

DEA Modélisation et Méthodes Mathématiques en Économie

**Cours de DEA
2004**

OPTIMISATION DES GRANDS SYSTÈMES

Guy Cohen

CERMICS-École Nationale des Ponts et Chaussées
et
INRIA

Table des Matières

1	Introduction	1
1.1	Sur les “Grands Systèmes”	1
1.1.1	Tentative de caractérisation	1
1.1.2	Quelques exemples typiques de grands systèmes	1
1.1.3	Idées et problématiques	2
1.1.4	Cadre du cours	3
1.2	Décomposition-coordination en optimisation : généralités	3
1.2.1	Intérêt de l’optimisation	3
1.2.2	Avantages de la décomposition	3
1.2.3	Coordination	3
1.2.4	Un exemple	4
1.3	Aperçu du cours	5
1.4	Prérequis du cours	6
1.5	Corrigé des exercices	6
1.5.1	Corrigé de l’Exercice 1.1	6
1.5.2	Corrigé de l’Exercice 1.2	6
1.5.3	Corrigé de l’Exercice 1.3	7
2	Présentation élémentaire des méthodes de décomposition-coordination	9
2.1	Premier modèle	9
2.1.1	Introduction du premier modèle	9
2.1.2	Décomposition “par les prix”	10
2.1.3	Décomposition “par les quantités”	13
2.1.4	Décomposition “par prédiction”	19
2.1.5	Une classification politico-économique des méthodes de décomposition-coordination	25
2.2	Deuxième modèle	26
2.2.1	Introduction du second modèle	26
2.2.2	Décomposition par les prix	27
2.2.3	Décomposition par les quantités	28
2.2.4	Décomposition “par prédiction”	29
2.3	Annexe : À propos des séries divergentes	30
2.4	Annexe : Exercices sur la dualité	30
2.5	Corrigé des exercices	33
2.5.1	Corrigé de l’Exercice 2.3	33
2.5.2	Corrigé de l’Exercice 2.6	33
2.5.3	Corrigé de l’Exercice 2.9	33
2.5.4	Corrigé de l’Exercice 2.13	34
2.5.5	Corrigé de l’Exercice 2.15	34
2.5.6	Corrigé de l’Exercice 2.16	34
2.5.7	Corrigé de l’Exercice 2.19	35
2.5.8	Corrigé de l’Exercice 2.21	35
2.5.9	Corrigé de l’Exercice 2.23	36
2.5.10	Corrigé de l’Exercice 2.24	36
2.5.11	Corrigé de l’Exercice 2.25	36

2.5.12	Corrigé de l'Exercice 2.27	37
2.5.13	Corrigé de l'Exercice 2.29	37
2.5.14	Corrigé de l'Exercice 2.30	38
2.5.15	Corrigé de l'Exercice 2.31	40
2.5.16	Corrigé de l'Exercice 2.32	41
3	Le Principe du Problème Auxiliaire	43
3.1	Introduction	43
3.2	Le Principe du Problème Auxiliaire en optimisation sur un ensemble admissible	44
3.2.1	Cadre général	44
3.2.2	Idée et algorithme de base	45
3.2.3	Utilisations	46
3.2.3.1	Algorithme “prox”	46
3.2.3.2	Algorithme du gradient et du gradient projeté	46
3.2.3.3	Algorithme de Newton	46
3.2.3.4	Décomposition	47
3.2.4	Étude de la convergence	48
3.2.4.1	Schéma général	48
3.2.4.2	Convergence de l'Algorithme 3.3	48
3.2.4.3	Un cas de convergence de toute la suite	53
3.2.4.4	Vitesse de convergence dans le cas \mathcal{J} fortement convexe	54
3.2.4.5	Quelques mots du cas sous-différentiable	55
3.2.5	Variantes	55
3.2.5.1	Augmentation de la fonction auxiliaire	55
3.2.5.2	Relaxation	56
3.2.5.3	Version séquentielle	56
3.2.5.4	Linéarisation partielle	57
3.2.6	Une application	57
3.2.6.1	Exposé du problème	57
3.2.6.2	Approche “ad hoc”	57
3.2.6.3	Approche par le PPA	58
3.3	Le Principe du Problème Auxiliaire en optimisation sous contraintes explicites	59
3.3.1	Le PPA et les problèmes de point selle	60
3.3.2	Le PPA et la décomposition par les prix	60
3.3.2.1	Obtention de l'algorithme général	60
3.3.2.2	Utilisations	61
3.3.2.3	Convergence	62
3.3.3	Le PPA et la décomposition par prédiction	67
3.3.3.1	Obtention de l'algorithme de point fixe	68
3.3.3.2	Utilisations	69
3.3.3.3	Convergence	70
3.3.3.4	Discussion de la condition géométrique	72
3.3.3.5	Obtention de l'algorithme d'Arrow-Hurwicz	73
3.4	Annexe : Monotonie, forte monotonie et forte convexité	76
3.5	Annexe : Propriété de Lipschitz	77
3.6	Annexe : Propriété de Dunn	78
3.7	Annexe : Problèmes sous contraintes et dualité	80
3.7.1	Formulation et relation d'ordre dans un espace de Hilbert	80
3.7.2	Convexité et croissance en relation avec un cône positif	81
3.7.3	Cône positif dual	82
3.7.4	Lagrangien et dualité	82
3.8	Corrigé des exercices	83
3.8.1	Corrigé de l'Exercice 3.4	83
3.8.2	Corrigé de l'Exercice 3.12	84
3.8.3	Corrigé de l'Exercice 3.13	84

3.8.4	Corrigé de l'Exercice 3.16	84
3.8.5	Corrigé de l'Exercice 3.21	84
3.8.6	Corrigé de l'Exercice 3.22	84
3.8.7	Corrigé de l'Exercice 3.25	85
3.8.8	Corrigé de l'Exercice 3.30	85
3.8.9	Corrigé de l'Exercice 3.32	85
3.8.10	Corrigé de l'Exercice 3.35	86
3.8.11	Corrigé de l'Exercice 3.37	86
3.8.12	Corrigé de l'Exercice 3.49	86
3.8.13	Corrigé de l'Exercice 3.50	86
3.8.14	Corrigé de l'Exercice 3.51	86
3.8.15	Corrigé de l'Exercice 3.52	87
4	Régularisation et Lagrangien augmenté	89
4.1	Lien entre Lagrangien augmenté et régularisation	89
4.1.1	Non convexité et convexité plus ou moins forte	89
4.1.2	Lagrangien augmenté	89
4.1.3	Régularisation de Yosida-Moreau	93
4.2	Un point de vue sur la régularisation	94
4.2.1	Bref rappel sur la transformée de Fenchel(-Legendre) et sur l'inf-convolution	94
4.2.2	Transmission de propriétés par l'inf-convolution	96
4.2.3	Zéro d'un opérateur et inf-convolution	97
4.2.4	Apport de la régularisation de Yosida-Moreau	98
4.3	Retour sur les Lagrangiens augmentés	99
4.3.1	Une nouvelle définition du Lagrangien augmenté	99
4.3.2	Propriétés du Lagrangien augmenté	101
4.4	Algorithme	102
4.4.1	Obtention de l'algorithme	102
4.4.2	Convergence	103
4.5	Corrigé des exercices	104
4.5.1	Corrigé de l'Exercice 4.2	104
4.5.2	Corrigé de l'Exercice 4.5	105
4.5.3	Corrigé de l'Exercice 4.6	105
4.5.4	Corrigé de l'Exercice 4.8	106
4.5.5	Corrigé de l'Exercice 4.12	106
4.5.6	Corrigé de l'Exercice 4.17	107
4.5.7	Corrigé de l'Exercice 4.21	107
	Bibliographie	109

Chapitre 1

Introduction

1.1 Sur les “Grands Systèmes”

1.1.1 Tentative de caractérisation

On peut parler de “*grand système*” ou encore de “*système complexe*” dans l’un des cas suivants :

- le système est décrit par un *grand nombre de variables et de contraintes*, ce qui entraîne beaucoup de calculs et/ou de stockage (la situation la plus typique étant celle où l’on doit utiliser la *programmation dynamique* pour résoudre le problème associé) ;
- le système global est fait de sous-systèmes interconnectés : il a donc une *structure spatiale complexe* (voir Figure 1.1), ce qui s’accompagne souvent d’une certaine *hétérogénéité* ;
- le système met en jeu des phénomènes à *plusieurs échelles de temps* ou changeant de façon brutale au cours du temps : dans ce cas, c’est la *structure temporelle* qui est *complexe* (exemple : fusée lançant ses étages successifs — Figure 1.2) ;
- le système est commandé simultanément par *plusieurs décideurs* agissant sur (voir Figure 1.3) l’ensemble du système ou sur des systèmes interconnectés avec
 - des informations différentes : en effet, la centralisation des informations peut être impossible ou économiquement prohibitive pour des systèmes très étendus géographiquement),
 - des objectifs concordants, ou bien différents, voire conflictuels.

1.1.2 Quelques exemples typiques de grands systèmes

On les rencontre notamment à travers les grands ensembles de production, de distribution et de services (eau, électricité, gaz, téléphone et autres télécommunications, etc.), les réseaux (de distribution, de transport), en économie (entreprises à filiales ou unités multiples), ...



Figure 1.1: Sous-systèmes interconnectés

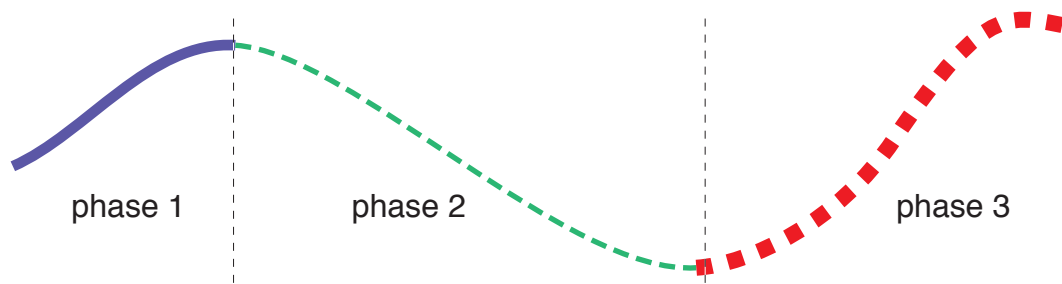


Figure 1.2: Trajectoire composite

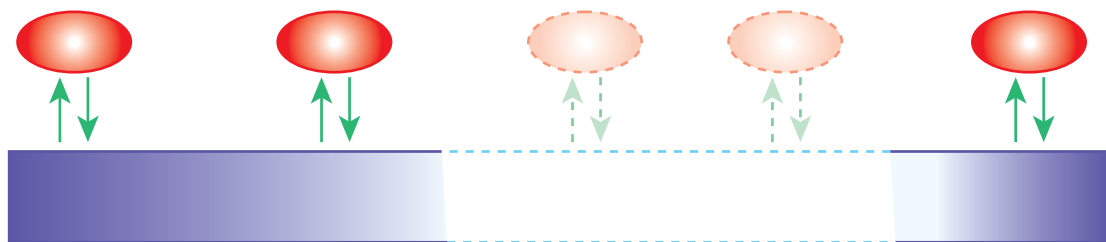


Figure 1.3: Décision multi-agents

1.1.3 Idées et problématiques

Pour contourner ces difficultés et exploiter les caractéristiques ci-dessus, on fait appel aux idées générales suivantes :

- idée de la *décomposition* pour isoler des sous-systèmes
 - plus petits,
 - plus homogènes ;
- idée de la *parallélisation* des tâches ;
- en situation stochastique, l'action de plusieurs décideurs avec des objectifs communs relève de la *commande stochastique décentralisée* et de la *théorie des équipes* ;
- lorsque les objectifs sont différents voire conflictuels, on entre dans le champ de la *théorie des équilibres et des jeux*.

