie que nous nous engageons, en essayant d'étendre un certain nombre de résultats de Cohen [1] au cas **probabiliste**, en nous inspirant des méthodes de gradient stochastique.

## Motivation

On considère un espace de probabilité  $(\Omega, \mathcal{A}, P)$  et on désigne par  $\omega$  un élément de  $\Omega$ . Considérons un processus stochastique discret d'état  $x$  à valeurs dans un espace de Hilbert  $X$  sur  $\mathbf{R}$ , soumis à une équation d'évolution du type

$$x_{t+1} = g(x_t, u_t, v_t(x_t, \omega), \omega, t)$$

$u_t$  et  $v_t$  étant des commandes respectivement en boucle ouverte et en boucle fermée, à valeurs dans des espaces de Hilbert réels  $U$  et  $V$

Un tel modèle mathématique peut, à première vue, paraître simple. Lorsqu'il s'agit de l'identifier à un processus réel, on s'aperçoit qu'il n'en est rien.

Ce système peut, par exemple, être un parc électrique contenant une composante hydraulique (retenues d'eau équipées de turbines), soumis à des aléas météorologiques (précipitations, fonte des neiges) et à une demande aléatoire en électricité. Le vecteur d'état sera la concaténation d'un grand nombre de variables d'état comme les stocks d'eau des différents lacs concernés, les stocks de combustibles des centrales thermiques, l'état des machines à l'instant considéré (pannes éventuelles, etc.). Remarquons que les aléas, résumés par l'espace probabilisé  $(\Omega, \mathcal{A}, P)$  sont de natures très variées.

Les commandes pourront être des décisions en boucle ouverte  $u$  comme le turbinage moyen hebdomadaire pour chaque lac, des décisions en boucle fermée  $v$  pour satisfaire instantanément le complément de demande électrique (par des éléments à faible constante de temps du système), des décisions de marche ou d'arrêt de groupes thermiques, etc.

On constate qu'en face d'un problème d'une telle complexité, il est nécessaire, non seulement de découpler les problèmes en échelles de temps, mais aussi de faire appel à une méthode de décomposition pour rendre la résolution du problème ("Satisfaire la demande au moindre coût") réalisable dans de bonnes conditions sur ordinateur.

Dans le cas le plus général de la commande optimale stochastique, le système doit satisfaire l'objectif

$$\min_{u \in U^{ad}, v \in V^{ad}} \int_{\Omega} j(x, u, v(x, \omega), \omega) P(d\omega)$$

$U^{ad}$  et  $V^{ad}$  étant des espaces admissibles pour les commandes, sous des contraintes déterministes de type  $\theta(u) \leq 0$  ou stochastiques comme  $\rho(v(x, \omega)) \leq 0$ , les opérateurs  $\theta$  et  $\rho$  étant à valeurs dans des espaces de Hilbert munis d'un ordre partiel " $\leq$ "

Pour notre exemple, le critère global à minimiser sera une combinaison de plusieurs termes : l'espérance du coût de combustible fourni par les centrales thermiques, un coût de défaillance associé à la satisfaction plus ou moins complète de la demande,

un coût final par lac destiné à imposer un volume minimal d'eau au début de chaque nouvelle année hydraulique, etc.

Les contraintes de type  $u \in U^{ad}$  seront, par exemple, des limitations sur les turbinés : turbiné minimum pour assurer l'approvisionnement en eau de l'aval, turbiné maximum.

Les variables  $v$  et  $x$  ont le même type de dépendance en  $\omega$ . Les contraintes sur  $x$  seront, par exemple, des contraintes physiques de stocks ou plus simplement des équations d'évolution. Tout comme les contraintes de type  $v \in V^{ad}$ , ce sont évidemment des contraintes presque-sûres.

Dans tous les cas où elles s'expriment par l'appartenance des solutions à des ensembles admissibles, nous appellerons ce type de contraintes, **contraintes implicites**, par opposition aux contraintes obtenues par la donnée explicite d'un opérateur comme  $\theta$  ou  $\rho$ . De plus, ces contraintes seront généralement dites **non-couplantes** car nous supposons qu'elles n'introduisent pas de liaisons particulières entre les sous-systèmes.

Dans notre exemple, des contraintes explicites du type  $\theta(u) \leq 0$  pourront apparaître au niveau des groupes hydrauliques fournissant le turbiné total d'un lac. De même, par transformation de contraintes plus subtiles mais moins faciles à gérer comme  $P(\chi(u) > \eta) \leq \epsilon$ , on peut obtenir des contraintes "en espérance" similaires à  $\theta(u) \leq 0$ . Un cas évident de contrainte explicite  $\rho(v(x, \omega)) \leq 0$  est la contrainte presque-sûre de satisfaction de la demande en électricité. Avec la terminologie précédente, on pourra parler de contraintes explicites **couplantes** ou **non-couplantes** suivant qu'elles font, ou non, coopérer les sous-systèmes entre eux. Dans le cas d'une vallée contenant deux lacs en cascades, l'équation d'évolution du lac aval tiendra compte du turbiné du lac amont et sera donc une contrainte **explicite couplante**.

La commande optimale stochastique n'est pas le seul objet de ce mémoire. En optimisation des investissements, notamment dans les dimensionnements de réseaux (électriques, téléphoniques, informatiques), on trouve des situations similaires, où  $u$  sera la variable de dimensionnement et  $v$  la variable de décision de routage. L'aléa est ici issu d'une demande externe à satisfaire et de pannes éventuelles. Ces applications sont traitées dans Minoux [1] pour les réseaux téléphoniques et Dodu et al. [1] pour les réseaux électriques. On peut aussi trouver dans Ermoliev [1,2] des applications au choix d'emplacements pour des bâtiments publics (piscine, école, service administratif, etc.). Dans ce dernier cas, seules des variables en boucle ouverte sont recherchées.

La plupart des exemples décrits plus haut sont couramment traités par un algorithme dérivé de la méthode dite du "gradient stochastique", introduit par Robbins & Monro [1] en approximation stochastique. Cette méthode ne permettant pas de calculer des variables de commande en boucle fermée, on recherche éventuellement une approximation de celles-ci, au prix d'une perte d'optimalité, sous la forme de feedbacks locaux (Quadrat [1]) dont les paramètres sont, quant à eux, des variables

en boucle ouverte.

Par exemple, dans la recherche de la suite des turbinés hebdomadaires optimaux au niveau d'un lac, on posera  $v_t = \alpha_t x_t + \beta_t$  (feedback local sur le stock  $x_t$ ) et on déterminera les suites de paramètres  $\alpha_t, \beta_t, t = 1, \dots, T$  en boucle ouverte.

Les contraintes "presque-sûres" sont, quant à elles, généralement traitées par pénalisation (coût de défaillance, par exemple). Compte tenu du caractère souvent continu des aléas (si l'on exclut les aléas de pannes), il nous semble malheureusement difficile de faire mieux pour le moment.

## Décomposition et coordination

L'apport original de ce mémoire est d'envisager la décomposition / coordination des problèmes d'optimisation stochastiques sous l'angle du "Principe du Problème Auxiliaire" (Cohen [1])<sup>1</sup>. D'autres approches ont été proposées (voir Benveniste et al. [1] ou Looze & Sandell [1]). Par exemple, décomposer un problème stochastique exprimé avec le formalisme de la Programmation Dynamique est particulièrement difficile. Dans le cas le plus simple, (problème linéaire quadratique gaussien) on ne peut décomposer qu'au prix de manipulations matricielles assez lourdes venant du fait que l'état est en réalité une distribution de probabilité c'est-à-dire, dans le cas présent, la matrice de covariance de la loi gaussienne associée. On passe donc de  $N$  variables à  $\frac{N(N+1)}{2}$  sous-problèmes en voulant décomposer "naturellement" le problème global. De plus, les  $\frac{N(N+1)}{2}$  sous-problèmes obtenus ne peuvent plus être réinterprétés comme des problèmes d'optimisation stochastique.

L'idée originale ici est donc plutôt de tenter une symbiose entre l'approche variationnelle des méthodes de décomposition / coordination basées sur le principe du problème auxiliaire et la méthode du gradient stochastique.

En voici une description rapide dans un cas simple (variables  $u$  seules). Supposons donnée une décomposition des variables de commandes. Cela définit  $N$  sous-systèmes qui coopèrent à la satisfaction minimale d'un certain critère en espérance, et qui sont éventuellement couplés par des contraintes déterministes. A l'étape  $k$ , en suivant la méthode du gradient stochastique, on tire un aléa,  $\omega^{k+1}$  suivant la loi  $P$ , indépendant des tirages précédents. Cela permet de calculer une approximation stochastique du gradient du critère par rapport à chaque variable  $u_i, i = 1, \dots, N$ . On peut alors résoudre séparément les  $N$  sous-problèmes déterministes issus de l'application du principe du problème auxiliaire. Lorsque des contraintes explicites existent, on peut mener la décomposition, soit à l'aide de l'introduction d'une variable duale **déterministe** par contrainte (méthode "par les prix"), soit à l'aide d'un schéma d'allocation des ressources (méthode "par les quantités"), soit enfin en mettant en oeuvre une méthode mixte (méthode "par prédiction").

<sup>1</sup>On trouvera une introduction au Principe du Problème Auxiliaire à la section 3 du chapitre 1.

## Organisation du mémoire

Afin de présenter l'organisation du mémoire, nous allons donner une classification des problèmes d'optimisation stochastique et préciser quels sont les problèmes que nous pensons avoir résolus et quels sont ceux qui, pour nous, restent ouverts.

### Classe A : contraintes déterministes implicites découplées et commande $u$ seule

Le problème à résoudre est

$$\min_{u \in U^{ad}} E\{j(u, \omega)\} = \int_{\Omega} j(u, \omega) P(d\omega)$$

Nous sommes dans le cas où un algorithme de gradient stochastique est directement applicable. On suppose dans toute la suite l'existence d'une décomposition de  $U$  (et  $U^{ad}$ ) sous la forme  $\prod_{i=1}^N U_i$ , (et  $\prod_{i=1}^N U_i^{ad}$ ). L'objet du chapitre 2 est de montrer qu'un mariage heureux entre Principe du Problème Auxiliaire et gradient stochastique est réalisable. Des contraintes explicites sur  $u$  peuvent se présenter, à condition qu'elles soient sous une forme découplée. D'autres problèmes n'ayant pas a priori cette forme peuvent toute fois s'y ramener, comme

$$\min_{u \in U^{ad}} E\left\{ \min_{v \in V^{ad}(u, \omega)} g(u, v(u, \omega), \omega) \right\}$$

à condition que soit satisfaite la propriété suivante

$$\forall u \in U^{ad}, V^{ad}(u, \omega) \neq \emptyset \text{ p.s.}$$

qui empêche l'apparition via  $v$  de contraintes non déterministes sur  $u$ , et aussi à condition que la minimisation "interne" en  $v$  soit naturellement décomposée.

### Classe B : contraintes déterministes explicites couplantes et commande $u$ seule

On considère le même problème que précédemment, mais avec des contraintes déterministes couplantes comme

$$\theta(u) \leq 0$$

Dans le cas déterministe, ce sont des techniques de dualité qui permettent de prendre en compte le couplage par les contraintes. On applique le Principe du Problème Auxiliaire à la recherche des point-selles du Lagrangien. Nous traitons les trois méthodes de décomposition / coordination :

- dans la coordination par les prix, à chaque étape de l'algorithme, on itère sur les "prix", c'est-à-dire les multiplicateurs de Lagrange associés aux contraintes couplantes. On remarquera que même dans le cadre de l'optimisation stochastique sous contraintes déterministes, ces variables duales sont elles-mêmes déterministes, ce qui laisse a priori la possibilité que l'on puisse les

approximer par un algorithme de gradient stochastique. De tels algorithmes ont déjà été étudiés dans Kushner & Clark [1] pour des contraintes “en espérance” Au chapitre 3, des résultats de convergence sont obtenus pour des algorithmes de type “Arrow-Hurwicz Stochastique” Nous montrerons, de plus, que des algorithmes de type “Uzawa Stochastique” ne peuvent converger vers la bonne solution.

- Au chapitre 4, nous entreprendrons une étude similaire pour la décomposition “par les quantités” Les résultats de ce chapitre sont, à notre connaissance, originaux.
- Au chapitre 5, nous proposerons un test numérique d’une méthode de décomposition “par prédiction” et nous énoncerons une conjecture qui nous semble raisonnable, bien que non démontrée.

## Classes C et D : contraintes stochastiques découplées ou non et éventuellement variable $v$

Nous sommes, par exemple, dans le cas de contraintes presque-sûres découplées :

$$\min E\{j(u, \omega)\} \text{ sous } u \in U^{ad}(\omega)$$

Tout ce que nous pourrons proposer, au chapitre 6, c’est soit une méthode de pénalisation exacte, soit une transformation de la contrainte si c’est une contrainte en probabilité. Dans les deux cas, on pourra, dans certaines conditions se ramener à des problèmes de classe A ou B.

Nous aborderons aussi dans ce chapitre le problème de la décomposition par rapport à des variables de type  $v(x, \omega)$ .

Enfin, le chapitre 1 est consacré à des rappels destinés à rendre la lecture de ce mémoire plus aisée et contient, notamment, une introduction à la Méthode du Problème Auxiliaire de G. Cohen. Chaque chapitre se termine par une comparaison avec des travaux antérieurs et par un test numérique des algorithmes proposés. La bibliographie est au chapitre 7

# Chapitre 1

## Notations et résultats connus

Nous allons résoudre des problèmes d'optimisation convexe stochastique. Nous rappelons donc ici quelques résultats d'optimisation convexe et de théorie des probabilités, ainsi qu'un certain nombre d'hypothèses qui seront faites assez systématiquement dans la suite. Sans donner exhaustivement les démonstrations des résultats annoncés, nous indiquons parfois les arguments fondamentaux permettant de les démontrer.

Enfin, nous présentons (section 1.3) le Principe du Problème Auxiliaire introduit par G. Cohen, en montrant, sur un cas particulier déterministe, sa traduction algorithmique et la méthode de démonstration de convergence dont nous nous sommes inspirés tout au long de ce mémoire.

Pour cette brève synthèse, nous nous sommes inspirés de Aubin[1,2], Faurre[1], Schwartz[1,2] et Céa[1] pour l'optimisation convexe, de Métivier[1,2], Guikman & Skorokhod[1] et Neveu[1] pour le cadre probabiliste, et bien sûr de Cohen [1] pour le Principe du Problème Auxiliaire.

### 1.1 Optimisation Convexe

Nous nous plaçons dans un espace de Hilbert réel  $U$ , de dimension quelconque, identifié à son dual  $U'$  et muni d'un produit scalaire noté  $(\cdot, \cdot)$ .

Les topologies forte et faible sont les topologies canoniques associées respectivement à la norme  $\|\cdot\| = (\cdot, \cdot)^{1/2}$  et au produit scalaire.

Nous considérons des problèmes d'optimisation convexe sous contraintes : on minimise la fonctionnelle convexe  $J$  sur le fermé  $U^{ad}$  de  $U$ . On introduit un autre espace de Hilbert,  $\mathcal{C}$ , dit "espace des contraintes", lui aussi identifié à son dual  $\mathcal{C}'$ , de produit scalaire  $\langle \cdot, \cdot \rangle$  et norme associée aussi notée  $\|\cdot\|$ .

On définit une contrainte par la donnée d'un opérateur  $\theta(\cdot)$  de  $U^{ad}$  dans  $\mathcal{C}$  et d'un cône convexe fermé  $C$  de  $\mathcal{C}$ .

Ainsi posé, le problème est <sup>(1)</sup> :

$$\left\{ \min_{u \in U^{ad}} J(u) \mid \theta(u) \in -C \right\} \quad (1.1)$$

Dans notre cas,  $J$  est de plus l'espérance d'une fonctionnelle  $j(u, \omega)$  définie du produit  $\Omega \times U$  dans  $\mathbf{R}$ .  $\Omega$  est l'ensemble des réalisations de l'espace probabilisé  $(\Omega, \mathcal{A}, P)$ ,  $\mathcal{A}$  étant la  $\sigma$ -algèbre associée et  $P$  la loi de probabilité.

$$J(u) = \int_{\Omega} j(u, \omega) P(d\omega) \quad (1.2)$$

### 1.1.1 Optimisation Libre

**Définition 1** Une fonctionnelle  $J$  définie sur  $U$  est dite **convexe** (resp. **strictement convexe**) si  $\forall t \in ]0, 1[$ ,

$$\forall (u, v) \in U^2 \text{ [resp. } u \neq v], \quad J(tu + (1-t)v) \leq \text{ [resp. } < ] tJ(u) + (1-t)J(v)$$

On appelle **épigraphe** et **domaine** de  $J$  respectivement les ensembles :

$$\text{Epi}J = \{(v, y) \in U \times \mathbf{R} \mid y \geq J(v)\}, \quad \text{Dom}J = \{v \in U \mid J(v) < +\infty\}$$

$J$  est dite **propre** si son domaine est non vide et si elle n'est pas identique à  $-\infty$ .

**Définition 2**  $J$  est dite **semi-continue inférieurement** (ou **s.c.i.**) si

$$\forall u \in U, \quad J(u) \leq \liminf_{v \rightarrow u} J(v)$$

#### Résultat 1

- (i) L'épigraphe et le domaine d'une fonctionnelle convexe sont convexes.
- (ii) L'épigraphe d'une fonctionnelle s.c.i. est fermé.
- (iii) Toute fonctionnelle convexe, s.c.i. et propre est minorée par une fonctionnelle affine.

Les points (i) et (ii) découlent immédiatement des définitions. Le point (iii) est une application du théorème de séparation de Hahn-Banach appliqué à  $\text{Epi}J$

**Définition 3** Une fonctionnelle  $J$  est dite **coercive sur**  $U^{ad}$  si  $U^{ad}$  est borné ou si

$$\lim_{\substack{u \in U^{ad}, \\ \|u\| \rightarrow +\infty}} J(u) = +\infty$$

<sup>1</sup>pour les contraintes égalité,  $C = \{0\}$

**Résultat 2** Soit  $U$  un Hilbert,  $J : U \rightarrow \mathbf{R}$  une fonctionnelle convexe s.c.i., et  $U^{ad} \subset U$  un convexe fermé. Si  $J$  est coercive sur  $U^{ad}$ , alors  $J$  admet au moins un minimum sur  $U^{ad}$ . Si  $J$  est strictement convexe, alors ce minimum est unique.

En effet, soit  $(u^k)_{k \in \mathbf{N}}$  une suite contenue dans  $U^{ad}$ , et telle que  $J(u^k)$  tend vers  $\inf_{u \in U^{ad}} J(u)$ . Une telle suite est bornée car  $J$  est coercive sur  $U^{ad}$ . Elle admet donc une sous-suite  $u^{\phi(k)}$  faiblement convergente vers un point  $u^* \in U^{ad}$  (car  $U^{ad}$  est convexe fermé, donc faiblement fermé). De plus,  $J$  étant s.c.i.,  $J(u^*) \leq \liminf J(u^{\phi(k)}) = \inf_{u \in U^{ad}} J(u)$ . L'unicité est immédiate.

**Définition 4**  $J$  est dite **fortement convexe de rapport  $a$**  [ ou  $a$ -convexe ], si  $\forall t \in [0, 1]$ ,

$$\forall (u, v) \in U^2, \quad J(tu + (1-t)v) \leq tJ(u) + (1-t)J(v) - \frac{a}{2}t(1-t)\|u - v\|^2$$

**Définition 5** Une fonctionnelle  $J$  est dite **Gâteaux-différentiable** (ou simplement **différentiable**) au point  $u \in U$  s'il existe une forme linéaire continue notée  $J'(u) : U \rightarrow \mathbf{R}$  telle que :

$$\forall v \in U, \quad \lim_{h \rightarrow 0^+} \frac{J(u + hv) - J(u)}{h} = (J'(u), v)$$

$J$  est dite **différentiable sur  $U$**  si elle l'est pour tout  $u \in U$

**Définition 6** On appelle **sous-différentiel** de  $J$  au point  $u \in \text{Dom} J$ , l'ensemble

$$\partial J(u) = \{r \in U' \mid \forall v \in U, J(v) - J(u) \geq (r, v - u)\}$$

$r \in \partial J(u)$  est un **sous-gradient** de  $J$  au point  $u$ .

Si  $\partial J(u) \neq \emptyset$ ,  $J$  est dite **sous-différentiable en  $u$** .



**Résultat 3**

- (i) Toute fonctionnelle convexe  $J : U \rightarrow \mathbf{R}$  est dérivable à droite sur l'intérieur de son domaine, c'est-à-dire que la limite suivante (aussi appelée dérivée directionnelle) existe et est finie, pour tout  $v \in U$

$$DJ(u)[v] = \lim_{h \rightarrow 0^+} \frac{J(u + hv) - J(u)}{h}$$

- (ii) de plus, si  $J$  est sous-différentiable en  $u$ , on a l'équivalence entre

$$r \in \partial J(u) \text{ et } (r, v) \leq DJ(u)[v]$$

- (iii) finalement,

$$\forall v \in U, DJ(u)[v] = \max_{r \in \partial J(u)} (r, v)$$

et si  $\partial J(u)$  est un singleton,  $J$  est différentiable en  $u$ .

**Résultat 4**

- (i) Si  $J$  est différentiable, l' $\alpha$ -convexité implique que

$$\forall (u, v) \in U^2, J(v) - J(u) - (J'(u), v - u) \geq \frac{\alpha}{2} \|u - v\|^2$$

ou à

$$\forall (u, v) \in U^2, (J'(v) - J'(u), v - u) \geq \alpha \|u - v\|^2$$

- (ii) Une fonctionnelle  $J$ ,  $\alpha$ -convexe est strictement convexe et coercive.  
 (iii) Si  $J$  est sous-différentiable, (i) reste vrai en remplaçant  $J'(\cdot)$  par tout  $r(\cdot) \in \partial J(\cdot)$ . On dit alors que le sous-différentiel est fortement monotone de rapport  $\alpha$  ou encore  $\alpha$ -monotone.

**Définition 7 (croissance de la dérivée)**

- (i) une fonctionnelle  $J$  est dite Lipchitzienne de rapport  $L_J$  sur  $U$  si

$$\forall (u, v) \in U^2, |J(u) - J(v)| \leq L_J \|u - v\|$$

- (ii)  $J$  est dite à sous-gradient linéairement borné (de paramètres  $c_1$  et  $c_2$ ) si

$$\forall u \in \text{Dom} J, \forall r \in \partial J(u), \|r\| \leq c_1 \|u\| + c_2$$

On résumera cette dernière définition par  $J$  est SGLB( $c_1, c_2$ ).

- (iii)  $J$  est dite à dérivée Lipchitzienne de constante  $B$  si

$$\forall (u, v) \in U^2, |J'(u) - J'(v)| \leq B \|u - v\|$$

**Résultat 5**

- Si  $J$  est sous-différentiable, alors 7-(i) est équivalent à

$$\forall u \in \text{Dom}J, \forall r \in \partial J(u), \|r\| \leq L_J$$

- D'autre part, 7-(ii) implique 7-(i) sur tout borné de  $U$
- Enfin,  $J$  SGLB( $c_1, c_2$ ) est équivalent à (Zhu[1])

$$\forall u, v \in \text{Dom}J, |J(u) - J(v)| \leq [c_1 \max(\|u\|, \|v\|) + c_2] \|u - v\|$$

On utilisera presque systématiquement la définition 7-(ii) sous la forme équivalente

$$\|r\| \leq c_1 \|u - u^*\| + c_2,$$

pour d'autres valeurs des constantes  $c_1$  et  $c_2$ ,  $u^*$  étant un point de  $U^{ad}$ , solution du problème de minimisation.

**Résultat 6** Si la fonctionnelle  $J : U \rightarrow \mathbf{R}$  est à dérivée Lipchitzienne de constante  $B$ , alors

$$\forall (u, v) \in U^2, J(v) - J(u) \leq (J'(u), v - u) + \frac{B}{2} \|v - u\|^2$$

**Résultat 7** Conditions d'optimalité.

Soit  $U$  un espace de Hilbert. Soient  $J_1, J_2$  et  $J_3$  fonctionnelles sur  $U$ , toutes convexes propres et  $J_2$  sous-différentiable. Soit  $\phi$  une fonction réelle convexe, croissante et sous-différentiable. Soit  $U^{ad}$  un convexe fermé de  $U$ . Si  $J_2(\cdot) = \phi(J_3(\cdot))$ , alors pour  $u^* \in U^{ad}$ , les quatre conditions suivantes sont équivalentes.

$$u^* = \text{Arg} \min_{v \in U^{ad}} J_1(v) + J_2(v) \quad (1.3)$$

$$\exists r^* \in \partial J_2(u^*), \forall v \in U^{ad}, J_1(v) - J_1(u^*) + (r^*, v - u^*) \geq 0 \quad (1.4)$$

$$\forall v \in U^{ad}, \forall r \in \partial J_2(v), J_1(v) - J_1(u^*) + (r, v - u^*) \geq 0 \quad (1.5)$$

$$\exists f^* \in \partial \phi(J_3(u^*)), \forall v \in U^{ad}, J_1(v) - J_1(u^*) + f^*(J_3(v) - J_3(u^*)) \geq 0 \quad (1.6)$$

**Résultat 8** *Projection.*

(i) L'opérateur de projection  $\Pi_{U^{ad}}$  sur un convexe fermé  $U^{ad} \subset U$  est défini par

$$\Pi u = \text{Arg} \min_{v \in U^{ad}} \frac{1}{2} \|v - u\|^2$$

et caractérisé par (application de 1.5)

$$\forall v \in U^{ad}, (v - \Pi(u), v - u) \geq 0$$

ou (application de 1.4)

$$\forall v \in U^{ad}, (u - \Pi(u), v - \Pi(u)) \leq 0$$

de plus,

$$\forall (u, v) \in U^2, \|\Pi(v) - \Pi(u)\| \leq \|v - u\|$$

(ii) Si  $U^{ad}$  est un sous-espace de  $U$ , alors  $\Pi_{U^{ad}}$  est linéaire.

**1.1.2 Dualité**

En plus de  $U$ , on introduit un autre espace de Hilbert,  $\mathcal{C}$ , dit "espace des contraintes", lui aussi identifié à son dual  $\mathcal{C}'$ , de produit scalaire  $\langle \cdot, \cdot \rangle$  et norme associée aussi notée  $\|\cdot\|$ .

On définit une contrainte par la donnée d'un opérateur  $\theta(\cdot)$  de  $U^{ad}$  dans  $\mathcal{C}$  et d'un cône  $C$  de  $\mathcal{C}$ .

Ainsi posé, le problème est :<sup>(2)</sup>

$$\left\{ \min_{u \in U^{ad}} J(u) \mid \theta(u) \in -C \right\} \quad (1.7)$$

**Définition 8** Soit  $C$  un sous-ensemble de  $\mathcal{C}$ , espace de Hilbert. On dit que  $C$  est un cône de sommet 0 si

$$\forall q \in C, \forall t \in \mathbf{R}^+, t \cdot q \in C$$

Par définition,  $-C = \{q \in C \mid -q \in C\}$ . Dans la suite, on suppose  $C$  convexe, fermé d'intérieur non vide et tel que  $C \cap (-C) = \{0\}$ . On définit alors la relation d'ordre " $\leq$ ", compatible avec la structure d'espace vectoriel :

$$\forall p, q \in C, p \leq q \text{ ssi } p - q \in -C$$

Cette définition donne un sens au problème (1.7).

<sup>2</sup>pour les contraintes égalité,  $C = \{0\}$

**Définition 9** On appelle cône conjugué de  $C$ , l'ensemble  $C'$  défini par

$$C' = \{q' \in C' \mid \langle q', q \rangle \geq 0, \forall q \in C\}$$

Nous rappelons ici les généralisations des notions de convexité et de sous-différentiabilité pour la notion d'ordre introduite sur  $C$ .

**Définition 10**  $\theta : U \rightarrow C$  est dite  $C$ -conexe si

$$\forall u, v \in U, \forall t \in [0, 1], \theta(t.u + (1-t).v) - t.\theta(u) - (1-t).\theta(v) \in -C$$

$\theta$  est dite  $C$ -sous-différentiable en  $u$  si l'ensemble

$$\partial\theta(u) = \{\tau \in \mathcal{L}(U, C) \mid \forall v \in U, \tau.(v-u) - (\theta(v) - \theta(u)) \in -C\}$$

est non vide. Cet ensemble est appelé  $C$ -sous-différentiel de  $\theta$ .

Si  $\theta$  est  $C$ -conexe, la fonctionnelle conexe  $f_p(\cdot) := \langle p, \theta(\cdot) \rangle$  est conexe si  $p \in C'$ , et dans ce cas elle est dite **régulièrement** sous-différentiable si  $\partial f_p(u)$  se réduit aux éléments de la forme  $\tau^!p$  avec  $\tau \in \partial\theta(u)$ , le signe " $!$ " désignant l'opérateur adjoint.

On remarque que si  $C = \{0\}$  (contraintes égalités),  $\theta$   $C$ -convexe équivaut à  $\theta$  affine.

**Définition 11** Le Lagrangien associé au problème (1.7) est par définition

$$L(u, p) := J(u) + \langle p, \theta(u) \rangle \quad (1.8)$$

On appellera "conditions de Slater" les conditions suivantes

- contraintes inégalité :  $\theta(U^{ad}) \cap \text{Int}(-C) \neq \emptyset$
- contraintes égalité :  $0 \in \text{Int}(\theta(U^{ad}))$

$\text{Int}$  désignant l'intérieur d'un ensemble.

**Résultat 9** *Théorème du point-selle.*

On suppose  $J$  convexe, s.c.i. et coercive sur  $U^{ad}$

- si la condition de Slater est satisfaite, alors  $u^* \in U^{ad}$  est solution de (1.7) si et seulement si

$$\exists p^* \in C', \forall (u, p) \in U^{ad} \times C', L(u^*, p) \leq L(u^*, p^*) \leq L(u, p^*)$$

ou si et seulement si  $\forall u \in U^{ad}$ ,

$$J(u^*) \leq J(u) + \langle p^*, \theta(u) \rangle, \langle p^*, \theta(u^*) \rangle = 0, \theta(u^*) \in -C, p^* \in C'$$

- si  $J$  est sous-différentiable et  $\theta$  est régulièrement  $C$ -sous-différentiable, alors  $u^* \in U^{ad}$  est solution de (1.7) si et seulement si  $\exists p^* \in C'$  tel que

$$\exists r^* \in \partial J(u^*), \exists r^* \in \partial \theta(u^*), (r^* + (r^*)^* p^*, u - u^*), \forall u \in U^{ad}$$

et

$$\langle p^*, \theta(u^*) \rangle = 0, \theta(u^*) \in -C$$

Les conditions ci-dessus sont les conditions de Kuhn et Tucker pour le problème (1.7). On est parfois amené à introduire le problème dual de (1.7), ainsi que les ensembles  $\hat{U}(p)$  et  $\hat{P}(u)$  pour  $u \in U^{ad}$  et  $p \in C'$  :

**Définition 12** On note  $U^*$  l'ensemble des solutions du problème (1.7) et  $P^*$  l'ensemble des solutions du problème dual suivant

$$\max_{p \in C'} G(p) \text{ avec } G(p) = \min_{u \in U^{ad}} L(u, p) \text{ si } p \in C' \text{ et } G(p) = -\infty \text{ sinon (1.9)}$$

$G(\cdot)$  est dite fonction duale de  $J$ . De plus, on note

$$\hat{U}(p) = \text{Arg min}_{u \in U^{ad}} L(u, p), \hat{P}(u) = \text{Arg max}_{p \in C'} L(u, p)$$

$G$  est concave et

$$\forall (p^*, u^*) \in P^* \times U^*, U^* \subset \hat{U}(p^*) \text{ et } P^* \subset \hat{P}(u^*)$$

On dit que le Lagrangien  $L$  est stable en  $u$  si

$$\forall p^* \in P^*, \hat{U}(p^*) = U^*$$

**Résultat 10** Solutions primales et duales de (1.7) et propriétés de  $\hat{U}$

- l'ensemble des points-selles de  $L$  sur  $U^{ad} \times C'$  est égal à  $U^* \times P^*$
- si  $\hat{U}(p)$  existe et est borné, alors  $G$  est sous-différentiable et  $\partial G(p)$  est l'enveloppe convexe fermée des  $\theta(u)$  pour  $u \in \hat{U}(p)$ .
- si  $\hat{U}(p)$  est un singleton,  $G$  est différentiable il suffit en fait que  $\theta(\hat{U}(p))$  soit réduit à un singleton.
- si  $J$  (et donc  $L$ ) est strictement convexe en  $u$  alors  $L$  est stable en  $u$ .

### 1.1.3 Lagrangien augmenté

Nous présentons ici brièvement la définition et quelques résultats concernant les Lagrangiens augmentés. Une étude très complète de cette technique se trouve dans Zhu[1, p. 84 et suiv.]

**Définition 13** On considère le programme

$$\{ \min_{u \in U^{ad}} J(u) \mid \theta(u) \in -C \}$$

On définit comme suit le **Lagrangien augmenté** associé à ce problème :

- si  $C = \{0\}$  (cas égalité), alors

$$L_c(u, p) := J(u) + \langle p, \theta(u) \rangle + \frac{c}{2} \|\theta(u)\|^2$$

- sinon (cas inégalité),

$$L_c(u, p) := \min_{\chi \in -C} [J(u) + \langle p, \theta(u) - \chi \rangle + \frac{c}{2} \|\theta(u) - \chi\|^2]$$

La définition, dans le cas de contraintes inégalité, n'est pas très opératoire. Les résultats suivants, extraits de Zhu[1], donnent plus d'information sur les manipulations des Lagrangiens augmentés.

**Résultat 11** Dans le cas de contraintes inégalité, on peut encore écrire  $L_c$  sous la forme suivante

$$L_c(u, p) = J(u) + \Phi(\theta(u), p)$$

avec

$$\Phi : C \times C' \rightarrow \mathbf{R}, \quad \Phi(\theta, p) = \min_{\chi \in -C} \langle p, \theta - \chi \rangle + \frac{c}{2} \|\theta - \chi\|^2$$

ou encore

$$\Phi(\theta, p) = \max_{q \in C'} \langle q, \theta \rangle - \frac{1}{2c} \|q - p\|^2$$

$\Phi$  a les propriétés suivantes :

- $\Phi$  est convexe en  $\theta$ , concave en  $p$  et différentiable en  $\theta$  et  $p$

$$\Phi'_\theta(\theta, p) = \Pi_{C'}(p + c\theta), \quad \Phi'_p(\theta, p) = \frac{1}{c} [\Pi_{C'}(p + c\theta) - p]$$

- $\Phi$  peut encore s'écrire

$$\Phi(\theta, p) = \langle \Phi'_\theta(\theta, p), \theta \rangle - \frac{c}{2} \|\Phi'_p(\theta, p)\|^2 = \frac{1}{2c} [\|\Pi_{C'}(p + c\theta)\|^2 - \|p\|^2]$$

- enfin, on a les majorations suivantes

$$\forall p \in C', \forall \theta \in -C, \Phi(\theta, p) \leq 0 \quad \text{et} \quad \forall \theta \in C, \forall p \in C', \Phi(\theta, p) \geq \langle p, \theta \rangle$$

Tout l'intérêt de l'introduction de Lagrangiens augmentés provient du résultat suivant.

**Résultat 12** Sous la condition de Slater, le Lagrangien augmenté a un point-selle sur  $U^{ad} \times C'$ . Plus précisément, les Lagrangiens  $L$  et  $L_c$  ont le même ensemble de points-selle  $U^* \times P^*$ , respectivement sur les ensembles admissibles  $U^{ad} \times C'$  et  $U^{ad} \times C'$ . De plus,  $L_c$  est stable en  $u$  : en notant  $U_c^*$  l'ensemble des solutions de  $\min_{u \in U^{ad}} L_c(u, p^*)$ , on a

$$\forall p^* \in P^*, u \in U_c^* \implies u \in U^*$$

### 1.1.4 Lemmes techniques

**Définition 14** La suite  $(\epsilon^k)_{k \in \mathbf{N}}$  est une  $\sigma$ -suite si

$$\sum_{k=1}^{\infty} \epsilon^k = \infty \quad \text{et} \quad \sum_{k=1}^{\infty} (\epsilon^k)^2 < \infty$$

**Lemme 1** Soit  $(y^k)_{k \in \mathbb{N}}$  et  $(\alpha^k)_{k \in \mathbb{N}}$  deux suites de réels positifs telles que  $\alpha^k$  soit le terme d'une série convergente et

$$y^{k+1} \leq \sup_{l \leq k+1} \alpha^l y^l + \mu^k$$

les  $\mu^k$  étant bornés. Alors  $(y^k)_{k \in \mathbb{N}}$  est une suite bornée.

La démonstration de ce lemme se trouve dans Cohen [1, p. 146]

**Lemme 2** Soit  $J$ , fonctionnelle lipschitzienne sur tout borné du Hilbert  $U$ . Soient  $(u^k)_{k \in \mathbb{N}}$  et  $(\epsilon^k)_{k \in \mathbb{N}}$  deux suites à valeurs dans  $U$  et  $\mathbb{R}^+$  respectivement, telles que :

$$\exists \delta > 0, \quad \forall k \in \mathbb{N}, \quad \|u^{k+1} - u^k\| < \delta \epsilon^k$$

avec

$$\sum_{k \in \mathbb{N}} \epsilon^k = \infty \quad \text{et} \quad \exists \mu, \quad \sum_{k \in \mathbb{N}} \epsilon^k |J(u^k) - \mu| < \infty$$

alors

$$\lim_{k \rightarrow \infty} J(u^k) = \mu$$

La démonstration de ce lemme se trouve dans Cohen [1, p. 144].

Ce lemme, très utilisé dans la suite, est justifié par le contre-exemple suivant :

Soit une suite **bornée**  $(u^k)_{k \in \mathbb{N}}$  à valeurs dans  $\mathbb{R}^+$ . Même si

$$\sum_k \epsilon^k u^k = l < +\infty \quad \text{et} \quad \sum_k \epsilon^k = +\infty,$$

on ne peut rien conclure sur une limite éventuelle de  $u^k$ .<sup>(3)</sup>

En effet, prenons

$$u^k = 1 \text{ lorsque } k = n^2 \text{ et } = \frac{1}{k} \text{ sinon ; } \epsilon^k = \frac{1}{k}$$

alors

$$\sum_{k \in \mathbb{N}} \epsilon^k u^k < \sum_{k \in \mathbb{N}} \left(\frac{1}{k}\right)^2 + \sum_{n \in \mathbb{N}} \left(\frac{1}{n}\right)^2 < +\infty$$

<sup>3</sup>mis à part le fait que  $(u^k)$  admette au moins 0 comme point d'accumulation.



**Lemme 3** *Considérons une suite  $(u^k)_{k \in \mathbf{N}}$  dans  $U^{\text{ad}}$  convexe fermé de  $U$ ,  $U^*$  un sous-ensemble de  $U^{\text{ad}}$  et  $K$  une fonctionnelle, pour lesquels nous faisons les hypothèses suivantes :*

- (i) *tout point d'accumulation de  $(u^k)_{k \in \mathbf{N}}$  dans la topologie faible appartient à  $U^*$*
- (ii)  *$K$  est fortement convexe de  $U$  dans  $\mathbf{R}$  et admet une dérivée continue de  $U$  muni de la topologie faible dans  $U'$  muni de la topologie faible-\**
- (iii) *pour tout  $u^* \in U^*$ , la suite  $\psi^k = K(u^*) - K(u^k) - (K'(u^*), u^* - u^k)$  admet une limite pouvant dépendre de  $u^*$*

*Alors  $(u^k)_{k \in \mathbf{N}}$  converge faiblement vers une limite dans  $U^*$*

La démonstration se trouve dans Cohen [1, p. 276].

**Lemme 4** *Considérons une suite  $(u^k)_{k \in \mathbf{N}}$  à valeurs dans  $U^{\text{ad}}$  convexe fermé de  $U$ ,  $U^* \subset U^{\text{ad}}$ , l'ensemble des solutions de (1.7),  $P^* \subset C'$  l'ensemble des solutions de (1.9) et  $(u^*, p^*) \in U^* \times P^*$  pour lesquels nous faisons les hypothèses suivantes :*

- (i)  $L(u^k, p^*) \rightarrow L(u^*, p^*)$
- (ii)  $L$  est stable et s.c.i. en  $u$ .

*Alors tout point d'accumulation  $\bar{u}$  de la suite appartient à  $U^*$  et  $(\bar{u}, p^*)$  est point-selle de  $L$ .*

## 1.2 Modèle probabiliste

Nous introduisons ici le modèle probabiliste minimal permettant de décrire les algorithmes que nous proposerons dans la suite.

### 1.2.1 Définitions et Propriétés

#### Définition 15

##### Base stochastique

On appelle **base stochastique** un quadruplet  $(\Omega, (\mathcal{F}^k)_{k \in \mathbb{N}}, \mathbf{F}, \mathbf{P})$ , dans lequel  $\Omega$  est un ensemble dont on note  $\tilde{\omega}$  un élément,  $\mathbf{F}$  est une  $\sigma$ -algèbre de  $\Omega$ ,  $\mathbf{P}$  est une loi de probabilité sur  $(\Omega, \mathbf{F})$  et  $\mathbf{F}^k$  une suite croissante (au sens de l'inclusion des ensembles) de sous- $\sigma$ -algèbres de  $\mathbf{F}$ . On supposera en général que  $\mathbf{F}$  est la  $\sigma$ -algèbre engendrée par la réunion des  $\mathbf{F}^k$

##### Processus stochastique

Soit  $(X, \mathcal{B})$  un espace mesurable. Un **processus stochastique** sur l'espace de probabilité  $(\Omega, \mathbf{F}, \mathbf{P})$  d'espace d'état  $(X, \mathcal{B})$  est une application  $X$  de  $\mathbb{N} \times \Omega$  dans  $X$  telle que pour tout  $k \in \mathbb{N}$ , l'application  $\tilde{\omega} \rightarrow X(k, \tilde{\omega})$  est mesurable par rapport à  $\mathbf{F}$ . Les applications  $X(\cdot, \tilde{\omega}) : k \rightarrow X(k, \tilde{\omega})$ ,  $\tilde{\omega} \in \Omega$  sont appelées **trajectoires** du processus.

##### Processus adapté

Soit  $(\Omega, (\mathcal{F}^k)_{k \in \mathbb{N}}, \mathbf{F}, \mathbf{P})$ , une base stochastique et  $X$  un processus stochastique sur  $(\Omega, \mathbf{F}, \mathbf{P})$ , d'espace d'état  $(X, \mathcal{B})$ . On dit que  $X$  est **adapté** à la filtration  $\mathbf{F}$  (ou **non-anticipatif**) si pour tout  $k \in \mathbb{N}$ , l'application  $\tilde{\omega} \rightarrow X(k, \tilde{\omega})$  est  $\mathbf{F}^k$ -mesurable.

#### Définition 16 Convergence d'une suite de variables aléatoires.

Soit  $(X^k)_{k \in \mathbb{N}}$  une suite de variables aléatoires à valeurs dans un Hilbert  $U$ , l'espace probabilisé associé étant  $(\Omega, \mathbf{F}, \mathbf{P})$ .

On dit que la suite converge vers 0

- en moyenne quadratique si  $\lim_{k \rightarrow \infty} \mathbf{E} \|X^k\|^2 = 0$
- en probabilité si  $\forall \epsilon > 0$ ,  $\lim_{k \rightarrow \infty} \mathbf{P}(\|X^k\| > \epsilon) = 0$
- presque-sûrement si  $\forall \epsilon > 0$ ,  $\lim_{k \rightarrow \infty} \mathbf{P}(\sup_{m \geq k} \|X^m\| \geq \epsilon) = 0$

#### Propriété 1 Propriété de l'espérance conditionnelle.

Soient  $\phi_1(X)$  et  $\phi_2(Y)$  des variables aléatoires fonctions respectivement de  $X$  et  $Y$ , on a :

$$\mathbf{E}(\phi_1(X) \cdot \phi_2(Y)) = \mathbf{E}(\phi_1(X) \mathbf{E}(\phi_2(Y)/X))$$

**Définition 17** *Martingale.*

Soit  $(X^k)_{k \in \mathbb{N}}$  un processus aléatoire adapté à une filtration  $\mathbf{F}$ . On dit que la suite est une martingale si

$$\forall k \in \mathbb{N}, \mathbf{E}^k X^{k+1} = X^k$$

**Résultat 13** *Inégalité de Doob.*

Si  $(X^k)_{k \in \mathbb{N}}$  est une martingale, alors

$$\forall \epsilon, P\left(\sup_{1 \leq k \leq m} |X^k| \geq \epsilon\right) \leq \mathbf{E} \frac{|X^m|^2}{\epsilon^2}$$

On en déduit le corollaire (inégalité de Kolmogorov) :

si  $(X^k)_{k \in \mathbb{N}}$  est une suite de variables indépendantes et  $\mathbf{E}X^k = 0$ , alors

$$P\left(\sup_{1 \leq k \leq m} |\chi^1 + \chi^2 + \dots + \chi^k| > \epsilon\right) \leq \frac{1}{\epsilon^2} \sum_k \mathbf{E}(\chi^k)^2$$

**Définition 18** *Propriété caractéristique Définition des quasi-martingales.*

Un processus stochastique  $(X_k)$  adapté à la filtration  $\mathbf{F}$ , tel que :

$$\inf_{k \in \mathbb{N}} \mathbf{E}(X_k) > -\infty$$

et tel que,

$$\sum_{k \in \mathbb{N}} \mathbf{E}(1_{G_k}(X_{k+1} - X_k)) < \infty$$

avec

$$G_k = \{\tilde{\omega} \in \Omega \mid \mathbf{E}^k(X_{k+1} - X_k) > 0\}$$

est appelé une  $(\mathbf{F}, \mathbf{P})$  quasi-martingale.

**Résultat 14** *Convergence p.s. des quasimartingales.*

Soit  $(X^k)_{k \in \mathbb{N}}$  une  $(\mathbf{F}, \mathbf{P})$  - quasi-martingale. Alors  $(X^k)_{k \in \mathbb{N}}$  converge p.s. vers une variable aléatoire mesurable  $X^\infty$  telle que

$$\mathbf{E}|X^\infty| \leq \liminf_k \mathbf{E}|X^k| < \infty$$

**1.2.2 Modèle issu des algorithmes stochastiques**

Tous les algorithmes étudiés dans ce mémoire contiennent des itérations du type  $\omega^{k+1} = A(\omega^k, \omega^{k+1})$ ,  $\omega^{k+1}$  étant un tirage suivant la loi  $P$ , indépendant des tirages précédents.

Le modèle probabiliste construit par ces tirages successifs est

$$(\Omega, \mathbf{F}, \mathbf{P}) = (\Omega, \mathcal{A}, P)^{\otimes \mathbb{N}}$$

Un élément  $\tilde{\omega}$  de  $\Omega$  est

$$\tilde{\omega} = (\omega^1, \dots, \omega^k, \dots)$$

A tout  $\tilde{\omega}$  correspond donc une réalisation de l'algorithme obtenue par la suite des  $\omega^k$ , tirages indépendants de loi  $P$ . La suite  $u^k$  construite par l'algorithme est donc à considérer comme un processus stochastique sur  $(\Omega, \mathbf{F}, \mathbf{P})$ , d'espace d'état  $(U, \mathcal{B}_U)$ .

On note  $\mathbf{F}^k$  la sous-tribu engendrée par les tirages  $\omega^1, \dots, \omega^k$  et  $\mathbf{E}^k = \mathbf{E}(\dots / \mathbf{F}^k)$ , l'espérance conditionnelle associée. Avec ces définitions, on peut prendre pour  $\mathbf{F}$  la tribu engendrée par  $\bigcup_{k \in \mathbf{N}} \mathbf{F}^k$

Soient  $g$  et  $G$  deux fonctions mesurables telles que

$$\int_{\Omega} g(u, \omega) P(d\omega) = G(u)$$

En appliquant les règles de l'espérance conditionnelle, on aura par exemple

$$\mathbf{E}^k(g(u^k, \omega^{k+1})) = G(u^k)$$

Les résultats de convergence presque-sûre ou de convergence en moyenne quadratique qui seront démontrés plus loin seront donc pris au sens de la loi de probabilité  $\mathbf{P}$ . Le point de vue intuitif de la convergence  $\mathbf{P}$ -presque sûre de la suite engendrée par l'algorithme stochastique sera le suivant : "presque toutes les fois que l'on appliquera l'algorithme, la suite de valeurs calculées convergera vers la solution"

### 1.3 Principe du Problème Auxiliaire

Avec la terminologie de l'introduction, nous exposerons ce principe dans le cas déterministe sous-différentiable pour des contraintes implicites du type  $u \in U^{ad}$ . Pour les autres cas, nous renvoyons à Cohen [1] ou aux chapitres suivants qui contiennent quelques extensions.

Basé sur une méthode de point fixe, le principe du problème auxiliaire consiste à remplacer le problème

$$\min_{u \in U^{ad}} J(u) \quad (1.10)$$

avec  $U^{ad}$  convexe fermé d'un Hilbert  $U$  et  $J$  fonctionnelle à valeurs réelles, sous-différentiable et convexe, par la séquence de problèmes auxiliaires suivante :

$$\left[ \min_{u \in U^{ad}} K(u) + (\epsilon^k r^k - K'(u^k), u) \rightarrow u^{k+1} \right] \quad (1.11)$$

$K$  étant une fonctionnelle fortement convexe choisie par le concepteur d'algorithmes et  $r^k$  étant un élément du sous-différentiel de  $J$  en  $u^k$

Avec des hypothèses adéquates, la suite  $(u^k)_{k \in \mathbf{N}}$  ainsi engendrée converge faiblement vers la solution  $u^*$  du problème (1.10).

Après avoir rappelé un résultat de Cohen [1], nous donnerons une interprétation de l'algorithme et de la technique de démonstration utilisée, dans le cas où le  $K$  choisi est une fonctionnelle quadratique.

On considère  $J$  une fonctionnelle convexe s.c.i., sur un espace de Hilbert  $U$ , à valeurs dans  $\mathbf{R}$ . On cherche à réaliser le programme (1.10). On appelle  $U^*$ , l'ensemble supposé non vide, des solutions et on suppose que le convexe fermé  $U^{ad}$  est décomposable sous la forme :

$$U^{ad} = U_1^{ad} \times \dots \times U_N^{ad}$$

Afin d'éclater ce problème en  $N$  problèmes avec  $u = (u_1, \dots, u_N)$ , on choisit une fonctionnelle  $K : U \rightarrow \mathbf{R}$ , appelée critère auxiliaire, **additive par rapport à la décomposition des variables de commande**, c.a.d :

$$K(u) = \sum_{i=1}^{i=N} K_i(u_i)$$

L'algorithme est le suivant :

**Algorithme 1**

- soit  $u^0 \in U^{ad}$ , soit la suite de réels positifs  $\epsilon^k$
- connaissant  $u^k = (u_1^k, \dots, u_N^k)$  on calcule  $u^{k+1}$  par

$$\min_{u \in U^{ad}} K(u) + (\epsilon^k r^k - K'(u^k), u) \quad (1.12)$$

avec  $r^k \in \partial J(u^k)$ .

Cet algorithme est décomposé et revient à résoudre éventuellement en parallèle

$$\min_{u_i \in U_i^{ad}} K_i(u_i) + (\epsilon^k r_i^k - K'_i(u_i^k), u_i), \quad i = 1, \dots, N \quad (1.13)$$

**Théorème 1** On suppose :

- $J$  convexe, propre, sci, coercive sur  $U^{ad}$
- $K$  est fortement convexe, différentiable, de  $U$  dans  $\mathbf{R}$ .

Alors, la solution de (1.12) existe et est unique. On la note  $u^{k+1}$

Si de plus,

- $J$  est  $SGLB(c_1, c_2)(^a)$  et  $(\epsilon^k)_{k \in \mathbf{N}}$  est une  $\sigma$ -suite,
- ou si  $\epsilon^k = \rho^k / \|r^k\|$ , avec  $\rho^k$   $\sigma$ -suite

Alors :

$$\lim_{k \rightarrow \infty} J(u^k) = J(u^*),$$

la suite  $(u^k)_{k \in \mathbf{N}}$  est bornée et tout point d'accumulation appartient à  $U^*$

Enfin, si  $J$  est fortement convexe, alors :

$$U^* = \{u^*\} \quad \text{et} \quad \lim_{k \rightarrow \infty} \|u^k - u^*\| = 0$$

<sup>a</sup>voir la définition 7

La démonstration se trouve dans Cohen [1] p.145 et suivantes.

### 1.3.1 Quelques réflexions sur le cas $K$ quadratique

Nous proposons ici une interprétation de cet algorithme et de la technique de démonstration utilisée pour prouver la convergence dans le cas où  $K$  est quadratique.

#### C'est presque une méthode Quasi-Newton

L'inéquation variationnelle associée à une itération (1.11) est :

$$(K'(u^{k+1}) - K'(u^k) + \epsilon^k r^k, u - u^{k+1}) \geq 0, \quad \forall u \in U^{\text{ad}}$$

Notamment, si  $U^{\text{ad}} = U$ , cette inéquation se réduit à l'équation :

$$K'(u^{k+1}) - K'(u^k) + \epsilon^k r^k = 0$$

Si l'on suppose  $K$  quadratique,

$$K(u) = \frac{1}{2}(\mathbf{K}u, u)$$

ce qui est possible, sans une grande perte de généralité, puisque l'utilisateur peut choisir cette fonctionnelle, la condition d'optimalité devient :

$$\mathbf{K} \cdot (u^{k+1} - u^k) + \epsilon^k r^k = 0$$

Comme  $\mathbf{K}$  est inversible si  $K$  est fortement convexe,

$$u^{k+1} = u^k - \epsilon^k \mathbf{K}^{-1} \cdot r^k \tag{1.14}$$

itération qui n'est pas sans rappeler les algorithmes de type Quasi-Newton. Si  $J$  admet une dérivée seconde,  $\mathbf{K}$  peut, par exemple être une approximation bloc-diagonale du Hessian de  $J$ .<sup>(4)</sup>

#### La fonction de Lyapunov $\psi$ est une distance particulière

La fonction de Lyapunov introduite par Cohen pour prouver la convergence de ce type d'algorithmes est la suivante :

$$\psi^k = K(u^*) - K(u^k) - (K'(u^k), u^* - u^k)$$

Si  $K$  est quadratique,  $\psi$  se simplifie en :

$$\psi^k = \frac{1}{2}(\mathbf{K} \cdot (u^* - u^k), u^* - u^k) = \frac{1}{2} \|u^* - u^k\|_{\mathbf{K}}^2$$

ce qui,  $\mathbf{K}$  étant un opérateur défini positif, définit une distance issue de la norme induite par  $\mathbf{K}$ , équivalente à la norme euclidienne.

<sup>4</sup>alors  $K$  dépendra de l'indice  $k$ . Ce cas est traité dans Cohen [1, p. 109 et suiv.]

La technique de démonstration est alors la suivante . après avoir retranché  $u^*$  de chaque membre de la condition d'optimalité (1.14), élevé au carré, et développé,

$$\|u^{k+1} - u^*\|_K^2 = \|u^k - u^*\|_K^2 + (\epsilon^k)^2 \|K^{-1}.r^k\|_K^2 - 2\epsilon^k(u^k - u^*, r^k)$$

c'est-à-dire, en termes de  $\psi$  :

$$\psi^{k+1} = \psi^k + \frac{(\epsilon^k)^2}{2} \|K^{-1}.r^k\|_K^2 - \epsilon^k(u^k - u^*, r^k)$$

L'hypothèse de convexité de  $J$  permet d'écrire :

$$\psi^{k+1} \leq \psi^k + \frac{(\epsilon^k)^2}{2} \|K^{-1}.r^k\|_K^2 + \epsilon^k(J(u^*) - J(u^k))$$

$u^*$  étant solution de (1.10), le dernier terme de l'inégalité précédente est négatif.

Comme  $J$  est  $SGLB(c_1, c_2)^{(5)}$ ,

$$\|r^k\| \leq c_1 \|u^k - u^*\| + c_2, \quad c_i > 0$$

et  $K$  est définie positive, il existe des constantes positives  $m_i$  telles que

$$\psi^{k+1} \leq \psi^k + (\epsilon^k)^2(m_1\psi^k + m_2) + \epsilon^k(J(u^*) - J(u^k))$$

D'après le lemme 1, la suite  $\psi^k$  est donc bornée.

Pour conclure,

$$0 \leq \sum_{k=1}^m \epsilon^k(J(u^k) - J(u^*)) \leq \psi(u^1) - \psi(u^{m+1}) + \sum_{k=1}^m (\epsilon^k)^2(m_1\psi^k + m_2) < \infty$$

car  $(\epsilon^k)^2$  est le terme d'une série convergente.

Et d'après (1.14), comme  $\psi^k$  est bornée (et donc  $\|u^k - u^*\|_K$  aussi ainsi que  $\|r^k\|$ ),

$$\|u^{k+1} - u^k\| \leq \epsilon^k.m_3$$

Comme  $J$  est lipchitzienne sur le borné contenant la suite  $(u^k)_{k \in \mathbf{N}}$ , le lemme 2 montre enfin que

$$\lim_k J(u^k) = J(u^*)$$

$J$  étant s.c.i, on déduit que tout point d'accumulation faible de la suite  $(u^k)$  est solution du problème (1.10). Enfin, puisque  $K$  est quadratique, et donc  $K'$  est linéaire, on peut appliquer le lemme 3 et conclure que  $(u^k)_{k \in \mathbf{N}}$  tend faiblement vers  $u^*$

Dans le cas où  $U^{ad} \neq U$ ,  $u^{k+1}$  est caractérisé par

$$u^{k+1} = \Pi_{U^{ad}}^K(u^k - \epsilon^k K^{-1}.r^k) \quad (1.15)$$

$\Pi_{U^{ad}}^K$  étant la projection sur  $U^{ad}$  au sens de la norme induite par  $K$ .

La démonstration de convergence est alors exactement la même car  $\Pi_{U^{ad}}^K$  est non expansive et  $u^* \in U^{ad}$

<sup>5</sup> hypothèse exprimée par rapport à  $u^*$



## Chapitre 2

# Contraintes Déterministes Découplées

Nous nous limitons ici aux problèmes sans contraintes ou avec contraintes déterministes implicites découplées, problèmes de la forme :

$$\min_{u \in U^{ad}} \int_{\Omega} j(u, \omega) P(d\omega) = J(u) \quad (2.1)$$

On appelle  $U^*$ , l'ensemble supposé non vide des solutions du problème (2.1) et on note  $u^*$  un des ses éléments.

On s'inspire simultanément des algorithmes de gradient stochastique et de la méthode du problème auxiliaire pour proposer l'algorithme suivant :

### 2.1 Algorithme de décomposition stochastique

#### Algorithme 2

- soit  $u^0 \in U^{ad}$
- connaissant  $u^k = (u_1^k, \dots, u_N^k)$ , on tire suivant la loi  $P$ , un  $\omega^{k+1} \in \Omega$  indépendant des tirages précédents, et on calcule  $u^{k+1}$  par

$$\min_{u \in U^{ad}} K(u) + (\epsilon^k r^k - K'(u^k), u) \quad (2.2)$$

avec  $r^k$  appartenant au sous-différentiel de  $j$  au point  $(u^k, \omega^{k+1})$ .

Si  $K$  est additive, cet algorithme est décomposé et revient à résoudre en parallèle

$$\min_{u_i \in U_i^{ad}} K_i(u_i) + (\epsilon^k r_i^k - K'_i(u_i^k), u_i), \quad i = 1, \dots, N \quad (2.3)$$

**Théorème 2** On suppose :

1.  $j(\cdot, \omega)$  convexe, propre, sci, coercive sur  $U^{ad}$ , pour tout  $\omega \in \Omega$  ;  $j(u, \cdot)$  mesurable sur  $\Omega$ , pour tout  $u \in U^{ad}$
2.  $K : U \rightarrow \mathbf{R}$  fortement convexe de rapport  $b > 0$ .

Alors, la solution de (2.2) existe et est unique. On la note  $u^{k+1}$

Si de plus, pour tout  $\omega$ ,  $j(\cdot, \omega)$  est à sous-gradient linéairement borné de paramètres  $c_1$  et  $c_2$  et si  $(\epsilon^k)_{k \in \mathbf{N}}$  est une  $\sigma$ -suite

Alors :

$$\lim_{k \rightarrow \infty} J(u^k) = J(u^*)$$

p.s. au sens de  $\mathbf{P}$  sur  $\Omega$ , la suite  $(u^k)_{k \in \mathbf{N}}$  est bornée p.s. et tout point d'accumulation appartient à  $U^*$ . Si  $K$  a une dérivée continue dans  $U'$  muni de la topologie faible-\*,  $(u^k)_{k \in \mathbf{N}}$  converge faiblement p.s. vers un point de  $U^*$

Enfin, si  $J$  est fortement convexe, alors :

$$U^* = \{u^*\} \text{ et } \lim_{k \rightarrow \infty} \|u^k - u^*\| = 0 \text{ (P - p.s.)}$$

<sup>a</sup>Voir les hypothèses du lemme 3

## 2.2 Convergence de l'algorithme

### 2.2.1 Inéquations variationnelles et fonction de Lyapunov

L'inéquation variationnelle associée au problème (2.2), est :

$$\forall v \in U^{ad}, (K'(u^{k+1}) - K'(u^k) + \epsilon^k r^k, v - u^{k+1}) \geq 0 \quad (2.4)$$

Celle associée au problème (2.1) est :

$$\forall v \in U^{ad}, \forall \bar{v} \in \partial J(v), (\bar{v}, v - u^*) \geq 0$$

On définit la fonction de Lyapunov  $\psi^k$  par

$$\psi(u^k) = K(u^*) - K(u^k) - (K'(u^k), u^* - u^k)$$

Nous noterons parfois pour les besoins de la démonstration  $\psi^k$  de la façon plus explicite suivante :  $\psi(u^k(\bar{\omega}))$ . Les propriétés de  $\psi^k$  nous permettrons de démontrer le résultat annoncé. Elles proviennent directement du lemme suivant.

## 2.2.2 Majorations préliminaires

**Lemme 5** *D'après la définition de  $\psi^k$ , la convexité de rapport  $b$  de  $K$  et les inéquations variationnelles caractérisant  $u^{k+1}$  et  $u^*$ , nous avons :*

$$\psi^k \geq \frac{b}{2} \|u^k - u^*\|^2 \quad (2.5)$$

il existe des constantes positives  $c'_1$  et  $c'_2$  telles que

$$\|u^{k+1} - u^k\| \leq \epsilon^k (c'_1 \sqrt{\psi^k} + c'_2) \quad (2.6)$$

et des constantes positives  $c''_1$  et  $c''_2$  telles que

$$\psi^{k+1} - \psi^k \leq \epsilon^k (r^k, u^* - u^k) + (\epsilon^k)^2 (c''_1 \psi^k + c''_2) \quad (2.7)$$

En effet, d'une part, la forte convexité de rapport  $b$  de  $K$  implique (2.5) :

$$\psi(u) = K(u^*) - K(u) - (K'(u), u^* - u) \geq \frac{b}{2} \|u^* - u\|^2$$

D'autre part,  $\psi^{k+1} - \psi^k$  s'écrit

$$\underbrace{K(u^k) - K(u^{k+1}) - (K'(u^k), u^k - u^{k+1})}_{A_1} + \underbrace{(K'(u^k) - K'(u^{k+1}), u^* - u^{k+1})}_{A_2}$$

$A_1$  est négatif à cause de la convexité de  $K$

D'après l'inéquation variationnelle (2.4) appliquée en  $v = u^*$ ,

$$(K'(u^{k+1}) - K'(u^k), u^* - u^{k+1}) \geq 0$$

on a

$$\psi^{k+1} - \psi^k \leq A_2 \leq (\epsilon^k r^k, u^* - u^{k+1}) = \underbrace{(\epsilon^k r^k, u^* - u^k)}_{B_1} + \underbrace{(\epsilon^k r^k, u^k - u^{k+1})}_{B_2}$$

Il reste donc à majorer  $B_2$  pour avoir (2.7). En appliquant l'inéquation variationnelle (2.4) à  $v = u^k$  et l'inégalité de Schwarz,

$$\overbrace{\epsilon^k \|r^k\| \|u^{k+1} - u^k\|}^{\text{Schwarz}} \geq B_2 \geq \underbrace{(K'(u^k) - K'(u^{k+1}), u^k - u^{k+1})}_{\text{Ineq. var.}} \geq \overbrace{b \|u^k - u^{k+1}\|^2}^{\text{Forte monotonie}}$$

en divisant les membres extrêmes de cette inégalité par  $\|u^{k+1} - u^k\|$  (supposé non nul sans perte de généralité),

$$\|u^k - u^{k+1}\| \leq \epsilon^k \|r^k\| / b$$

par hypothèse *SGLB*, et (2.5), il existe des constantes positives  $c'_i$  telles que

$$\|u^{k+1} - u^k\| \leq \epsilon^k (c_1' \sqrt{\psi^k} + c_2)$$

Nous pouvons maintenant majorer  $B_2$  :

$$B_2 = (\epsilon^k r^k, u^k - u^{k+1}) \leq (\epsilon^k)^2 \|r^k\|^2 / b$$

En utilisant  $(a + b)^2 \leq 2a^2 + 2b^2$ , il existe des constantes  $c_1''$  telles que

$$B_2 \leq (\epsilon^k)^2 (c_1'' \psi^k + c_2'')$$

En définitive, en "recollant les morceaux", (2.7) est prouvé.

### 2.2.3 Convergence de la suite $\psi^k$

$y^k := \mathbf{E}(\psi^k)$  est une suite positive bornée.

Prenons l'espérance conditionnelle sachant  $\mathbf{F}^k$  de (2.7). D'après les considérations du chapitre 1,

$$\mathbf{E}^k \psi^{k+1} - \psi^k \leq \epsilon^k (\mathbf{E}^k r^k, u^* - u^k) + (\epsilon^k)^2 (c_1'' \psi^k + c_2'')$$

On note

$$\bar{r}^k := \mathbf{E}^k r^k = \int_{\Omega} r(u^k, \omega) P(d\omega)$$

$\bar{r}^k$  est un élément de  $\partial J(u^k)$  et donc

$$(\bar{r}^k, u^* - u^k) \leq J(u^*) - J(u^k)$$

soit

$$\mathbf{E}^k \psi^{k+1} - \psi^k \leq \epsilon^k (J(u^*) - J(u^k)) + (\epsilon^k)^2 (c_1'' \psi^k + c_2'') \quad (2.8)$$

Comme par définition,  $J(u^*) \leq J(u^k)$ ,

$$\mathbf{E}^k \psi^{k+1} - \psi^k \leq \alpha_k \psi^k + \beta_k \quad (2.9)$$

avec  $\alpha_k$  et  $\beta_k$  deux suites positives sommables ( $(\epsilon^k)_{k \in \mathbf{N}}$  est une  $\sigma$ -suite).

En prenant maintenant l'espérance au sens de  $\mathbf{P}$ ,

$$\forall k \in \mathbf{N}, \quad y^{k+1} - y^k \leq \alpha_k y^k + \beta_k$$

et d'après le lemme 1,  $(y^k)_{k \in \mathbf{N}}$  est une suite bornée. On note  $M$  l'un de ses majorants.

$\psi^k$  converge p.s. au sens de  $\mathbf{P}$

$\psi^k$  a la propriété caractéristique des quasi-martingales (définition propriété 18). En effet,  $(\psi^k)_{k \in \mathbf{N}}$  vérifie  $\inf_{k \in \mathbf{N}} \mathbf{E}(\psi^k) > -\infty$  ( $y^k$  est positive) et d'après (2.8)

$$\mathbf{E}^k(\psi^{k+1} - \psi^k) \leq \alpha_k \psi^k + \beta_k$$

on déduit, comme  $y^k$  est bornée par  $M$  :

$$\sum_{k \in \mathbf{N}} \mathbf{E}(\mathbf{E}^k \psi^{k+1} - \psi^k)^+ \leq \sum_{k \in \mathbf{N}} (\alpha_k y^k + \beta_k) \leq \sum_{k \in \mathbf{N}} (M \alpha_k + \beta_k) < \infty$$

On a utilisé la propriété 1 (espérance conditionnelle) pour écrire :

$$\mathbf{E}(1_{G_k}(\psi^{k+1} - \psi^k)) = \mathbf{E}(1_{G_k} \mathbf{E}^k(\psi^{k+1} - \psi^k))$$

Donc  $\psi^k$  est une quasi-martingale. D'après le résultat 14, elle converge presque-sûrement vers une variable aléatoire d'espérance finie.

### 2.2.4 Limites de $J(u^k)$ et de $u^k$

$(u^k)_{k \in \mathbf{N}}$  admet p.s. des points d'accumulation faible.

Comme  $\psi^k$  converge p.s., vers une variable aléatoire d'espérance finie, il existe un ensemble  $\Omega_0$  de mesure nulle tel que

$$\forall \tilde{\omega} \in \Omega \setminus \Omega_0, \lim_{k \rightarrow \infty} \psi(u^k(\tilde{\omega})) < \infty$$

et donc en vertu de (2.5),

$$\|u^k - u^*\|^2 \leq \frac{2}{b} \sup_k \psi^k < \infty$$

Soit un tel  $\tilde{\omega}$ , la suite  $u^k(\tilde{\omega})$  est donc bornée et admet d'après le théorème de compacité faible des points d'accumulation faible dans  $U^{sd}$

$(J(u^k))_{k \in \mathbf{N}}$  converge p.s. vers  $J(u^*)$ .

Ce résultat est une application du lemme 2 qui est déterministe.