1.1.4 Cadre du cours

Dans ce cours, on se limite à la *théorie de la décomposition-coordination en optimisation* : cette situation se caractérise par

- la présence d'un *seul décideur* avec toute l'information disponible (décision *centralisée*) ;
- la définition *un seul objectif* à optimiser ;
- on se limite enfin à la situation *déterministe*.

1.2 Décomposition-coordination en optimisation : généralités

1.2.1 Intérêt de l'optimisation

Du fait de leur taille, les grands systèmes sont soumis à une *économie d'échelle* : pour de tels systèmes, 1% d'économie peut représenter beaucoup en valeur absolue.

1.2.2 Avantages de la décomposition

- Les sous-systèmes étant *plus petits* que le système global sont *plus faciles à traiter* (dix sous-systèmes de dimension 1, c'est plus facile à traiter qu'un sous-système de dimension 10 : question de volume total des calculs, de contrôle des calculs — par exemple, éviter l'usage de la mémoire virtuelle).
- Les sous-systèmes étant *plus homogènes* sont plus faciles
 - à comprendre ;
 - à traiter (*conditionnement* d'un sous-problème meilleur que le conditionnement global) ;
 - à simplifier (idée de l'*agrégation*).
- On a la possibilité d'*adapter l'algorithme de résolution* à chaque type de sous-problème ("croisement d'algorithmes").
- On a la possibilité de *paralléliser les calculs* (parallélisme MIMD : Multiple Instructions Multiple Data).
- Enfin, la résolution par décomposition constitue un modèle mathématique de *décentralisation des décisions*.

1.2.3 Coordination

La coordination intervient à un niveau hiérarchique supérieur comme illustré par la Figure 1.4. La théorie de la décomposition-coordination est au carrefour de

- l'Économie mathématique ;
- la Recherche Opérationnelle ;
- l'Automatique ;
- l'Informatique (calcul parallèle et réparti).

La coordination

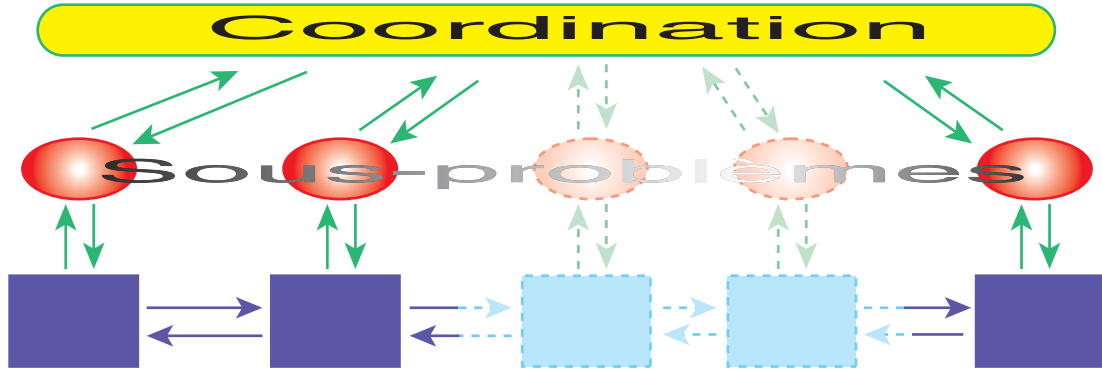


Figure 1.4: Coordination

- est nécessaire parce qu'on ne renonce pas à la recherche de l'*optimum collectif* ;
- est un *processus itératif* ;
- est un “*méta-algorithme*” d'optimisation : chaque itération est une “question-réponse” entre la coordination et les sous-problèmes.

1.2.4 Un exemple

On considère le problème suivant où toutes les variables sont scalaires :

$$\min_{u_1, u_2, y_1, y_2} \left(\underbrace{(u_1 - 7)^2 + y_1^2}_{J_1(u_1, y_1)} + \underbrace{(u_2 - 8)^2 + y_2^2}_{J_2(u_2, y_2)} \right) \quad (1.1a)$$

sous

$$y_1 = \underbrace{u_1 + u_2}_{P_1(u_1, u_2)} \quad \text{et} \quad y_2 = \underbrace{u_1 + u_2}_{P_2(u_1, u_2)} . \quad (1.1b)$$

On peut essayer d'obtenir la solution par le processus itératif suivant : à l'étape k , la coordination communique (ou “*prédit*”) u_1^k au sous-système 2 et u_2^k au sous-système 1, et les sous-problèmes à résoudre sont alors :

$$\min J_i(u_i, y_i) \quad \text{avec} \quad y_i = P_i(u_i, u_j^k), \quad i = 1, 2, \quad j \neq i, \quad (1.2)$$

ce qui définit deux nouvelles valeurs de u_1 et u_2 utilisées à l'étape suivante comme nouvelles “prédictions” u_1^{k+1} et u_2^{k+1} . Ceci est illustré par la Figure 1.5.

Explicitement, cela revient à la récurrence suivante :

$$u_1^{k+1} = \frac{1}{2}(7 - u_2^k), \quad u_2^{k+1} = \frac{1}{2}(8 - u_1^k). \quad (1.3)$$

On montre facilement que cette récurrence est contractante et que son point fixe est

$$u_1^\infty = 2, \quad u_2^\infty = 3, \quad y_1^\infty = 5, \quad y_2^\infty = 5. \quad (1.4)$$

Exercice 1.1. Retrouver la récurrence (1.3), montrer qu'elle est contractante, et retrouver le point d'équilibre (1.4).

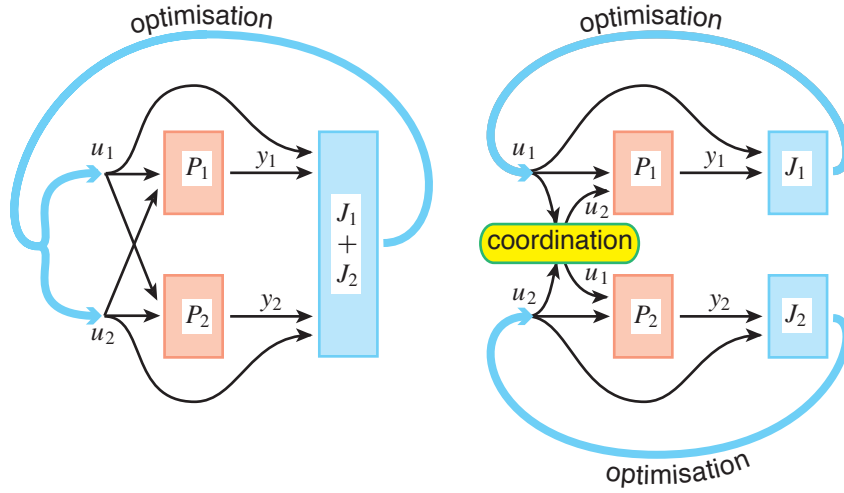


Figure 1.5: Un exemple

La question est alors de savoir si on peut affirmer que le point d'équilibre du processus itératif (1.4) fournit la solution du problème global (1.1).

En fait, la solution du problème (1.1) est

$$u_1^\# = 1, \quad u_2^\# = 2, \quad y_1^\# = 3, \quad y_2^\# = 3. \quad (1.5)$$

Trouver l'erreur !

Exercice 1.2. Écrire et résoudre les conditions d'optimalité du problème (1.1).

En fait, en définissant

$$G_i(u_i, u_j) = J_i(u_i, P_i(u_i, u_j)), \quad i = 1, 2, \quad j \neq i, \quad (1.6)$$

on réalise qu'on n'a pas résolu le problème d'optimisation posé, mais plutôt l'équilibre de Nash

$$\forall u_i, \quad G_i(u_i^\infty, u_j^\infty) \leq G_i(u_i, u_j^\infty) \quad i = 1, 2, \quad j \neq i. \quad (1.7)$$

Exercice 1.3. Démontrer cette affirmation en traduisant la méthode de coordination ci-dessus avec les notations (1.6), puis en se plaçant à l'équilibre. Expliquer en quoi les conditions d'optimalité de (1.7) diffèrent de celles du problème (1.2), une fois celui-ci réexprimé en termes des fonctions G_i .

1.3 Aperçu du cours

Le cours comprend essentiellement deux parties :

1. dans la *première partie*, on n'a aucune préoccupation de généralité ni de rigueur mathématique ; on cherche plutôt, sur deux modèles simples, à introduire les idées sur la décomposition-coordination, à développer les interprétations, etc. ;
2. dans la *deuxième partie*, on montre une théorie générale basée sur le “*principe du problème auxiliaire*” : dans ce cadre, il devient possible

- de rendre compte de la plupart des méthodes à partir d'un minimum de principes de base ;
- de lever certaines restrictions (hypothèses de séparabilité ou d'additivité des fonctions-coût et des contraintes) qui paraissaient essentielles dans la première partie ;
- d'étudier la convergence des algorithmes itératifs de coordination dans un cadre unifié.

1.4 Prérequis du cours

On suppose une certaine familiarité avec les théories et notions de base suivantes :

- analyse convexe ;
- théorie des conditions nécessaires d'optimalité du premier ordre (Lagrange-Kuhn-Tucker), et éventuellement du second ordre ;
- théorie des conditions suffisantes (cas convexe, Lagrangien, dualité et fonction duale, point-selle, etc.) ;
- interprétation marginaliste des multiplicateurs de Lagrange-Kuhn-Tucker (fonction perturbation, i.e. coût optimal comme fonction du second membre des contraintes).