L'équation (2.8) se réécrit

$$\epsilon^k (J(u^k) - J(u^*)) \leq \psi^k (1 + \alpha_k) - \mathbf{E}^k(\psi^{k+1}) + \beta_k$$

en prenant l'espérance au sens de  $\mathbf{P}$ ,

$$\epsilon^k \mathbf{E}(J(u^k) - J(u^*)) \leq y^k (1 + \alpha_k) - y^{k+1} + \beta_k$$

soit par sommation sur  $k$  :

$$\sum_{k=1}^n \epsilon^k \mathbf{E}(J(u^k) - J(u^*)) \leq \sum_{k=1}^n (y^k (1 + \alpha_k) - y^{k+1} + \beta_k)$$

soit, comme  $y^k$  est bornée par un réel  $M$ , en réarrangeant les termes :

$$\sum_{k=1}^n \epsilon^k \mathbf{E}(J(u^k) - J(u^*)) \leq \sum_{k=1}^n (M \alpha_k + \beta_k) + y^1 - y^{n+1} \leq M' < \infty$$

Les termes  $J(u^k) - J(u^*)$  étant positifs, cela implique que :

$$\sum_{k \in \mathbf{N}} \epsilon^k (J(u^k) - J(u^*)) < \infty \quad \mathbf{P} - p.s.$$

On appellera  $\Omega_1$  l'ensemble de mesure nulle sur lequel cette majoration n'est pas vérifiée.

D'autre part,  $\psi^k$  bornée p.s. implique, d'après (2.6)

$$\|u^{k+1} - u^k\| \leq \epsilon^k M'$$

presque-sûrement (à  $\Omega_0$  près). Pour tout  $\tilde{\omega}$  n'appartenant ni à  $\Omega_0$  ni à  $\Omega_1$ , les conditions du lemme 2 sont vérifiées (en particulier,  $J$  est lipchitzienne sur le borné contenant  $(u^k)_{k \in \mathbf{N}}$ ) : pour cet  $\tilde{\omega}$ , la suite  $J(u^k(\tilde{\omega}))$  converge vers  $J(u^*)$ .

**Tout point d'accumulation faible de la suite  $(u^k(\tilde{\omega}))_{k \in \mathbf{N}}$  appartient à  $U^*$ .** Pour  $\tilde{\omega} \in \Omega \setminus \Omega_0 \cup \Omega_1$  (1), on note  $\tilde{u}(\tilde{\omega})$  un point d'accumulation faible de  $(u^k)$ , et  $(u^{\phi(k)})$  une sous-suite convergeant vers  $\tilde{u}$ .  $\tilde{u} \in U^{ad}$  car  $U^{ad}$  est convexe fermé donc faiblement fermé. Par faible semi-continuité inférieure de  $J$ ,

$$J(\tilde{u}) \leq \liminf_{k \rightarrow \infty} J(u^{\phi(k)}) = J(u^*)$$

et donc  $\tilde{u}(\tilde{\omega}) \in U^*$

**$(u^k)_{k \in \mathbf{N}}$  converge faiblement P-p.s.**

Pour cela, nous utilisons le lemme 3, en supposant que  $K$  admet une dérivée faiblement continue de  $U$  muni de la topologie faible, dans  $U'$  muni de la topologie faible.\* Les trois conditions du lemme 3 sont donc satisfaites.

Le cas où  $J$  est fortement convexe permet de conclure la démonstration du théorème. Nous l'envisageons dans la section suivante.

\* ces ensembles pouvant dépendre de  $u^*$

### 2.3 Cas particulier : J fortement convexe.

Si  $J$  est fortement convexe de rapport  $\alpha > 0$ , on a (résultat 4)

$$\forall u \in U, \quad J(u) - J(u^*) \geq \frac{\alpha}{2} \|u - u^*\|^2$$

Or on a montré que :

$$\forall \tilde{\omega} \in \Omega \setminus \Omega_0 \cup \Omega_1, \quad \lim_{k \rightarrow \infty} J(u^k(\tilde{\omega})) = J(u^*)$$

donc

$$\lim_{k \rightarrow \infty} \|u^k(\tilde{\omega}) - u^*\|^2 = 0$$

Le théorème est entièrement démontré.

#### Une variante intéressante de l'algorithme 2 :

$J$  étant fortement convexe, on peut s'intéresser à une variante de l'algorithme précédent qui consiste à effectuer de "grands pas" dans l'évaluation de la suite  $u^k$ . En fait, ce nouvel algorithme ne converge pas p.s. mais en **moyenne quadratique vers une boule** aussi petite que l'on veut, centrée en  $u^*$ .

Nous le présenterons pour un  $K$  quelconque mais nous démontrerons la convergence dans le cas simple où  $K(u) = \frac{1}{2}\|u\|^2$  et  $U^{ad} = U$ , la généralisation étant immédiate.

L'intérêt de cet algorithme est de donner rapidement une bonne estimation de  $u^*$ , estimation que l'on peut prendre comme point de départ d'un algorithme à "petits pas"  $\epsilon^k$ . Cette idée se trouve déjà dans Bertran[1].

#### Algorithme 3

- soit  $u^0 \in U^{ad}$
- connaissant  $u^k = (u_1^k, \dots, u_N^k)$ , on tire suivant la loi  $P$ , un  $\omega^{k+1} \in \Omega$  indépendant des tirages précédents, et on calcule  $u^{k+1}$  par

$$\min_{u \in U^{ad}} K(u) + (\epsilon r^k - K'(u^k), u) \quad (2.10)$$

avec  $r^k$  appartenant au sous-différentiel de  $j$  au point  $(u^k, \omega^{k+1})$ .

Si  $K(u) = \frac{1}{2}\|u\|^2$  et  $U^{ad} = U$ , alors  $u^{k+1} = u^k - \epsilon r^k$ . Avec  $\psi^k = \frac{1}{2}\|u^k - u^*\|^2$ , on vérifie immédiatement que  $\epsilon$

$$\psi^{k+1} \leq \psi^k + \epsilon^2(c_1 + c_2\psi^k) - \epsilon(r^k, u^k - u^*)$$

L'espérance conditionnelle sachant  $\mathbf{F}^k$  conduit donc (après avoir introduit  $\mathbf{r}^k \in \partial J(u^k)$  et exploité la définition de sous-différentiabilité) à

$$\mathbf{E}^k \psi^{k+1} \leq \psi^k + \epsilon^2(c_1 + c_2 \psi^k) - \epsilon(J(u^k) - J(u^*))$$

Comme  $J$  est fortement convexe, on a

$$\mathbf{E}^k \psi^{k+1} \leq \psi^k + \epsilon^2(c_1 + c_2 \psi^k) - \epsilon a \psi^k$$

La suite  $y^k := \mathbf{E} \psi^k = \frac{1}{2} \mathbf{E} \|u^k - u^*\|^2$ , est donc telle que

$$y^{k+1} \leq y^k(1 - \beta) + \gamma, \quad \beta = \epsilon a - \epsilon^2 c_2, \quad \gamma = \epsilon^2 c_1$$

En ayant supposé  $\epsilon$  suffisamment petit pour que  $\beta$  soit strictement positif et en exploitant récursivement cette inégalité, on a immédiatement

$$0 \leq y^k \leq [y^0 - \frac{\gamma}{\beta}](1 - \beta)^k + \frac{\gamma}{\beta}$$

D'où l'on déduit le résultat annoncé.

On remarquera que la vitesse de convergence vers la boule

$$B(u^*, \frac{\epsilon c_1}{a - \epsilon c_2})$$

est géométrique. D'un point de vue pratique, plus  $\epsilon$  sera petit, plus l'algorithme sera lent, mais meilleure sera l'estimation...

Nous comparerons cette variante avec l'algorithme principal dans la section "tests numériques" de ce chapitre.

## 2.4 Application à la méthode des moindres carrés

On peut envisager de choisir un opérateur auxiliaire  $K$  qui dépende de l'itération. Cela permet, en effet, d'introduire des algorithmes de quasi-Newton stochastiques qui auront un meilleur comportement pour des problèmes test comme celui de Syski et Ruszczyński[1] <sup>(2)</sup>. **L'évolution de  $K^k$  doit néanmoins, en un certain sens, être plus lente que celle imposée par les tirages successifs indépendants  $\omega^{k+1}$**  Voici une illustration simple de ce propos <sup>(3)</sup>, fondée sur un argument d'équation différentielle (ou o.d.e. pour "ordinary differential equation") moyenne <sup>(4)</sup> :

Supposons que  $j(u, \omega)$  est la fonctionnelle quadratique

$$j(u, \omega) = \frac{1}{2} \langle u, A(\omega).u \rangle + \langle b(\omega), u \rangle$$

<sup>2</sup>voir les sections "tests numériques"

<sup>3</sup>due à Cohen

<sup>4</sup>voir section suivante les résultats de Kushner & Clark et de Métivier & Priouret



$A$  et  $b$  étant respectivement une matrice définie positive et un vecteur aléatoires. La solution du problème de minimisation associé est

$$u^* = E[A]^{-1}E[b]$$

En adoptant comme critère auxiliaire la fonctionnelle  $K(u) = \frac{1}{2}(u, B(\omega)u)$  avec  $B(\omega)$  matrice bloc-diagonale extraite directement de  $A(\omega)$ , la solution du problème e auxiliaire est

$$u^{k+1} = u^k - \epsilon^k B(\omega^{k+1})^{-1} [A(\omega^{k+1})u^k + b(\omega^{k+1})]$$

Le résultat d'o.d.e. énoncé à la section suivante (Kuhner & Clark[1] ou Ljung[1]), nous apprend que le point d'équilibre éventuel de cette récurrence ne peut être autre que la solution de

$$E[B(\omega)^{-1}(A(\omega)u + b(\omega))] = 0$$

Or cette solution, compte tenu du produit de probabilités non indépendantes n'est pas en général égale à  $u^*$

Pourtant, il existe des algorithmes de moindres carrés qui, comme le fait remarquer Ljung[2] peuvent être vus comme des algorithmes de Newton stochastiques. En le suivant, supposons maintenant que  $j(u, \omega)$  est  $C^2$  et remarquons qu'un algorithme de Newton stochastique peut s'écrire

$$u^{k+1} = u^k - \epsilon^k [M^k]^{-1} j'_u(u^k, \omega^{k+1}) \quad (2.11)$$

avec  $M^k$  obtenu par l'approximation stochastique du problème

$$\text{Trouver } M \text{ tel que } M = E[j''(u^*, \omega)]$$

c'est-à-dire

$$M^{k+1} = M^k + \epsilon^k [j''(u^k, \omega^{k+1}) - M^k] \quad (2.12)$$

Alors, l'o.d.e. associée au système (2.11), (2.12) est

$$\dot{u} = M^{-1} E[j'_u(u, \omega)], \quad \dot{M} = E[j''_u(u, \omega)] - M$$

et admet comme points stationnaires  $\bar{u}$  et  $\bar{M}$  tels que

$$E[j'(\bar{u}, \omega)] = 0, \text{ et } \bar{M} = E[j''_u(\bar{u}, \omega)]$$

qui sont bien solution du problème de départ.

Le critère auxiliaire serait, dans ce cas,  $K^k(u) = \frac{1}{2}(M^k u, u)$  et vérifierait

$$K^{k+1}(u) = (1 - \epsilon^k) K^k(u) + \epsilon^k (j''_k, u, u)$$

avec  $j''_k$  une approximation par exemple bloc-diagonale de  $j''_{uu}(u^k, \omega^{k+1})$ . Pour  $M^k$ , cela donnerait

$$M^{k+1} = M^k - \epsilon^k [M^k - j''_k]$$

## 2.5 Variantes de l'algorithme principal

Nous donnons ici deux variantes destinées à exploiter la forme particulière du problème de départ.

### 2.5.1 $J$ est de la forme $J_1 + J_2$ , avec $J_2$ additive

On suppose que  $J_2$  est additive par rapport à la décomposition des variables de commande.

$$J = J_1 + J_2 \text{ avec } j_2(u, \omega) = \sum_{i=1}^N j_{2,i}(u_i, \omega)$$

On suppose que  $j_1$  et  $j_2$  satisfont aux mêmes hypothèses que  $j(\cdot, \cdot)$  dans le théorème 2. La variante de l'algorithme 2 est la suivante :

**Algorithme 4** Afin de préserver la décomposition de  $J_2$ , on remplace le problème auxiliaire (2.2) par

$$\min_{u \in U^{ad}} K(u) + \epsilon^k j_2(u, \omega^{k+1}) + (\epsilon^k r_1^k - K'(u^k), u)$$

avec  $r_1^k$  appartenant au sous-différentiel de  $j_1$  au point  $(u^k, \omega^{k+1})$ .

#### Adaptation de la preuve.

Nous notons  $j = j_1 + j_2$  et nous utilisons la même fonction de Lyapunov  $\psi(u^k)$ . L'inéquation variationnelle associée au nouveau problème auxiliaire est, par convexité de  $j_2$ ,

$$\forall v \in U^{ad}, (K'(u^{k+1}) - K'(u^k) + \epsilon^k r_1^k, v - u^{k+1}) + \epsilon^k [j_2(v, \omega^{k+1}) - j_2(u^{k+1}, \omega^{k+1})] \geq 0$$

Le lemme de majoration 5 reste vrai à l'exception de (2.7) qui devient :

Il existe des constantes positives  $c_1, c_2$  et  $c_3$  telles que

$$\psi^{k+1} - \psi^k \leq \epsilon^k (j(u^*, \omega^{k+1}) - j(u^k, \omega^{k+1})) + (\epsilon^k)^2 (c_1 \psi^k + c_2 \psi^{k+1} + c_3)$$

En effet,  $\psi^{k+1} - \psi^k$  s'écrit

$$\underbrace{K(u^k) - K(u^{k+1}) - (K'(u^k), u^k - u^{k+1})}_{A_1} + \underbrace{(K'(u^k) - K'(u^{k+1}), u^* - u^{k+1})}_{A_2}$$

$A_1$  est négatif à cause de la convexité de  $K$

D'après l'inéquation variationnelle appliquée en  $v = u^*$ ,

$$\psi^{k+1} - \psi^k \leq A_2 \leq \epsilon^k [(r_1^k, u^* - u^{k+1}) + j_2(u^*, \omega^{k+1}) - j_2(u^{k+1}, \omega^{k+1})]$$

par convexité de  $j_1$ ,

$$\psi^{k+1} - \psi^k \leq \epsilon^k (j(u^*, \omega^{k+1}) - j(u^k, \omega^{k+1})) + \epsilon^k [(r_1^k, u^k - u^{k+1}) + j_2(u^k, \omega^{k+1}) - j_2(u^{k+1}, \omega^{k+1})]$$

D'après le résultat 5, comme  $j_2$  est *SGLB*,

$$|j_2(u^k, \omega^{k+1}) - j_2(u^{k+1}, \omega^{k+1})| \leq [\alpha \max(\|u^k\|, \|u^{k+1}\|) + \beta] \|u^{k+1} - u^k\|$$

et on majore  $\|u^{k+1} - u^k\|$  comme pour le lemme 5 en appliquant l'inéquation variationnelle à  $v = u^k$ . Soit, après avoir majoré  $\max(\|u^k\|, \|u^{k+1}\|)$  par  $\|u^k\| + \|u^{k+1}\|$  et  $\|u\|$  par  $\|u^*\| + \|u - u^*\|$ ,

$$\|u^{k+1} - u^k\| \leq \epsilon^k [c'_1 \sqrt{\psi(u^k)} + c'_2 \sqrt{\psi(u^{k+1})} + c'_3]$$

Il reste

$$\psi^{k+1} - \psi^k \leq \epsilon^k (j(u^*, \omega^{k+1}) - j(u^k, \omega^{k+1})) + (\epsilon^k)^2 (c_1 \psi^k + c_2 \psi^{k+1} + c_3)$$

On est conduit, en prenant l'espérance conditionnelle sachant  $\mathbf{F}^k$ , à

$$\mathbf{E}^k \psi(u^{k+1}) - \psi^k \leq \epsilon^k (J(u^*) - J(u^k)) + (\epsilon^k)^2 (c_1 \psi^k + c_2 \mathbf{E}^k \psi^{k+1} + c_3)$$

Le terme  $(J(u^*) - J(u^k))$  est négatif, et l'espérance au sens de  $\mathbf{P}$  de l'inéquation précédente donne (avec la notation  $y^k = \mathbf{E}\psi(u^k)$ )

$$y^{k+1} \leq y^k + (\epsilon^k)^2 (c_1 y^k + c_2 y^{k+1} + c_3)$$

A partir de ce point, la suite de la démonstration est la même que celle du théorème 2.

### 2.5.2 $J$ est la composée d'une fonction croissante et d'une fonctionnelle additive

On suppose que  $J$  est définie par

$$j(u, \omega) = \phi\left(\sum_{i=1}^N j_{2,i}(u_i, \omega), \omega\right)$$

avec  $\phi$  fonction convexe et globalement lipchitzienne de constante  $L_\phi$  uniformément en  $\omega$ , et donc sous-différentiable. En particulier, ses sous-gradients sont uniformément bornés en norme par  $L_\phi$ .

**Algorithme 5** Au lieu de différencier complètement  $j$ , on remplace le problème local (2.2) par

$$\min_{u \in U^{\text{ad}}} K(u) - (K'(u^k), u) + (\epsilon^k \chi^k, j_2(u, \omega^{k+1}))$$

avec  $\chi^k \in \partial\phi(j_2(u^k, \omega^{k+1}), \omega^{k+1})$ .

#### Adaptation de la preuve.

Nous utilisons à nouveau la même fonction de Lyapunov  $\psi(u^k)$ .

L'inéquation variationnelle associée au nouveau problème auxiliaire est

$$\forall v \in U^{\text{ad}}, (K'(u^{k+1}) - K'(u^k), v - u^{k+1}) + \epsilon^k < \chi^k, j_2(v, \omega^{k+1}) - j_2(u^{k+1}, \omega^{k+1}) > \geq 0$$

En posant  $v = u^*$  dans cette inéquation, avec les mêmes arguments que plus haut,

$$\psi^{k+1} - \psi^k \leq \langle \chi^k, j_2(u^*, \omega^{k+1}) - j_2(u^{k+1}, \omega^{k+1}) \rangle$$

Par convexité de  $\phi$ ,

$$\begin{aligned} \psi^{k+1} - \psi^k &\leq \epsilon^k [\phi(j_2(u^*, \omega^{k+1}), \omega^{k+1}) - \phi(j_2(u^k, \omega^{k+1}), \omega^{k+1})] \\ &\quad + \epsilon^k \langle \chi^k, j_2(u^k, \omega^{k+1}) - j_2(u^{k+1}, \omega^{k+1}) \rangle \end{aligned}$$

En posant  $v = u^k$  dans l'inéquation variationnelle, on majore  $\|u^{k+1} - u^k\|$  comme pour le lemme 5, en utilisant les propriétés supplémentaires

$$\|\chi^k\| \leq L_\phi \quad \text{et} \quad |j_2(u^k, \omega^{k+1}) - j_2(u^{k+1}, \omega^{k+1})| \leq [\alpha \max(\|u^k\|, \|u^{k+1}\|) + \beta] \|u^{k+1} - u^k\|$$

Il reste, comme pour la variante précédente,

$$\psi^{k+1} - \psi^k \leq \epsilon^k (j(u^*, \omega^{k+1}) - j(u^k, \omega^{k+1})) + (\epsilon^k)^2 (c_1 \psi^k + c_2 \psi^{k+1} + c_3)$$

et la suite de la démonstration est la même.

Cette dernière variante sera citée dans le chapitre 6 pour le traitement des contraintes stochastiques par pénalisation exacte.

## 2.6 Comparaison avec des résultats antérieurs

Nous discutons ici les hypothèses faites et les résultats obtenus dans les principales publications sur les algorithmes stochastiques actuellement disponibles. Remarquons tout d'abord que dans la plupart des papiers traitant d'algorithmes stochastiques, les hypothèses de différentiabilité de  $J$  et de dérivée lipchitzienne sont faites. La convexité, en revanche, n'est pas toujours requise, mais son absence conduit bien entendu à des résultats locaux. Notre hypothèse *SGLB* permet de relâcher les premières mais nous oblige à supposer la convexité de  $J$  au moins dans  $U^{ad}$  pour avoir un résultat global.

### 2.6.1 Dodu, Goursat, Hertz, Quadrat, Viot (1981)

Hypothèses :

1.  $U = \mathbf{R}^r$ ,  $j(u, \omega)$  convexe et continue,  $U^{ad}$  est un convexe fermé.
2.  $\forall r \in \partial j(u, \omega)$ ,  $\|r\| \leq q$
3.  $J(u) - J(u^*) \geq c \cdot \min_{u^* \in U^*} \|u - u^*\|^2 =: c l^2(u)$
4.  $\epsilon^k \rightarrow 0$  et  $\sum_k \epsilon^k = \infty$
5. les tirages  $\omega^{k+1}$  suivent la loi  $P$  et sont indépendants entre eux.

**Théorème 3** Sous les hypothèses précédentes,  $y^k := \mathbf{E}l^2(u) \rightarrow 0$  et

$$\text{si } \epsilon^k = \frac{1}{ck + \frac{q^2}{30c}} \text{ alors } \forall k, y^k \leq \frac{1}{\frac{c^2}{q^2}k + \frac{1}{30}}$$

**Théorème 4** Sous les hypothèses précédentes, et si  $\sum_k (\epsilon^k)^2 < \infty$ , alors  $\lim_k l(u) = 0$ , P-p.s. De plus,

$$\mathbf{P}(\sup_{k \geq m} l^2(u^k) \geq \epsilon) \leq \frac{1}{\epsilon} \frac{A(m)}{\frac{c^2}{q^2}m + \frac{1}{30}}$$

avec  $\lim_{m \rightarrow 0} A(m) = 1$

#### Remarques

Les hypothèses 2. et 3. sont moins générales que les nôtres et ne sont pas vérifiables simultanément de manière globale (sur  $U$ ). Ce paradoxe disparaît si l'on suppose  $U^{ad}$  borné.

### 2.6.2 Bertran (1975)

#### Hypothèses :

- 1  $U$  Hilbert,  $U^{ad}$  est un convexe fermé de  $U$
- 2  $J$  fortement convexe et Fréchet différentiable.
- 3 le gradient de  $J$ ,  $\nabla J$  est un opérateur lipchitzien.
- 4  $r^k = \nabla J(u^k) + Z(u^k)$  avec  $Z(u^k)$  réalisation d'un "bruit" sur la mesure de  $\nabla J(u^k)$ . Ces réalisations suivent une loi  $P$  et sont indépendantes entre elles.
- 5  $E[\Pi_{U^{ad}}[u - \epsilon(\nabla J(u) + Z(u))] = \Pi_{U^{ad}}(u - \epsilon \nabla J(u))](\epsilon)$   
et  $E(\|Z(u)\|) \leq B^2$
- 6  $(\epsilon^k)_{k \in \mathbb{N}}$  est une  $\sigma$ -suite.

L'algorithme associé est

$$u^{k+1} = \Pi_{U^{ad}}[u^k - \epsilon^k(\nabla J(u^k) + Z(u^k))]$$

#### **Théorème 5** (concaténation des théorèmes 2.2.2.1 à 2.2.2.3)

*Sous les hypothèses précédentes,  $u^k$  converge fortement p.s. et en moyenne quadratique vers  $u^*$ , unique solution du problème.<sup>(a)</sup>*

<sup>a</sup>puisque  $J$  strictement convexe

#### Remarques

- $E$  est relative à  $(\Omega, \mathcal{A}, P)$
- 2. et 3. sont vérifiées par des fonctionnelles quadratiques. Nous n'avons pas réussi à exhiber de fonctionnelles non quadratiques qui vérifient simultanément ces deux hypothèses. Une modification de 2. en "J seulement strictement convexe" conduit à une hypothèse sur  $(\epsilon^k)_{k \in \mathbb{N}}$  assez technique et difficile à vérifier.
- 3. et 5. impliquent notre hypothèse " $j(\omega)$  est SGLB" :

$$\|r^k\| \leq \|r^*\| + \|\nabla J(u^k) - \nabla J(u^*) + Z(u^k) - Z(u^*)\|$$

donc

$$\|r^k\| \leq c_1 \|u^k - u^*\| + c_2$$

- la première partie de 5. semble difficile à vérifier dans la pratique si  $\Pi$  est non linéaire. La seconde partie est plus restrictive que notre hypothèse SGLB.

<sup>b</sup>espérance au sens de  $P$

### 2.6.3 Gladishev (1965)

(cité dans Bertran [1])

**Hypothèses :**

1.  $U = \mathbf{R}^r$ ,  $U^{ad} = U$
2.  $J$  Fréchet différentiable et admet un minimum unique  $u^*$
3. le gradient de  $J$ ,  $\nabla J$  vérifie

$$\forall \epsilon, \inf_{\|u - u^*\| < \epsilon} (u - u^*, \nabla J(u)) > 0$$

4.  $r^k = \nabla J(u^k) + Z(u^k)$  avec  $Z(u^k)$  réalisation d'un "bruit" sur la mesure de  $\nabla J(u^k)$ . Ces réalisations sont supposées indépendantes entre elles.
5. limitation du bruit :

$$\exists d > 0, \forall u \in U, E(\|\nabla J(u) + Z(u)\|^2) \leq d(1 + \|u - u^*\|^2)$$

6.  $(\epsilon^k)_{k \in \mathbf{N}}$  est une  $\sigma$ -suite.

L'algorithme associé est

$$u^{k+1} = u^k - \epsilon^k (\nabla J(u^k) + Z(u^k))$$

**Théorème 6** *Sous les hypothèses précédentes,  $u^k$  converge fortement p.s. et en moyenne quadratique vers  $u^*$  unique solution du problème.*

**Remarques**

- le résultat est démontré en dimension finie.
- $E$  est relative à  $(\Omega, \mathcal{A}, P)$
- 3. est un peu technique (difficile à vérifier sur une fonctionnelle quelconque), mais  $J$  strictement convexe implique 3.
- notre hypothèse *SGLB* implique 5.

### 2.6.4 Kushner & Clark (1978)

L'approche étant différente de celles décrites précédemment (et de la nôtre), nous allons présenter une discussion plus approfondie. Le problème de départ est aussi la minimisation d'une fonctionnelle  $J$ , mais celle-ci est non nécessairement convexe. L'algorithme est le suivant

$$u^{k+1} = u^k + \epsilon^k h(u^k) + \epsilon^k \beta^k + \epsilon^k \chi^k$$

**Hypothèses :**<sup>(6)</sup>

- $U = U^{ad} = \mathbf{R}^r$  ( $U^{ad}$  peut être un convexe fermé. Dans ce cas, on projette l'algorithme sur ce fermé et le théorème ci-dessous reste valable).
- $(u^k)_{k \in \mathbf{N}}$  est une suite bornée p.s. (au besoin, on la projette sur un fermé borné).
- $h$  est continue sur  $U$
- $\epsilon^k \rightarrow 0$  et  $\sum_k \epsilon^k = \infty$
- $(\beta^k)_{k \in \mathbf{N}}$  est une suite bornée P-p.s., convergeant p.s. vers 0.
- la suite  $(\chi^k)_{k \in \mathbf{N}}$  est telle que

$$\forall \epsilon > 0, \lim_{k \rightarrow \infty} P(\sup_{n > k} |\sum_{i=k}^n \epsilon^i \chi^i| \geq \epsilon) = 0$$

**Théorème 7** (résultat 2.5.1, p.38)

Sous les hypothèses précédentes, il existe un ensemble de mesure nulle  $\Omega_0$  tel que

si  $u(t) = u^*$  est une solution localement asymptotiquement stable de l'équation différentielle ordinaire (o.d.e.)  $\dot{u} = h(u)$ , de domaine d'attraction  $DA(u^*)$ .

alors pour tout  $\tilde{\omega} \in \Omega_0$ ,

si la suite  $u^k(\tilde{\omega})$  appartient infiniment souvent à un compact inclus dans  $DA(u^*)$ , alors  $u^k(\tilde{\omega})$  converge vers  $u^*$

En particulier, si  $u(t) = u^*$  est la seule trajectoire bornée solution de l'o.d.e., alors  $u^k$  converge p.s. vers  $u^*$ .

**Application à notre situation :**

Supposons que  $U$  est de dimension finie. Dans le cas simple où  $K(u) = \frac{1}{2}\|u\|^2$  et  $U = U^{ad}$ , la condition d'optimalité de  $u^{k+1}$  est

$$u^{k+1} = u^k - \epsilon^k r^k$$

on peut écrire cet algorithme sous la forme proposée par Kushner et Clark,

$$u^{k+1} = u^k + \epsilon^k(-\tilde{r}^k) + \epsilon^k(\tilde{r}^k - r^k)$$

avec  $\tilde{r}^k = \int_{\Omega} r(u^k, \omega) P(d\omega)$ , à condition de supposer  $\tilde{r}^k$  lipchitz, afin d'assurer l'existence de solutions à l'o.d.e. associée

$$\dot{u} = -\tilde{r}(u)$$

<sup>6</sup>nous présentons ici les hypothèses les plus simples de la référence. Kushner propose d'autres hypothèses, moins restrictives mais plus techniques et plus difficiles à vérifier sur  $\beta$  et  $\chi$ .



Remarquons que cette hypothèse est relativement forte puisqu'elle implique que  $J$  est Gâteaux-différentiable à dérivée lipchitzienne.

Si l'on prend  $\psi(u) := \frac{1}{2}\|u - u^*\|^2$  comme fonction de Lyapunov pour étudier cette o.d.e., on constate que

$$\frac{d}{dt}\psi(u) = \langle \dot{u}, u - u^* \rangle = \langle -\bar{r}(u), u - u^* \rangle$$

Si l'on suppose  $J$  strictement convexe, alors

$$0 > J(u^*) - J(u) > \langle -\bar{r}(u), u - u^* \rangle, \quad \forall u \neq u^*$$

On en déduit que  $\psi(u)$  est positive et strictement décroissante le long des trajectoires de l'o.d.e. Donc  $u^*$  est la seule solution stable de l'algorithme. On peut donc appliquer le résultat annoncé si les hypothèses sur  $\beta^k$  et  $\chi^k$  sont vérifiées.

On constate qu'il suffit de poser  $\beta^k = 0$ , p.s.<sup>(7)</sup> et  $\chi^k = \bar{r}^k - r^k$ .  
En effet, dans ce cas,  $\mathbf{E}^k \chi^k = 0$  et donc

$$H^k := \sum_{p=1}^k \epsilon^p \chi^p$$

est une martingale pour laquelle on peut appliquer l'inégalité de Doob <sup>(8)</sup> :

$$\forall \epsilon > 0, \quad P(\sup_{m \geq k} |H^m - H^{k-1}| \geq \epsilon) \leq \left(\frac{A}{\epsilon}\right)^2 \mathbf{E} \sum_{i=k}^{\infty} \epsilon^i$$

avec

$$\mathbf{E}^k(\|\chi^k\|^2) = \int_{\Omega} \|r(u^k, \omega)\|^2 P(d\omega) - \left\| \int_{\Omega} r(u^k, \omega) P(d\omega) \right\|^2 \leq A$$

On constate qu'il a été nécessaire de faire, tout au long de cette courte application, un certain nombre d'hypothèses sur  $J$  plus restrictives que celles que nous faisons habituellement pour nos algorithmes. Néanmoins, cette méthodologie, développée par Ljung puis Kushner & Clark, permet, comme on le montrera dans les deux chapitres suivants, d'éliminer rapidement des "candidats algorithmes" pour lesquels une étude complète pourrait être beaucoup plus coûteuse qu'une simple résolution d'o.d.e. De plus cette technique permet, comme le montrent Métivier & Priouret [1], de considérer des algorithmes où les réalisations du bruit ne sont pas indépendantes entre elles, cas que nous ne traitons pas.

En conclusion, l'originalité de notre preuve réside en la non forte convexité de  $J$  et l'obtention d'un résultat de convergence faible p.s.

<sup>7</sup>Ces variables servent en particulier lors d'approximations de dérivées par des différences finies  
<sup>8</sup>résultat 13

## 2.7 Tests Numériques

Les sections "tests numériques" de ce mémoire sont constituées des sorties de quelques programmes BASIC <sup>(9)</sup> permettant la recherche de solutions au problème

$$\left\{ \min_{u \in U} E j(u, \omega) \mid \theta(u) = c \right\}$$

Notre exemple est choisi dans  $U = \mathbf{R}^4$ , avec  $\Omega = [0, 1]^4$ . La loi de probabilité  $P$  est uniforme sur  $[0, 1]^4$  et les  $\omega_i$  sont considérés comme indépendants.

La fonction test  $j(u, \omega)$  a l'expression suivante

$$j(u, \omega) = \frac{1}{2}[(u_1 - \omega_1)^2 + \omega_2(u_2 - 1)^2 + (\omega_3 u_3 + \omega_4 u_4 - 1)^2]$$

La contrainte égalité déterministe associée est  $\theta(u) = c$  avec

$$\theta(u) = u_1 + u_2 + u_3 + u_4, \quad c = 6$$

Nous allons tout d'abord considérer le problème de l'optimisation libre, c'est-à-dire sans la contrainte  $\theta$ . Dans un deuxième temps, nous considérerons la contrainte  $\theta$  comme une contrainte implicite qui définit  $U^{ad}$

### Remarques :

- nous avons choisi une loi  $P$  uniforme car c'est la loi de probabilité la plus diffuse et donc celle sur laquelle on doit tester la robustesse des algorithmes.
- du point de vue du "bruit",  $j$  contient un terme avec bruit additif, un terme avec bruit multiplicatif et un mélange des deux.
- $j$  satisfait aux conditions des théorèmes de convergence démontrés. Elle est strictement convexe ainsi que ses réalisations en  $\omega_1$  et  $\omega_2$ .  
En revanche, sa composante en  $\omega_3$  et  $\omega_4$  n'a aucune de ses réalisations strictement convexe, puisqu'elles s'annulent sur toute une droite de  $\mathbf{R}^2$ .

### 2.7.1 Optimisation Libre : $U^{ad} = U$

Nous présentons tout d'abord dans les pages suivantes les résultats des algorithmes 2 ("Optimisation Libre") et 3 ("Libre Fortement Convexe") dans la résolution du problème d'optimisation libre.

Le système est décomposé en quatre sous-systèmes indépendants

$$u = (u_1, u_2, u_3, u_4)$$

L'espérance du critère est

$$J(u) = \frac{1}{2}[u_1^2 - u_1 + \frac{1}{2}(1 - u_2)^2 + \frac{1}{2}u_3 u_4 + \frac{1}{3}(u_3^2 + u_4^2) - u_3 - u_4] + \frac{2}{3}$$

<sup>9</sup>sur ATARI ST 1040

Les valeurs optimales sont :

$$\begin{aligned} u_1^* &= 0,5, & u_2^* &= 1 \\ u_3^* &= u_4^* = \frac{6}{7} \simeq 0,857 \end{aligned}$$

et le critère optimal est  $J(u^*) = \frac{19}{168} \simeq 0,113$

Les points de départ sont  $u_i^0 = 2$ . Les paramètres  $\epsilon^k$  donnant les meilleurs résultats sont indiqués sur les graphes. Les deux algorithmes ont bien sûr été testés sur les mêmes tirages aléatoires, mais l'algorithme 3 étant plus rapide (grands pas en  $u$ ), nous l'avons testé sur 250 itérations au lieu de 500 pour l'algorithme 2. Les figures ci-dessous donnent une réalisation de chaque algorithme, représentée dans les plans  $(u_1, u_2)$  et  $(u_3, u_4)$ . L'optimum est le petit cercle blanc.