Pour toutes ces notions, on pourra consulter le cours “Convexité et Optimisation” [9] du même auteur, ainsi que les références bibliographiques qui y sont mentionnées.

1.5 Corrigé des exercices

1.5.1 Corrigé de l'Exercice 1.1

En remplaçant y_1 dans l'expression de J_1 par $u_1 + u_2$, en remplaçant ensuite u_2 par la “prédiction” u_2^k , puis en annulant la dérivée de l'expression ainsi obtenue par rapport à u_1 au point $u_1 = u_1^{k+1}$, on obtient la formule (1.3) pour u_1^{k+1} . On procède d'une façon analogue pour obtenir u_2^{k+1} .

La récurrence (1.3) est une relation affine entre u^k et u^{k+1} dont la partie linéaire fait intervenir la matrice

$$\begin{pmatrix} 0 & -1/2 \\ -1/2 & 0 \end{pmatrix}$$

dont les valeurs propres sont de module $1/2$ inférieur à 1. La relation affine est donc contractante. Le point d'équilibre est obtenu en posant $u^{k+1} = u^k$ dans (1.3) et en résolvant le système en u ainsi obtenu : la solution est (1.4).

1.5.2 Corrigé de l'Exercice 1.2

On a deux possibilités : soit on écrit les conditions de Lagrange/Kuhn-Tucker pour le Lagrangien du problème (1.1) en ayant dualisé les contraintes qui relient y à u , soit on substitue à chaque y_i son expression en fonction des deux u_i dans J_i (voir (1.6)) et on écrit la stationnarité du critère ainsi défini (c'est-à-dire $G_1 + G_2$) par rapport à (u_1, u_2) . Cette seconde façon de procéder est évidemment plus simple et conduit à un système de deux équations (stationnarité de $G_1 + G_2$ en u_1 et u_2) à deux inconnues (u_1, u_2) . En résolvant ce système, on obtient la solution (1.5).

1.5.3 Corrigé de l'Exercice 1.3

La définition des fonctions G_i par (1.6) découle de l'élimination des variables intermédiaires y_i au profit des variables de décision u_i . Puisque la coordination fixe, dans le problème i , la valeur a priori de la variable u_j pour $j \neq i$, l'optimisation de chaque G_i se fait bien par rapport à u_i exclusivement (voir (1.2)), comme l'indiquent aussi les inégalités (1.7) qui se trouvent donc vérifiées à l'équilibre (c'est-à-dire quand les valeurs des u_i fixées a priori sont retrouvées comme solutions des sous-problèmes).

Ce point d'équilibre annule donc les dérivées partielles de *chacun* des G_i par rapport au u_i correspondant, contrairement à ce qui devrait se passer lorsque l'on considère la solution de (1.1). En effet, comme on l'a vu dans le corrigé précédent, ce problème revient à minimiser la *somme* $G_1 + G_2$ par rapport à (u_1, u_2) . Par conséquent, la solution (1.5) est caractérisée par l'annulation des dérivées partielles de cette somme — et non pas de chaque G_i — par rapport aux deux u_i .

Chapitre 2

Présentation élémentaire des méthodes de décomposition-coordination

2.1 Premier modèle

2.1.1 Introduction du premier modèle

On considère le problème

$$\min_{u_i \in U_i^{\text{ad}}} \sum_{i=1}^N J_i(u_i) \quad (2.1a)$$

sous

$$\sum_{i=1}^N \Theta_i(u_i) = \vartheta . \quad (2.1b)$$

Ici,

- $u_i \in \mathbb{R}^{n_i}$,
- U_i^{ad} est un sous-ensemble convexe fermé de \mathbb{R}^{n_i} ,
- $J_i : \mathbb{R}^{n_i} \rightarrow \mathbb{R}$,
- $\Theta_i : \mathbb{R}^{n_i} \rightarrow \mathbb{R}^m$,
- $\vartheta \in \mathbb{R}^m$ est un second membre de contraintes donné.

Par commodité, on notera

- $u = (u_1^\top, \dots, u_N^\top)^\top$ où $^\top$ symbolise la transposition ;
- par conséquent $u \in \mathbb{R}^n$ avec $n = \sum_{i=1}^N n_i$;
- $J(u) = \sum_{i=1}^N J_i(u_i)$, par conséquent $J : \mathbb{R}^n \rightarrow \mathbb{R}$;
- $\Theta(u) = \sum_{i=1}^N \Theta_i(u_i)$, par conséquent $\Theta : \mathbb{R}^n \rightarrow \mathbb{R}^m$.

Par souci de simplicité, le cas de contraintes (2.1b) de type égalité sera essentiellement considéré. Lorsque des différences notables le justifieront, on évoquera le cas de contraintes inégalité.¹

Une interprétation économique possible de ce problème est la suivante :

¹Le signe = est remplacé par le signe \leq coordonnée par coordonnée.

- N unités de production coopèrent au sein d'une même entreprise ; u_i représente les variables de décision de l'unité i , $J_i(u_i)$ son coût ;
- les contraintes (2.1b), qui sont les contraintes couplantes,² peuvent être interprétées comme des objectifs de production (le vecteur ϑ est un vecteur de m quantités à produire) et/ou comme des contraintes sur les ressources disponibles (ϑ représente des quantités de ressources à consommer par les unités), et/ou encore comme des équations d'équilibre des échanges de produits intermédiaires entre les unités.

2.1.2 Décomposition “par les prix”

Intuitivement, une façon de décentraliser la gestion des unités est de faire payer les consommations (ou de rémunérer les productions) de chaque unité suivant un certain système de prix, et de laisser chaque unité optimiser son propre bilan. Ce bilan est composé du coût $J_i(u_i)$ et du coût induit par la coordination $\langle p, \Theta_i(u_i) \rangle$, où p est un vecteur des prix et $\langle \cdot, \cdot \rangle$ est le produit scalaire dans \mathbb{R}^m . Le problème de gestion de l'unité i se formule donc

$$\min_{u_i \in U_i^{\text{ad}}} (J_i(u_i) + \langle p, \Theta_i(u_i) \rangle) . \quad (2.2)$$

L'objectif de la coordination est alors d'assurer l'objectif de production global, autrement dit, d'assurer le respect de la contrainte (2.1b) par le choix d'un bon système de prix p .

Soit $\widehat{U}_i(p)$ l'ensemble des solutions (supposé non vide) de (2.2) (et soit $\widehat{U}(p) = \widehat{U}_1(p) \times \cdots \times \widehat{U}_N(p)$). Trois ou quatre questions naturelles se posent alors.

Questions 2.1.

1. S'il existe $p^\sharp \in \mathbb{R}^m$ tel qu'il existe $\widehat{u} \in \widehat{U}(p^\sharp)$ vérifiant (2.1b), alors est-ce que \widehat{u} est solution du problème (2.1) ? La question est en effet de savoir comment la coordination peut reconnaître que l'optimum du problème global est atteint par la solution trouvée (par morceaux) par les sous-problèmes.
2. Existe-t-il une telle valeur de p^\sharp (à supposer que la réponse à la première question est positive) ? C'est la question de la *coordonnabilité* par les prix : existe-t-il un tel instrument de coordination ?
3. Quel algorithme permet de trouver p^\sharp (à supposer que la réponse aux deux premières questions est positive) ?
4. Une autre question pratique se pose dans le cas où $\widehat{U}(p)$ n'est pas réduit à un singleton : tout $\widehat{u} \in \widehat{U}(p^\sharp)$ est-il optimal,³ et sinon comment sélectionner le bon ?

Le lemme suivant, dont la démonstration est laissée en exercice, apporte une réponse aux deux premières questions.

Lemme 2.2. *Si il existe p^\sharp et $\widehat{u} \in \widehat{U}(p^\sharp)$ vérifiant (2.1b), alors la paire (\widehat{u}, p^\sharp) est un point selle du Lagrangien*

$$L(u, p) = J(u) + \langle p, \Theta(u) - \vartheta \rangle \quad (2.3)$$

sur $U^{\text{ad}} \times \mathbb{R}^m$, et donc \widehat{u} est une solution de (2.1).

Réciproquement, si L a un point selle (u^\sharp, p^\sharp) sur $U^{\text{ad}} \times \mathbb{R}^m$, alors u^\sharp appartient à $\widehat{U}(p^\sharp)$ et vérifie (2.1b).

²Les contraintes $u_i \in U_i^{\text{ad}}$, qui peuvent représenter un modèle de chaque unité ou des contraintes de fonctionnement, sont, elles, découplées.

³À la question 2.1-1, on a supposé seulement l'existence d'un “bon” \widehat{u} .

Exercice 2.3. Démontrer le lemme ci-dessus.

Commentaires 2.4.

- (a) Le lemme ci-dessus montre que la réussite de la coordination par les prix est liée à l'existence *nécessaire* d'un point selle du Lagrangien du problème global (voir cependant le commentaire 2.4-b ci-dessous). Moyennant cette existence, le problème (2.1) — de toute façon toujours équivalent au problème du min max de L — est remplacé par le problème du max min. Dans la phase de minimisation à p fixé, on tire alors profit du fait que $L(u, p)$ (voir (2.3)) s'écrit comme une somme de $L_i(u_i, p)$ (le terme $-\langle p, \vartheta \rangle$ étant affecté arbitrairement à l'un des termes de la somme). En effet :

$$J(u) + \langle p, \Theta(u) - \vartheta \rangle = \sum_{i=1}^N (J_i(u_i) + \langle p, \Theta_i(u_i) \rangle) - \langle p, \vartheta \rangle .$$

L'existence d'un point selle de L ne peut généralement être garantie que moyennant des hypothèses de convexité, de qualification des contraintes, et de comportement à l'infini des fonctions J_i si les U_i^{ad} ne sont pas bornés. Il faut considérer cette constatation de la nécessité de l'existence d'un point selle du Lagrangien global comme une assez sensible limitation du champ d'utilisation de la méthode. L'utilisation d'une technique de Lagrangien augmenté⁴ permet d'étendre ce champ (voir l'Exemple 2.5 ci-dessous) mais il faut surmonter, du point de vue de la décomposition, la perte d'additivité du terme supplémentaire apporté par le Lagrangien augmenté. On renvoie au Chapitre 4 où ce problème est traité.