Pour l'algorithme 3, notons que le Hessien de  $J$  est

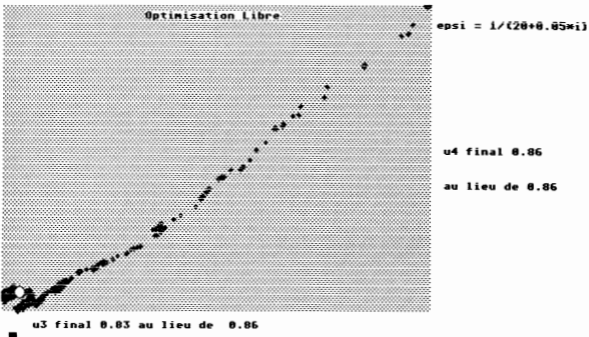
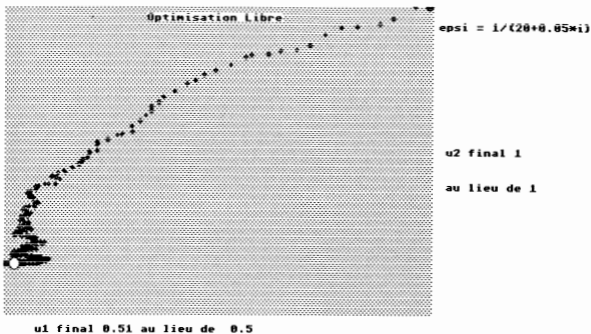
$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{1}{3} \end{pmatrix}$$

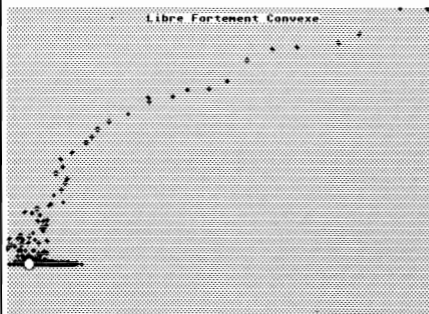
ses valeurs propres sont  $1, \frac{1}{2}, \frac{7}{12}$  et  $\frac{1}{12}$ . On peut donc appliquer l'algorithme 3 avec  $\alpha = \frac{1}{12}$ .

### Conclusion

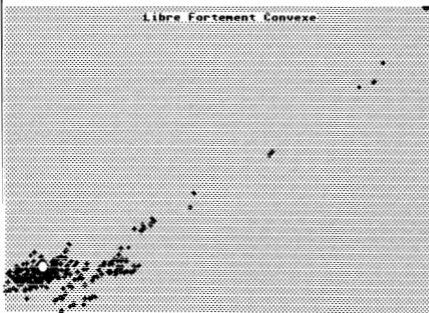
Les deux algorithmes convergent bien sur l'exemple proposé. Cela n'a rien d'étonnant puisqu'il s'agit simplement d'un gradient stochastique. L'algorithme 3 est plus rapide que l'algorithme 2, et c'est bien ce que la théorie annonce. En revanche, il est plus "brouillon" au voisinage de la solution, ce qui est également logique puisque les pas ne tendent pas vers 0. D'un point de vue méthodologique, il est très recommandable d'utiliser d'abord l'algorithme 3 pour avoir une estimation de la solution. En effet, il n'est pas rare de voir converger 2 "avant" la solution lorsque la suite  $\epsilon^k$  est mal choisie.

Nous avons constaté expérimentalement que  $\epsilon^k = \frac{1}{\alpha + \beta k}$  avec  $\frac{1}{\alpha}$  de l'ordre du  $\epsilon$  optimal pour l'algorithme 3 et  $\beta \in [10^{-2}, 10^{-1}]$  est un choix judicieux.



 $\rho = 1/10$  $u_2$  final 1

au lieu de 1

 $u_1$  final 0.51 au lieu de 0.5 $\rho = 1/10$  $u_4$  final 0.9

au lieu de 0.86

 $u_3$  final 0.85 au lieu de 0.86

**2.7.2  $\theta$  vue comme contrainte implicite définissant  $U^{ad}$** 

Il s'agit en fait d'un algorithme de sous-gradient projeté sur

$$U^{ad} = \{u \in \mathbf{R}^4 \mid u_1 + u_2 + u_3 + u_4 = 6\}$$

Les valeurs optimales sont :

$$u_1^* = \frac{14}{15} \simeq 0,933 \quad u_2^* = \frac{28}{15} \simeq 1,867$$

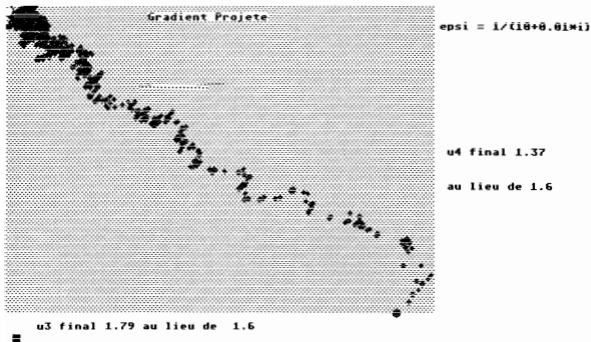
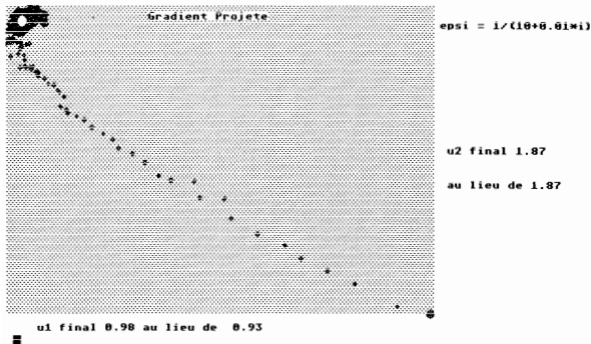
$$u_3^* = u_4^* = \frac{8}{5} = 1,6$$

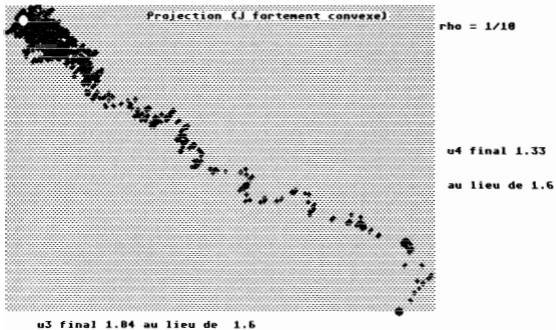
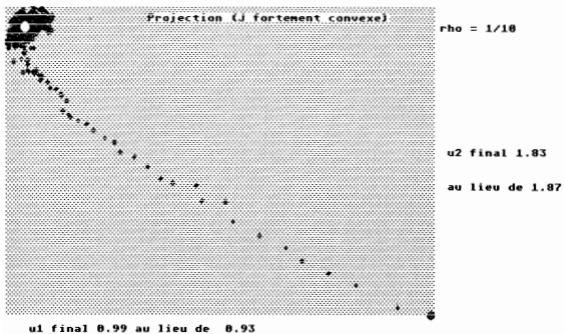
et le critère optimal est  $J(u^*) = \frac{43}{60} \simeq 0,716$

Nous donnons à nouveau les résultats des deux algorithmes, cette fois sur 1000 itérations pour les deux.

**Conclusion**

Les deux algorithmes se comportent particulièrement bien, comme on pourra le voir en comparant avec les algorithmes de décomposition / coordination des chapitres suivants.







## Chapitre 3

### Décomposition par les prix

Nous résolvons le problème suivant :

$$\min_{u \in U^*} \int_{\Omega} j(u, \omega) P(d\omega) = J(u) \quad \text{sous } \theta(u) \in -C \quad (3.1)$$

Avec  $\Omega$ , ensemble des réalisations aléatoires,  $P$  étant la loi de probabilité associée. On appelle  $U^*$ , l'ensemble des solutions du problème (3.1). On suppose qu'il existe un point-selle  $(u^*, p^*)$  à ce problème.

Le Lagrangien associé est :

$$L(u, p) = J(u) + \langle p, \theta(u) \rangle$$

Nous allons proposer deux algorithmes pour les cas suivants :

$J$  est strictement convexe : le Lagrangien ordinaire permettra de trouver une suite  $(u^k)_{k \in \mathbf{N}}$  qui convergera faiblement p.s. vers  $u^*$  et de prouver que la suite  $(p^k)_{k \in \mathbf{N}}$  associée est bornée p.s. Nous étudierons une variante particulière dans le cas où  $J$  est fortement convexe.

$J$  est seulement convexe : le Lagrangien augmenté nous permettra d'exhiber une suite double  $(u^k, p^k)_{k \in \mathbf{N}}$  dont les points d'accumulation faibles p.s.  $(\bar{u}, \bar{p})$  sont tels que  $\bar{u} \in U^*$  et  $\bar{p} \in \hat{P}(u^*)$ ,  $\forall u^* \in U^*$ .

### 3.1 Tentative d'extension directe

Avant d'étudier les algorithmes des sections suivantes, montrons qu'une extension trop directe d'un algorithme classique comme celui de Uzawa[1] n'est pas acceptable en stochastique.

#### Algorithme 6 (tentative)

- soit  $u^0 \in U^{ad}$  et  $p^0 \in C'$ ,
- connaissant  $u^k$  et  $p^k$ , on tire suivant la loi  $P$ , un  $\omega^{k+1} \in \Omega$ , indépendant des tirages précédents, et on calcule
  - $u^{k+1}$  par
 
$$\min_{u \in U^{ad}} j(u, \omega^{k+1}) + \langle p^k, \theta(u) \rangle$$
  - et  $p^{k+1}$  par  $p^{k+1} = \Pi(p^k + \epsilon^k \theta(u^{k+1}))$

$\Pi$  étant la projection sur  $C'$ , cône conjugué de  $C$ .

Etudions cette extension dans un cas linéaire quadratique avec contraintes affines :

$$N = 2, \quad U_i^{ad} = U_i, \quad \theta(u) = \theta_1 u_1 + \theta_2 u_2, \quad C = \{0\}$$

$$j(u, \omega) = \frac{1}{2}(A_1(\omega)u_1, u_1) + (b_1(\omega), u_1) + \frac{1}{2}(A_2(\omega)u_2, u_2) + (b_2(\omega), u_2)$$

Les solutions du problème de départ sont caractérisées par le système d'équations

$$E[A_i]u_i^* + E[b_i] + \theta_i^* p^* = 0, \quad i = 1, 2; \quad \theta_1 u_1^* + \theta_2 u_2^* = 0$$

ce sont donc

$$u_i^* = -(E[A_i])^{-1}(E[b_i] + \theta_i^* p^*)$$

$$p^* = -[\theta_1 E[A_1]^{-1} \theta_1^* + \theta_2 E[A_2]^{-1} \theta_2^*]^{-1} (\theta_1 E[A_1]^{-1} E[b_1] + \theta_2 E[A_2]^{-1} E[b_2])$$

Les conditions d'optimalité associées à l'algorithme sont

$$A_i(\omega^{k+1})u_i^{k+1} + b_i(\omega^{k+1}) + \theta_i^* p^k = 0, \quad i = 1, 2$$

$$p^{k+1} - p^k = \epsilon^k (\theta_1 u_1^{k+1} + \theta_2 u_2^{k+1})$$

ce qui conduit à l'équation d'évolution suivante pour  $p$  :

$$p^{k+1} - p^k = -\epsilon^k [\theta_1 (A_1(\omega^{k+1}))^{-1} b_1(\omega^{k+1}) + \theta_2 (A_2(\omega^{k+1}))^{-1} b_2(\omega^{k+1}) \\ + (\theta_1 (A_1(\omega^{k+1}))^{-1} \theta_1^* + \theta_2 (A_2(\omega^{k+1}))^{-1} \theta_2^*) p^k]$$

Le point stationnaire associé  $\bar{p}$ , à supposer qu'il soit stable, est donc solution de

$$\theta_1 E[A_1^{-1} b_1] + \theta_2 E[A_2^{-1} b_2] + (\theta_1 E[A_1^{-1}] \theta_1^* + \theta_2 E[A_2^{-1}] \theta_2^*) \bar{p} = 0$$

et donc,  $\bar{p}$  est, compte tenu de l'espérance de produit de variables non indépendantes, différent de  $p^*$ . On constatera d'ailleurs l'allure "chahutée" de cet algorithme dans la section "tests numériques"

### 3.2 J est strictement convexe : Lagrangien ordinaire

On propose l'algorithme suivant :

#### Algorithme 7

- soit  $u^0 \in U^{ad}$  et  $p^0 \in C'$ ,
- connaissant  $u^k$  et  $p^k$ , on tire suivant la loi  $P$ , un  $\omega^{k+1} \in \Omega$ , indépendant des tirages précédents, et on calcule
  - $u^{k+1}$  par

$$\min_{u \in U^{ad}} K(u) + (\epsilon^k r^k - K'(u^k), u) + \epsilon^k \langle p^k, \theta(u) \rangle \quad (3.2)$$

avec  $r^k$  appartenant au sous-différentiel de  $j$  au point  $(u^k, \omega^{k+1})$ .

- et  $p^{k+1}$  par
 
$$p^{k+1} = \Pi(p^k + \epsilon^k \theta(u^{k+1})) \quad (3.3)$$

$\Pi$  étant la projection sur  $C'$ , cône conjugué de  $C$ .

Si  $K(u) = \frac{1}{2}\|u\|^2$ , cet algorithme est de type Arrow-Hurwicz.

Si  $K$  et  $\theta$  sont additives, il est décomposé. Dans le cas où  $\theta$  n'est pas additive, une variante consiste à résoudre, au lieu de (3.2), le problème auxiliaire :

$$\min_{u \in U^{ad}} K(u) + (\epsilon^k r^k - K'(u^k), u) + \epsilon^k \langle p^k, \tau^k \cdot u \rangle \quad (3.4)$$

avec  $\tau^k \in \partial\theta(u^k)$ . Il faut évidemment supposer alors que  $\theta$  est  $C$ -sous-différentiable. Compte tenu de l'indépendance probabiliste entre  $\omega^{k+1}$  et  $u^k$ , cette modification de l'algorithme ne change pas les conclusions du théorème suivant. Nous renvoyons à Cohen[1, p. 124] pour le traitement de cette variante dans le cas déterministe.

Remarque :

L'hypothèse  $J$  strictement convexe faite dans le théorème suivant pourrait être remplacée par l'hypothèse plus générale mais moins facilement vérifiable "L stable en u". Sans cette dernière, il est certainement très difficile d'obtenir un résultat de convergence en u en utilisant un Lagrangien ordinaire.

**Théorème 8** En plus de l'existence d'un point-selle au problème (3.1), on suppose :

- $j(\cdot, \omega)$  convexe, propre, sci, pour tout  $\omega \in \Omega$  ;  $j(u, \cdot)$  mesurable sur  $\Omega$ , pour tout  $u \in U^{ad}$
- $J(\cdot)$  strictement convexe.
- $\theta(\cdot)$  lipschitzienne de rapport  $L_\theta$ .
- $K$  fonctionnelle fortement convexe de rapport  $b > 0$ , différentiable.

Alors, la solution de (3.4) existe et est unique. On la note  $u^{k+1}$

Si de plus, pour tout  $\omega$ ,  $j(\cdot, \omega)$  est à sous-gradient linéairement borné de paramètres  $c_1$  et  $c_2$ , et si  $(\epsilon^k)_{k \in \mathbb{N}}$  est une  $\sigma$ -suite

Alors :

les suites  $(u^k)_{k \in \mathbb{N}}$  et  $(p^k)_{k \in \mathbb{N}}$  sont bornées p.s. et  $(u^k)_{k \in \mathbb{N}}$  converge faiblement P-p.s. vers  $u^*$ <sup>a</sup>.

Enfin, si  $J$  est fortement convexe, la convergence vers  $u^*$  est forte.

<sup>a</sup>D'après la stricte convexité de  $J$ ,  $U^* = \{u^*\}$ ; singleton.

### 3.2.1 Inéquations variationnelles et fonction de Lyapunov

L'inéquation variationnelle associée au problème (3.4) est :

$$(K'(u^{k+1}) - K'(u^k) + \epsilon^k r^k, u - u^{k+1}) + \epsilon^k \langle p^k, \theta(u) - \theta(u^{k+1}) \rangle \geq 0, \quad \forall u \in U^{ad} \quad (3.5)$$

Les inégalités de point-selle associées au problème (3.1) sont

$$\forall (v, q) \in U^{ad} \times C', \quad J(u^*) + \langle q, \theta(u^*) \rangle \leq J(u^*) + \langle p^*, \theta(u^*) \rangle \leq J(v) + \langle p^*, \theta(v) \rangle \quad (3.6)$$

On introduit la fonction de Lyapunov

$$\psi^k = K(u^*) - K(u^k) - (K'(u^k), u^* - u^k) + \frac{1}{2} \|p^k - p^*\|^2$$

Le théorème résulte des variations de la suite  $(\psi^k)_{k \in \mathbb{N}}$ .

## 3.2.2 Majorations préliminaires

**Lemme 6** Par définition de  $\psi^k$ ,

$$\|p^k - p^*\|^2 \leq 2\psi^k, \quad \|u^k - u^*\|^2 \leq \frac{2}{b}\psi^k \quad (3.7)$$

Il existe des constantes positives  $c'_1$  et  $c'_2$  telles que

$$\|u^{k+1} - u^k\| \leq \epsilon^k (c'_1 \sqrt{\psi^k} + c'_2) \quad (3.8)$$

Par définition de  $p^{k+1}$  et  $p^*$ ,

$$\|p^{k+1} - p^*\|^2 \leq \|p^k - p^*\|^2 + (\epsilon^k)^2 L_\theta^2 \|u^{k+1} - u^*\|^2 + 2\epsilon^k \langle p^k - p^*, \theta(u^{k+1}) - \theta(u^*) \rangle > \quad (3.9)$$

Enfin, il existe des constantes positives  $c''_1$ ,  $c''_2$  et  $c''_3$  telles que,

$$\psi^{k+1} - \psi^k \leq \epsilon^k [(\langle r^k, u^* - u^k \rangle) + \langle p^*, \theta(u^*) - \theta(u^k) \rangle] + (\epsilon^k)^2 (c''_1 \psi^k + c''_2 + c''_3 \psi^{k+1}) \quad (3.10)$$

Ce lemme se démontre de manière analogue au lemme 5.

(3.7) découle immédiatement de la définition de  $\psi$ .

Par application de l'inégalité de Schwarz,

$$\epsilon^k \|r^k\| \cdot \|u^{k+1} - u^k\| \geq (\epsilon^k r^k, u^k - u^{k+1})$$

et en utilisant le caractère lipchitzien de  $\theta$ ,

$$\epsilon^k L_\theta \|p^k\| \|u^{k+1} - u^k\| \geq \epsilon^k \langle p^k, \theta(u^k) - \theta(u^{k+1}) \rangle$$

d'où, avec l'inéquation variationnelle (3.5) en  $u = u^k$ ,

$$\begin{aligned} b \|u^{k+1} - u^k\|^2 &\leq (K'(u^k) - K'(u^{k+1}), u^k - u^{k+1}) \\ &\leq \epsilon^k \|u^{k+1} - u^k\| [\|r^k\| + L_\theta \|p^k\|] \end{aligned}$$

enfin, en divisant les extrémités de gauche et de droite par  $\|u^{k+1} - u^k\|$  supposé non nul sans perte de généralité,

$$b \|u^{k+1} - u^k\| \leq \epsilon^k (\|r^k\| + L_\theta \|p^k\|)$$

ce qui, par hypothèse *SGLB*, d'après (3.7) et après avoir majoré  $\|p^k\|$  par  $\|p^k - p^*\| + \|p^*\|$ , implique l'existence de constantes positives  $c'_1$  et  $c'_2$  qui satisfont (3.8).

(3.9) provient de l'équation

$$\forall k, \quad p^* = \Pi(p^* + \epsilon^k \theta(u^*))$$

que l'on a retranchée de la définition de  $p^{k+1}$  (3.3), élevée au carré et développée en tenant compte du caractère lipchitz de  $\theta$  et de  $\Pi$ .

Pour monter (3.10), on évalue  $\psi^{k+1} - \psi^k$  qui s'écrit :

$$\begin{aligned} & \overbrace{K(u^k) - K(u^{k+1}) - (K'(u^k), u^k - u^{k+1})}^{A_1} \\ & + \underbrace{(K'(u^k) - K'(u^{k+1}), u^* - u^{k+1})}_{A_2} + \underbrace{\frac{1}{2}(\|p^{k+1} - p^*\|^2 - \|p^k - p^*\|^2)}_{A_3} \end{aligned}$$

$A_1$  est négatif car  $K$  est convexe.

A l'aide de l'inéquation variationnelle exprimée en  $u = u^*$ ,

$$(K'(u^{k+1}) - K'(u^k) + \epsilon^k r^k, u^* - u^{k+1}) + \epsilon^k < p^k, \theta(u^*) - \theta(u^{k+1}) > \geq 0$$

on majore  $A_2$  :

$$A_2 \leq \epsilon^k [(r^k, u^* - u^{k+1}) + < p^k, \theta(u^*) - \theta(u^{k+1}) >]$$

$A_3$  est, d'après (3.9), majoré par

$$A_3 \leq \frac{1}{2}(\epsilon^k)^2 L_\theta^2 \|u^{k+1} - u^*\|^2 + \epsilon^k < p^k - p^*, \theta(u^{k+1}) - \theta(u^*) >$$

En définitive

$$A_1 + A_2 + A_3 \leq \epsilon^k (r^k, u^* - u^{k+1}) + \epsilon^k < p^*, \theta(u^*) - \theta(u^{k+1}) > + \frac{1}{2}(\epsilon^k)^2 L_\theta^2 \|u^{k+1} - u^*\|^2$$

$$\begin{aligned} & A_1 + A_2 + A_3 \leq \epsilon^k [(r^k, u^* - u^k) + < p^*, \theta(u^*) - \theta(u^k) >] \\ & + \epsilon^k (\|r^k\| + L_\theta \|p^*\|) \|u^{k+1} - u^k\| + \frac{1}{2}(\epsilon^k)^2 L_\theta^2 \|u^{k+1} - u^*\|^2 \end{aligned}$$

Compte tenu de (3.7), (3.8) et de l'hypothèse " $j$  est  $SGLB$ ", il existe des constantes positives  $c_1''$ ,  $c_2''$  et  $c_3''$  telles que (3.10), c.a.d.

$$\psi^{k+1} - \psi^k \leq \epsilon^k [(r^k, u^* - u^k) + < p^*, \theta(u^*) - \theta(u^k) >] + (\epsilon^k)^2 (c_1'' \psi^k + c_2'' + c_3'' \psi^{k+1})$$

### 3.2.3 Convergence de la suite $\psi^k$

$y^k := \mathbf{E}(\psi^k)$  est une suite positive bornée.

Prenons l'espérance conditionnelle sachant  $\mathbf{F}^k$  de (3.10). Comme  $u^k$  et  $\omega^{k+1}$  sont indépendants,

$$\mathbf{E}^k \psi^{k+1} - \psi^k \leq \epsilon^k [(\mathbf{E}^k r^k, u^* - u^k) + < p^*, \theta(u^*) - \theta(u^k) >] + (\epsilon^k)^2 (c_1'' \psi^k + c_2'' + c_3'' \mathbf{E}^k \psi^{k+1})$$

$r^k := \mathbf{E}^k r^k = \int_{\Omega} r(u^k, \omega) P(d\omega)$  est un élément de  $\partial J(u^k)$  et donc

$$(r^k, u^* - u^k) \leq J(u^*) - J(u^k)$$

soit

$$\mathbf{E}^k \psi^{k+1} - \psi^k \leq \epsilon^k [(J(u^*) - J(u^k)) + \langle p^*, \theta(u^*) - \theta(u^k) \rangle] + (\epsilon^k)^2 (c_1'' \psi^k + c_2'' + c_3'' \mathbf{E}^k \psi^{k+1})$$

ou, en termes de Lagrangien  $L$ ,

$$\mathbf{E}^k \psi^{k+1} - \psi^k \leq \epsilon^k [L(u^*, p^*) - L(u^k, p^*)] + (\epsilon^k)^2 (c_1'' \psi^k + c_2'' + c_3'' \mathbf{E}^k \psi^{k+1}) \quad (3.11)$$

D'après l'inégalité de point-selle (problème (3.1)), on a  $L(u^*, p^*) - L(u^k, p^*) \leq 0$ ,

$$\mathbf{E}^k \psi^{k+1} - \psi^k \leq \alpha_k \psi^k + \beta_k + \gamma_k \mathbf{E}^k \psi^{k+1} \quad (3.12)$$

avec  $\alpha_k, \beta_k$  et  $\gamma_k$  suites positives sommables ( $(\epsilon^k)_{k \in \mathbf{N}}$  est une  $\sigma$ -suite).

En prenant maintenant l'espérance au sens de  $\mathbf{P}$ ,

$$\forall k \in \mathbf{N}, \quad y^{k+1} - y^k \leq \alpha_k y^k + \beta_k + \gamma_k y^{k+1}$$

et d'après le lemme 1,  $(y^k)_{k \in \mathbf{N}}$  est une suite bornée. On note  $M$  l'un de ses majorants.

$\psi^k$  converge p.s. au sens de  $\mathbf{P}$

$\psi^k$  est une quasi-martingale (propriété caractéristique 18).

En effet,  $(\psi^k)_{k \in \mathbf{N}}$  vérifie  $\inf_{k \in \mathbf{N}} \mathbf{E}(\psi^k) > -\infty$  ( $y^k$  est positive) et d'après (3.11)

$$\mathbf{E}^k(\psi^{k+1} - \psi^k) \leq \alpha_k \psi^k + \beta_k + \gamma_k \mathbf{E}^k \psi^{k+1}$$

on déduit, comme  $y^k$  est bornée par  $M$  :

$$\sum_{k \in \mathbf{N}} \mathbf{E}(\mathbf{E}^k \psi^{k+1} - \psi^k)^+ \leq \sum_{k \in \mathbf{N}} (\alpha_k y^k + \beta_k + \gamma_k y^{k+1}) \leq \sum_{k \in \mathbf{N}} (M[\alpha_k + \gamma_k] + \beta_k) < \infty$$

Donc  $\psi^k$  est une quasi-martingale et, d'après le résultat 14, elle converge presque-sûrement vers une variable aléatoire d'espérance finie.

### 3.2.4 Limites de $L(u^k, p^*)$ et de $u^k$

La suite  $(u^k)_{k \in \mathbf{N}}$  admet p.s. des points d'accumulation faible.

Comme  $\psi^k$  converge p.s., vers une variable aléatoire d'espérance finie, il existe un ensemble  $\Omega_0$  de mesure nulle tel que

$$\forall \tilde{\omega} \in \Omega \setminus \Omega_0, \quad \lim_{k \rightarrow \infty} \psi(u^k(\tilde{\omega})) < \infty$$

et donc en vertu de (3.7),

$$\|u^k - u^*\|^2 \leq \frac{2}{b} \sup_k \psi^k < \infty$$

Soit un tel  $\tilde{\omega}$ , la suite  $u^k(\tilde{\omega})$  est bornée et admet d'après le théorème de compacité faible des points d'accumulation faible dans  $U^{\text{ad}}$  Il en est de même pour  $p^k$

$(L(u^k, p^*))_{k \in \mathbf{N}}$  converge p.s. vers  $L(u^*, p^*)$ .

Ce résultat est une application du lemme 2 qui est déterministe.

L'inéquation (3.11) se réécrit

$$\epsilon^k (L(u^k, p^*) - L(u^*, p^*)) \leq \psi^k (1 + \alpha_k) - (1 - \gamma_k) \mathbf{E}^k \psi^{k+1} + \beta_k$$

en prenant l'espérance au sens de  $\mathbf{P}$ ,

$$\epsilon^k \mathbf{E}[L(u^k, p^*) - L(u^*, p^*)] \leq y^k (1 + \alpha_k) - y^{k+1} (1 - \gamma_k) + \beta_k$$

soit par sommation sur  $k$  :

$$\sum_{k=1}^n \epsilon^k \mathbf{E}[L(u^k, p^*) - L(u^*, p^*)] \leq \sum_{k=1}^n (y^k (1 + \alpha_k) - y^{k+1} (1 - \gamma_k) + \beta_k)$$

soit, comme  $(y^k)_{k \in \mathbf{N}}$  est bornée par  $M$ , en réarrangeant les termes :

$$\sum_{k=1}^n \epsilon^k \mathbf{E}[L(u^k, p^*) - L(u^*, p^*)] \leq \sum_{k=1}^n (M(\alpha_k + \gamma_k) + \beta_k) + y^1 - y^{n+1} \leq M' < \infty$$

Les termes  $L(u^k, p^*) - L(u^*, p^*)$  étant positifs, on a

$$\sum_{k \in \mathbf{N}} \epsilon^k [L(u^k, p^*) - L(u^*, p^*)] < \infty \quad \mathbf{P} - p.s.$$

On appellera  $\Omega_1$  l'ensemble de mesure nulle sur lequel cette majoration n'est pas vérifiée.

D'autre part,  $\psi^k$  bornée p.s. implique, d'après (3.8)

$$\|u^{k+1} - u^k\| \leq \epsilon^k M''$$

presque-sûrement (à  $\Omega_0$  près). Pour tout  $\tilde{\omega}$  n'appartenant ni à  $\Omega_0$  ni à  $\Omega_1$ , les conditions du lemme 2 sont vérifiées (y compris le caractère lipchitzien de  $L(\cdot, p^*)$  sur le borné contenant la suite  $(u^k(\tilde{\omega}))_{k \in \mathbf{N}}$ ). Donc pour cet  $\tilde{\omega}$ , la suite  $L(u^k(\tilde{\omega}), p^*)$  converge vers  $L(u^*, p^*)$ .

$(u^k)_{k \in \mathbf{N}}$  converge faiblement P-p.s.

Pour cela, nous utilisons le lemme 4 qui s'applique immédiatement.

Pour  $\tilde{\omega} \in \Omega \setminus \Omega_0 \cup \Omega_1$  (1), on note  $\tilde{u}(\tilde{\omega})$  un point d'accumulation faible de  $(u^k)$ , et  $(u^{\phi(k)})$  une sous-suite convergeant vers  $\tilde{u}$ . Par faible semi-continuité inférieure de  $L$  en  $u$ ,

$$L(\tilde{u}, p^*) \leq \liminf_{k \rightarrow \infty} L(u^{\phi(k)}, p^*) = L(u^*, p^*)$$

<sup>1</sup> ces ensembles pouvant dépendre de  $u^*$



et donc  $\bar{u}(\bar{\omega}) \in \hat{U}(p^*)$ .  $J$  étant strictement convexe,  $\hat{U}(p^*)$  est un singleton. Comme de plus,  $\hat{U}(p^*)$  contient  $U^*$ , lui-même réduit au singleton  $\{u^*\}$ , on en déduit que  $\bar{u} = u^*$ . La suite  $(u^k)_{k \in \mathbf{N}}$  n'admet qu'un seul point d'accumulation. Elle converge donc faiblement  $\mathbf{P}$ -p.s. vers  $u^*$ .

Lorsque  $J$  est fortement convexe, on démontre exactement comme pour le théorème 2 que  $(u^k)_{k \in \mathbf{N}}$  converge fortement vers  $u^*$ .

Remarque :

On peut démontrer un résultat analogue lorsque "la contrainte  $\theta$  est connue en espérance", c'est-à-dire lorsque le niveau de contrainte est "bruité". Dans ce cas,  $\theta(u) = E\tau(u, \omega)$  et le problème local devient

$$\min_{u \in U^{\text{ad}}} K(u) + (\epsilon^k r^k - K'(u^k), u) + \epsilon^k < p^k, \tau(u, \omega^{k+1}) >$$

Si l'on suppose les mêmes hypothèses sur  $\tau$ , uniformément en  $\omega$ , que sur  $\theta$ , les conclusions sont les mêmes.

### 3.3 Variante lorsque $J$ est fortement convexe

On propose ici une variante de l'algorithme précédent, pour laquelle on effectue des "grands pas" pour l'évaluation de  $p^k$ . Il faut alors supposer que  $p^k$  est contenu dans  $B(0, M)$ , une boule centrée en 0 et de rayon  $M$  suffisamment grand pour que  $p^* \in B(0, M)$ .

Nous comparerons cet algorithme avec la variante principale (algorithme 7) dans la section "tests numériques" de ce chapitre.

#### Algorithme 8

- soit  $\rho > 0$ , soit  $u^0 \in U^{\text{ad}}$  et  $p^0 \in C' \cap B(0, M)$ ,
- connaissant  $u^k$  et  $p^k$ , on tire suivant la loi  $P$ , un  $\omega^{k+1} \in \Omega$ , indépendant des tirages précédents, et on calcule
  - $u^{k+1}$  par

$$\min_{u \in U^{\text{ad}}} K(u) + (\epsilon^k r^k - K'(u^k), u) + \epsilon^k \langle p^k, \theta(u) \rangle \quad (3.13)$$

avec  $r^k$  appartenant au sous-différentiel de  $j$  au point  $(u^k, \omega^{k+1})$ .

- $p^{k+1}$  par  $p^{k+1} = \Pi(p^k + \rho\theta(u^{k+1}))$

$\Pi$  étant maintenant la projection sur  $B(0, M) \cap C'$

Une première justification heuristique de cet algorithme, dans le cas  $U = U^{\text{ad}}$ ,  $C = \{0\}$  et  $\theta$  affine, peut être obtenue par un argument de "perturbations singulières" :

les équations récurrentes de l'algorithme sont

$$u^{k+1} = u^k - \epsilon^k r^k - \epsilon^k \theta^t p^k$$

$$p^{k+1} = \Pi(p^k + \rho\theta(u^{k+1}))$$

$\Pi$  étant maintenant la projection sur  $B(0, M)$  seule. Nous avons donc une variable "rapide"  $p^k$  et une variable "lente"  $u^k$ . La deuxième récurrence peut donc être considérée à  $u^k = u$  fixé et la mesure invariante associée est une mesure de Dirac sur le point

$$\tilde{p} = M \frac{\theta u}{\|\theta u\|}$$

En substituant maintenant cette valeur limite dans la première équation,

$$u^{k+1} = u^k - \epsilon^k r^k - \epsilon^k M \theta^t \frac{\theta u^k}{\|\theta u^k\|}$$

Or cette récurrence n'est pas sans analogie avec une itération de sous-gradient avec pénalisation "exacte" sur la contrainte, c'est-à-dire une pénalisation de la forme

$M\|\theta(u)\|$  où  $M \geq p^*$  pour au moins un  $p^*$ . On est donc amené à penser que les variables  $u^k$  convergent vers la solution du problème de départ.

Nous énonçons et démontrons donc le théorème suivant.

**Théorème 9** Avec les mêmes hypothèses que le théorème précédent excepté que  $J(\cdot)$  est maintenant fortement convexe de rapport  $a > 0$  et que  $(\epsilon^k)_{k \in \mathbb{N}}$  est décroissante, si  $\rho < \frac{2a}{L_\theta^2}$ , alors la suite  $(u^k)_{k \in \mathbb{N}}$  converge fortement P-p.s. vers  $u^*$ , solution unique de (3.1).

### 3.3.1 Inéquations variationnelles et fonction de Lyapunov

L'inéquation variationnelle associée au problème auxiliaire est encore (3.5) :

$$(K'(u^{k+1}) - K'(u^k) + \epsilon^k r^k, u - u^{k+1}) + \epsilon^k \langle p^k, \theta(u) - \theta(u^{k+1}) \rangle \geq 0, \quad \forall u \in U^{ad}$$

Les inégalités de point-selle associées au problème (3.1) sont

$$\forall (v, q) \in U^{ad} \times C', \quad J(u^*) + \langle q, \theta(u^*) \rangle \leq J(u^*) + \langle p^*, \theta(u^*) \rangle \leq J(v) + \langle p^*, \theta(v) \rangle$$

On introduit la nouvelle fonction de Lyapunov

$$\psi^k = K(u^*) - K(u^k) - (K'(u^k), u^* - u^k) + \frac{\epsilon^k}{2\rho} \|p^k - p^*\|^2$$

Le théorème résulte des variations de la suite  $(\psi^k)_{k \in \mathbb{N}}$  le long des trajectoires de l'algorithme.

### 3.3.2 Majorations préliminaires

**Lemme 7** Par définition de  $p^{k+1}$  et  $p^*$ ,

$$\frac{\epsilon^{k+1}}{\rho} \|p^{k+1} - p^*\|^2 \leq \frac{\epsilon^k}{\rho} \|p^k - p^*\|^2 + \epsilon^k \rho L_\theta^2 \|u^{k+1} - u^*\|^2 + 2\epsilon^k \langle p^k - p^*, \theta(u^{k+1}) - \theta(u^*) \rangle \quad (3.14)$$

Il existe des constantes positives  $c_1$  telles que,

$$\|u^{k+1} - u^k\| \leq \epsilon^k (c_1 \psi^k + c_2) \quad (3.15)$$

Enfin, il existe des constantes positives  $c_3$  telles que,

$$\begin{aligned} \psi^{k+1} - \psi^k &\leq \epsilon^k [(r^k, u^* - u^k) + \langle p^*, \theta(u^*) - \theta(u^k) \rangle] \\ &\quad + (\epsilon^k)^2 (c_3 \psi^k + c_4) + \frac{1}{2} \epsilon^k \rho L_\theta^2 \|u^{k+1} - u^*\|^2 \end{aligned} \quad (3.16)$$

(3.14) provient de l'équation (vraie puisque  $p^* \in B(0, M)$ )

$$\forall \rho, p^* = \Pi(p^* + \rho\theta(u^*))$$

que l'on a retranchée de la définition de  $p^{k+1}$ , élevée au carré et développée en tenant compte du caractère lipchitz de  $\theta$  et de  $\Pi$ . On a, de plus, utilisé le fait que  $(\epsilon^k)_{k \in \mathbb{N}}$  est décroissante.

(3.15) se démontre comme suit :

Par application de l'inégalité de Schwarz,

$$\epsilon^k \|r^k\| \|u^{k+1} - u^k\| \geq (\epsilon^k r^k, u^k - u^{k+1})$$

et en utilisant le caractère lipchitzien de  $\theta$ ,

$$\epsilon^k L_\theta \|p^k\| \|u^{k+1} - u^k\| \geq \epsilon^k \langle p^k, \theta(u^k) - \theta(u^{k+1}) \rangle$$

d'où, avec l'inéquation variationnelle en  $u = u^k$ ,

$$\begin{aligned} b \|u^{k+1} - u^k\|^2 &\leq (K'(u^k) - K'(u^{k+1}), u^k - u^{k+1}) \\ &\leq \epsilon^k \|u^{k+1} - u^k\| [\|r^k\| + L_\theta \|p^k\|] \end{aligned}$$

enfin, en divisant les extrémités de gauche et de droite par  $\|u^{k+1} - u^k\|$  supposé non nul sans perte de généralité, et comme  $\|p^k\| \leq M$ ,

$$b \|u^{k+1} - u^k\| \leq \epsilon^k (\|r^k\| + L_\theta \|p^k\|) \leq \epsilon^k (\|r^k\| + M L_\theta)$$

ce qui, par hypothèse *SGLB*, et par définition de  $\psi$  implique l'existence de constantes positives  $c_1$  et  $c_2$  qui satisfont (3.15).

Pour montrer (3.16), on évalue  $\psi^{k+1} - \psi^k$  qui s'écrit :

$$\begin{aligned} &\overbrace{K(u^k) - K(u^{k+1}) - (K'(u^k), u^k - u^{k+1})}^{A_1} \\ &+ \underbrace{(K'(u^k) - K'(u^{k+1}), u^* - u^{k+1})}_{A_2} + \underbrace{\frac{\epsilon^{k+1}}{2\rho} \|p^{k+1} - p^*\|^2 - \frac{\epsilon^k}{2\rho} \|p^k - p^*\|^2}_{A_3} \end{aligned}$$

$A_1$  est majoré par  $-\frac{b}{2} \|u^{k+1} - u^k\|^2$  car  $K$  est fortement convexe de rapport  $b$ .

A l'aide de l'inéquation variationnelle exprimée en  $u = u^*$ ,

$$(K'(u^{k+1}) - K'(u^k) + \epsilon^k r^k, u^* - u^{k+1}) + \epsilon^k \langle p^k, \theta(u^*) - \theta(u^{k+1}) \rangle \geq 0$$

on majore  $A_2$  :

$$A_2 \leq \epsilon^k [(r^k, u^* - u^{k+1}) + \langle p^k, \theta(u^*) - \theta(u^{k+1}) \rangle]$$

$A_3$  est, d'après (3.14), majoré par

$$A_3 \leq \frac{1}{2} \epsilon^k \rho L_\theta^2 \|u^{k+1} - u^*\|^2 + \epsilon^k < p^k - p^*, \theta(u^{k+1}) - \theta(u^*) >$$

On a donc

$$A_1 + A_2 + A_3 \leq \epsilon^k [(r^k, u^* - u^k) + < p^*, \theta(u^*) - \theta(u^k) >]$$

$$- \frac{b}{2} \|u^{k+1} - u^k\|^2 + \frac{1}{2} \epsilon^k \rho L_\theta^2 \|u^{k+1} - u^*\|^2 + \epsilon^k (r^k, u^k - u^{k+1}) + \epsilon^k < p^*, \theta(u^k) - \theta(u^{k+1}) >$$

Pour  $\alpha > 0$ , une majoration élémentaire conduit, comme  $\|p^*\| \leq M$ , à

$$\begin{aligned} & \epsilon^k [(r^k, u^{k+1} - u^k) + < p^*, \theta(u^k) - \theta(u^{k+1}) >] \\ & \leq \|u^{k+1} - u^k\| \quad \epsilon^k (\|r^k\| + L_\theta M) \\ & \leq \frac{\alpha}{2} \|u^{k+1} - u^k\|^2 + \frac{(\epsilon^k)^2}{2\alpha} (\|r^k\| + L_\theta M)^2 \end{aligned}$$

Soit, par hypothèse *SGLB* sur  $J$  et pour  $\alpha \leq b$ , il existe des constantes  $c$ , telles que (3.16) soit vérifiée.

### 3.3.3 Convergence de la suite $\psi^k$

$y^k := E(\psi^k)$  est bornée

Prenons l'espérance conditionnelle sachant  $\mathbf{F}^k$  de (3.16). Comme  $u^k$  et  $\omega^{k+1}$  sont indépendants,

$$\begin{aligned} \mathbf{E}^k \psi^{k+1} - \psi^k & \leq \epsilon^k [(E^k r^k, u^* - u^k) + < p^*, \theta(u^*) - \theta(u^k) >] \\ & \quad + (\epsilon^k)^2 (c_3 \psi^k + c_4) + \frac{1}{2} \epsilon^k \rho L_\theta^2 \|u^{k+1} - u^*\|^2 \end{aligned}$$

$\bar{r}^k := E^k r^k = \int_{\Omega} r(u^k, \omega) P(d\omega)$  est un élément de  $\partial J(u^k)$  et donc, par forte convexité,

$$(\bar{r}^k, u^* - u^k) \leq J(u^*) - J(u^k) - \frac{a}{2} \|u^k - u^*\|^2$$

Comme de plus

$$L(u^*, p^*) - L(u^k, p^*) \leq -\frac{a}{2} \|u^k - u^*\|^2,$$

il reste

$$\begin{aligned} \mathbf{E}^k \psi^{k+1} - \psi^k & \leq (\epsilon^k)^2 (c_3 \psi^k + c_4) + \epsilon^k \left( \frac{\rho}{2} L_\theta^2 \|u^k - u^*\|^2 - a \|u^k - u^*\|^2 \right) \\ & \quad + \epsilon^k a (\|u^{k+1} - u^*\|^2 - \|u^k - u^*\|^2) \end{aligned}$$

Si  $\rho$  est suffisamment petit, c'est-à-dire

$$\rho < \frac{2a}{L_\theta^2}$$

on a

$$\mathbf{E}^k \psi^{k+1} - \psi^k \leq (\epsilon^k)^2 (c_3 \psi^k + c_4) + \epsilon^k a (\|u^{k+1} - u^*\|^2 - \|u^k - u^*\|^2)$$

Or

$$\begin{aligned} a \epsilon^k (\|u^{k+1} - u^*\|^2 - \|u^k - u^*\|^2) &= a \epsilon^k (u^{k+1} - u^k, u^{k+1} + u^k - 2u^*) \\ &\leq \frac{a}{2} \|u^{k+1} - u^k\|^2 + \frac{(\epsilon^k)^2}{2} \|(u^{k+1} - u^*) + (u^k - u^*)\|^2 \end{aligned}$$

En majorant le premier terme d'après (3.15) et le deuxième en application de  $(a+b)^2 \leq 2(a^2 + b^2)$ , il existe des constantes  $c_i$  telles que

$$\mathbf{E}^k \psi^{k+1} - \psi^k \leq (\epsilon^k)^2 (c'_3 \psi^k + c'_4 \psi^{k+1} + c'_5)$$

En prenant maintenant l'espérance au sens de  $\mathbf{P}$ ,

$$\forall k \in \mathbf{N}, y^{k+1} \leq y^k + (\epsilon^k)^2 (c'_3 y^k + c'_4 y^{k+1} + c'_5)$$

Par application du lemme 1, on en déduit que  $(y^k)_{k \in \mathbf{N}}$  est bornée.

Le reste de la démonstration est le même que celui de la démonstration précédente, à la différence près que l'on obtient directement

$$\lim_{k \rightarrow +\infty} \|u^k - u^*\|^2 = 0$$

Remarques :

On retrouve la même majoration sur  $\rho$  que dans les résultats déterministes exposés dans Cohen [1].

On pourrait se demander si la **forte** convexité de  $J$  est nécessaire. Il semble que ce soit le cas. En effet, la fonction duale associée  $G(p) = \min_u \mathcal{L}(u, p)$ , tout en restant différentiable, n'aurait plus forcément un gradient lipchitzien, ce qui imposerait l'utilisation de "petits pas" pour les itérations en  $p$ .

### 3.4 J non strictement convexe : Lagrangien augmenté

Pour le problème (3.1), on recherche maintenant des points-selles du Lagrangien augmenté  $L_c(u, p)$  sur  $U^{ad} \times C'$  dans le but d'appliquer le résultat 12.

On rappelle que (résultat 11)

$$L_c(u, p) = J(u) + \Phi(\theta(u), p)$$

avec

- $\Phi(\theta, p) = \frac{1}{2c} [\|\Pi(p + c\theta)\|^2 - \|p\|^2]$
- $\Phi'_p(\theta(u), p) = \frac{1}{c} [\Pi(p + c\theta(u)) - p]$
- $\Phi'_\theta(\theta(u), p) = \Pi(p + c\theta(u))$

$\Pi$  étant la projection sur  $C'$

Afin de permettre la décomposition du problème avec Lagrangien augmenté, on propose, en suivant Cohen & Zhu[1], l'algorithme suivant :

#### Algorithme 9

- soit  $u^0 \in U^{ad}$  et  $p^0 \in C'$ ,
- connaissant  $u^k$  et  $p^k$ , on tire, suivant la loi  $P$ , un  $\omega^{k+1} \in \Omega$  indépendant des tirages précédents, et on calcule  $u^{k+1}$  par

$$\min_{u \in U^{ad}} K(u) + (\epsilon^k r^k - K'(u^k), u) + \epsilon^k < q^k, \theta(u) > \quad (3.17)$$

avec  $q^k = \Phi'_\theta(\theta(u^k), p^k) = \Pi(p^k + c\theta(u^k))$  et  $r^k \in \partial j(u^k, \omega^{k+1})$ .  
et  $p^{k+1}$  par

$$p^{k+1} = p^k + \epsilon^k \Phi'_p(\theta(u^{k+1}), p^k) \quad (3.18)$$

On notera que (3.18) s'écrit encore

$$p^{k+1} = p^k + \frac{\epsilon^k}{c} (q^{k+\frac{1}{2}} - p^k) \quad \text{avec} \quad q^{k+\frac{1}{2}} = \Pi(p^k + c\theta(u^{k+1}))$$

Remarque :

Par définition de  $p^k$ , si l'on ne peut initialiser  $p^0$  dans  $C'$ , la suite  $(p^k)_{k \in \mathbf{N}}$  n'appartient pas nécessairement à  $C'$ , contrairement à  $q^k$

**Théorème 10** *On fait les mêmes hypothèses sur  $J$ ,  $j$ ,  $K$  et  $\theta$  que pour le théorème précédent, excepté la stricte convexité de  $J$*

*Alors la suite  $(u^k)_{k \in \mathbb{N}}$  est bornée p.s. et tous ses points d'accumulation appartiennent à  $U^*$ . La suite  $(p^k)_{k \in \mathbb{N}}$  est bornée p.s. Tout point d'accumulation  $\bar{p}$  est tel que  $\bar{p} \in \hat{P}(u^*)$ ,  $\forall u^* \in U^*$ .*

### 3.4.1 Inéquations variationnelles et fonction de Lyapunov

L'inéquation variationnelle associée à (3.17) est

$$(K(u^{k+1}) - K(u^k) + \epsilon^k r^k, u - u^{k+1}) + \epsilon^k < q^k, \theta(u) - \theta(u^{k+1}) > \geq 0, \quad \forall u \in U^{ad}$$

et celle associée au problème (3.1) est :

$$\forall (u, p) \in U^{ad} \times C', J(u^*) + \Phi(\theta(u^*), p) \leq J(u^*) + \Phi(\theta(u^*), p^*) \leq J(u) + \Phi(\theta(u), p^*)$$

On choisit la fonction de Lyapunov

$$\psi^k = K(u^*) - K(u^k) - (K'(u^k), u^* - u^k) + \frac{1}{2} \|p^k - p^*\|^2$$

### 3.4.2 Majorations préliminaires

**Lemme 8** *Il existe des constantes positives  $a_i$  telles que*

$$\begin{aligned} \|p^k - p^*\|^2 &\leq 2\psi^k & \|u^k - u^*\|^2 &\leq \frac{2}{b}\psi^k \\ \|q^k\| &\leq (a_1 + a_2\sqrt{\psi^k}) & \|p^k - q^{k+\frac{1}{2}}\| &\leq (a_3 + a_4\sqrt{\psi^k} + a_5\sqrt{\psi^{k+1}}) \end{aligned} \quad (3.19)$$

*Il existe des constantes positives  $c'_i$  telles que*

$$\|u^{k+1} - u^k\| \leq \epsilon^k (c'_1\sqrt{\psi^k} + c'_2), \quad \|p^{k+1} - p^k\| \leq \epsilon^k (c'_3\sqrt{\psi^k} + c'_4\sqrt{\psi^{k+1}} + c'_5), \quad (3.20)$$

$$\|q^{k+\frac{1}{2}} - q^k\| \leq \epsilon^k (c'_6\sqrt{\psi^k} + c'_7)$$

*Il existe des constantes positives  $c''_1, c''_2, c''_3$  telles que*

$$\psi^{k+1} - \psi^k \leq \epsilon^k [(r^k, u^* - u^k) + \Phi(\theta(u^*), p^k) - \Phi(\theta(u^k), p^*)] + (\epsilon^k)^2 [c''_1\psi^k + c''_2 + c''_3\psi^{k+1}] \quad (3.21)$$

(3.19) vient immédiatement des définitions des quantités majorées et de  $\psi$ . En particulier, on majore  $\|q^{k+\frac{1}{2}} - p^k\|$  de la façon suivante :<sup>(2)</sup>

$$\|q^{k+\frac{1}{2}} - p^k\| \leq 2\|p^k\| + c\|\theta(u^{k+1})\|$$

et comme  $\theta$  est lipchitz,

$$\leq 2\|p^k - p^*\| + 2\|p^*\| + cL_\theta\|u^{k+1} - u^*\| + c\|\theta(u^*)\|$$

<sup>2</sup>On rappelle que  $p^k$  n'appartient pas nécessairement à  $C'$



Par définition de  $\psi$  il existe des constantes positives telles que

$$\|q^{k+\frac{1}{2}} - p^k\| \leq a_3 + a_4\sqrt{\psi^k} + a_5\sqrt{\psi^{k+1}}$$

(3.20) se montre exactement comme (3.8) (inéquation variationnelle du problème auxiliaire exprimée en  $u = u^k$ , inégalité de Schwarz et  $\theta$  lipchitzienne), en utilisant la majoration de  $\|q^k\|$  de (3.19).

Ensuite,  $\|q^{k+\frac{1}{2}} - q^k\|$  se majore par

$$\begin{aligned} \|q^{k+\frac{1}{2}} - q^k\| &\leq \|\Pi(p^k + c\theta(u^{k+1})) - \Pi(p^k + c\theta(u^k))\| \\ &\leq c\|\theta(u^{k+1}) - \theta(u^k)\| \leq L_\theta\|u^{k+1} - u^k\| \end{aligned}$$

et  $\|p^{k+1} - p^k\|$  se majore en utilisant la définition de  $p^{k+1}$  et la majoration de  $\|q^{k+\frac{1}{2}} - p^k\|$  (3.19).

Démontrons (3.21) :

$\psi^{k+1} - \psi^k$  vaut :

$$\begin{aligned} &\overbrace{K(u^k) - K(u^{k+1}) - (K'(u^k), u^k - u^{k+1})}^{A_1} \\ &+ \underbrace{(K'(u^k) - K'(u^{k+1}), u^* - u^{k+1})}_{A_2} + \underbrace{\frac{1}{2}(\|p^{k+1} - p^*\|^2 - \|p^k - p^*\|^2)}_{A_3} \end{aligned}$$

$A_1$  est négatif car  $K$  est convexe.

Pour  $A_2$ , d'après l'inéquation caractérisant  $u^{k+1}$ , exploitée en  $u = u^*$ ,

$$\begin{aligned} A_2 &= (K'(u^k) - K'(u^{k+1}), u^* - u^{k+1}) \leq (\epsilon^k r^k, u^* - u^k) + (\epsilon^k r^k, u^k - u^{k+1}) \\ &\quad + \epsilon^k < q^k, \theta(u^*) - \theta(u^k) > + \epsilon^k < q^k, \theta(u^{k+1}) - \theta(u^k) > \end{aligned}$$

D'après la convexité de  $\Phi$  en  $\theta$  (résultat 11),

$$< q^k, \theta(u^*) - \theta(u^k) > \leq \Phi(\theta(u^*), p^k) - \Phi(\theta(u^k), p^k)$$

car  $q^k = \Phi'_\theta(\theta(u^k), p^k)$ .

Soit encore,

$$\begin{aligned} A_2 &\leq (\epsilon^k r^k, u^* - u^k) + \epsilon^k(\Phi(\theta(u^*), p^k) - \Phi(\theta(u^k), p^k)) \\ &\quad + \epsilon^k(\|r^k\| + \|q^k\|L_\theta) \|u^{k+1} - u^k\| \end{aligned}$$

D'autre part,

$$A_3 = \frac{\epsilon^k}{c}(q^k - p^k, p^k - p^*) + \frac{\epsilon^k}{c}(q^{k+\frac{1}{2}} - q^k, p^k - p^*) + \frac{1}{2}\left(\frac{\epsilon^k}{c}\right)^2\|q^{k+\frac{1}{2}} - p^k\|^2$$

puis, par concavité de  $\Phi$  en  $p$ , en notant que  $\Phi'_p(\theta(u^k), p^k) = \frac{q^k - p^k}{c}$ ,

$$\frac{\epsilon^k}{c}(q^k - p^k, p^k - p^*) \leq \epsilon^k(\Phi(\theta(u^k), p^k) - \Phi(\theta(u^k), p^*))$$

et donc

$$A_3 \leq \epsilon^k(\Phi(\theta(u^k), p^k) - \Phi(\theta(u^k), p^*)) + \frac{\epsilon^k}{c}\|q^{k+\frac{1}{2}} - q^k\| \cdot \|p^k - p^*\| + \frac{1}{2}\left(\frac{\epsilon^k}{c}\right)^2\|q^{k+\frac{1}{2}} - p^k\|^2$$

En utilisant l'hypothèse de *SGLB* pour  $r^k$  et les majorations (3.19) et (3.20), il existe des constantes positives  $c_i''$  telles que

$$\begin{aligned} \psi^{k+1} - \psi^k &\leq A_1 + A_2 + A_3 \\ &\leq \epsilon^k[(r^k, u^* - u^k) + \Phi(\theta(u^*), p^k) - \Phi(\theta(u^k), p^*)] + (\epsilon^k)^2[c_1''\psi^k + c_2'' + c_3''\psi^{k+1}] \end{aligned}$$

### 3.4.3 Convergence de la suite $\psi^k$

$y^k := E(\psi^k)$  est une suite positive bornée.

Prenons l'espérance conditionnelle sachant  $\mathbf{F}^k$  de (3.21). Comme  $u^k$  et  $\omega^{k+1}$  sont indépendants,

$$\begin{aligned} \mathbf{E}^k \psi^{k+1} - \psi^k &\leq \epsilon^k[(\mathbf{E}^k r^k, u^* - u^k) + \Phi(\theta(u^*), p^k) - \Phi(\theta(u^k), p^*)] \\ &\quad + (\epsilon^k)^2(c_1''\psi^k + c_2'' + c_3''\mathbf{E}^k \psi^{k+1}) \end{aligned}$$

Il est clair que  $\bar{r}^k := \mathbf{E}^k r^k = \int_{\Omega} r(u^k, \omega)P(d\omega)$  est un élément de  $\partial J(u^k)$  et donc

$$(\bar{r}^k, u^* - u^k) \leq J(u^*) - J(u^k)$$

soit

$$\begin{aligned} \mathbf{E}^k \psi^{k+1} - \psi^k &\leq \epsilon^k[(J(u^*) - J(u^k)) + \Phi(\theta(u^*), p^k) - \Phi(\theta(u^k), p^*)] \\ &\quad + (\epsilon^k)^2(c_1''\psi^k + c_2'' + c_3''\mathbf{E}^k \psi^{k+1}) \end{aligned}$$

ou, en termes de Lagrangien  $L_c$ ,

$$\mathbf{E}^k \psi^{k+1} - \psi^k \leq \epsilon^k[L_c(u^*, p^k) - L_c(u^k, p^*)] + (\epsilon^k)^2(c_1''\psi^k + c_2'' + c_3''\mathbf{E}^k \psi^{k+1}) \quad (3.22)$$

En combinant les deux inégalités de point-selle du Lagrangien augmenté, on a

$$L_c(u^*, p^k) - L_c(u^k, p^*) \leq 0,$$

et donc

$$\mathbf{E}^k \psi^{k+1} - \psi^k \leq \alpha_k \psi^k + \beta_k + \gamma_k \mathbf{E}^k \psi^{k+1} \quad (3.23)$$

avec  $\alpha_k, \beta_k$  et  $\gamma_k$  suites positives sommables ( $(\epsilon^k)_{k \in \mathbf{N}}$  est une  $\sigma$ -suite).

En prenant maintenant l'espérance au sens de  $\mathbf{P}$ ,

$$\forall k \in \mathbf{N}, \quad y^{k+1} - y^k \leq \alpha_k y^k + \beta_k + \gamma_k y^{k+1}$$

et d'après le lemme 1,  $(y^k)_{k \in \mathbf{N}}$  est une suite bornée. On note  $M$  l'un de ses majorants.

$\psi^k$  converge p.s. au sens de  $\mathbf{P}$

$\psi^k$  est une quasi-martingale (propriété 18).

En effet,  $(\psi^k)_{k \in \mathbf{N}}$  vérifie  $\inf_{k \in \mathbf{N}} E(\psi^k) > -\infty$  ( $y^k$  est positive) et

$$E^k(\psi^{k+1} - \psi^k) \leq \alpha_k \psi^k + \beta_k + \gamma_k E^k \psi^{k+1}$$

on déduit, comme  $y^k$  est bornée par  $M$  :

$$\sum_{k \in \mathbf{N}} E(E^k \psi^{k+1} - \psi^k)^+ \leq \sum_{k \in \mathbf{N}} (\alpha_k y^k + \beta_k + \gamma_k y^{k+1}) \leq \sum_{k \in \mathbf{N}} (M[\alpha_k + \gamma_k] + \beta_k) < \infty$$

Donc  $\psi^k$  est une quasi-martingale et, d'après le résultat 14, elle converge presque-sûrement vers une variable aléatoire d'espérance finie.

### 3.4.4 Limites de $L_c(u^*, p^k) - L_c(u^k, p^*)$ et de $u^k$

$(u^k)_{k \in \mathbf{N}}$  admet p.s. des points d'accumulation faible.

Comme  $\psi^k$  converge p.s. vers une variable aléatoire d'espérance finie, il existe un ensemble  $\Omega_0$  de mesure nulle tel que

$$\forall \tilde{\omega} \in \Omega \setminus \Omega_0, \lim_{k \rightarrow \infty} \psi(u^k(\tilde{\omega})) < \infty$$

et donc en vertu de (3.19),

$$\|u^k - u^*\|^2 \leq \frac{2}{b} \sup_k \psi^k < \infty$$

Soit un tel  $\tilde{\omega}$ , la suite  $u^k(\tilde{\omega})$  est bornée et admet d'après le théorème de compacité faible des points d'accumulation faible dans  $U^{sd}$

$(L_c(u^*, p^k) - L_c(u^k, p^*))_{k \in \mathbf{N}}$  converge p.s. vers 0

(application du lemme 2)

L'équation (3.22) se réécrit

$$e^k (L_c(u^k, p^*) - L_c(u^*, p^k)) \leq \psi^k (1 + \alpha_k) - (1 - \gamma_k) E^k \psi^{k+1} + \beta_k$$

en prenant l'espérance au sens de  $\mathbf{P}$ ,

$$e^k E[L_c(u^k, p^*) - L_c(u^*, p^k)] \leq y^k (1 + \alpha_k) - y^{k+1} (1 - \gamma_k) + \beta_k$$

soit par sommation sur  $k$  :

$$\sum_{k=1}^n e^k E[L_c(u^k, p^*) - L_c(u^*, p^k)] \leq \sum_{k=1}^n (y^k (1 + \alpha_k) - y^{k+1} (1 - \gamma_k) + \beta_k)$$

soit, comme  $(y^k)_{k \in \mathbf{N}}$  est bornée par  $M$ , en réarrangeant les termes :

$$\sum_{k=1}^n \epsilon^k \mathbf{E}[L_c(u^k, p^*) - L_c(u^*, p^k)] \leq \sum_{k=1}^n (M(\alpha_k + \gamma_k) + \beta_k) + y^1 - y^{n+1} \leq M' < \infty$$

Les termes  $L_c(u^k, p^*) - L_c(u^*, p^k)$  étant positifs, on a

$$\sum_{k \in \mathbf{N}} \epsilon^k [L_c(u^k, p^*) - L_c(u^*, p^k)] < \infty \quad \mathbf{P} - p.s.$$

On appellera  $\Omega_1$  l'ensemble de mesure nulle sur lequel cette majoration n'est pas vérifiée.

D'autre part,  $\psi^k$  bornée p.s. implique, d'après (3.20) et (3.19) on a presque-sûrement (à  $\Omega_0$  près)

$$\|u^{k+1} - u^k\| \leq \epsilon^k M'' \quad , \quad \|p^{k+1} - p^k\| \leq \epsilon^k M'' \quad , \quad \|p^k\| \leq M'''$$

Pour tout  $\tilde{\omega}$  n'appartenant ni à  $\Omega_0$  ni à  $\Omega_1$ , les conditions du lemme 2 sont vérifiées, notamment le caractère lipchitzien de  $L_c$  sur le borné contenant  $(u^k, p^k)_{k \in \mathbf{N}}$ . Pour  $\tilde{\omega}$ , la suite  $L_c(u^k(\tilde{\omega}), p^*) - L_c(u^*, p^k)$  converge vers 0

Cela implique notamment que

$$\lim_k L_c(u^k, p^*) = L_c(u^*, p^*), \quad \lim_k L_c(u^*, p^k) = L_c(u^*, p^*)$$

**Les points d'accumulation de  $(u^k)_{k \in \mathbf{N}}$  appartiennent à  $U^*$  et ceux de  $(p^k)_{k \in \mathbf{N}}$  appartiennent à  $\bigcap_{u^* \in U^*} \hat{P}(u^*)$ .**

En effet,  $(u^k)$  et  $(p^k)$  sont deux suites bornées. Elles admettent donc des points d'accumulation faible p.s. Soit  $(\tilde{u}, \tilde{p})$  un couple de points d'accumulation et  $\phi(k)$  la suite d'indices de la sous-suite de  $(u^k, p^k)$  qui converge faiblement vers ce couple.  $L_c$  étant faiblement s.c.i. en  $u$  et faiblement s.c.s en  $p$ , d'une part,

$$L_c(\tilde{u}, p^*) \leq \liminf_{k \rightarrow \infty} L_c(u^{\phi(k)}, p^*) = L_c(u^*, p^*)$$

puisque  $U^{ad}$  est faiblement fermé,  $\tilde{u} \in U^{ad}$  et donc  $\tilde{u} \in \hat{U}(p^*)$  et par stabilité de  $L_c$  en  $u$ ,  $\tilde{u} \in U^*$

D'autre part,

$$L_c(u^*, \tilde{p}) \geq \limsup_{k \rightarrow \infty} L_c(u^*, p^{\phi(k)}) = L_c(u^*, p^*)$$

ce qui implique que

$$\tilde{p} \in \hat{P}(u^*)$$

ce dernier résultat étant vrai pour tout  $u^* \in U^*$

Ce qui précède est valable pour tout  $\tilde{\omega} \in \Omega \setminus \Omega_0 \cup \Omega_1$ .

### 3.5 Comparaison avec des résultats antérieurs

Nous comparons ici avec les résultats de Kushner & Clark [1] et Bertran [1] qui sont les seules publications, à notre connaissance, qui proposent des solutions aux problèmes de minimisation stochastique sous contraintes explicites. Remarquons à nouveau que des hypothèses de différentiabilité de  $J$  et de dérivée lipchitzienne sont faites dans ces deux références.

#### 3.5.1 Bertran (1975)

Nous sommes un peu embarrassé pour comparer notre approche de l'algorithme de Uzawa stochastique avec celle de Bertran. Il semble qu'il démontre la convergence des quantités primales et duales, alors que nous avons montré théoriquement et expérimentalement que ni les unes ni les autres n'ont de chance de converger (voir la section "tests numériques") ! La raison de cette différence vient des hypothèses sur le bruit qu'introduit Bertran. Ces hypothèses sont difficilement vérifiables dans la réalité et ne sont de toute façon pas vérifiées dans le contre-exemple linéaire quadratique que nous avons donné. Nous restons sceptiques quant à l'utilisation d'un algorithme d'Uzawa dans un cadre non déterministe. De plus la technique de démonstration fait intervenir un formalisme assez compliqué qui nous paraît démesuré avec le but recherché.

En revanche, pour l'algorithme de Arrow-Hurwicz stochastique, nous aboutissons à des conclusions comparables (un peu plus fortes), mais avec des hypothèses nettement moins restrictives. La convergence des quantités duales vers le multiplicateur déterministe optimal semble poser autant de problème à Bertran qu'à nous. A la lumière de ces remarques, nous ne citerons que le résultat sur l'algorithme de Arrow-Hurwicz.

#### Hypothèses :

1.  $U$  Hilbert,  $U^{ad}$  est un convexe fermé de  $U$
2.  $J$  est  $C^2$  et fortement convexe.
3. le gradient de  $J$ ,  $\nabla J$  est un opérateur lipchitzien.
4. hypothèses de qualification de la contrainte  $\theta$
5.  $r^k = \nabla J(u^k) + Z^k$  avec  $Z^k$  réalisation d'un "bruit" sur la mesure de  $\nabla J(u^k)$ . Ces réalisations suivent la loi  $P$  et sont indépendantes entre elles.
6.  $E[\theta(u + Z^k) = \theta(u)]$  et  $E(\|Z(u)\|) \leq B^2$ .
7.  $(\epsilon^k)_{k \in \mathbb{N}}$  est une  $\sigma$ -suite.

#### Résultat (théorème 5.5.1)

Sous les hypothèses précédentes,  $(u^k)_{k \in \mathbb{N}}$  admet une sous-suite qui converge en moyenne

quadratique vers  $u^*$ , unique solution du problème.

### Remarques

Avec les hypothèses précédentes, nous savons montrer la convergence forte presque-sûre de la suite  $(u^k)_{k \in \mathbf{N}}$  même lorsque  $J$  n'est que sous-différentiable. D'autre part, l'hypothèse 6. nous semble supposer une certaine forme de linéarité pour  $\theta$ .

### 3.5.2 Kushner & Clark (1978)

L'approche est à nouveau celle de l'o.d.e. Les cas considérés sont l'algorithme de Arrow-Hurwicz et des méthodes de pénalisation, dans un espace de dimension finie.

#### Arrow-Hurwicz : théorème 5.2.1

L'algorithme est similaire au nôtre, à la différence près que les variables  $u^k$  et  $p^k$  sont projetées dans des boules de rayons finis afin d'assurer la bornitude. Les hypothèses sur le bruit sont équivalentes à celles du théorème 2.3.1.  $\theta$  est supposée  $C^1$  et convexe.  $J$  est supposée strictement convexe. L'énoncé du théorème est complexe, compte-tenu du formalisme employé pour montrer la convergence, mais conclut sur la convergence de la suite  $(u^k, p^k)$  vers le point-selle du problème de départ.