- (b) En plus des solutions de (2.2), intéressons nous aussi aux points stationnaires⁵ dont on peut garantir que ce sont des minima locaux si, de plus, la dérivée seconde de (2.2) est définie positive. En supposant qu'une famille $\{\widehat{u}_i\}_{i=1,\dots,N}$ de tels points stationnaires — correspondant à une certaine valeur de p — vérifient aussi globalement la contrainte (2.1b), alors on montre facilement que ces points vérifient les conditions nécessaires d'optimalité du premier ordre — voire du second ordre pour des minima locaux — pour le problème global (2.1) : ils peuvent donc éventuellement constituer des solutions de ce problème, même si le Lagrangien global n'a pas de point selle.

Il y a cependant dans ce cas une difficulté pour évaluer algorithmiquement la valeur correspondante de p : une telle valeur de p peut n'avoir aucune propriété caractéristique vis-à-vis de la fonctionnelle duale définie par

$$\psi(p) = \min_{u \in U^{\text{ad}}} L(u, p) \quad (2.4)$$

comme l'illustre l'Exemple 2.5 que l'on trouvera plus loin, ce qui rend difficile sa recherche par un algorithme itératif.

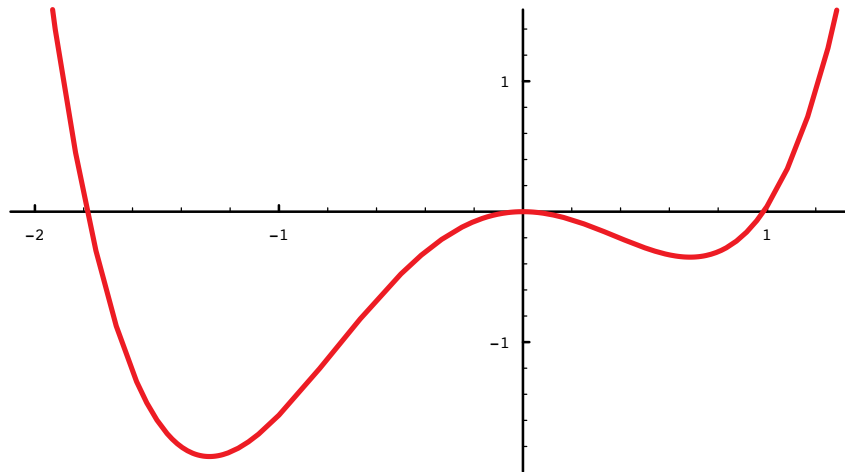
- (c) Supposant à nouveau l'existence d'un point selle (u^\sharp, p^\sharp) du Lagrangien du problème global, on considère ici la question de la non unicité de $\widehat{u}(p^\sharp)$. En général, dans ce cas, tout point de $\widehat{U}(p^\sharp)$ ne vérifie pas les contraintes (2.1b), et donc a fortiori n'est pas une solution du problème global. Il ne suffit donc pas de trouver un “bon” prix p^\sharp pour avoir résolu le problème.⁶ Cette difficulté est évitée si on fait une hypothèse garantissant l'unicité de $\widehat{u}(p)$ (par exemple J strictement convexe⁷) ou si le Lagrangien est “stable en u ”, i.e. *tout* $u \in \widehat{U}(p^\sharp)$ est une solution ; ceci est en particulier garanti avec l'usage du Lagrangien augmenté.

⁴Dans le cas d'une contrainte égalité comme (2.1b), on rajoute à l'expression (2.3) de L un terme $c \|\sum_{i=1}^N \Theta_i(u_i) - \vartheta\|^2/2$ où $c \in \mathbb{R}^+$.

⁵dans le cas le plus simple où J_i et Θ_i sont — éventuellement, deux fois — différentiables, et où $U^{\text{ad}} = \mathbb{R}^n$

⁶Voir Exercice 2.30 en fin de chapitre.

⁷En fait, pour démontrer la convergence de l'algorithme d'Uzawa dont il est question plus loin, il faut même supposer la *forte* convexité de J .

Figure 2.1: Graphe de $L(\cdot, 0,768)$

- (d) Dans le cas où les contraintes (2.1b) sont des contraintes inégalité $\Theta(u) \leq \vartheta$, on sait que le multiplicateur p doit être non négatif ($p \in (\mathbb{R}^+)^m$). Une réponse positive à la question 2.1-1 n'est alors possible qu'en rajoutant une condition supplémentaire d'"écarts complémentaires", c'est-à-dire que le test d'optimalité pour des solutions $\hat{u}_i(p)$ des sous-problèmes est maintenant

$$\Theta(\hat{u}(p)) \leq \vartheta \text{ et } \langle p, \Theta(\hat{u}(p)) - \vartheta \rangle = 0. \quad (2.5)$$

Cette condition d'écarts complémentaires s'interprète intuitivement comme le fait que "les prix ne doivent pas être plus forts que le respect des contraintes globales ne l'exige".

- (e) L'avantage pratique de la décomposition par les prix est de faire disparaître un certain nombre de contraintes au niveau des sous-problèmes (par opposition à la méthode par les quantités décrite ci-dessous). L'inconvénient correspondant est que les solutions $\hat{u}_i(p)$ des sous-problèmes ne respecteront la contrainte globale (2.1b) qu'une fois que le système de prix aura convergé vers sa valeur optimale au cours de l'algorithme décrit ci-dessous. C'est pourquoi on parle de "*méthode non admissible*" à propos de la coordination par les prix.

Exemple 2.5. Il s'agit ici d'un exemple en dimension 1, ne faisant donc pas intervenir la notion de décomposition, mais cela n'a pas d'importance pour le phénomène à illustrer. Soit à résoudre

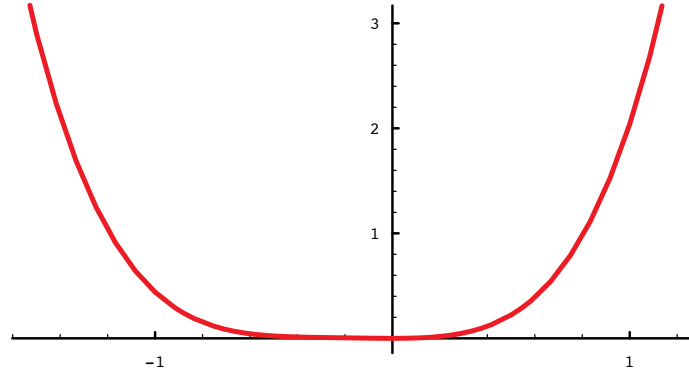
$$\min J(u) \text{ avec } J(u) = u^4 + 0,8u^3 - 1,76u^2 - 0,768u$$

sous la contrainte $u = 0$ (solution évidente !). Le Lagrangien $L(u, p) = J(u) + pu$ est stationnaire en $u^\sharp = 0$ pour $p^\sharp = 0,768$ mais u^\sharp correspond à un maximum local de $L(\cdot, p^\sharp)$, et cette fonction a deux autres points stationnaires, dont un minimum absolu (voir Figure 2.1). Par ailleurs, ψ (voir (2.4)), toujours concave comme enveloppe inférieure de fonctions affines, est maximale en $p = 0$:⁸ $p^\sharp = 0,768$ est donc une valeur "quelconque" vis-à-vis de cette fonction.

Notons de plus sur cet exemple que l'utilisation du Lagrangien augmenté $L_c(u, p) = J(u) + pu + cu^2/2$ avec par exemple $c = 4$ résout toutes ces difficultés puisque $u^\sharp = 0$ devient le minimum unique de $L(\cdot, p^\sharp)$ avec $p^\sharp = 0,768$ (voir Figure 2.2) et $\psi_c(p) = \min_u L_c(u, p)$ est maximale en p^\sharp .

Exercice 2.6. Reprendre la question des contraintes inégalité évoquée au commentaire 2.4-d et montrer la nécessité des conditions (2.5) pour pouvoir démontrer l'équivalent du Lemme 2.2 dans ce cas.

⁸Voir Exercice 2.29 en fin de chapitre.

Figure 2.2: Graphe de $L_c(\cdot, 0,768)$ pour $c = 4$

Algorithmes. Avec l’interprétation économique utilisée précédemment, si à l’itération k de coordination, avec les prix courant p^k , la “demande” globale $\Theta(u^k)$ des unités de production (où $u^k \in \widehat{U}(p^k)$) excède l’“offre” ϑ , il faut augmenter les prix, cette augmentation étant par exemple directement proportionnelle à l’écart constaté entre demande et offre, d’où la formule itérative

$$p^{k+1} = p^k + \rho^k (\Theta(u^k) - \vartheta) , \quad (2.6)$$

où ρ^k est un “pas” positif.

Du point de vue mathématique, cet algorithme est en fait l’algorithme d’Uzawa qui peut encore s’interpréter comme un algorithme de gradient (ou de sous-gradient) pour maximiser la fonctionnelle concave ψ (voir (2.4)). Une condition suffisante de différentiabilité de ψ est que $\widehat{U}(p)$ soit toujours réduit à un singleton. Dans le cas contraire, en plus de la non-différentiabilité de ψ , on est confronté au problème déjà évoqué au commentaire 2.4-c quant à la convergence des u^k vers de “bonnes” solutions du problème global⁹.

D’autres formules itératives (gradients conjugués, méthode de Newton, etc.) peuvent également être considérées.

Dans le cas de contraintes inégalité, le second membre de (2.6) doit subir une opération de projection pour maintenir p^{k+1} non négatif. La formule devient

$$p_j^{k+1} = \max \left(0, p_j^k + \rho^k \left(\sum_{i=1}^N \Theta_{ji}(u_i^k) - \vartheta_j \right) \right) , \quad j = 1, \dots, m .$$

2.1.3 Décomposition “par les quantités”

Cette méthode de décomposition, également connue sous le nom d’“*allocation de ressources*”, de “*méthode directe*” ou “*primale*”, est duale de la méthode par les prix. Au lieu d’agir à l’aide d’un système de prix sur la formation des décisions des sous-problèmes, la coordination agit par l’intermédiaire de quantités, en distribuant la ressource disponible (ou en répartissant l’objectif de production) entre les unités de production. Le sous-problème i se formule alors ainsi :

$$\min_{u_i \in U_i^{\text{ad}}} J_i(u_i) \quad (2.7a)$$

sous

$$\Theta_i(u_i) = \theta_i , \quad (2.7b)$$

⁹Voir à nouveau l’Exercice 2.30 en fin de chapitre.

où le vecteur $\theta = (\theta_1^\top, \dots, \theta_N^\top)^\top$ appartient au sous-espace de $(\mathbb{R}^m)^N$ défini par la contrainte

$$\sum_{i=1}^N \theta_i = \vartheta. \quad (2.8)$$

On appellera “*allocation*” tout vecteur θ vérifiant cette contrainte. Le but de la coordination est donc de trouver la meilleure allocation au sens du critère global (2.1a), lorsque les u_i sont obtenus via la résolution de (2.7) (on notera $\widehat{U}_i(\theta_i)$ l’ensemble de ces solutions et $\widehat{U}(\theta) = \widehat{U}_1(\theta_1) \times \dots \times \widehat{U}_N(\theta_N)$).

Questions 2.7.

1. Existe-t-il une valeur θ^\sharp telle qu’il existe $\widehat{u} \in \widehat{U}(\theta^\sharp)$ qui soit solution du problème global. C’est la question de la *coordonnabilité* par allocation de ressources.
2. Si il existe une telle allocation, comment la caractériser au niveau de la coordination ? Dans le cas de la méthode par les prix, la réponse était fournie par la vérification de la contrainte globale (2.1b) par les solutions $\widehat{u}(p)$ des sous-problèmes (2.2) (plus une condition d’écarts complémentaires dans le cas de contraintes inégalité). Ici, ce test est ipso facto vérifié par la construction même des sous-problèmes et par la contrainte (2.8) portant sur l’allocation de ressources θ (c’est pourquoi cette approche est qualifiée d’“*admissible*”). Il faut donc trouver un autre test d’optimalité.
3. Quel algorithme permet de trouver θ^\sharp (à supposer qu’il existe) ?
4. Que se passe-t-il en cas de non unicité des solutions des sous-problèmes ?

Du point de vue mathématique, il est facile de voir que le problème global (2.1) est équivalent au problème suivant :

$$\min_{\theta} \sum_{i=1}^N \varphi_i(\theta_i) \text{ sous (2.8) ,} \quad (2.9)$$

où les fonctions φ_i sont définies par

$$\varphi_i(\theta_i) = \begin{cases} J_i(\widehat{u}_i(\theta_i)) & \text{si } \widehat{u}_i(\theta_i) \text{ est une solution de (2.7),} \\ +\infty & \text{si (2.7) n’a pas de solution.} \end{cases} \quad (2.10)$$

L’équivalence entre (2.1) d’une part et (2.9) d’autre part est précisée par le lemme suivant.

Lemme 2.8. *Si θ^\sharp est solution de (2.9), toute solution $\widehat{u}(\theta^\sharp) = (\widehat{u}_1^\top(\theta_1^\sharp), \dots, \widehat{u}_N^\top(\theta_N^\sharp))^\top$ est solution de (2.1). Réciproquement, si u^\sharp est solution de (2.1), alors*

$$\theta^\sharp = (\Theta_1^\top(u_1^\sharp), \dots, \Theta_N^\top(u_N^\sharp))^\top \quad (2.11)$$

est solution de (2.9).

Exercice 2.9. Donner une démonstration de ce lemme.

Ce lemme répond à la question de l’existence de “bons” paramètres de coordination.

Commentaires 2.10. On notera que contrairement au cas de la méthode par les prix, la coordonnabilité par allocation est pratiquement assurée dès que le problème global a des solutions, sans supposer que son Lagrangien admet un point selle.

Venons en maintenant à la question 2.7-2 de la caractérisation de “bons” paramètres de coordination. De façon duale à la méthode par les prix, on ne peut ici obtenir une caractérisation d’une bonne allocation qu’en faisant appel à une notion de “prix marginaux” ou de “sensibilité” des fonctions φ_i aux variations des θ_i , c’est-à-dire, mathématiquement parlant, à une notion de “dérivée” de φ_i (gradient, sous-gradient, dérivée directionnelle, etc.).

Supposons pour l’instant, et idéalement, que les φ_i sont différentiables. Une allocation sera optimale si et seulement si le gradient de $\sum_i \varphi_i$ est orthogonal à la variété affine définie par (2.8).

Lemme 2.11. *Un vecteur g de $\mathbb{R}^{N \times m}$ composé de sous-vecteurs $g_i, i = 1, \dots, N$, dans \mathbb{R}^m est orthogonal à la variété affine \mathcal{A} définie par (2.8) si et seulement si $g_i = g_j$ pour tous i, j . La projection d’un vecteur x de $\mathbb{R}^{N \times m}$ sur \mathcal{A} est le vecteur composé des sous-vecteurs i égaux à*

$$x_i - \frac{1}{N} \left(\sum_{j=1}^N x_j - \vartheta \right), \quad i = 1, \dots, N.$$

Démonstration. Un vecteur g est orthogonal à \mathcal{A} si, pour tous vecteurs θ et θ' dans \mathcal{A} , g est orthogonal à $\theta - \theta'$. Comme

$$\langle g, \theta - \theta' \rangle = \sum_{i=1}^N \langle g_i, \theta_i - \theta'_i \rangle,$$

l’égalité des g_i est une condition suffisante pour que cette expression soit nulle. Réciproquement, pour tous (i, j) et tout g supposé orthogonal à \mathcal{A} , prenons θ and θ' telles que

$$\begin{aligned} \theta_k &= \theta'_k = 0 \text{ si } k \neq i \text{ et } k \neq j, \\ \theta_i &= \vartheta + g_i - g_j \text{ et } \theta_j = g_j - g_i, \\ \theta'_i &= \vartheta + g_j - g_i \text{ et } \theta'_j = g_i - g_j. \end{aligned}$$

Observons que ces θ et θ' appartiennent bien à \mathcal{A} et que

$$\langle g, \theta - \theta' \rangle = 2 \|g_i - g_j\|^2,$$

et donc ce produit scalaire est nul seulement si $g_i = g_j$.

Pour calculer la projection $\Pi_{\mathcal{A}}(x)$ de x sur \mathcal{A} , on doit résoudre le problème d’optimisation sous contraintes suivant :

$$\min_{\theta} \frac{1}{2} \|\theta - x\|^2 \text{ sous (2.8) .}$$

Soit p un multiplicateur associé à la contrainte (2.8). En formant le Lagrangien et en dérivant par rapport à θ , on obtient: pour $i = 1, \dots, N$,

$$\theta_i^\# - x_i + p^\# = 0.$$

En sommant pour tout i , on a

$$p^\# = \frac{1}{N} \sum_{i=1}^N (x_i - \theta_i^\#) = \frac{1}{N} \left(\sum_{i=1}^N x_i - \vartheta \right),$$

en tenant compte de (2.8). Donc,

$$(\Pi_{\mathcal{A}})_i(x) = \theta_i^\# = x_i - p^\# = x_i - \frac{1}{N} \left(\sum_{i=1}^N x_i - \vartheta \right).$$

□

D'après ce lemme, on peut dire qu'une allocation sera optimale si et seulement si le gradient de $\sum_i \varphi_i$ a toutes ses coordonnées égales. Ceci traduit l'idée intuitive que, pour que l'on ne puisse plus améliorer une allocation en redistribuant différemment la ressource, il faut que toutes les utilités marginales soient égales.

Comment cette vérification peut-elle être faite sur la base des informations fournies par la résolution des sous-problèmes (2.7) ? La théorie de la dualité introduit des multiplicateurs de Lagrange p_i associés aux contraintes (2.7b) ; les multiplicateurs optimaux intervenant dans les conditions suffisantes (théorie globale de la dualité basée sur l'existence d'un point selle du Lagrangien) ou nécessaires (théorie locale dans le cas différentiable) peuvent s'interpréter, au signe près, comme les sensibilités des fonctions φ_i aux points θ_i (nous y reviendrons). On peut alors énoncer les lemmes suivants qui apportent une réponse à la question 2.7-2.

On suppose ici que $U_i^{\text{ad}} = \mathbb{R}^{n_i}$ pour tout i et que toutes les dérivées requises existent.

Lemme 2.12. *Si, pour une certaine allocation θ , et pour $i = 1, \dots, N$, le couple $(u_i^\#, p_i^\#) \in \mathbb{R}^{n_i} \times \mathbb{R}^m$ vérifient les conditions nécessaires du premier ordre*

$$J'_i(u_i^\#) + \left(\Theta'_i(u_i^\#) \right)^\top p_i^\# = 0, \quad (2.12a)$$

$$\Theta_i(u_i^\#) = \theta_i, \quad (2.12b)$$

et si tous les $p_i^\#$ sont égaux à une valeur commune $p^\#$, alors $(u^\#, p^\#)$ vérifie des conditions analogues pour le problème global (2.1) ($u^\#$ est la concaténation des $u_i^\#$ alors que $p^\#$ est la valeur commune des $p_i^\#$).

Exercice 2.13. Donner une démonstration de ce lemme.

On revient maintenant à l'hypothèse que U_i^{ad} est un sous-ensemble convexe fermé quelconque de \mathbb{R}^{n_i} et la différentiabilité des fonctions n'est plus requise.

Lemme 2.14. *Si pour $i = 1, \dots, N$, le Lagrangien du problème (2.7) admet un point selle $(u_i^\#, p_i^\#)$ sur $U_i^{\text{ad}} \times \mathbb{R}^m$ et si tous les $p_i^\#$ sont égaux à une valeur commune $p^\#$, alors $(u^\#, p^\#)$ est un point selle du Lagrangien du problème global (2.1) sur $U^{\text{ad}} \times \mathbb{R}^m$, et donc $u^\#$ est solution de ce problème.*

Exercice 2.15. Donner une démonstration de ce lemme.

Concernant les conditions nécessaires (respectivement, suffisantes) du second ordre,¹⁰ il n'est pas possible en général d'énoncer un résultat équivalent, en raison notamment du fait que le produit cartésien des sous-espaces des \mathbb{R}^{n_i} tangents aux contraintes (2.7b) est inclus dans le — mais n'est pas égal au — sous-espace de \mathbb{R}^n tangent à la contrainte (2.1b). Ceci est illustré par l'exemple suivant (l'exercice consiste à démontrer les affirmations contenues dans cet exemple).

Exemple et Exercice 2.16. Considérons un cas où

- $N = 2$,
- $u_i = (x_i, y_i) \in \mathbb{R}^2, i = 1, 2$,
- $U_i^{\text{ad}} = \mathbb{R}^2$,
- $J_1 = J_2 = J : \mathbb{R}^2 \rightarrow \mathbb{R}$, avec $J(x, y) = -x^2 - y^2 - 4xy$,

¹⁰Elles font intervenir la condition que la dérivée seconde en u du Lagrangien au point vérifiant les conditions du premier ordre soit semi-définie positive (respectivement, définie positive) sur le plan tangent aux contraintes en ce point.