#### Pénalisation : théorèmes 5.1.1 et 5.4.1

En plus des hypothèses habituelles sur les bruits, on suppose que  $J$  est  $C^1$  et  $\theta'(u)$  surjective. Dans les deux cas, on conclut que  $\lim_k \theta(u^k) \in -C$ .

En conclusion, l'originalité de notre approche semble être sa simplicité et sa plus grande généralité :  $J$  non strictement convexe ni différentiable, lagrangien augmenté, espace de Hilbert.

### 3.6 Tests numériques

On teste ici les trois algorithmes proposés dans ce chapitre.

On rappelle que

$$j(u, \omega) = \frac{1}{2}[(u_1 - \omega_1)^2 + \omega_2(u_2 - 1)^2 + (\omega_3 u_3 + \omega_4 u_4 - 1)^2]$$

$$\theta(u) = u_1 + u_2 + u_3 + u_4, \quad c = 6$$

$$J(u) = \frac{1}{2}[u_1^2 - u_1 + \frac{1}{2}(1 - u_2)^2 + \frac{1}{2}u_3 u_4 + \frac{1}{3}(u_3^2 + u_4^2) - u_3 - u_4] + \frac{2}{3}$$

Pour la décomposition par les prix, on choisit de prendre la décomposition naturelle de  $u$  en quatre sous-systèmes.

$$u = (u_1, u_2, u_3, u_4)$$

Les valeurs optimales sont :

$$u_1^* = \frac{14}{15} \simeq 0,933 \quad u_2^* = \frac{28}{15} \simeq 1,867$$

$$u_3^* = u_4^* = \frac{8}{5} = 1,6$$

$$p^* = -\frac{13}{30} \simeq -0,433$$

et le critère optimal est  $J(u^*) = \frac{43}{60} \simeq 0,716$

Les trois algorithmes sont testés sur 1000 itérations. Les valeurs des  $\epsilon^k$  optimisés sont indiquées sur les figures. Les points de départ sont

$$u_1 = u_3 = 5, \quad u_2 = u_4 = -2, \quad p = 2$$

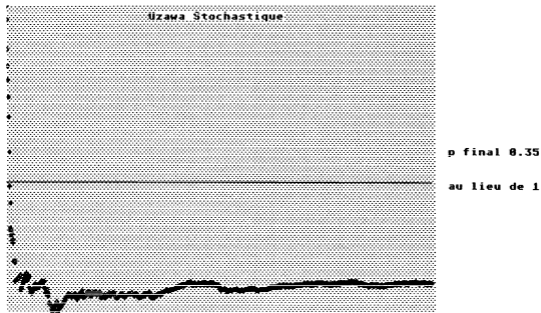
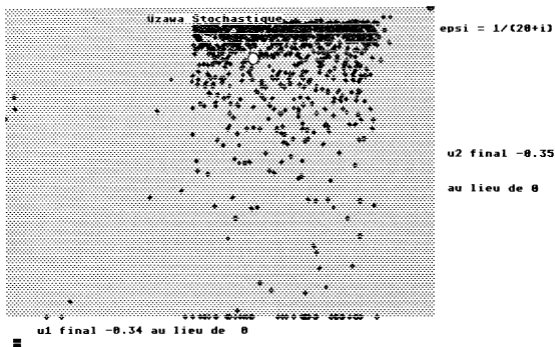
L'algorithme 6 est appelé "Uzawa Stochastique", l'algorithme principal (7) est "Arrow-Hurwicz" et la variante pour  $J$  fortement convexe (algorithme 8 : grands pas sur  $p$ ) s'appelle "Arrow Fortement Convexe"

On constate que l'algorithme "Uzawa Stochastique" ne converge pas... En fait nous ne l'avons mis en œuvre que pour la partie en  $u_1$  et  $u_2$  car il est **impossible** de résoudre le problème de point-selle associé à l'itération  $k$  pour le morceau en  $(u_3, u_4)$ . En effet, les conditions d'optimalité sont

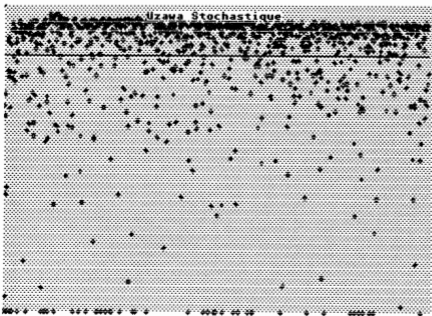
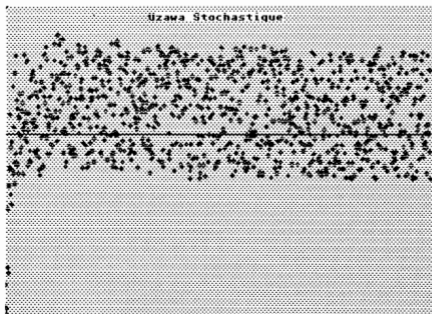
$$\omega_3^2 u_3^{k+1} + \omega_3 \omega_4 u_4^{k+1} + p^{k+1} = 0$$

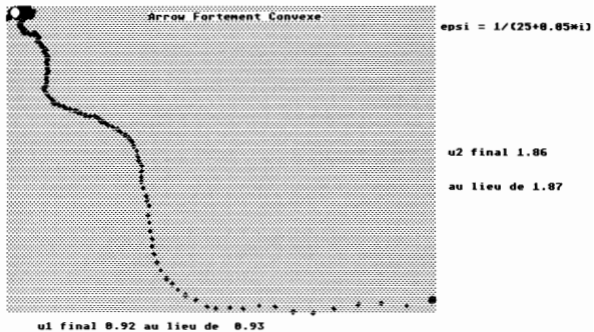
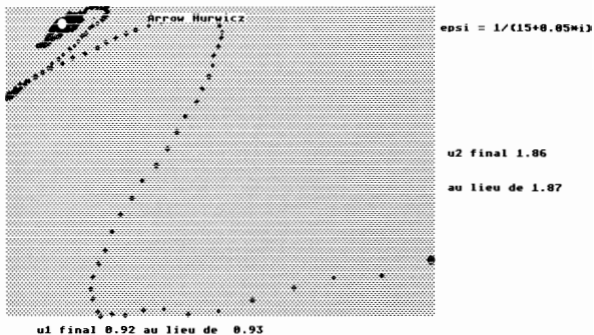
$$\omega_3 \omega_4 u_3^{k+1} + \omega_4^2 u_4^{k+1} + p^{k+1} = 0$$

et le déterminant associé est nul !









Arrow Hurwicz

$$\text{epsi} = 4 / (25 + 0.05w_i)$$

u4 final 1.24

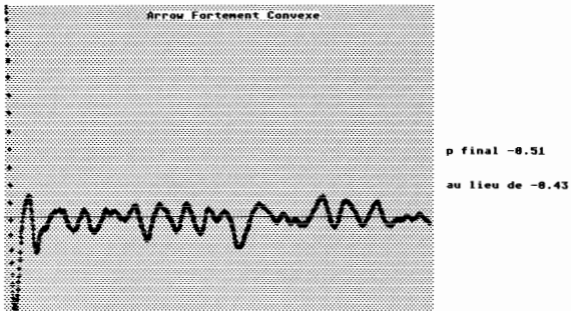
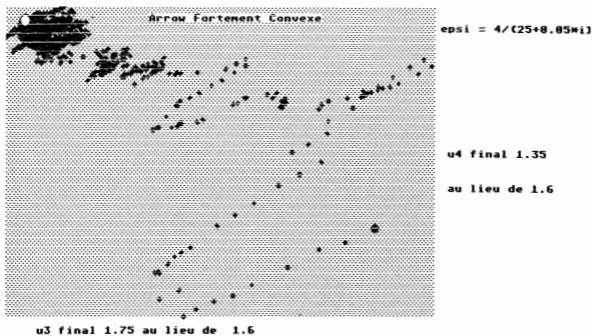
au lieu de 1.6

u3 final 1.77 au lieu de 1.6

Arrow Hurwicz

p final -0.53

au lieu de -0.43



## Chapitre 4

### Décomposition par les quantités

Le problème initial que l'on désire résoudre est le suivant :

$$\min_{u \in U} J(u) \quad \text{sous} \quad \sum_{i=1}^N \theta_i(u_i) = d \quad (4.1)$$

avec  $J(u) = \mathbf{E}j(u, \omega)$

L'espace de Hilbert  $U$  est décomposé en  $U_1 \times \dots \times U_N$ .  $J$  est une fonctionnelle convexe, s.c.i sur  $U$  (voir les théorèmes suivants pour des hypothèses plus précises). Pour tout  $i$ ,  $\theta_i : U_i \rightarrow \mathcal{C}$  est supposée affine continue.

Dans un premier temps, on supposera  $J$  additive par rapport à la décomposition des variables  $u_i$ , c'est-à-dire

$$J(u) = \sum_{i=1}^N J_i(u_i)$$

On suppose qu'il existe une solution  $(u_1^*, \dots, u_N^*; p^*)$  du problème de recherche de point-selle sur  $U \times \mathcal{C}$  et on pose  $c_i^* := \theta_i(u_i^*)$ .

On définira une **allocation** par la donnée d'un vecteur de la variété linéaire  $\Sigma$  de  $C^N$  définie par

$$\Sigma = \{(c_1, \dots, c_N) \in C^N \mid \sum_{i=1}^N c_i = d\}$$

Le principe de la décomposition par les quantités consiste à allouer à chaque unité  $i$ , une quantité  $c_i$  (la  $i$ ème composante d'un vecteur  $c$  de  $\Sigma$ ), ce qui conduit à la résolution en parallèle (décomposition) des problèmes locaux

$$\{ \min_{u_i \in U_i} J_i(u_i) \mid \theta_i(u_i) = c_i \} \quad (4.2)$$

La coordination consistera en une mise à jour des  $c_i$  en fonction de la comparaison des "utilités marginales" associées aux problèmes locaux, c'est-à-dire en comparant les "sensibilités" (au sens de la diminution du critère local) en  $c_i$  des fonctionnelles  $\phi_i$  définies par

$$\phi_i(c_i) = J_i(\bar{u}_i(c_i)) \text{ si } \bar{u}_i(c_i) \text{ est solution de (4.2)} \quad (4.3)$$

$$\phi_i(c_i) = +\infty \text{ si (4.2) n'admet pas de solution}$$

ou sous une forme plus condensée,

$$\phi_i(c_i) = \inf_{u_i} \sup_{p_i} J_i(u_i) + \langle p_i, \theta_i(u_i) - c_i \rangle$$

Nous montrerons plus loin que les variables duales  $\bar{p}_i(c_i)$  obtenues en résolvant (4.2) sont justement (au signe près) de bon candidats pour évaluer ces "sensibilités" (sous-gradients).

On peut donc voir la résolution de (4.1) comme équivalente à celle de

$$\{ \min_{c_1, \dots, c_N} \sum_{i=1}^N \phi_i(c_i) \mid c \in \Sigma \} \quad (4.4)$$

Les deux lemmes suivants (Cohen[1, p. 23]) montrent l'équivalence mathématique entre ces problèmes.

**Lemme 9** Si  $c = (c_1, \dots, c_N)$  est solution de (4.4), alors le vecteur issu de la concaténation des  $\bar{u}_i(c_i)$  (solutions de (4.2) avec  $c_i$  comme second membre de la contrainte) est solution de (4.1).

Réciproquement, si  $u_i^*$  est la  $i$ ème composante d'une solution  $u^*$  de (4.1), alors  $\bar{c}_i(u_i^*) = \theta_i(u_i^*)$  est solution de (4.4).

On appellera alors allocation **optimale** pour le problème (4.1), un vecteur  $c^*$  solution du problème (4.4).

**Lemme 10** *Si, pour tout  $i$ , le Lagrangien*

$$L_i(u_i, p_i) := J_i(u_i) + \langle p_i, \theta_i(u_i) - c_i \rangle$$

*admet un point-selle  $(\bar{u}_i, \bar{p}_i)$  sur  $U_i \times C'$  et si tous les  $\bar{p}_i$  sont égaux (soit  $q$  leur valeur commune),*

*alors le vecteur  $(\bar{u}_1, \dots, \bar{u}_N, q)$  est un point-selle de*

$$L(u, p) = \sum_{i=1}^N J_i(u_i) + \langle p, \sum_{i=1}^N \theta_i(u_i) - c \rangle$$

*sur  $U \times C'$  et donc  $\bar{u} = (\bar{u}_1, \dots, \bar{u}_N)$  est solution de (4.1).*

L'interprétation intuitive de ce lemme est que l'optimum est atteint quand toutes les utilités marginales (au signe près, les  $p_i$ ) des problèmes locaux sont égales.

Cette technique de décomposition-coordination étant assez peu traitée dans la littérature en dehors du cas linéaire-quadratique, nous allons tout d'abord envisager le cas où  $J$  est **déterministe et sous-différentiable** avant de proposer une extension stochastique. Le lemme suivant permettra de donner un sens aux commentaires intuitifs faits plus haut, sur la "sensibilité" ou les "utilités marginales"

**Lemme 11** *Pour tout  $i$ , supposons que*

- (i)  $J_i$  convexe, sci et coercive.
- (ii)  $\theta_i$  affine continue.
- (iii)  $c_i$  appartient l'intérieur de  $\theta_i(\text{Dom} J_i)$ .

*Alors  $\phi_i$  est convexe,  $c_i$  est un point intérieur à  $\text{Dom} \phi_i$  et  $\phi_i$  est sous-différentiable en  $c_i$  (elle est donc continue en  $c_i$  et donc s.c.i., car convexe, pour la topologie faible).*

*D'autre part, si  $(u_i, p_i)$  est un point-selle associé au calcul de  $\phi_i(c_i)$  (voir (4.2)), alors on a*

$$-p_i \in \partial \phi_i(c_i)$$

Remarque : l'hypothèse [(iii)] est vérifiée pour tout  $c_i$ , en particulier si  $\text{Dom} J_i = U$  et  $\theta_i$  est surjective.

Ce lemme signifie que  $-p_i$  est une direction de montée pour  $\phi_i$ .

Démonstration <sup>(1)</sup>:

### Convexité

L'application  $\Phi$  de  $U \times \mathcal{C}$  dans  $\bar{\mathbf{R}}$  définie par

$$\Phi(u, c) = \sup_p J(u) + \langle p, \theta(u) - c \rangle$$

est conjointement convexe et s.c.i. en  $u$  et  $c$ , comme enveloppe supérieure de fonctions convexes et s.c.i. en  $(u, c)$ . D'après Ekeland-Temam [1, lemme 2.1 page 50], l'application  $\phi$ , qui s'écrit aussi

$$\phi(c) = \inf_u \Phi(u, c)$$

est convexe en  $c$ .

### s.c.i.

Les hypothèses  $[(i)]$  à  $[(iii)]$  assurent l'existence d'un point-selle dans le calcul de  $\phi(c)$  (la dernière étant une hypothèse de qualification des contraintes dans le cas égalité).

D'après Ekeland-Temam [1, définition 2.1 et proposition 2.1 page 51], cela assure aussi que  $\phi$  est sous-différentiable en ce point  $c$ .

### sous-gradient

Par définition, un sous-gradient  $q$  de  $\phi$  au point  $c$  est tel que

$$\forall h \in \mathcal{C}, \quad \phi(c+h) \geq \phi(c) + \langle q, h \rangle$$

Par construction,

$$\phi(c) = J(u(c)),$$

où  $u(c)$  vérifie <sup>(2)</sup>

$$\forall v \in U, \quad J(u(c)) + \langle p(c), \theta(u(c)) - c \rangle \leq J(v) + \langle p(c), \theta(v) - c \rangle$$

c'est-à-dire, comme  $\theta(u(c)) = c$ ,

$$\forall v \in U, \quad \phi(c) \leq J(v) + \langle p(c), \theta(v) - c \rangle$$

Soit  $h \in \mathcal{C}$ . Notons  $u(c+h)$  la quantité primale associée au calcul de  $\phi(c+h)$ . A nouveau, on a  $J(u(c+h)) = \phi(c+h)$ .

L'inéquation ci-dessus étant valable avec  $v = u(c+h)$ , on conclut que

$$\phi(c) \leq \phi(c+h) + \langle p(c), h \rangle$$

c.a.d

$$\phi(c+h) \geq \phi(c) + \langle -p(c), h \rangle$$

<sup>1</sup>dans la suite, nous omettons l'indice  $i$ .

<sup>2</sup>inégalité de droite du problème de point-selle, avec  $p(c)$  la variable duale associée



Le raisonnement pouvant être fait pour tout  $h \in \mathcal{C}$  <sup>(3)</sup>,

$$-p(c) \in \partial\phi(c)$$

Donc si l'on veut résoudre (4.4) avec, par exemple, un algorithme de sous-gradient projeté sur  $\Sigma$ ,

$$c^{k+1} = \Pi[c^k - \epsilon^k q^k]$$

il suffira, à chaque itération, de résoudre le problème de point-selle associé à l'allocation  $c^k$  et de poser  $q^k = -p(c^k)$ .

La projection sur  $\Sigma$  est facile à calculer ( $\Pi$  est linéaire et  $c^k \in \Sigma$ ). L'expression ci-dessus devient

$$c_i^{k+1} = c_i^k + \epsilon^k [p_i^k - \frac{1}{N} \sum_{j=1}^N p_j^k]$$

D'un point de vue économique, les  $p_i(c^k)$  seront donc interprétés au signe près comme des utilités marginales. L'itération de coordination remet à jour les quantités allouées en fonction des écarts locaux d'utilité marginale par rapport à la moyenne.

---

<sup>3</sup>si une solution existe. Sinon,  $\phi(c+h) = +\infty$  par définition.

## 4.1 Situation déterministe

L'algorithme proposé est le suivant :

### Algorithme 10

- on suppose connus  $u^k$  et  $p^k$
- au niveau global, l'unité de coordination annonce une allocation  $c^k = (c_1^k, \dots, c_N^k)$  telle que  $\sum_j c_j^k = d$
- chaque unité locale  $i$  résout

$$\left\{ \min_{u_i \in U_i} J_i(u_i) \mid \theta_i(u_i) = c_i^k \right\}$$

ce qui donne un point-selle local  $(u_i^{k+1}, p_i^{k+1})$

- l'étape de coordination est alors une itération de sous-gradient projeté sur  $\Sigma$ .

$$c^{k+1} = \Pi[c^k + \epsilon^k p^{k+1}]$$

c'est-à-dire, compte tenu de la définition de  $\Sigma$

$$c_i^{k+1} = c_i^k + \epsilon^k [p_i^{k+1} - q^{k+1}]$$

avec  $q^{k+1} := \frac{1}{N} \sum_{j=1}^N p_j^{k+1}$

### 4.1.1 Convergence de l'algorithme déterministe

#### Fonction de Lyapunov.

On va considérer l'évolution de la fonctionnelle

$$\psi(c) = \frac{1}{2} \|c - c^*\|^2 = \frac{1}{2} \sum_{i=1}^N \|c_i - c_i^*\|^2$$

le long des trajectoires de l'algorithme.

#### Evaluation de $\psi(c^{k+1}) - \psi(c^k)$ .

De la définition de  $c_i^{k+1}$ , il vient :

$$c_i^{k+1} - c_i^* = c_i^k - c_i^* + \epsilon^k (p_i^{k+1} - q^{k+1})$$

en élevant au carré et en sommant sur  $i$ , on a en fonction de  $\psi$

$$\psi(c^{k+1}) - \psi(c^k) = \epsilon^k \sum_{i=1}^N \langle p_i^{k+1} - q^{k+1}, c_i^k - c_i^* \rangle + \frac{(\epsilon^k)^2}{2} \sum_{i=1}^N \|p_i^{k+1} - q^{k+1}\|^2$$

Comme  $-p_i^{k+1}$  est un sous-gradient de  $\phi_i$  au point  $(c_i^k)$ ,

$$\langle p_i^{k+1}, c_i^k - c_i^* \rangle \leq \phi_i(c_i^*) - \phi_i(c_i^k)$$

donc

$$\begin{aligned} \psi(c^{k+1}) - \psi(c^k) &\leq \epsilon^k \sum_{i=1}^N \langle q^{k+1}, c_i^* - c_i^k \rangle + \epsilon^k \sum_{i=1}^N [\phi_i(c_i^*) - \phi_i(c_i^k)] \\ &\quad + \frac{(\epsilon^k)^2}{2} \sum_{i=1}^N \|p_i^{k+1} - q^{k+1}\|^2 \end{aligned}$$

Enfin, comme  $\sum_i c_i^k = d = \sum_i c_i^*$ ,

$$\psi(c^{k+1}) - \psi(c^k) \leq \epsilon^k \sum_{i=1}^N [\phi_i(c_i^*) - \phi_i(c_i^k)] + \frac{(\epsilon^k)^2}{2} \sum_{i=1}^N \|p_i^{k+1} - q^{k+1}\|^2$$

On remarquera que le terme  $\sum_{i=1}^N \|p_i^{k+1} - q^{k+1}\|^2$  s'apparente (au facteur  $\frac{1}{N}$  près) à une variance statistique (ici, somme des carrés des écarts à la moyenne). Ce terme rend donc compte de la dispersion des différents problèmes locaux et peut d'ailleurs servir au test de convergence.

Si l'on suppose que chaque problème local n'est pas trop mal conditionné, c'est-à-dire, par exemple que

$$\forall i, \exists c_{1,i}, c_{2,i} > 0, \|p_i^{k+1} - p^*\| \leq c_{1,i} \|c_i^k - c_i^*\| + c_{2,i}$$

ce qui signifie que les multiplicateurs associés aux contraintes varient au plus linéairement en fonction du second membre de ces contraintes, on a alors :

$$\exists c'_1, c'_2 > 0, \quad 0 \leq \psi(c^{k+1}) - \psi(c^k) + \epsilon^k \sum_{i=1}^N [\phi_i(c_i^*) - \phi_i(c_i^k)] + (\epsilon^k)^2 [c'_1 \psi(c^k) + c'_2] \quad (4.5)$$

On remarquera que cette hypothèse sur  $p^k$  est l'analogue de l'hypothèse *SGLB* habituellement faite sur  $J$ . La différence est qu'ici il serait assez difficile à vérifier dans la pratique que  $\phi$  est *SGLB*. Une condition suffisante sur  $J$  et  $\theta$  pour que cette majoration de  $\|p^{k+1} - p^*\|$  soit vérifiée est que  $J$  soit quadratique et fortement convexe et que  $\theta$  soit surjective. L'hypothèse sur  $J$  est très forte et ne sera pas utilisée pour les algorithmes suivants. Néanmoins, supposons que ces conditions soient remplies :

En posant

$$J_i(u_i) = \frac{1}{2}(u_i, A_i u_i) + (b_i, u_i),$$

la condition d'optimalité associée à

$$\left\{ \min_{u_i \in U_i} J_i(u_i) \mid \theta_i(u_i) = c_i^k \right\}$$

est

$$A_i u_i^{k+1} + \theta_i^* p^{k+1} + b_i = 0, \quad \theta_i u_i^{k+1} = c_i^k$$

En supposant  $A_i$  inversible et  $\theta_i$  surjective (ce qui implique que  $(\theta_i A_i^{-1} \theta_i^*)^{-1}$  existe), on déduit de ces équations que

$$p^{k+1} = -(\theta_i A_i^{-1} \theta_i^*)^{-1} [c_i^k + \theta_i A_i^{-1} b_i]$$

et donc

$$\|p_i^{k+1}\| \leq \alpha \|c^k\| + \beta, \quad \alpha > 0, \quad \beta > 0$$

$(\psi(c^k))_{k \in \mathbf{N}}$  est une suite bornée.

Comme  $c_i^*$  est solution optimale de (4.4),

$$0 \leq \psi(c^{k+1}) \leq \psi(c^k) + (\epsilon^k)^2 [c'_1 \psi(c^k) + c'_2]$$

En appliquant le lemme 1, on en déduit que  $(\psi(c^k))_{k \in \mathbf{N}}$  est une suite bornée. Il en est donc de même pour  $(c^k)_{k \in \mathbf{N}}$  et aussi pour la suite  $(p^k)_{k \in \mathbf{N}}$  d'après les considérations précédentes.

On notera  $M$  un majorant de  $(\psi(c^k))_{k \in \mathbf{N}}$ .

**De plus,**  $\lim_k \phi(c^k) = \phi(c^*)$

En réécrivant (4.5),

$$0 \leq \epsilon^k \sum_{i=1}^N [\phi_i(c_i^k) - \phi_i(c_i^*)] \leq \psi(c^k) - \psi(c^{k+1}) + (\epsilon^k)^2 [c'_1 \psi(c^k) + c'_2]$$

et en sommant sur  $k \in \mathbf{N}$ , comme  $\psi(c^k)$  est majoré par  $M$  et  $(\epsilon^k)_{k \in \mathbf{N}}$  est une  $\sigma$ -suite,

$$0 \leq \sum_{k \in \mathbf{N}} \epsilon^k \sum_{i=1}^N [\phi_i(c_i^k) - \phi_i(c_i^*)] < \infty$$

et comme  $(p^k)_{k \in \mathbf{N}}$  est bornée, il existe un  $M'$  positif tel que

$$\|c^{k+1} - c^k\| \leq \epsilon^k M',$$

on peut alors appliquer le lemme 2, en remarquant que  $\phi$  est, d'après l'hypothèse sur  $p$ , lipchtzienne sur tout borné de  $C$ , pour conclure que

$$\lim_{k \rightarrow \infty} \phi(c^k) = \phi(c^*)$$

**Les points d'accumulation de  $(c^k)_{k \in \mathbf{N}}$  sont solutions de (4.4).**

En effet, la suite  $(c^k)_{k \in \mathbf{N}}$  est bornée et admet donc des points d'accumulation faibles. Par un raisonnement déjà indiqué plusieurs fois, comme  $\phi$  est s.c.i., ces points d'accumulations sont solutions de (4.4)

**La suite  $(u^k)_{k \in \mathbf{N}}$  est bornée et ses points d'accumulation sont solutions de (4.1).**

En effet, par définition de  $\phi$  (4.3),

$$\liminf_k \phi(c^k) = \phi(c^*) \quad \text{ssi} \quad \liminf_k J_i(u_i^k) = J_i(u_i^*)$$

ce qui implique, chacun des  $J_i$  étant coercif et chacun des  $\theta_i$  étant continue, la bornitude de la suite  $(u^k)_{k \in \mathbf{N}}$ , puis l'existence de points d'accumulations.

Enfin, ceux-ci sont solutions, par semi-continuité inférieure des  $J_i$ .

On peut donc énoncer le

**Théorème 11** *On suppose que*

- pour tout  $i$ ,  $J_i$  est convexe, s.c.i. et coercive, et  $\theta_i$  est linéaire continue.
- $\phi_i$  est SGLB (ce qui est impliqué par  $J_i$  quadratique fortement convexe et  $\theta_i$  surjective).
- $(c^k)_{k \in \mathbb{N}}$  est une  $\sigma$ -suite.

*Alors la suite  $(u^k)_{k \in \mathbb{N}}$  engendrée par l'algorithme est bornée et ses points d'accumulation faible sont solutions de (4.1).*

Remarque :

Il faut noter que l'algorithme considéré ici est un algorithme où la vitesse d'évolution des variables  $u^k$  n'est pas contrôlée par un "petit pas" C'est la raison pour laquelle les hypothèses faites sur  $J$  sont très fortes. Lorsque l'on considèrera des algorithmes à petits pas en  $u$ , comme ceux que nous avons étudiés dans les chapitres précédents, c'est sur  $K_i$  que portera l'hypothèse de forte convexité, et il suffira à  $J$  d'être SGLB.

## 4.2 Discussion sur une extension stochastique

### 4.2.1 L'extension directe

Comme au chapitre 3, c'est par un argument d'o.d.e. qu'on démontre que l'extension directe de l'algorithme précédent ne converge pas vers la bonne solution, pour autant qu'elle converge.

#### Algorithme 11 (tentative)

- on suppose connus  $u^k, p^k$
- au niveau global, l'unité de coordination annonce une allocation  $c^k = (c_1^k, \dots, c_N^k)$  telle que  $\sum_j c_j^k = d$ , et un aléa,  $\omega^{k+1}$  est tiré suivant la loi  $P$ , indépendamment des aléas précédents.
- chaque unité locale  $i$  résout

$$\left\{ \min_{u_i \in U_i} j(u_i, \omega^{k+1}) \mid \theta_i(u_i) = c_i^k \right\}$$

ce qui donne un point-selle local  $(u_i^{k+1}, p_i^{k+1})$

- l'étape de coordination est alors

$$c_i^{k+1} = c_i^k + \epsilon^k [p_i^{k+1} - q^{k+1}]$$

$$\text{avec } q^{k+1} = \frac{1}{N} \sum_{j=1}^N p_j^{k+1}$$

La fonctionnelle  $\phi$ , toujours donnée par (4.3) à l'expression suivante

$$\phi_i(c_i) = \mathbf{E} j_i(\bar{u}_i(c_i), \omega) = J_i(\bar{u}_i(c_i)) \text{ si } \bar{u}_i(c_i) \text{ est solution de (4.2)}$$

$$\phi_i(c_i) = +\infty \text{ si pas de solution pour (4.2)}$$

On remarque que  $p^{k+1}$  n'est pas un sous-gradient de  $\phi$  au point  $u^k$ , mais une "approximation" stochastique de ce sous-gradient. Ce n'est pas un sous-gradient stochastique. En effet, la définition de  $\phi$  fait intervenir, non seulement une espérance, mais un *Min* d'espérance. La situation est similaire au cas des algorithmes de décomposition par les prix où la fonction duale est un *Max* d'espérance.

En fait, les sauts des  $u_i^k$  successifs imposés par les  $\omega^{k+1}$  étant "grands" (par rapport à la vitesse d'évolution imposée par les  $\epsilon^k$ ), on constate que ces variables  $u_i^k$  ne peuvent converger. Voir d'ailleurs la section "tests numériques" à la fin du chapitre, pour une illustration imagée du comportement chaotique de l'algorithme.

Etudions cette extension sur un cas linéaire quadratique avec contraintes affines :

$$N = 2, \quad \theta(u) = \theta_1 u_1 + \theta_2 u_2, \quad C = \{0\}$$

$$j(u, \omega) = \frac{1}{2}(A_1(\omega)u_1, u_1) + (b_1(\omega), u_1) + \frac{1}{2}(A_2(\omega)u_2, u_2) + (b_2(\omega), u_2)$$

Les solutions du problème de départ sont (voir chapitre 3)

$$u_i^* = -(E[A_i])^{-1}(E[b_i] + \theta_i^t p^*), \quad c_i^* = \theta_i u_i^*, \quad i = 1, 2$$

$$p^* = -[\theta_1 E[A_1]^{-1} \theta_1^t + \theta_2 E[A_2]^{-1} \theta_2^t]^{-1} (\theta_1 E[A_1]^{-1} E[b_1] + \theta_2 E[A_2]^{-1} E[b_2])$$

Les conditions d'optimalité associées à l'algorithme sont ( $i = 1, 2$ )

$$u_i^{k+1} = -(A_i^k)^{-1} b_i^k - (A_i^k)^{-1} \theta_i^t p_i^{k+1}, \quad p_i^{k+1} = -(\theta_i (A_i^k)^{-1} \theta_i^t)^{-1} (\theta_i (A_i^k)^{-1} b_i^k + c_i^k)$$

$$c_1^{k+1} - c_1^k = \frac{c_1^k}{2} \left( -(\theta_1 (A_1^k)^{-1} \theta_1^t)^{-1} [c_1^k + \theta_1 (A_1^k)^{-1} b_1^k] + (\theta_2 (A_2^k)^{-1} \theta_2^t)^{-1} [-c_1^k + \theta_2 (A_2^k)^{-1} b_2^k] \right)$$

$$\text{et } c_2^k = -c_1^k.$$

Ceci conduit à l'équation suivante pour  $\bar{c}_1$  :

$$-\left[ \overbrace{E[(\theta_1 A_1^{-1} \theta_1^t)^{-1}]}^{f_1} + \overbrace{E[(\theta_2 A_2^{-1} \theta_2^t)^{-1}]}^{f_2} \right] \bar{c}_1 \\ - \underbrace{E[(\theta_1 A_1^{-1} \theta_1^t)^{-1} \theta_1 (A_1)^{-1} b_1]}_{g_1} + \underbrace{E[(\theta_2 A_2^{-1} \theta_2^t)^{-1} \theta_2 (A_2)^{-1} b_2]}_{g_2} = 0$$

On voit apparaître des termes tels que  $g_1$  ou  $g_2$  qui n'apparaissent pas dans les formules donnant  $c^*$ . Ceci implique que  $\bar{c} \neq c^*$  en général.

Vérifions-le sur l'exemple numérique simple suivant :

$$U = \mathbf{R}^2, \quad \Omega = [0, 1], \quad A_1 = \omega, \quad b_1 = -(\omega)^2, \quad \theta_1(u_1) = u_1, \quad \theta_2(u_2) = -2u_2$$

Le problème est donc

$$\left\{ \min_{u \in \mathbf{R}^2} J(u_1, u_2) = E \left[ \frac{1}{2} \omega (u_1^2 + u_2^2) - \omega^2 (u_1 + u_2) \right] \mid u_1 - 2u_2 = 0 \right\}$$

Les solutions sont

$$u^* = \left( \frac{4}{5}, \frac{2}{5} \right), \quad p^* = \frac{-1}{15}, \quad c^* = \left( \frac{4}{5}, \frac{-4}{5} \right)$$

Les  $f_i$  et  $g_i$  limites de l'algorithme sont

$$f_1 = \frac{1}{2}, \quad f_2 = \frac{1}{8}, \quad g_1 = \frac{1}{3}, \quad g_2 = \frac{1}{6}$$

ce qui conduit à  $\bar{c}_1 = \frac{4}{15} \neq \frac{4}{5}$  !

### 4.2.2 Une extension “freinée” en $u$

Nous proposons ici une variante qui opère un “freinage” de l’algorithme précédent. Nous verrons que ce freinage est nécessaire mais non suffisant à l’extension stochastique. On remarquera à nouveau l’analogie avec les méthodes “par les prix” où il est aussi nécessaire d’utiliser une fonction auxiliaire et des pas  $\epsilon^k$  assurant le freinage en  $u$ .