- $\diamond m = 1, \vartheta = 0,$
- $\diamond \Theta_1 = -\Theta_2 = \Theta : \mathbb{R}^2 \rightarrow \mathbb{R}$
- $\diamond \Theta(x, y) = x + y.$

Pour l'allocation $\theta_1 = \theta_2 = 0$, le point $\hat{x}_i = \hat{y}^i = \hat{p}_i = 0, i = 1, 2,$ ⁽¹¹⁾ vérifie les conditions nécessaires du premier ordre et suffisantes du second ordre en ce qui concerne le problème (2.7) mais ce n'est pas le cas pour le problème (2.1). En fait, le point $0 \in \mathbb{R}^4$ n'est pas un minimum (même local) du problème global (il n'y en a pas et l'infimum vaut $-\infty$ ⁽¹²⁾).

L'explication de cette difficulté est la suivante :

- l'espace tangent aux contraintes $\Theta_1(u_1) = 0$ et $\Theta_2(u_2) = 0$ est le sous-espace de \mathbb{R}^4 des vecteurs de la forme $(a, -a, b, -b)^\top$ pour lesquels on vérifie facilement que la dérivée seconde de $J_1(u_1) + J_2(u_2)$ est définie positive ;
- le sous-espace tangent à la contrainte globale $\Theta_1(u_1) + \Theta_2(u_2)$ est le sous-espace de \mathbb{R}^4 des vecteurs de la forme $(a, b, c, a + b - c)^\top$ qui est plus grand que le précédent. Sur ce sous-espace, la dérivée seconde de $J_1(u_1) + J_2(u_2)$ n'est pas définie positive (essayer par exemple le vecteur $(a, 0, 0, a)^\top$).

Commentaires 2.17.

- (a) L'existence de multiplicateurs optimaux dans les sous-problèmes (2.7) est soumise à une condition de qualification des contraintes qui peut revêtir, par exemple dans le cas convexe, la forme $\theta_i \in \text{int } \Theta_i(U_i^{\text{ad}})$ (voir par exemple [9, Eq. (6.27a)]). Il est évident que la satisfaction d'une telle condition dépend d'abord des valeurs θ_i particulières choisies, et c'est la question même de l'existence de solutions primales admissibles qui est d'abord en cause (voir le point 2.17-d ci-dessous). De plus, l'unicité de multiplicateurs optimaux dans les sous-problèmes (pour les conditions nécessaires du premier ordre par exemple dans le cas simple où $U_i^{\text{ad}} = \mathbb{R}^{n_i}$) est liée à l'indépendance linéaire des gradients des fonctions-contraintes Θ_i , ou, autrement dit, à la surjectivité des applications linéaires $\Theta'_i(u_i) : \mathbb{R}^{n_i} \rightarrow \mathbb{R}^m$, pour tout i et pour tout u_i rencontré. Ces conditions sont bien plus difficiles à assurer que les conditions correspondantes pour le problème global (celles-ci ne portent que sur $\sum_i \Theta_i : \mathbb{R}^n \rightarrow \mathbb{R}^m$ et non sur chaque Θ_i séparément).
- (b) Revenons à la question de la différentiabilité des fonctions φ_i . En fait, il est pratiquement très difficile, sauf cas très particulier (problème à contraintes linéaires et à coût quadratique par exemple) de garantir la différentiabilité de ces fonctions, et corrélativement l'unicité des multiplicateurs optimaux p_i ⁽¹³⁾. Dans le cas convexe, on sait que $-p_i$ est un sous-gradient de φ_i (voir Exercice 2.31, question 2). Dans des cas plus généraux, $-p_i$ — celui issu des conditions nécessaires d'optimalité du premier ordre — peut encore être relié à une certaine notion de “gradient généralisé” de φ_i . Il existe de toute façon une difficulté pratique à la vérification de l'égalité mutuelle des multiplicateurs fournis par la résolution des sous-problèmes lorsque ces multiplicateurs ne sont pas uniques : l'égalité n'aura lieu que pour un certain jeu (un certain choix) parmi les multiplicateurs optimaux des sous-problèmes. Voir à ce sujet l'Exercice 2.32.

Cette difficulté est essentiellement duale de celle rencontrée dans la méthode par les prix lorsque $\hat{U}(p)$ n'est pas réduit à un point unique (voir commentaire 2.4-c). Plus généralement, il s'agit de la

¹¹ noter que $\hat{p}_1 = \hat{p}_2$

¹² mais on pourrait éventuellement changer cet état de choses en modifiant la fonction coût à l'extérieur d'une boule centrée à l'origine par exemple, tout en conservant la difficulté illustrée ici

¹³ On peut garantir l'unicité des multiplicateurs optimaux par exemple sous les hypothèses que J'_i est Lipschitzien, que $\Theta'_i(u_i)$ — comme application linéaire de \mathbb{R}^{n_i} dans \mathbb{R}^m — est surjective pour tout u_i et que $U_i^{\text{ad}} = \mathbb{R}^{n_i}$.

difficulté suivante : lorsqu'une fonction n'est pas différentiable, l'utilisation d'algorithmes de sous-gradient ou de "gradient généralisé" adaptés fait éventuellement converger l'argument (ici θ) vers une valeur optimale, mais rarement les sous-gradients (ici $-p$) vers les valeurs "optimales" (c'est-à-dire vers 0 dans le cas sans contrainte, vers un sous-gradient orthogonal à la contrainte dans le cas avec contrainte — qui est ici la contrainte d'appartenance à la variété affine des "allocations"). L'exception est constituée par les méthodes de type "faisceaux" ("bundle" en Anglais) ou certaines méthodes de "plans tangents".

- (c) Les solutions des sous-problèmes correspondant à n'importe quelle allocation vérifie la contrainte globale (2.1b). Ceci signifie que les itérées successives construites par tout algorithme de coordination seront des solutions *admissibles*.
- (d) Cet avantage apparent comporte cependant des contreparties sérieuses. Notons d'abord que pour certaines allocations, certains sous-problèmes peuvent ne pas avoir de solution : c'est ce qui arrive si $\theta_i \notin \Theta_i(U_i^{\text{ad}})$ (interférence des contraintes (2.7b) avec les contraintes $u_i \in U_i^{\text{ad}}$)¹⁴. Du point de vue algorithmique, il sera difficile de caractériser explicitement le domaine de définition de chaque $\varphi_i(\{\theta_i \mid \varphi_i(\theta_i) < +\infty\})$ et donc d'y maintenir le θ_i correspondant.
- (e) Un obstacle plus sérieux encore surgit s'il y a "trop" de contraintes (2.1b) par rapport au nombre de variables de décisions. Dans le cas de contraintes égalité, il est logique de supposer $n \geq m$. Cependant, cette condition n'exclut pas que $n_i < m$ pour certains i , de sorte que certains sous-problèmes ne pourront être résolus que si le θ_i correspondant est constamment maintenu dans l'image $\Theta_i(U_i^{\text{ad}})$ qui est alors un sous-ensemble d'intérieur vide de \mathbb{R}^m .

Pour pallier ces inconvénients, certains auteurs ont proposé de remplacer les contraintes (2.7b) par une pénalisation du critère (2.7a), mais il s'agit là de méthodes approchées.

Algorithmes. L'algorithme le plus simple et qui, de plus, se prête bien à une compréhension intuitive, est l'algorithme du gradient (ou sous-gradient) projeté pour résoudre (2.9). Cet algorithme peut être justifié en particulier dans le cas convexe (modulo le fait de maintenir les itérés θ_i^k successives dans le domaine des fonctions φ_i). Ayant obtenu les multiplicateurs p_i^k (ceux de la théorie locale de la dualité — conditions nécessaires — qui coïncident avec ceux de la théorie globale si les Lagrangiens ont un point selle), en utilisant la formule de projection du Lemme 2.11, on passe à la nouvelle allocation θ^{k+1} par la formule

$$\theta_i^{k+1} = \theta_i^k + \varepsilon^k \left(p_i^k - \frac{1}{N} \sum_{j=1}^N p_j^k \right), \quad i = 1, \dots, N, \quad (2.13)$$

qui suppose que θ_i^k vérifie déjà (2.8). La suite $\{\varepsilon^k\}$ vérifie¹⁵

$$\varepsilon^k > 0, \quad \sum_{k \in \mathbb{N}} \varepsilon^k = +\infty, \quad \sum_{k \in \mathbb{N}} (\varepsilon^k)^2 < +\infty.$$

Le signe positif de ε^k (alors qu'il s'agit de minimiser dans (2.9)) s'explique par le fait que c'est $-p_i$ qui est un sous-gradient de φ_i en θ_i .

Intuitivement, la coordination redistribue la ressource proportionnellement à l'écart constaté entre chaque utilité marginale et la moyenne arithmétique de celles-ci. L'algorithme s'arrête (théoriquement) lorsque toutes les utilités marginales sont devenues égales. En complément au commentaire 2.17-c ci-dessus, il convient donc de dire que la méthode produit des solutions primales toujours admissibles mais des solutions duales non admissibles — ou non consistantes : c'est la situation duale de la méthode par les prix où le vecteur des prix était le même pour tous les sous-problèmes.

¹⁴Voir encore à ce sujet l'Exercice 2.32

¹⁵Voir l'Annexe, §2.3, sur cette technique de pas en séries divergentes en fin de chapitre.

Commentaires 2.18. Pour le cas des contraintes inégalité, le plus simple, algorithmiquement, est de transformer (2.7b) en inégalité et de maintenir (2.8) sous forme d'égalité. Alors, la formule (2.13) est encore valable. Pour l'énoncé du Lemme 2.8, l'équation (2.11) doit être remplacée par

$$\theta_i^\sharp = \Theta_i(u_i^\sharp) - \frac{1}{N} \left(\sum_{j=1}^N \Theta_j(u_j^\sharp) - \vartheta \right), \quad i = 1, \dots, N. \quad (2.14)$$

Exercice 2.19. Redémontrer le Lemme 2.8 dans ce cas.

Des stratégies plus sophistiquées (du type approximations successives “internes” ou “externes” des φ_i) ont été proposées dans la littérature pour résoudre le problème (2.9) *dans le cas convexe*. Par exemple, on peut représenter une forme approchée des fonctions φ_i comme l'enveloppe supérieure d'un certain nombre d'hyperplans d'appui : à l'étape k , au point θ_i^k , on a $\varphi_i(\theta_i^k) = J_i(u_i^{k+1})$; le sous-gradient correspondant étant $-p_i^{k+1}$, un hyperplan d'appui est représenté par la fonction affine

$$f_i^k(\theta_i) = J_i(u_i^{k+1}) + \langle p_i^{k+1}, \theta_i^k - \theta_i \rangle.$$

La remise à jour de θ_i^k consiste à rechercher le minimum, pour les allocations admissibles, de la somme pour $i = 1, \dots, N$, des enveloppes supérieures, pour $l = 0, \dots, k$, des f_i^l , ce qui revient à résoudre le programme linéaire

$$\begin{aligned} & \min_{\theta, y \in \mathbb{R}^N} \sum_{i=1}^N y_i \\ \text{sous} \\ & \sum_{i=1}^N \theta_i = \vartheta ; \quad y_i \geq J_i(u_i^{l+1}) + \langle p_i^{l+1}, \theta_i^l - \theta_i \rangle, \\ & \quad i = 1, \dots, N, \quad l = 0, \dots, k. \end{aligned}$$

2.1.4 Décomposition “par prédiction”

Dans le cadre du premier modèle (problème (2.1)), cette méthode est moins bien connue que les deux précédentes. Elle a été introduite par Mesarovic et associés [22] dans un cadre plus proche du deuxième modèle ci-dessous sous le nom de “interaction prediction principle”, et nous montrerons qu'il s'agit bien de la même méthode que celle exposée ci-après. C'est pourquoi nous avons conservé l'appellation originale qui n'est pourtant pas très heureuse.

Comme pour les méthodes précédentes, nous en donnons d'abord une présentation intuitive. Celle-ci nécessite l'introduction d'une nouvelle décomposition, ce qui est un fait nouveau par rapport à ce qui précède : cet élément nouveau caractérise cette méthode. En effet, à la décomposition de l'espace \mathbb{R}^n des décisions, nous allons associer une décomposition de l'espace \mathbb{R}^m des contraintes. Afin de conserver initialement une symétrie dans le rôle joué par chaque sous-problème, nous supposons que \mathbb{R}^m est lui aussi décomposé en un produit de exactement N sous-espaces \mathbb{R}^{m_i} avec $\sum_{i=1}^N m_i = m$. Ceci n'est évidemment pas toujours possible (par exemple si $m < N$). Nous écarterons donc ces cas pour l'instant, mais nous lèverons cette restriction plus tard. Nous noterons Θ_{ji} l'application de \mathbb{R}^{n_i} dans \mathbb{R}^{m_j} obtenue par composition de l'application Θ_i de \mathbb{R}^{n_i} dans \mathbb{R}^m et de la projection de \mathbb{R}^m dans \mathbb{R}^{m_j} parallèlement au produit des autres sous-espaces composants.

Sur le plan de l'interprétation, cette décomposition de l'espace des contraintes revient à “allouer” ces contraintes aux sous-problèmes, autrement dit à rendre responsable chaque unité de production de

certain types de fabrications (ou de la gestion d'un certain nombre de ressources) représentées par les coordonnées correspondantes du vecteur ϑ . À ce propos, ϑ_i désignera désormais la projection de ϑ sur \mathbb{R}^{m_i} , ce qui ne devra pas être confondu avec la notation θ_i introduite au §2.1.3 et qui désignait un élément de \mathbb{R}^m et non de \mathbb{R}^{m_i} .

Chaque unité i recevra donc une contrainte à valeurs dans \mathbb{R}^{m_i} . Elle devra assurer le complément à ϑ_i des productions (ou la consommation complémentaire des ressources) par rapport au total, noté w_i , de ce qui aura été “prédit” (ou annoncé) pour les autres unités. Comme l'activité de l'unité i aura aussi pour effet de fabriquer des produits de types ϑ_j pour $j \neq i$ (ou de consommer des ressources de ces types) en quantités $\Theta_{ji}(u_i)$ (lorsque le niveau d'activité est fixé par la décision u_i), un certain système de prix μ_j (également annoncé à l'avance) permettra d'incorporer ces quantités au bilan de l'unité i . Finalement, pour des prédictions w_i du niveau de production ou de consommation des autres unités j quant aux ressources de type i (respectivement, des prédictions μ_j des prix marginaux des produits ou des ressources de type j) avec $j = 1, \dots, N$ et $j \neq i$, le sous-problème i se formule

$$\min_{u_i \in U_i^{\text{ad}}} \left(J_i(u_i) + \underbrace{\sum_{j \neq i} \langle \mu_j, \Theta_{ji}(u_i) \rangle}_{\text{essentiel}} \right) \quad (2.15a)$$

sous

$$\Theta_{ii}(u_i) = \vartheta_i - w_i. \quad (2.15b)$$

La présence des termes marqués *essentiel* est en effet essentielle. L'exemple du §1.2.4 a montré que sans ces termes, il n'est pas possible de construire une méthode de coordination valide. Pour les justifier pleinement, on doit se poser à nouveau l'équivalent des questions 2.1, c'est-à-dire

1. celle de la caractérisation de l'optimalité, vis-à-vis du problème global (2.1), des solutions obtenues en résolvant les sous-problèmes (2.15) en utilisant uniquement les éléments d'information issus de la résolution de ces sous-problèmes ;
2. celle de l'existence de paramètres de coordination (en l'occurrence ici w et μ) conduisant à cette optimalité ;
3. enfin celle des algorithmes permettant de trouver ces “bons” paramètres.

Les réponses aux deux premières questions sont apportées par les lemmes suivants. Dans le lemme qui suit, on suppose à nouveau que $U^{\text{ad}} = \mathbb{R}^n$ et que les fonctions sont différentiables.

Lemme 2.20. *Si les conditions nécessaires du premier ordre¹⁶ sont vérifiées dans les sous-problèmes (2.15) pour $i = 1, \dots, N$, par des couples (\hat{u}_i, \hat{p}_i) , et si*

$$w_i = \sum_{j \neq i} \Theta_{ij}(\hat{u}_j), \quad (2.16a)$$

$$\mu_i = \hat{p}_i, \quad (2.16b)$$

alors (\hat{u}, \hat{p}) vérifie aussi ces conditions nécessaires pour le problème (2.1). Réciproquement, si (u^\sharp, p^\sharp) vérifie les conditions nécessaires du premier ordre pour le problème global, alors en prenant les “prédictions” w^\sharp et μ^\sharp , obtenues à partir de (u^\sharp, p^\sharp) et des équations (2.16), pour définir les sous-problèmes (2.15), les couples (u_i^\sharp, p_i^\sharp) vérifient aussi ces conditions nécessaires pour ces sous-problèmes.

Exercice 2.21. Donner une démonstration de ce lemme.

¹⁶analogues à celles évoquées au Lemme 2.12

La première partie de cet énoncé fournit le test d'optimalité évoqué dans la première question ci-dessus. La réciproque fournit une réponse à l'existence de bons paramètres de coordination. La démonstration du lemme (laissée en exercice) montre l'utilité des termes marqués essentiel dans (2.15a).

Au lieu de s'appuyer sur la théorie des conditions nécessaires d'optimalité du premier ordre, nous nous penchons maintenant vers celle des conditions suffisantes. Dans le lemme qui suit, on revient à U^{ad} comme sous-ensemble convexe fermé de \mathbb{R}^n et la différentiabilité des fonctions n'est plus requise.

Lemme 2.22. *Si pour $i = 1, \dots, N$, (\hat{u}_i, \hat{p}_i) est un point selle du Lagrangien du problème (2.15) sur $U_i^{\text{ad}} \times \mathbb{R}^{m_i}$ et si les équations (2.16) sont satisfaites, alors (\hat{u}, \hat{p}) est un point selle du Lagrangien du problème (2.1) sur $U^{\text{ad}} \times \mathbb{R}^m$ et donc \hat{u} est solution de ce problème. Réciproquement, si (u^\sharp, p^\sharp) est un point selle du Lagrangien du problème global, il en est de même de (u_i^\sharp, p_i^\sharp) pour ce qui concerne le problème (2.15) à condition de définir (w, μ) à partir de (u^\sharp, p^\sharp) et des équations (2.16).*

Exercice 2.23. Donner une démonstration de ce lemme.

Comme dans la méthode par allocation de ressources, il n'est pas possible d'énoncer des résultats équivalents pour les conditions nécessaires ou suffisantes du second ordre (théorie locale de la dualité), ni dans un sens ni dans l'autre, comme le montrent les deux exemples qui suivent. Ceci tient essentiellement au fait que le produit cartésien des sous-espaces des \mathbb{R}^{m_i} tangents aux contraintes (2.15b) est a priori différent du sous-espace de \mathbb{R}^m tangent aux contraintes (2.1b) : il n'y a ici aucune relation d'inclusion ni dans un sens, ni dans l'autre.

Exemple et Exercice 2.24. On reprend l'Exemple 2.16 en rajoutant une contrainte

$$\min (-x_1^2 - y_1^2 - 4x_1y_1 - x_2^2 - y_2^2 - 4x_2y_2) \quad (2.17a)$$

sous

$$x_1 + y_1 - x_2 - y_2 = 0, \quad (2.17b)$$

$$x_1 + y_1 - x_2 - 2y_2 = 0. \quad (2.17c)$$

On rappelle la décomposition de u en $u_1 = (x_1, y_1)$ et $u_2 = (x_2, y_2)$. La première contrainte est "allouée" au sous-problème 1 et la seconde au sous-problème 2. Par application de (2.15), on obtient ici :

sous-problème 1

$$\min_{x_1, y_1} (-x_1^2 - y_1^2 - 4x_1y_1 + \mu_2(x_1 + y_1)) \quad (2.17d)$$

sous

$$x_1 + y_1 = -w_1, \quad (2.17e)$$

sous-problème 2

$$\min_{x_2, y_2} (-x_2^2 - y_2^2 - 4x_2y_2 + \mu_1(-x_2 - y_2)) \quad (2.17f)$$

sous

$$-x_2 - 2y_2 = -w_2. \quad (2.17g)$$

En prenant $(w, \mu) = (0, 0) \in \mathbb{R}^2 \times \mathbb{R}^2$, on trouve que $(\hat{u}_1, \hat{p}_1) = (\hat{u}_2, \hat{p}_2) = (0, 0) \in \mathbb{R}^2 \times \mathbb{R}$ vérifient les conditions nécessaires du premier ordre et suffisantes du second ordre pour les sous-problèmes 1 et 2 respectivement.

Pourtant le point $\hat{u} = 0 \in \mathbb{R}^4$ n'est pas un minimum local du problème global (là encore il n'y en a pas) et il ne vérifie pas, avec $\hat{p} = 0 \in \mathbb{R}^2$, des conditions du second ordre pour le problème global.

Exemple et Exercice 2.25. On considère un problème de même dimension que ci-dessus :

$$\min (x_1^2 + y_1^2 + 3x_1y_1 + x_2^2 + y_2^2 - 3x_2y_2) \quad (2.18a)$$

sous

$$x_1 + y_1 + x_2 + y_2 = 0, \quad (2.18b)$$

$$x_1 - y_1 + x_2 - y_2 = 0. \quad (2.18c)$$

et la même décomposition en sous-problèmes que dans l'exemple précédent. On vérifie que $u^\# = 0 \in \mathbb{R}^4$ est le seul minimum du problème (le critère restreint au sous-espace de dimension 2 défini par les contraintes est strictement convexe). De plus la paire $(u^\#, p^\#)$ (avec $p^\# = 0 \in \mathbb{R}^2$) vérifie les conditions nécessaires du premier ordre, et suffisantes du second ordre.