#### Algorithme 12 (tentative)

- on suppose connus  $u^k$  et  $p^k$
- au niveau global, l’unité de coordination annonce une allocation  $c^k = (c_1^k, \dots, c_N^k)$  telle que  $c^k \in \Sigma$  et l’on procède à un tirage aléatoire  $\omega^{k+1}$  suivant la loi  $P$ , indépendant des tirages précédents.
- chaque unité locale  $i$  résout

$$\left\{ \min_{u_i \in U_i} \frac{1}{2} \|u_i^k - u_i\|^2 + \epsilon^k (r_i^k, u_i) \mid \theta_i(u_i) = c_i^k \right\}$$

avec  $r_i^k \in \partial j_i(u_i^k, \omega^{k+1})$ , ce qui donne un point-selle local  $(u_i^{k+1}, p_i^{k+1})$

- l’étape de coordination est alors

$$c_i^{k+1} = c_i^k + \epsilon^k [p_i^{k+1} - q^{k+1}]$$

avec  $q^{k+1} = \frac{1}{N} \sum_{j=1}^N p_j^{k+1}$

On pourrait tout de suite envisager un cas plus général où l’on remplace  $\frac{1}{2} \|u_i^k - u_i\|^2$  par  $K_i(u_i) - (K_i'(u_i^k), u_i)$ . C’est ce que nous ferons pour l’algorithme suivant.

En attendant, supposons que, pour tout  $i$ ,  $\theta_i$  est affine

Les conditions d’optimalité associées à l’algorithme sont donc

$$u_i^{k+1} = u_i^k - \epsilon^k r_i^k - \theta_i^t p_i^{k+1}, \quad \theta_i \cdot u_i^{k+1} = c_i^k$$

$$c_i^{k+1} = c_i^k + \epsilon^k [p_i^{k+1} - \frac{1}{N} \sum_{j=1}^N p_j^{k+1}]$$

Appliquons  $\theta_i$  à la première équation puis exploitons la deuxième puis la troisième (appliquée en  $k-1$ ). Il reste

$$\theta_i \cdot \theta_i^t \cdot p_i^{k+1} = -\epsilon^k \theta_i \cdot r_i^k - \epsilon^{k-1} [p_i^k - \frac{1}{N} \sum_{j=1}^N p_j^k]$$

soit, en supposant  $\theta_i \cdot \theta_i^t$  inversible ( $\theta_i$  surjective), et en notant  $M_i = \theta_i \cdot \theta_i^t$  et  $h_i^k = M_i^{-1} \theta_i \cdot r_i^k$ , nous sommes conduits au système d’équations récurrentes

$$\forall i, \quad p_i^{k+1} = -\epsilon^k h_i^k - \epsilon^{k-1} M_i^{-1} \left[ p_i^k - \frac{1}{N} \sum_{j=1}^N p_j^k \right]$$



Comme  $\lim_{k \rightarrow \infty} r^k = 0$ , ce système d'équations implique que si la suite  $(h_i^k)_{k \in \mathbf{N}}$  est bornée (ce qui se produira si la suite  $(u^k)_{k \in \mathbf{N}}$  l'est, en particulier si elle converge), alors  $\lim_k p_i^k = 0$ . Or  $p^*$  n'a pas de raison particulière d'être nul.

En revanche, on constate que l'o.d.e associé au système donne bien les bonnes conditions d'optimalité pour  $u^k$  et  $c^k$  en supposant que  $\lim_k p_i^k = 0$ . En effet, en éliminant  $p_i^{k+1}$  entre les deux premières équations, on a

$$\Pi_i u_i^{k+1} = \Pi_i u_i^k - \epsilon^k \Pi_i r_i^k, \quad \text{avec } \Pi_i = \mathbf{I} - \theta_i'(\theta_i \theta_i')^{-1} \theta_i$$

$\Pi_i$  est le projecteur orthogonal sur le noyau de  $\theta_i$ . En utilisant un argument d'o.d.e. (au moins dans le cas différentiable où  $r_i^k = j_i'(u_i^k, \omega^{k+1})$ ) on voit que le  $J_i'(\bar{u}_i)$  "limite" appartient à l'orthogonal du noyau de  $\theta_i$  (c'est-à-dire  $\Pi_i J_i'(u_i^k) = 0$ ). D'autre part, par construction,  $\theta_i \bar{u}_i = \bar{c}_i$  et  $\sum_i c_i^{k+1} = \sum_i c_i^k = \dots = 0$  d'où  $\sum_i \bar{c}_i = 0$ .

Ainsi, même si l'on peut penser que cet algorithme devrait avoir asymptotiquement un bon comportement au moins en  $u$  et  $c$ , il se trouve que numériquement, il ne donne pas toujours satisfaction. En effet, supposons que les variables  $p^k$  convergent rapidement vers 0. A partir d'un certain rang  $k_0$ , les variables  $c_i$  ne vont plus évoluer (au "bruit numérique" près). Dans ce cas, pour  $k \geq k_0$ , les équations récurrentes associées à l'algorithme seront

$$u_i^{k+1} = u_i^k - \epsilon^k r_i^k, \quad \theta_i u_i^{k+1} = c_i^k \simeq c_{i,0}$$

$$c_i^{k+1} \simeq c_i^k \simeq c_{i,0}$$

La coordination disparaît. Les variables  $u^k$  convergeront donc vers les solutions des problèmes décentralisés

$$\{\min_i J_i(u_i) \mid \theta_i u_i = c_{i,0}\}$$

Or même si  $\sum_i c_{i,0} = c$ , il est tout de même peu probable que  $c_{i,0} = c_i^*$

Cet algorithme n'étant donc pas satisfaisant, on propose, dans la section suivante, une méthode de "normalisation" de  $p^k$  aboutissant à un algorithme dont on montrera que les variables  $u^k$  et  $c^k$  convergent vers les bonnes valeurs et les variables  $p^k$  restent bornées, mais ne tendent pas nécessairement vers 0.

Le lecteur pourra trouver toutes ces propositions / contre-exemples un peu longues. Nous avons pensé qu'elles méritaient d'être mentionnées, pour montrer que les extensions "naturelles" du cas déterministe au cas stochastique ne sont pas toujours les meilleures...

4.2.3 Une extension “freinée” en  $u$  et “normalisée” en  $p$ 

## Algorithme 13

- on suppose connus  $u^k$  et  $p^k$
- au niveau global, l'unité de coordination annonce une allocation  $c^k = (c_1^k, \dots, c_N^k)$  telle que  $c^k \in \Sigma$  et l'on procède à un tirage aléatoire  $\omega^{k+1}$  suivant la loi  $P$ , indépendant des tirages précédents.
- chaque unité locale  $i$  résout

$$\left\{ \min_{u_i \in U_i} K_i(u_i) + (\epsilon^k r_i^k - K'_i(u_i^k), u_i) \mid \epsilon^k \theta_i(u_i) = \epsilon^k c_i^k \right\}$$

avec  $r_i^k \in \partial_{j_i}(u_i^k, \omega^{k+1})$ , ce qui donne un point-selle local  $(u_i^{k+1}, p_i^{k+1})$

- l'étape de coordination est alors

$$c_i^{k+1} = c_i^k + \mu^k [p_i^{k+1} - q^{k+1}],$$

$$\text{avec } q^{k+1} = \frac{1}{N} \sum_{j=1}^N p_j^{k+1}$$

La “normalisation” de  $p^k$  consiste à introduire le facteur  $\epsilon^k$  dans la contrainte locale.

Supposons que, pour tout  $i$ ,  $\theta_i$  est affine,  $K_i(u_i) = \frac{1}{2} \|u_i\|^2$  et  $\mu^k = \epsilon^k$ . Les conditions d'optimalité associées à l'algorithme sont alors

$$u_i^{k+1} = u_i^k - \epsilon^k [r_i^k + \theta_i^* p_i^{k+1}], \quad \theta_i \cdot u_i^{k+1} = c_i^k$$

$$c_i^{k+1} = c_i^k + \epsilon^k [p_i^{k+1} - \frac{1}{N} \sum_{j=1}^N p_j^{k+1}]$$

L'argument d'o.d.e confirme que cet algorithme est un bon candidat pour résoudre le problème (4.1). En effet, la “normalisation” a consisté à remplacer le  $p^{k+1}$  précédent par  $\frac{p^{k+1}}{\epsilon^k}$ , ce qui, une fois  $p^{k+1}$  éliminé, ne change pas l'équation récurrente en  $u$  et conserve  $\sum_i c_i^{k+1} = \sum_i c_i^k$ . On obtient en plus l'équation limite  $\forall i, \bar{p}_i = q = \frac{1}{N} \sum_j \bar{p}_j$

Remarquons que nous n'avons pas besoin de supposer que  $J$  a la forme additive. Nous montrons la convergence de cet algorithme dans la section suivante.

## 4.3 Convergence de l'extension stochastique

Notations :

On utilisera dans la suite les notations suivantes :

$$K = \begin{pmatrix} K_1 & & \\ & \dots & \\ & & K_N \end{pmatrix}, \quad \theta = \begin{pmatrix} \theta_1 & & \\ & \dots & \\ & & \theta_N \end{pmatrix}$$

$$M = \theta K^{-1} \theta^t, \quad \|M^{-1}\| = \frac{1}{\delta}$$

En d'autres termes,  $\delta$  est la constante de forte monotonie de  $M$ .  $\|u\|_K$  sera la norme induite par l'opérateur  $K$  sur  $U$ , c'est-à-dire

$$\|u\|_K^2 := \frac{1}{2}(Ku, u)$$

On rappelle que  $\Pi$  est la projection sur  $\Sigma$ .

### Théorème 12

On suppose

- $J$  convexe, s.c.i. ; pour tout  $\omega \in \Omega$ ,  $j(\cdot, \omega)$  convexe, propre, s.c.i., à sous-gradient linéairement borné ;  $j(u, \cdot)$  mesurable sur  $\Omega$  pour tout  $u \in U$
- pour tout  $i$ ,  $\theta_i$  est affine et surjective.
- $K_i(u_i)$  quadratique, c.a.d  $K_i(u_i) = \frac{1}{2}(K_i u_i, u_i)$ , l'opérateur  $K_i$  étant auto-adjoint positif.
- $(\epsilon^k)_{k \in \mathbb{N}}$  est une  $\sigma$ -suite décroissante et  $\mu^k = \lambda \epsilon^{k+1}$  avec  $\lambda \leq \frac{\epsilon}{\tau}$  et  $\tau > 1$ .  
On suppose que la suite  $\frac{\epsilon^k}{\tau^k}$  est décroissante.

Alors la suite  $(u^k)_{k \in \mathbb{N}}$  engendrée par l'algorithme est bornée et ses points d'accumulation faible sont solutions de (4.1)

On remarquera que l'hypothèse sur la suite  $(\epsilon^k)_{k \in \mathbb{N}}$  est bien satisfaite pour les suites usuelles à série divergente comme  $\epsilon^k = \frac{1}{k}$ .

#### 4.3.1 Fonction de Lyapunov et Inéquations variationnelles

Pour démontrer ce théorème, nous allons étudier les variations de la fonctionnelle

$$\psi(u, c) = \frac{1}{2} \|u - u^*\|_K^2 + \frac{\epsilon^k}{2\mu^k} \|c - c^*\|^2$$

Les conditions d'optimalité associées à l'algorithme sont (on utilise la notation "globale" en  $u, p$  et  $c$ )

$$Ku^{k+1} = Ku^k - \epsilon^k r^k - \epsilon^k \theta^t p^{k+1}, \quad \theta u^{k+1} = c^k \quad (4.6)$$

$$c^{k+1} = c^k + \mu^k \Pi p^{k+1} \quad (4.7)$$

### 4.3.2 Majorations préliminaires

D'après la première condition d'optimalité, en multipliant les deux membres par  $u^{k+1} - u^*$ ,

$$\begin{aligned} \|u^{k+1} - u^*\|_K^2 &= (Ku^k - Ku^*, u^{k+1} - u^*) - \epsilon^k(r^k, u^k - u^*) \\ &\quad - \epsilon^k(r^k, u^{k+1} - u^k) - \epsilon^k(\theta^t p^{k+1}, u^{k+1} - u^*) \end{aligned}$$

L'équivalent vectoriel de l'identité classique  $ab = -\frac{1}{2}[(a-b)^2 - a^2 - b^2]$  donne

$$(Ku^k - Ku^*, u^{k+1} - u^*) = \frac{1}{2}\|u^{k+1} - u^*\|_K^2 + \frac{1}{2}\|u^k - u^*\|_K^2 - \frac{1}{2}\|u^{k+1} - u^k\|_K^2$$

Dans le même ordre d'idée,

$$|-\epsilon^k(r^k, u^{k+1} - u^k)| \leq \frac{1}{2}(\epsilon^k)^2 \|K^{-1}r^k\|_K^2 + \frac{1}{2}\|u^{k+1} - u^k\|_K^2$$

et donc

$$\frac{1}{2}\|u^{k+1} - u^*\|_K^2 \leq \frac{1}{2}\|u^k - u^*\|_K^2 - \epsilon^k(r^k, u^k - u^*) - \epsilon^k(\theta^t p^{k+1}, u^{k+1} - u^*) + \frac{(\epsilon^k)^2}{2}\|K^{-1}r^k\|_K^2$$

D'autre part, le terme  $(\theta^t p^{k+1}, u^{k+1} - u^*)$  vaut, d'après la deuxième condition d'optimalité,  $\langle p^{k+1}, c^k - c^* \rangle$ .

Enfin, par hypothèse *SGLB* sur  $j$  et par forte monotonie de  $K$ , il existe des constantes positives  $\alpha_1$  et  $\beta_1$  telles que

$$\frac{1}{2}\|K^{-1}r^k\|_K^2 \leq \alpha_1 \psi(u^k, c^k) + \beta_1$$

En conclusion,

$$\frac{1}{2}\|u^{k+1} - u^*\|_K^2 \leq \frac{1}{2}\|u^k - u^*\|_K^2 - \epsilon^k(r^k, u^k - u^*) - \epsilon^k \langle p^{k+1}, c^k - c^* \rangle + (\epsilon^k)^2 [\alpha_1 \psi(u^k, c^k) + \beta_1]$$

D'après l'équation de coordination, on a

$$\frac{\epsilon^k}{2\mu^k} \|c^{k+1} - c^*\|^2 \leq \frac{\epsilon^k}{2\mu^k} \|c^k - c^*\|^2 + \frac{1}{2}\epsilon^k \mu^k \|\Pi p^{k+1}\|^2 + \epsilon^k \langle \Pi p^{k+1}, c^k - c^* \rangle$$

On remarque, d'après la définition de  $\mu^k$  que

$$\frac{\epsilon^k}{\mu^k} = \lambda \frac{\epsilon^k}{\epsilon^{k+1}}$$

est une suite décroissante.

D'autre part,

$$\langle \Pi p^{k+1}, c^k - c^* \rangle = \langle p^{k+1}, c^k - c^* \rangle$$

car  $\Pi$  est un opérateur auto-adjoint,  $\Pi c^k = c^k$  et  $\Pi c^* = c^*$

On en déduit

$$\frac{\epsilon^{k+1}}{2\mu^{k+1}} \|c^{k+1} - c^*\|^2 \leq \frac{\epsilon^k}{2\mu^k} \|c^k - c^*\|^2 + \frac{1}{2} \epsilon^k \mu^k \|\Pi p^{k+1}\|^2 + \epsilon^k \langle p^{k+1}, c^k - c^* \rangle$$

En termes de  $\psi$ , les deux inéquations précédentes se résument à

$$\psi(u^{k+1}, c^{k+1}) \leq \psi(u^k, p^k) - \epsilon^k (r^k, u^k - u^*) + \frac{1}{2} \epsilon^k \mu^k \|\Pi p^{k+1}\|^2 + (\epsilon^k)^2 [\alpha_1 \psi(u^k, c^k) + \beta_1]$$

Pour nous ramener à des résultats déjà connus, il nous reste à trouver une majoration de  $\|\Pi p^{k+1}\|$ .

Pour ce faire, de la première condition d'optimalité, on déduit,  $K$  étant inversible,

$$u^{k+1} - u^k + \epsilon^k K^{-1} r^k + \epsilon^k K^{-1} \theta^t p^{k+1} = 0$$

en appliquant  $\theta$ , on a

$$c^k - c^{k-1} + \epsilon^k \theta K^{-1} r^k + \epsilon^k \theta K^{-1} \theta^t p^{k+1} = 0$$

Comme  $K$  est un opérateur bijectif et  $\theta$  est surjectif,  $M = \theta K^{-1} \theta^t$  est inversible. On en déduit  $p^{k+1}$

$$p^{k+1} = M^{-1} \frac{c^k - c^{k-1}}{\epsilon^k} - M^{-1} \theta K^{-1} r^k$$

Enfin, par définition de  $c^k$  en fonction de  $c^{k-1}$  et  $p^k$ ,

$$p^{k+1} = -M^{-1} \frac{\mu^{k-1}}{\epsilon^k} \Pi p^k - M^{-1} \theta K^{-1} r^k$$

$\Pi$  étant non-expansive,

$$\|\Pi p^{k+1}\| \leq \|p^{k+1}\| \leq \|M^{-1}\| \frac{\mu^{k-1}}{\epsilon^k} \|\Pi p^k\| + \|M^{-1} \theta K^{-1}\| \|r^k\|$$

Comme  $\mu^{k-1} = \frac{\epsilon^k}{\tau \|M^{-1}\|}$ ,

$$\|\Pi p^{k+1}\| \leq \frac{1}{\tau} \|\Pi p^k\| + \|M^{-1} \theta K^{-1}\| \|r^k\|$$

En notant  $z^k = \Pi p^k$  et  $a^k = \|M^{-1} \theta K^{-1}\| \sup_{l \leq k} \|r^l\|$ ,

$$z^{k+1} \leq \frac{1}{\tau} z^k + a^k$$

et donc

$$\begin{aligned} z^{k+1} &\leq \left(\frac{1}{r}\right)^{k+1} z^0 + a_k + \frac{1}{r} a^{k-1} + \dots + \left(\frac{1}{r}\right)^k a^0 \\ &\leq z^0 + \left(1 + \frac{1}{r} + \dots + \left(\frac{1}{r}\right)^k\right) a_k \leq z^0 + \left[1 + \frac{1 - \left(\frac{1}{r}\right)^{k+1}}{1 - \frac{1}{r}}\right] a^k \end{aligned}$$

Par application de l'hypothèse de *SGLB* sur  $j(u, \omega)$ , il existe donc des constantes positives  $\alpha$  et  $\beta$  telles que

$$\| |I| p^{k+1} \|^2 \leq \alpha_2 \sup_{l \leq k} \psi(u^l, c^l) + \beta_2$$

En définitive, en notant  $\beta_3 = \beta_1 + \beta_2$ , on aboutit à l'inéquation finale sur la variation de la fonction de Lyapunov  $\psi$

$$\psi(u^{k+1}, c^{k+1}) \leq \psi(u^k, c^k) - \epsilon^k (r^k, u^k - u^*) + \lambda \epsilon^k \epsilon^{k-1} \sup_{l \leq k} \psi(u^l, c^l) + (\epsilon^k)^2 [\alpha_1 \psi(u^k, c^k) + \beta_3]$$

### 4.3.3 Fin de la démonstration

Comme  $\epsilon^{k-1} \geq \epsilon^k$ , on reprend l'inéquation précédente sous la forme

$$\psi(u^{k+1}, c^{k+1}) \leq \psi(u^k, c^k) - \epsilon^k (r^k, u^k - u^*) + (\epsilon^{k-1})^2 [\lambda' \sup_{l \leq k} \psi(u^l, c^l) + \beta_3]$$

avec  $\lambda' = \lambda + \alpha_1$

La seule différence avec les majorations du chapitre 3 est le terme  $\lambda' \sup_{l \leq k} \psi(u^l, c^l)$ . Le lemme 1 restant applicable avec ce terme supplémentaire pour montrer la bornitude de la suite  $(y^k)_{k \in \mathbf{N}}$ , la suite de la démonstration suit donc fidèlement celle des théorèmes du chapitre 3.

## 4.4 Comparaison avec des résultats antérieurs

A notre connaissance, aucun résultat comparable à ceux de ce chapitre n'a été démontré précédemment, excepté dans le cas déterministe, fortement convexe, dans Cohen[1, p. 70].

Une étude des algorithmes déterministes de décomposition par les quantités a été faite, avec une approche économique, par G. Heal (Heal[1], 1968). On y suppose la forme additive de  $J$ . Les variables  $u_i$  et les fonctionnelles de coût  $J_i$  sont supposées  $C^1$  (en tant que fonctions du temps implicitement introduit par les itérations de l'algorithme). La démonstration de convergence utilise les propriétés des équations différentielles associées à l'algorithme.

## 4.5 Tests numériques

On teste ici les deux tentatives (algorithmes 11 et 12) et l'algorithme de décomposition stochastique 13.

On rappelle que

$$j(u, \omega) = \frac{1}{2}[(u_1 - \omega_1)^2 + \omega_2(u_2 - 1)^2 + (\omega_3 u_3 + \omega_4 u_4 - 1)^2]$$

$$\theta(u) = u_1 + u_2 + u_3 + u_4, \quad d = 1$$

$$J(u) = \frac{1}{2}[u_1^2 - u_1 + \frac{1}{2}(1 - u_2)^2 + \frac{1}{2}u_3 u_4 + \frac{1}{3}(u_3^2 + u_4^2) - u_3 - u_4] + \frac{2}{3}$$

Pour la décomposition par les quantités, on choisit de décomposer  $u$  en deux sous-systèmes

$$v_1 = (u_1, u_2), \quad v_2 = (u_3, u_4)$$

Les valeurs optimales sont :

$$u_1^* = \frac{14}{15} \simeq 0,933 \quad u_2^* = \frac{28}{15} \simeq 1,867$$

$$u_3^* = u_4^* = \frac{8}{5} = 1,6$$

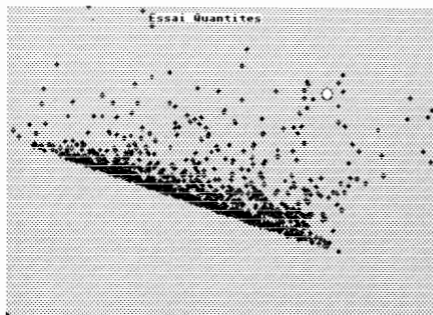
$$p^* = -\frac{13}{30} \simeq -0,433$$

et le critère optimal est  $J(u^*) = \frac{43}{60} \simeq 0,716$

Nous indiquons une réalisation de chaque algorithme sur 1000 itérations pour chaque sous-système. Les valeurs optimisées des  $\epsilon^k$  sont indiquées sur les figures. L'algorithme 11 est appelé "Essai Quantités" et ne converge absolument pas. L'algorithme 12 est "Quantités freinées en  $u$ " Il ne converge pas toujours vers les bonnes valeurs en  $u$  et  $c$ . En effet (voir discussion théorique), tout dépend de la vitesse à laquelle les variables  $p$  tendent vers 0. Sur le graphique correspondant, nous indiquons quelle est la solution (le petit cercle blanc) et vers quoi convergent les variables  $u$ . Si  $p$  est trop vite petit, ces variables convergent vers des optima locaux décentralisés (la coordination ne joue pas : l'optimum n'est pas global).

L'algorithme 13 ("Quantités :  $u$  freiné et  $p$  normé") converge en  $u$  et  $c$  et les variables  $p$  restent bornées au voisinage de la valeur optimale.



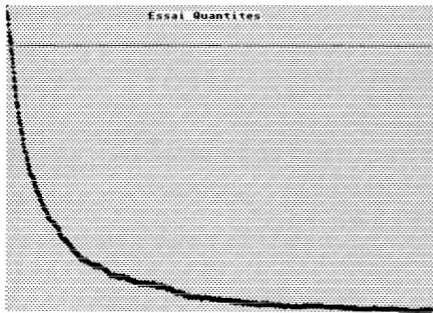


$\text{epsi} = 1/(20+0.1 \times i)$

u2 final 1.16

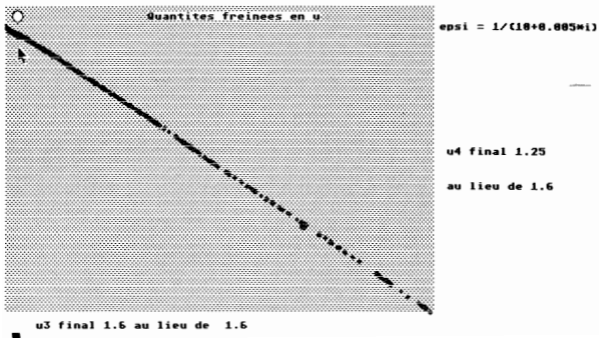
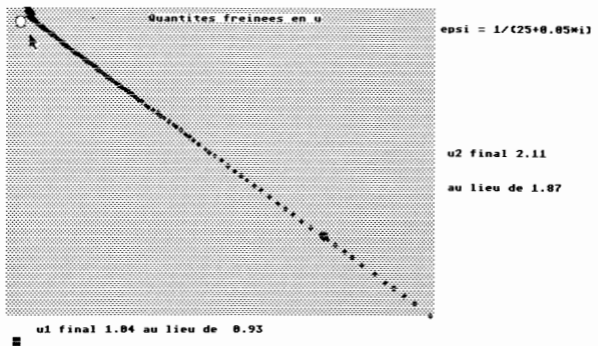
au lieu de 1.87

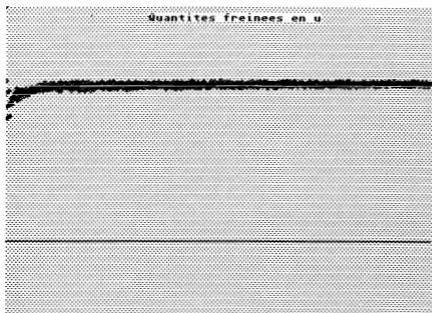
u1 final 0.34 au lieu de 0.93



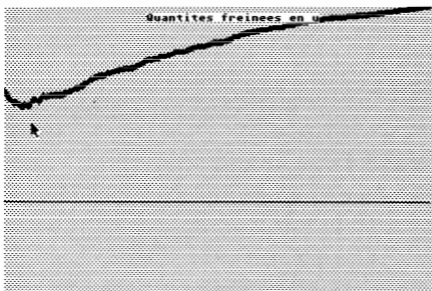
c1 final 1.5

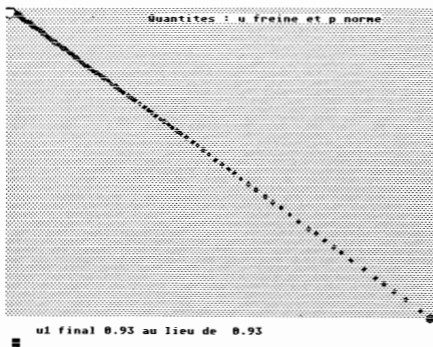
au lieu de 2.8





■

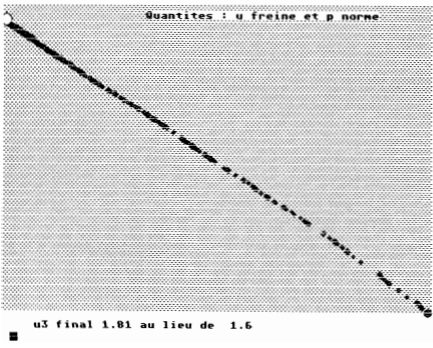




$$\text{epsi} = 1/(25+0.05 \times i)$$

u2 final 1.89

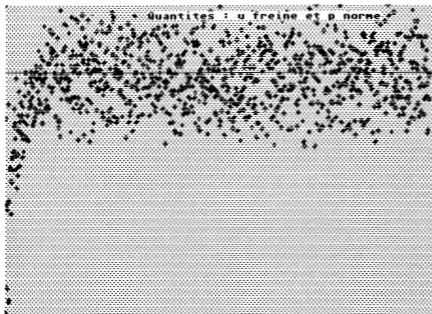
au lieu de 1.87



$$\text{epsi} = 1/(10+0.005 \times i)$$

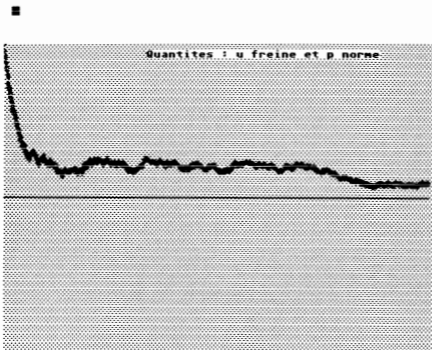
u4 final 1.37

au lieu de 1.6



pi final -0.44

au lieu de -0.43



ci final 2.82

au lieu de 2.8

## Chapitre 5

# Décomposition par prédiction

Les algorithmes de décomposition par prédiction, aussi appelés algorithmes à un niveau font intervenir une coordination décentralisée (il n'y a pas de niveau "supérieur").

Très rares sont les démonstrations de convergence de ce type d'algorithmes, excepté dans le cas linéaire quadratique (\*), cas que nous allons considérer. En revanche, ces algorithmes semblent se comporter particulièrement bien d'un point de vue numérique, ce qui justifie une tentative d'extension au cas stochastique. Nous montrons, à l'aide du raisonnement d'o.d.e. que l'extension proposée a (si elle converge) toutes les chances de converger vers des points limites qui sont solutions du problème de départ.

Dans la partie "Tests Numériques", nous appliquons le schéma général à un cas où l'un des systèmes considérés "gère" la contrainte de coordination tandis que l'autre ne fait qu'intégrer le prix issu de cette gestion dans son propre critère de gestion.

---

\*modulo des hypothèses de structures sur le critère et les contraintes que nous exposerons plus loin

## 5.1 Algorithmes à un niveau : Formulation générale

Nous résolvons le problème suivant :

$$\min_{u \in U^{ad}} \int_{\Omega} j(u, \omega) P(d\omega) = J(u) \quad \text{sous} \quad \theta(u) \in -C \quad (5.1)$$

Avec  $\Omega$ , ensemble des réalisations aléatoires,  $P$  étant la loi de probabilité associée. On appelle  $U^*$ , l'ensemble des solutions. On suppose qu'il existe un point-selle  $(u^*, p^*)$  à ce problème.

Nous allons introduire, en plus du critère auxiliaire  $K$ , une contrainte auxiliaire définie par un opérateur  $\Omega : U \rightarrow C$ .

On note  $(\rho^k)_{k \in \mathbb{N}}$  une  $\sigma$ -suite.

L'algorithme proposé est alors le suivant : <sup>(1)</sup>

### Algorithme 14

- soit  $u^0 \in U^{ad}$  et  $p^0 \in C'$ ,
- connaissant  $u^k$  et  $p^k$ , on tire suivant la loi  $P$ , un  $\omega^{k+1} \in \Omega$ , indépendant des tirages précédents, et (problème de point-selle) on résout le problème auxiliaire

$$\min_{u \in U^{ad}} K(u) + (e^{k,r^k} - K'(u^k), u) + \langle p^k, (e^k \theta'(u^k) - \Omega'(u^k)), u \rangle$$

(avec  $r^k$  appartenant au sous-différentiel de  $j$  au point  $(u^k, \omega^{k+1})$ ) sous la contrainte auxiliaire

$$\Omega(u) + \rho^k \theta(u^k) - \Omega(u^k) \in -C$$

ce qui donne  $(u^{k+1}, p^{k+1})$ .

La décomposition du problème auxiliaire nécessite celle du critère, et donc, comme d'habitude, le caractère additif de la fonctionnelle  $K$  par rapport à une décomposition de  $U$ . Mais elle nécessite aussi la décomposition des contraintes. Ceci suppose que l'on ait introduit une décomposition de l'espace des contraintes  $C$  en autant de composantes que pour l'espace  $U$  <sup>(2)</sup> et que l'on ait associé (par la numérotation) une composante de  $C$  à une composante de  $U$ . Il s'agit pratiquement d'"allouer" la gestion des contraintes auxiliaires aux sous-problèmes. Alors, pour que la décomposition ait lieu, il faut que  $\Omega$  soit "bloc-diagonale" (la composante  $\Omega_i(u)$  de  $\Omega(u)$  sur  $C_i$  ne doit dépendre que de  $u_i$ ). (Cohen [1, p. 28]).

<sup>1</sup>Nous exposons ici la variante où  $\theta$  et  $\Omega$  sont supposées différentiables

<sup>2</sup>en fait, il peut y en avoir moins certains sous-problèmes ne reçoivent pas de contraintes (voir l'exemple qui suit) - mais nous supposons ici pour faciliter l'exposé et pour permettre de parler de "Ω bloc-diagonale" qu'il y en a autant.

## 5.2 Résultat déterministe et discussion

G. Cohen [1, p. 71, théorème 5.1] a montré que lorsque  $C = \{0\}$ , si les opérateurs ont la forme :

$$J(u) = \frac{1}{2}(u - f, A(u - f)), \quad K(u) = \frac{1}{2}(u - g, B(u - g)),$$

$$\theta(u) = D(u - d), \quad \Omega(u) = E(u - e),$$

avec  $A$  et  $B$  auto-adjoints, linéaires et continus sur  $U$  ;  $D$  et  $E$  applications linéaires surjectives continues de  $U$  dans  $C$  ;  $f, g, d$  et  $e$  vecteurs de  $U$ ,

alors, si la condition de structure suivante

$$DA^{-1}E^t + EA^{-1}D^t - DA^{-1}BA^{-1}D^t \text{ est fortement monotone sur } C^*$$

est vérifiée, l'équivalent déterministe à pas fixes  $\epsilon$  et  $\rho$  de l'algorithme ci-dessus engendre une suite  $(u^k, p^k)$  qui converge fortement vers le point-selle  $(u^*, p^*)$  pour  $\epsilon$  et  $\rho$  assez petits.

L'état actuel de nos recherches ne nous permet pas d'énoncer une extension stochastique de ce résultat, mais nous nous proposons d'étudier l'extension avec la méthode de l'o.d.e. :

Nous conservons les définitions ci-dessus pour les opérateurs  $K, J, \Omega$  et  $\theta$ , avec, pour simplifier,  $d = f = g = 0$ . Nous notons  $A^k = A(\omega^{k+1})$  et  $\bar{A} = \mathbf{EA}(\omega)$ .

Les conditions d'optimalités satisfaites par le point-selle  $(u^{k+1}, p^{k+1})$  sont

$$B.(u^{k+1} - u^k) + \epsilon^k(A^k u^k + D^t p^k) - E^t.(p^{k+1} - p^k) = 0$$

$$E.(u^{k+1} - u^k) + \rho^k D u^k = 0$$

soit encore

$$\begin{pmatrix} B & E^t \\ E & 0 \end{pmatrix} \begin{pmatrix} u^{k+1} - u^k \\ p^{k+1} - p^k \end{pmatrix} = - \begin{pmatrix} \epsilon^k & 0 \\ 0 & \rho^k \end{pmatrix} \cdot \begin{pmatrix} A^k & D^t \\ D & 0 \end{pmatrix} \begin{pmatrix} u^k \\ p^k \end{pmatrix}$$

Or la matrice

$$\begin{pmatrix} B & E^t \\ E & 0 \end{pmatrix}$$

est inversible puisque  $B$  est auto-adjoint et  $E$  est surjective.

On note  $\mathbf{M}$  son inverse.

On peut donc écrire

$$\begin{pmatrix} \frac{u^{k+1} - u^k}{\epsilon^k} \\ \frac{p^{k+1} - p^k}{\rho^k} \end{pmatrix} = -\mathbf{M} \cdot \begin{pmatrix} A^k & D^t \\ D & 0 \end{pmatrix} \cdot \begin{pmatrix} u^k \\ p^k \end{pmatrix}$$



L'o.d.e. associée est

$$\begin{pmatrix} \dot{\bar{u}} \\ \dot{\bar{p}} \end{pmatrix} = -\mathbf{M} \begin{pmatrix} \bar{A} & D^t \\ D & 0 \end{pmatrix} \begin{pmatrix} \bar{u} \\ \bar{p} \end{pmatrix}$$

Les points stationnaires  $(\bar{u}, \bar{p})$  sont donc bien solution du problème de départ.

### 5.3 Tests Numériques

On considère le problème ( $N = 2$ ,  $U_1^{nd} = U$ )

$$\left\{ \min_{(u_1, u_2) \in \bar{U}_1 \times \bar{U}_2} J(u_1, u_2) \mid \theta_1 u_1 + \theta_2 u_2 = d \in C \right\} \quad (5.2)$$

$J$  est l'espérance de  $j(u, \omega)$ , les  $\theta_i$  sont affines.

L'algorithme que nous proposons est une application directe de l'algorithme précédent, pour laquelle la "décomposition" (ici immédiate) consiste à allouer la contrainte sur la ressource  $d$  au premier sous-système, c'est-à-dire que  $\Omega(u) = (\theta_1 \ 0) \cdot u$ .

Nous envisageons ici une version séquentielle de l'algorithme, afin de bénéficier à chaque instant des données disponibles les plus récentes.

#### Algorithme 15

- à l'étape  $k$ , on suppose connus  $u_1^k$ ,  $u_2^k$  et  $p^k$ . On tire un aléa  $\omega^{k+1}$ , indépendant des précédents, suivant la loi  $P$
- le problème auxiliaire résolu par l'unité 1 (qui est chargée de "gérer" la contrainte) est le problème de point-selle suivant :

$$\min_{u_1 \in \bar{U}_1} K_1(u_1) + (\epsilon^k r_1^k - K'_1(u_1^k), u_1) + (\epsilon^k - 1) \langle p^k, \theta_1 u_1 \rangle$$

sous la contrainte auxiliaire

$$\theta_1 u_1 = \theta_1 u_1^k + \rho^k (d - \theta_1 u_1^k - \theta_2 u_2^k)$$

avec  $r_1^k \in \partial_{u_1} j(u^k, \omega^{k+1})$ .

La solution de ce problème est  $(u_1^{k+1}, p^{k+1})$ .

- l'unité 2 résoud alors le problème auxiliaire

$$\left\{ \min_{u_2 \in \bar{U}_2} K_2(u_2) + (\epsilon^k r_2^k - K'_2(u_2^k), u_2) + \epsilon^k \langle p^{k+1}, \theta_2 u_2 \rangle \right.$$

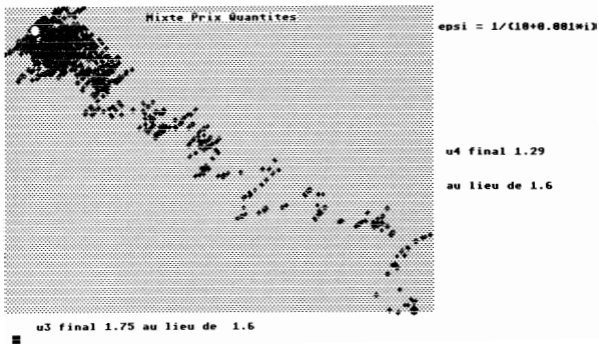
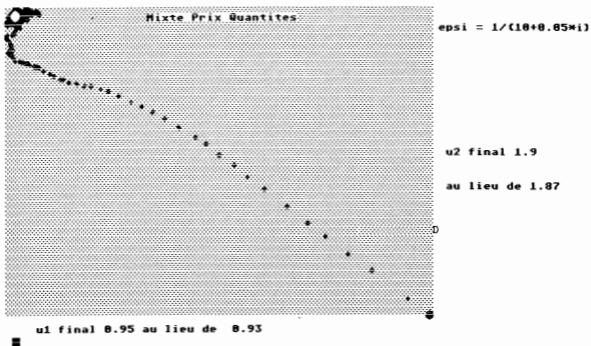
avec, en posant  $u^{k+\frac{1}{2}} = (u_1^{k+1}, u_2^k)$ ,  $r_2^k \in \partial_{u_2} j(u^{k+\frac{1}{2}}, \omega^{k+1})$

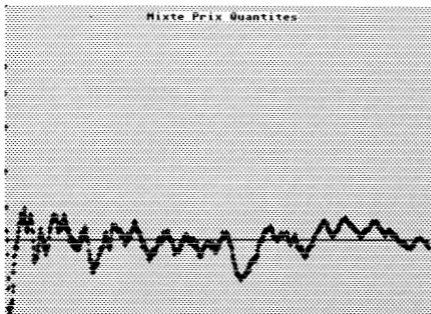
## Application :

Nous reprenons l'exemple numérique testé dans les chapitres précédents. On constate que pour une décomposition en deux sous-systèmes  $(u_1, u_2)$  et  $(u_3, u_4)$ , le deuxième sous-système est beaucoup plus "sensible" aux aléas que le premier (voir notamment les considérations de la section "Tests Numériques" du chapitre 2). En effet, la part de  $J$  associée au sous-système 2 est moins fortement convexe (les valeurs propres du Hessien sont plus faibles).

Compte-tenu du caractère perturbateur de la contrainte, on se propose donc de laisser le sous-système 1 plus stable gérer la contrainte, en espérant que le prix associé sera assez vite proche de la valeur optimale : on laissera le sous-système 2 en "optimisation libre", son critère tenant simplement compte du prix "prédit" par 1. On verra que cette politique donne d'assez bons résultats. Rappelons qu'à chaque itération, l'échange d'information est horizontal (algorithme à un seul niveau).

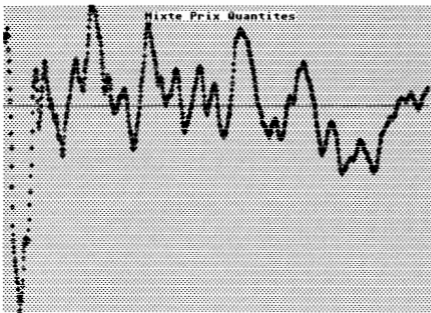
L'algorithme 15 ("Mixte Prix Quantités") est testé sur 1000 itérations et donne des résultats à ceux des algorithmes à deux niveaux (comparer avec les pages 75 à 77 et 101).





p final -0.47

au lieu de -0.43



u1+u2 final 2.85

au lieu de 2.8

## Chapitre 6

# Conclusion et Problèmes Ouverts

Après avoir montré que les classes de problèmes A et B peuvent être abordées par le Principe du Problème Auxiliaire version Stochastique, il nous reste à considérer les classes C et D : contraintes stochastiques découplées ou non.

### Problèmes de la Classe C

Dans la classe C, nous supposons que les contraintes sont stochastiques mais découplées :

$$\min E\{J(u, \omega)\} \text{ sous } u \in U^{ad}(\omega)$$

avec l'hypothèse que

$$U^{ad}(\omega) = U_1^{ad}(\omega) \times \dots \times U_N^{ad}(\omega)$$

Pratiquement, ce type de situation se produit par exemple lorsque des contraintes presque-sûres sur l'état du système (comme l'équation d'évolution) induisent implicitement des contraintes sur les variables en boucle ouverte  $u$ . En effet, prenons le cas d'un lac de stock  $x$  dont l'équation d'évolution est ( $t$  est par exemple un numéro de semaine)

$$x(t+1) = x(t) + a(t, \omega) - u(t)$$

$a$  étant les apports hydrauliques (aléatoires) de la période considérée et  $u$  la commande de turbiné (ici, en boucle ouverte).

Supposons que l'on veuille minimiser l'espérance une certaine fonction coût  $J(u)$ . Il est clair que des contraintes physiques comme

$$x \in [x_{\min}, x_{\max}], p.s.$$

induisent implicitement les contraintes presque-sûres

$$u(t) \in [x(t) - x_{\max} + a(t, \omega), x(t) - x_{\min} + a(t, \omega)], p.s.$$

Nous supposerons que l'ensemble des  $u$  satisfaisant les contraintes presque-sûres est non vide, au besoin en éliminant les aléas les plus rares (stock plein et apports gigantesques ou stocks nuls en permanence, pannes simultanées de nombreux éléments,

etc.). Enfin, on suppose que les contraintes sont découplées, ce qui par exemple, ne serait pas le cas avec des lacs en cascade (la loi d'évolution du lac inférieur et donc les contraintes induites implicitement sur son turbiné dépendent de la dynamique du lac supérieur).

Deux approches sont alors possibles :

- soit transforme la contrainte en une contrainte en probabilité

$$\min E\{j(u, \omega)\} \text{ sous } P(u \in U^{ad}(\omega)) > \eta$$

avec  $\eta$  proche de 1

- soit on a affaire à une loi de probabilité discrète sur un nombre fini d'aléas, cas pour lequel nous proposons un traitement particulier sous forme de conjecture.

### Contrainte en probabilité

Une transformation assez souvent décrite dans la littérature (Ermoliev [1,2]) et applicable aux contraintes de la forme ( $d$  est une distance)

$$P[d(u, U^{ad}) < \epsilon] > 1 - \eta$$

consiste à définir une fonction  $g(u, \omega)$  de la façon suivante

$$g(u, \omega) = \begin{cases} \eta & \text{si } d(u, U^{ad}(\omega)) < \epsilon \\ -1 + \eta, & \text{sinon} \end{cases}$$

On constate alors que  $P[d(u, U^{ad}) < \epsilon] > \eta$  peut être remplacé par  $G(u) := E(g(u, \omega)) > 0$ .

Si l'on a la chance (théoriquement inespérée, même si l'ont sait que l'opération "espérance" a souvent la propriété de "régulariser" les fonctions) d'obtenir une contrainte  $G(u)$  assez régulière (convexité), on retombe dans la Classe A ou B suivant que  $G(u)$  est couplante et les algorithmes des chapitres 2 et 3 sont applicables.

### Loi de probabilité discrète et nombre fini d'aléas

L'algorithme que nous proposons est une conjecture basée sur une analogie avec une méthode itérative d'intersections de contraintes (analogie déterministe de contraintes presque-sûres) proposée dans Bensoussan et al. [1] et étudiée aussi dans Céa [1] ( $J$  linéaire quadratique). Par rapport à l'algorithme 2, on remplace, à l'étape  $k$  la contrainte  $u \in U^{ad}$  par  $u \in U^{ad}(\omega^{k+1})$ . La loi de probabilité étant discrète, on de fortes chances de visiter tous les  $U^{ad}(\omega_j)$ ,  $m = 1, \dots, M$  en un nombre suffisant d'itérations, et on peut espérer que la quantité

$$v^k := \frac{\sum_{i=1}^k e^i u^i}{\sum_{i=1}^k e^i}$$

convergera, comme dans le cas déterministe vers la solution  $u^*$ . Remarquons que si  $K$  a la forme additive, le problème est encore décomposé. Bien que nous ayons des

éléments de généralisation du résultat de Céa dans le cas déterministe (hypothèses sur  $J$  plus faibles), nous ne savons pas comment aborder théoriquement la convergence dans le cas stochastique.

Une troisième approche envisageable serait une technique de pénalisation de la contrainte, et nous allons l'expliciter pour les problèmes de la Classe D, c'est-à-dire lorsque que les contraintes presque-sûres sont couplantes.

## Problèmes de Classe D

Dans ce cas, les méthodes de décomposition/coordination font habituellement appel à la dualité. Or si les multiplicateurs associés à des contraintes déterministes sont eux-mêmes déterministes, ce sont des variables aléatoires si les contraintes sont presque-sûres. Cela exclut donc a priori toute utilisation d'une méthode de gradient stochastique pour les évaluer. C'est pourquoi l'idée de mettre en œuvre une pénalisation exacte, au besoin après avoir, à nouveau, extrait de  $\Omega$  les aléas les moins probables, semble intéressante.

### Pénalisation exacte

Voyons ce que cela signifie dans le **cas où les aléas sont en nombre fini**.

On suppose que  $\Omega$  est un ensemble fini contenant  $M$  points  $\omega_m$ , de probabilités associées  $\pi_m$ . Nous traitons le cas d'une seule contrainte presque-sûre  $\rho(u, \omega)$ . Le problème s'énonce alors :

$$\min_{u \in U^{ad}} \sum_{m=1}^M \pi_m j(u, \omega_m) \quad \text{sous} \quad \{\rho(u, \omega_m) \leq 0, m = 1, \dots, M\}$$

Désignons par  $p_m$  les multiplicateurs optimaux (non négatifs) associés aux contraintes. Alors, on sait que si l'on choisit un scalaire  $\lambda$  tel que  $\lambda > \max_m p_m$  (Bertsekas [1]), on peut trouver la solution optimale en résolvant le problème "pénalisé" suivant :

$$\min_{u \in U^{ad}} \sum_{m=1}^M [\pi_m j(u, \omega_m) + \lambda \max(0, \rho(u, \omega_m))]$$

Afin de réinterpréter le problème ci-dessus comme celui de la minimisation de l'espérance d'une fonctionnelle, on met  $\pi_m$  en facteur dans le terme correspondant de la somme, ce qui revient, si l'on veut, à renormaliser les contraintes par leur probabilité. On obtient ainsi le problème équivalent

$$\min_{u \in U^{ad}} \sum_{m=1}^M \pi_m [j(u, \omega_m) + \lambda' \max(0, \rho(u, \omega_m))]$$

avec, comme condition sur  $\lambda'$ ,

$$\lambda' > \frac{\lambda}{\min_m \pi_m} > \frac{\max_m p_m}{\min_m \pi_m}$$

La difficulté de cette approche réside dans la présence des probabilités  $\pi_m$  au dénominateur, ce qui peut conduire à moins d'écartier les événements rares à des valeurs très élevées de  $\lambda$  si certaines probabilités approchent 0 (il n'y a aucune raison pour que les  $p_m$  associés soient petits). Cette difficulté est encore plus grande lorsque  $\Omega$  est infini, et dans ce cas, il n'existe pas encore de théorie bien établie de la pénalisation exacte pour un nombre infini de contraintes.

Supposons maintenant que le problème peut se poser sous la forme

$$\min_{u \in U^{\text{ad}}} E\{j(u, \omega) + \lambda \phi(\sum_{i=1}^N \rho_i(u_i, \omega))\}$$

avec  $\phi$  fonction de pénalisation, généralement non différentiable mais convexe ("valeur absolue" pour une contrainte égalité scalaire ou norme de la projection sur le cône positif dual pour une contrainte inégalité dans un espace de Hilbert) et globalement lipchitzienne. Dans ce cas, les variantes de l'algorithme principal du chapitre 2 pourraient être applicables et la convergence vers les bonnes valeurs aurait lieu pour peu que le paramètre  $\lambda$  soit choisi assez grand. C'est-à-dire, plus rigoureusement, que  $\lambda$  doit être, si l'on veut appliquer au moins dans le cas d'un nombre fini de contraintes des résultats déterministes connus de pénalisation exacte (voir discussion plus bas), plus grand que  $\sup_{\omega} p(\omega)$  si  $p(\omega)$  est le multiplicateur associé à la contrainte  $\rho(u, \omega) \in -C$ . Cela suppose l'existence d'un  $\lambda$  fini au prix de se débarrasser des aléas les moins probables auxquels risquent d'être associés des  $p(\omega)$  très grands.

En tout état de cause nous ne faisons que suggérer ce traitement possible des contraintes presque-sûres. Nous n'avons pu, faute de temps, en faire une étude théorique.



Décomposition par rapport aux variables de type  $v(u, \omega)$ 

Reprenons le modèle classique de la commande optimale.

$$x_{t+1} = f(x_t, u_t, \omega_t, t), \quad t = 0, \dots, T; \quad x_0 \text{ donnée} \quad (6.1)$$

où  $x$  est le vecteur d'état,  $u$  le vecteur de commande,  $\omega$  est un "bruit" aléatoire, et  $x_0$  peut-être une variable aléatoire.

On cherche à optimiser un critère s'exprimant comme une fonction de coût de  $x$ ,  $u$  et  $\omega$ . Écartons, pour rester simples, toute contrainte, d'état en particulier. Pour résoudre ce problème par des algorithmes s'apparentant au gradient stochastique, on se limite à des commandes  $u$  en "boucle ouverte", c'est-à-dire au plus sous forme de feedback local paramétré (loi de structure fixée a priori).

Dans ce contexte, la trajectoire  $\{u_t, t = 0, \dots, T\}$  s'apparente exactement à la variable  $u$  de la situation décrite dans l'introduction. La trajectoire  $\{x_t, t = 0, \dots, T\}$ , elle, est similaire à la variable  $v$  puisqu'elle dépend évidemment de  $\omega$ . L'ensemble  $V^{ad}(u, \omega)$  est défini par (6.1). C'est donc un singleton, puisque par hypothèse, nous n'avons pas pris en compte d'autres contraintes. Avec ce point de vue, ce problème tombe dans la classe A, puisque l'hypothèse  $\forall u, V^{ad}(u, \omega) \neq \emptyset$  p.s. est vérifiée, à la condition de résoudre pour chaque  $\omega$  le problème interne en  $v(u, \omega)$  (ou ici, en  $x$ ), et calculer ensuite le gradient du coût par rapport à  $u$  pour une valeur courante  $u^k$  et une réalisation (trajectoire)  $\omega^{k+1}$ . Ici, cela revient à "intégrer" ou "simuler" la dynamique puis à calculer l'état adjoint (multiplicateur associé à la contrainte (6.1)) en intégrant l'équation adjointe.

On voit donc ici que si l'on peut concevoir une décomposition / coordination portant sur  $u$  (en application des algorithmes du chapitre 2), on doit en revanche traiter la dynamique globalement à chaque étape de l'algorithme. Or, ce qui est le plus intéressant pratiquement, c'est de voir le système comme un ensemble de sous-systèmes interconnectés, et de coordonner des sous-problèmes de commande optimale en  $(x_i, u_i)$  décomposition simultanée de l'état et de la commande. C'est ce que l'on fait en optimisation déterministe dans la méthode de "prédiction des interactions" (G. Cohen [1]). Voilà une motivation importante pour une décomposition simultanée en  $u$  et  $v$ .

Reconsidérons alors le problème

$$\min_{u \in U^{ad}} E \left\{ \min_{v(u, \omega) \in V^{ad}(u, \omega)} g(u, v(u, \omega), \omega) \right\}$$

avec  $g$  différentiable en  $u$  et  $v$  pour simplifier. Supposons de plus que  $V^{ad}$  ne dépend ni de  $u$  ni de  $\omega$  <sup>(1)</sup>. Comme les conditions de type " $(\epsilon^k)$  est une  $\sigma$ -suite" ont pour rôle, intuitivement, de permettre l'effet de moyenne sur les réalisations  $\omega^{k+1}$  pour approximer l'espérance, comme par ailleurs les itérées  $v^k$  doivent pouvoir "suivre" les tirages  $\omega^{k+1}$ , on peut penser à l'algorithme "à deux vitesses" suivant. On se donne une fonctionnelle auxiliaire  $K^1(u) + K^2(v)$  (la forme séparée n'est pas indispensable)

<sup>1</sup>Nous acheminant vers une conclusion négative quant à la décomposition simultanée, dans un cas simple, nous évitons de considérer des situations trop compliquées.

et un nombre positif  $\rho$ . A l'étape  $k$ , on résoud

$$\min_{v \in V^{ad}} K^2(v) + (\rho g'_v(u^k, v^k, \omega^{k+1}) - (K^2)'(v^k), v)$$

dont la solution est  $v^{k+1}$ , et

$$\min_{u \in U^{ad}} K^1(u) + (\epsilon^k g'_u(u^k, v^{k+1}, \omega^{k+1}) - (K^1)'(u^k), u)$$

de solution  $u^{k+1}$

On montre alors sur l'exemple suivant que la simultanéité des décompositions semble vouée à l'échec.

On considère  $u \in ]-0,3, 0[ \subset \mathbf{R}$  et  $v \in \mathbf{R}$ . On suppose que

$$g(u, v, \omega) = \frac{1}{2}u^2 + v^2 + u\phi(v) - \frac{3}{4}\omega v$$

où  $\phi$  est une fonction à préciser et  $\omega$  est une variable aléatoire gaussienne centrée de variance 1. Calculons d'abord la solution du problème avec  $\phi = v^2$ .

La minimisation en  $v$  à  $u$  et  $\omega$  fixés donne  $w(u, \omega) = \frac{3u\omega}{8(1+u)}$  (minimum absolu de  $g$  en  $v$  pour toutes les valeurs de  $u$  dans la plage considérée). En reportant cette valeur dans  $g$ , puis en minimisant l'espérance, on trouve finalement la solution  $u^* = -0,25$  dont on vérifie que c'est bien un minimum absolu.

Appliquons alors l'algorithme ci-dessus à cet exemple avec, pour fonctionnelles auxiliaires,  $K^2(v) = \frac{1}{2}v^2$  et  $K^1(u) = \frac{1}{2}u^2$ . En supposant que les variables  $u^k$  restent dans la plage autorisée, on obtient les équations récurrentes suivantes :

$$v^{k+1} = v^k - \rho[2v^k(1+u^k) - \frac{3}{4}\omega^{k+1}]$$

$$u^{k+1} = u^k - \epsilon^k[u^k + (v^{k+1})^2]$$

Appliquons alors le théorème de l'équation différentielle moyenne, généralisé par Métivier et Priouré au cas non-linéaire.

**La première récurrence ci-dessus est considérée, à  $u$  fixé** (car intuitivement  $u$  évolue lentement par rapport à la vitesse d'évolution de  $v$ ), comme un processus récurrent en  $v$ , markovien linéaire et gaussien, stable si  $|1 - 2\rho(1+u^k)| < 1$ .

Cette dernière condition est vérifiée si  $u$  reste dans la plage autorisée et  $\rho < 1$ . La mesure invariante correspondante est une loi gaussienne centrée, de variance

$$\sigma_v^2(u) = \frac{9\rho}{64(1+u)(1-\rho(1+u))}$$

**La seconde récurrence en  $u$  se comporte alors asymptotiquement comme une équation différentielle** dont le second membre est l'espérance de l'expression en facteur du "petit pas"  $\epsilon^k$ , connaissant  $u$  et considérant  $v$  comme une variable

aléatoire régie par la mesure invariante calculée ci-dessus. On a donc l'équation différentielle :

$$\frac{du}{dt} = -[u + \sigma_\rho^2(u)]$$

Le tableau ci-dessous montre que cette équation admet un point d'équilibre stable (noté  $\hat{u}_2$ ) dans la plage considérée pour  $u$  dès que  $\rho \leq 0,7$  (ce qui justifie a posteriori l'hypothèse que  $u^k$  reste lui-même dans cette plage si l'on est parti assez près du point d'équilibre la simulation indiquée page suivante pour diverses valeurs de  $\rho$  confirme aussi cette hypothèse).

$\rho$	$\hat{u}_1$	$\hat{u}_2$	$\hat{u}_3$
0, 1	-0, 985713	-0, 0158487	9, 00156
0, 2	-0, 970861	-0, 0361481	4, 00701
0, 3	-0, 955234	-0, 0626133	2, 35118
0, 4	-0, 938558	-0, 09754	1, 5361
0, 5	-0, 920446	-0, 143585	1, 06403
0, 6	-0, 900322	-0, 202883	0, 769872
0, 66	-0, 885345	-0, 25	0, 635345
0, 7	-0, 877238	-0, 275679	0, 581488
0, 75	-0, 864088	-0, 316606	0, 514027
0, 8	-0, 849415	-0, 360202	0, 459617

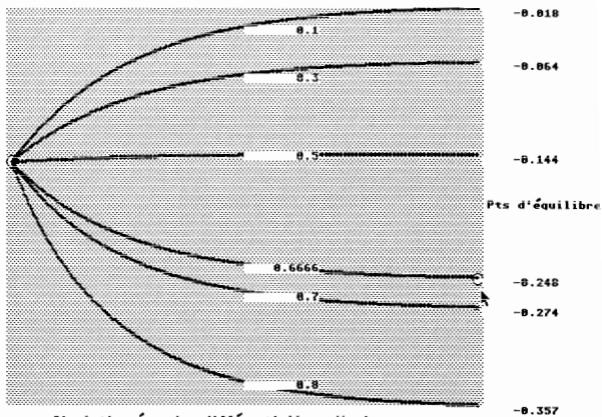
On constate néanmoins que le point d'équilibre obtenu est **fonction de  $\rho$**  et que ce n'est qu'accidentellement, pour  $\rho = \frac{2}{3}$ , qu'il est égal à l'optimum.

En général donc, l'algorithme ne converge pas vers l'optimum. Ceci est dû intuitivement au fait que la récurrence sur  $v^k$  ne converge pas assez rapidement par rapport à "l'évolution lente" de  $u^k$ , malgré le "grand pas"  $\rho$ . Bien sûr, si l'on répète une infinité de pas de la récurrence en  $v$ , **en maintenant  $\omega^{k+1}$  fixe**, avant de remettre à jour  $u^k$  et de réaliser un nouveau tirage de  $\omega$ , alors on convergera vers l'optimum puisque cela correspond à calculer exactement le  $w(u^k, \omega^{k+1})$  optimal à chaque étape. Une étude théorique intéressante pourrait être d'essayer de déterminer, sachant qu'on ne peut raisonnablement effectuer qu'un nombre fini de pas pour chaque récurrence en  $v$ , une évaluation de l'approximation que l'on fait en suivant telle ou telle stratégie ( $n$  itérations en  $v$  à chaque étape, un nombre croissant d'itérations, etc.).

Incidentement, on peut noter, en revenant à notre exemple, que si  $\phi$  est affine ( $\phi(v) = v$ , par exemple), alors l'algorithme précédent (si la valeur de  $\rho$  assure la stabilité) donne toujours la bonne solution  $u^*$ . Mais les problèmes pour lesquels ce phénomène se produit vérifient un "certainty equivalence principle", c'est-à-dire que la solution est la même que celle du "problème déterministe équivalent" obtenu en

remplaçant les bruits par leur moyenne. C'est le cas, par exemple, si on considère un problème de commande optimale en boucle ouverte avec la dynamique (6.1) affine en  $x$ ,  $u$  et  $\omega$  et une fonction coût quadratique en  $(x, u)$ . Cela cesse d'être vrai si par exemple, les matrices de dynamique se mettent à dépendre de  $\omega$ . La classe de ces problèmes est donc très restreinte et peu intéressante il suffit de résoudre le problème déterministe équivalent.

En conclusion de ce paragraphe, on doit, pour l'instant dresser un constat d'échec quant à la possibilité de dérouler des procédures de décomposition/coordination qui traitent au même niveau les variables en boucle ouverte et les variables en boucle fermée, du moins si l'on veut rester dans l'esprit des méthodes de gradient stochastique.



Simulation équation différentielle ordinaire  
pour différentes valeurs de  $\rho$

## Conclusion Générale

Pour clore ce mémoire, nous pensons, modulo les considérations négatives du paragraphe précédent, avoir montré qu'un grand nombre d'algorithmes classiques de décomposition/coordination peuvent quand même être généralisés au cas stochastique où les aléas ne créent pas de couplage entre les sous-systèmes, à condition de rester vigilant (en s'aidant au besoin de la technique d'o.d.e.) et de "freiner" les variables en boucle ouverte que l'on optimise. Le Principe du Problème auxiliaire ayant justement cet effet dans le cas déterministe non-différentiable, il est logique qu'il permette de construire un grand nombre de d'algorithmes efficaces en stochastique. Nous n'avons pas abordé le problème du choix des fonctionnelles auxiliaires  $K_i$ , mais nous avons noté l'analogie du Principe du Problème Auxiliaire avec une méthode de Quasi-Newton. Malgré l'existence de résultats sur l'"optimalité" du gradient stochastique quant à la vitesse de convergence (Dodu et al. [1]), on peut se demander si des techniques de type "gradient conjugué stochastique (tentative dans Syski et al.[1]) ne seraient pas possibles avec un bon choix itératif des  $K_i$ .

De toute manière, après avoir rappelé le grand nombre d'algorithmes que le concepteur peut créer avec cette théorie (le choix des  $K_i$  possibles est immense) nous pensons, comme l'auteur du Principe du Problème Auxiliaire, "qu'il faut considérer cette théorie comme un guide général et un ensemble de "garde-fous" à l'intérieur desquels on doit s'adapter aux problèmes concrets rencontrés"

# Chapitre 7

## Bibliographie

### Références de l'Introduction

K. J. Arrow, L. Hurwicz & H. Uzawa [1], "Studies in Linear and Non Linear Programming", Stanford University Press, CA (1972).

A. Benveniste, P. Bernhard & G. Cohen [1], "On the Decomposition of Stochastic Control Problems, 1rst IFAC Symposium on LSSTA, Udine, Italy, 1976.

G. Cohen [1], "Décomposition et Coordination en optimisation déterministe différentiable et non-différentiable". *Thèse de Doctorat ès-Sciences Mathématiques*, Université de Paris-IX Dauphine, 1984.

G. Cohen [2], "Optimization by Decomposition and Coordination : a Unified Approach", IEEE Trans. on Automatic Control, AC-23, pp. 222-232, 1978.

Cohen & Zhu [3], "Decomposition Coordination Methods in Large-Scale Optimization Problems. The Nondifferentiable Case and the use of Augmented Lagrangians. In: J. B. Cruz, Ed., "Advances in Large-Scale Systems, Theory and Applications", Vol 1, JAI Press Inc., Greenwich, Connecticut, 1984.

G. B. Dantzig & P. Wolfe [1], "The Decomposition Algorithm for Linear Program", *Econometrica*, Vol. 29, pp. 767-778, 1961.

J. C. Dodu, M. Goursat, A. Hertz, J. P. Quadrat & M. Viot [1], "Méthodes de gradient stochastique pour l'optimisation des investissements dans un réseau électrique", *Bulletin de la Direction des Etudes et Recherches EDF, Série C, No. 2*, pp. 133-164, 1981.

Y. Ermoliev [1], "Stochastic Quasigradient Methods and their Applications in Systems Optimization", IIASA, WP-81-2, January 1981, Laxenburg, Austria.

Y. Ermoliev & G. Leonardi [2], "Some proposals for stochastic facility location Models", IIASA, WP-80-176, Déc. 1980.

H. Kushner & D. S. Clark [1], "Stochastic approximation for constrained and unconstrained

systems", Appl. Math. Sc., Vol. 26, Springer Verlag, 1978.

D. P. Looze & N. R. Sandell Jr. [1], "Decomposition of Linear Decentralized Stochastic Control Problems" IFAC Workshop on Control and Management of Integrated Industrial Complexes, Toulouse, France, 1977

L. S. Lasdon [1], "Optimization Theory for Large Scale Systems", MacMillan Company, 1970.

M. Minoux & J. Y. Serreault [1], "Subgradient Optimization and Large-Scale Programming: an Application to Optimum Multicommodity Network Synthesis with Security Constraints", Revue RAIRO Recherche Opérationnelle, Vol. 15, No. 2, Mai 1981

M. D. Mesarovic, D. Macko & Y. Takahara [1], "Theory of Hierarchical Multilevel Systems", Academic Press, New York, 1970.

J. P. Quadrat [1], "On Optimal Control Problem of Large Systems", INRIA, 1982.

L. Robbins & S. Monro [1], "A Stochastic Approximation Method", Ann. Math. Stat., 22, 1951, pp. 400-407

Y. Takahara [1], "Multilevel Approach To Dynamic Organization", Syst. Res. Center, case Western Reserve University, Cleveland, Ohio, Rept SRC-50-c-64-18, 1964.

## 7.1 Références de la synthèse

J.P. Aubin[1], "Analyse Hilbertienne", Cours de l'Ecole Polytechnique, édition 1980.

J.P. Aubin[2], "Inclusions différentielles", Cours de l'Université de Dauphine, 1982.

J. Céa[1], "Optimisation, Théorie et Algorithmes", Dunod, 1971.

G. Cohen[1], "Décomposition et Coordination en optimisation déterministe différentiable et non différentiable". *Thèse de Doctorat ès-Sciences Mathématiques*, Université de Paris-IX Dauphine, 1984.

P. Faurre[1], "Notes d'Optimisation", Cours de l'Ecole Polytechnique, ed. 1983.

I. Guikman & A. Skorokhod[1], "Introduction à la théorie des processus aléatoires", Editions du MIR, 1977

M. Métivier[1], "Semimartingales", De Gruyter Studies in Mathematics, 1982.

M. Métivier [2], Creppel, G. Ruget, A. Benveniste : In: "Outils et modèles mathématiques pour l'analyse des systèmes", Vol. 1, I. D. Landau Ed., Editions du CNRS, 1981.

L. Schwartz[1], "Analyse Hilbertienne", Hermann, 1979.



L. Schwartz[2], "Analyse : Topologie générale et Analyse fonctionnelle", Hermann, 1970.

D. L. Zhu[1], "Optimisation sous-différentiable et méthodes de décomposition", *Thèse de Doctorat d'Ingénieur, Ecole des Mines*, 1982.

## 7.2 Références des chapitres suivants

A. Bensoussan, J. L. Lions & R. Temam, "Sur les méthodes de décomposition, de décentralisation et de coordination et applications" In: J. L. Lions, G. I. Marchouk, Eds., Sur les méthodes numériques en Sciences Physiques et Economiques, Dunod, 1974.

J. P. Bertran [1], "Optimisation Stochastique dans un Espace de Hilbert et Analyse Fonctionnelle", *Thèse de Doctorat ès-Sciences Mathématiques*, Université de Nancy-1, 1975.

D. P. Bertsekas [1], "Necessary and Sufficient Conditions for a Penalty Method to be Exact", *Mathematical Programming*, Vol. 9, pp. 87-99, 1975.

I. Ekeland & R. Temam [1], Convex Analysis and Variational Problems, North-Holland, 1976.

J. Céa [1], "Optimisation, Théorie et Algorithmes", Dunod, 1971.

G. M. Heal [1], "Planning without Prices", *Review of Economic Studies*, 1969, pp. 347-362.

H. Kushner & D. S. Clark [1], "Stochastic approximation for constrained and unconstrained systems", *Appl. Math. Sc.*, Vol. 26, Springer Verlag, 1978.

L. Ljung [1], "Analysis of Recursive Stochastic Algorithms", *IEEE Trans. on Automatic Control*, AC-22, No 2., 1977, pp. 551-575.

L. Ljung & T. Soderstrom [2], Theory and Practice of Recursive Identification, MIT Press, Cambridge, MA, 1983.

M. Métivier & P. Priouret [3], "Application of a Lemma of Kushner and Clark to General Classes of Stochastic Algorithms", *Rapport interne 97 du Centre de Mathématiques Appliquées, Ecole Polytechnique, Palaiseau*, 1983.

W. Syski & A. Ruszczynski [1], "Stochastic Approximation Method with Gradient Averaging for Unconstrained Problems", *IEEE trans. on Automatic Control*, Vol. AC 28, No. 12, December 1983.