La décomposition adoptée conduit aux sous-problèmes suivants en application de (2.15) :

sous-problème 1

$$\min_{x_1, y_1} (x_1^2 + y_1^2 + 3x_1y_1 + \mu_2(x_1 - y_1)) \quad (2.18d)$$

sous

$$x_1 + y_1 = -w_1, \quad (2.18e)$$

sous-problème 2

$$\min_{x_2, y_2} (x_2^2 + y_2^2 - 3x_2y_2 + \mu_1(x_2 + y_2)) \quad (2.18f)$$

sous

$$x_2 - y_2 = -w_2. \quad (2.18g)$$

En prenant $(w, \mu) = (0, 0)$, on s'aperçoit que les conditions du second-ordre ne sont vérifiées en $(0, 0) \in \mathbb{R}^2 \times \mathbb{R}$ dans aucun des deux sous-problèmes. Les points $u_i = 0 \in \mathbb{R}^2$ correspondent même à des maxima des sous-problèmes ci-dessus !

Commentaires 2.26.

- (a) La situation de l'Exemple 2.25 est évidemment beaucoup plus gênante que celle de l'Exemple 2.24 car cela signifie que la solution du problème global ne peut pas être trouvée dans ce cas par la formulation (2.15) des sous-problèmes si on ne se contente pas de satisfaire les conditions du premier ordre et si on insiste pour vérifier aussi celles du second ordre. En fait, les sous-problèmes (2.15) sont un outil pour résoudre ces conditions nécessaires du premier ordre. Le passage de la stationnarité à la minimalité nécessite implicitement de “convexifier” les sous-problèmes, mais il faut alors sortir du cadre élémentaire dans lequel on se situe ici. Dans le cas où des conditions suffisantes, issues de la théorie globale de la dualité, sont vérifiées (c'est en particulier le cas pour les problèmes convexes), les choses sont beaucoup plus simples.
- (b) Le cas des contraintes inégalité n'offre formellement aucune difficulté d'adaptation. L'étude de la convergence des algorithmes décrits ci-dessous est par contre plus difficile dans ce cas (elle n'a pu être réalisée à ce jour).
- (c) Les mêmes difficultés que précédemment existent évidemment en ce qui concerne la vérification du “test” (2.16) en cas de non unicité des solutions primales et duales des sous-systèmes.
- (d) La méthode est “non admissible” du point de vue primal en ce sens que, comme dans la coordination par les prix, les solutions primales des sous-systèmes ne vérifient en général la contrainte globale (2.1b) que pour une valeur particulière des paramètres de coordination. Par contre, par

type de production (ou de ressource), il n'existe qu'un seul prix (à l'inverse de la coordination par les quantités) ; c'est-à-dire que la méthode est admissible (ou consistante) du point de vue dual.

- (e) Dans le premier algorithme exposé ci-dessous, pour chaque ressource (ou production), le prix est élaboré par le sous-système ayant reçu la responsabilité de cette ressource : ici intervient la façon dont les contraintes (2.1b) ont été distribuées parmi les sous-problèmes. En écho au commentaire 2.17-e, il devient possible (et souhaitable) de s'arranger pour que $m_i \leq n_i$ pour $i = 1, \dots, N$. La situation est donc beaucoup moins contraignante qu'avec la coordination par les quantités.
- (f) Par contre, contrairement au cas de la coordination par les prix, une partie des contraintes (2.1b) se retrouve dans chaque sous-problème. Il existe donc un risque, si $\Theta_{ii}(U_i^{\text{ad}})$ n'est pas égal à \mathbb{R}^{m_i} , que certains sous-problèmes n'aient pas de solution admissible pour certaines "prédictions" w_i déterminant le second membre de (2.15b). Pratiquement, dans certaines applications, il peut être utile de remplacer les contraintes décomposées $u_i \in U_i^{\text{ad}}$ par une "pénalisation exacte" du critère afin de diminuer ce risque (on étend ainsi la minimisation à \mathbb{R}^{n_i} tout entier). Cependant, ceci conduira à des fonctions coût non différentiables (mais encore additives). La théorie correspondante n'est malheureusement pas disponible.
- (g) Revenons sur la restriction que l'espace \mathbb{R}^m des contraintes soit décomposé en N espaces composants exactement. S'il est décomposé en moins de N composants (par exemple si $m = 1$ et si $N > 1$), certains sous-problèmes ne recevront pas de contraintes du type (2.15b) (et n'auront donc pas à élaborer de multiplicateurs p_i). Il n'y a là, cependant, aucune difficulté particulière pour adapter la méthode qui a été décrite ci-dessus dans une situation plus symétrique.

Exercice 2.27. Se pencher sur la situation évoquée au commentaire 2.26-g ci-dessus et écrire en détail les sous-problèmes à définir à la place de (2.15) en considérant par exemple le cas $N = 2, m = 1$.

Algorithmes.

Algorithme de point fixe. L'algorithme le plus immédiatement suggéré par la méthode de décomposition-coordination elle-même consiste à partir d'une "prédiction" initiale (w^0, μ^0) , à résoudre les sous-problèmes (2.15), ce qui conduit, pour $i = 1, \dots, N$, à des solutions u_i^1 et p_i^1 (comme multiplicateur associé à (2.15b)), puis à repartir pour une nouvelle itération avec la nouvelle prédiction (w^1, μ^1) définie par les relations (2.16) où l'on a placé (u^1, p^1) au second membre, et ainsi de suite... Autrement dit, la remise à jour des paramètres à l'étape $k + 1$ de coordination se fait par les formules

$$w_i^{k+1} = \sum_{j \neq i} \Theta_{ij}(u_j^{k+1}), \quad (2.19a)$$

$$\mu_i^{k+1} = p_i^{k+1}. \quad (2.19b)$$

Contrairement à la situation rencontrée avec les deux méthodes précédentes, il n'y a donc pas dans cet algorithme, à proprement parler, un niveau hiérarchique supérieur chargé de la coordination. Celle-ci s'opère par échange direct d'informations entre les sous-problèmes.

Pour améliorer la convergence, on peut envisager une *sous-* ou une *sur-relaxation* sur le couple (w^k, μ^k) , c'est-à-dire qu'au lieu des formules (2.19), on utilise plutôt :

$$w_i^{k+1} = (1 - \varepsilon^k)w_i^k + \varepsilon^k \sum_{j \neq i} \Theta_{ij}(u_j^{k+1}), \quad (2.20a)$$

$$\mu_i^{k+1} = (1 - \rho^k)\mu_i^k + \rho^k p_i^{k+1}, \quad (2.20b)$$

où ε^k et ρ^k sont des nombres positifs. On parle de *sur-relaxation* si ces nombres sont plus grands que 1 et de *sous-relaxation* si ils sont plus petits que 1 (le cas où ils sont égaux à 1 revient à l'algorithme (2.19)).

Comme il est toujours difficile de garantir l'unicité des multiplicateurs optimaux notamment, et comme toute non unicité est, sur le plan algorithmique, généralement une source de difficultés, il n'est pas étonnant que la convergence de ces algorithmes n'ait pu être établie que dans le seul cas où l'on peut pratiquement assurer cette unicité, à savoir le cas de contraintes égalité affines, de fonctions coût quadratiques et en l'absence d'autres contraintes de type $u_i \in U_i^{\text{ad}}$.

Les conditions de convergence obtenues font apparaître des conditions structurelles, ou géométriques, sur le découpage de l'application $\Theta : \mathbb{R}^n \rightarrow \mathbb{R}^m$ en applications $\Theta_{ij} : \mathbb{R}^{n_j} \rightarrow \mathbb{R}^{m_i}$ qui mettent directement en cause la façon dont les contraintes (2.1b) ont été réparties entre les sous-problèmes (voir (2.15b)). Une règle pratique et intuitive est de procéder au découpage et à l'allocation des contraintes qui rendent la partie bloc-diagonale¹⁷ de $\Theta'(u)$ (dérivée de Θ en u) la plus "dominante" possible (autrement dit, d'affecter à un sous-problème i les contraintes où l'influence de u_i est plus forte que celle des autres u_j autour d'un point nominal u , en espérant que cette propriété est uniforme par rapport au point nominal u). Cependant, ce cas favorable n'est pas le seul possible et les conditions structurelles recouvrent d'autres cas intéressants qui seront examinés plus loin dans le cours.

L'algorithme ci-dessus, que nous appellerons "algorithme de point fixe" (éventuellement sur- ou sous-relaxé) a été le premier historiquement introduit, et il peut paraître à première vue assez naturel. Cependant, une analyse mathématique plus fine montre qu'il n'est pas aussi naturel qu'il y paraît et qu'un algorithme alternatif finalement plus conforme aux principes variationnels habituels est possible.

Algorithme de type Arrow-Hurwicz. Le problème (2.1) est équivalent à la formulation suivante :

$$\begin{aligned} & \min_{u_i \in U_i^{\text{ad}}, w_i \in \mathbb{R}^{m_i}} \sum_{i=1}^N J_i(u_i) \\ & \text{sous} \\ & \Theta_{ii}(u_i) = \vartheta_i - w_i, \quad i = 1, \dots, N, \\ & w_i = \sum_{j \neq i} \Theta_{ij}(u_j), \quad i = 1, \dots, N. \end{aligned} \tag{2.21}$$

Le Lagrangien de ce problème est

$$L(u, w; p, \mu) = \sum_{i=1}^N \left(J_i(u_i) + \langle p_i, \Theta_{ii}(u_i) + w_i - \vartheta_i \rangle + \sum_{j \neq i} \langle \mu_j, \Theta_{ji}(u_i) \rangle - \langle \mu_i, w_i \rangle \right).$$

La décomposition par prédiction peut être vue comme le fait de fixer les paramètres de coordination (w, μ) , puis de résoudre les conditions nécessaires du premier ordre (ou les inégalités du point selle si celui-ci existe) par rapport au couple (u, p) , ce qui conduit aux sous-problèmes (2.15), et finalement de remettre à jour (w, μ) . On a

$$\left(\frac{\partial L}{\partial w_i} \right)^\top = p_i - \mu_i, \quad \left(\frac{\partial L}{\partial \mu_i} \right)^\top = \sum_{j \neq i} \Theta_{ij}(u_j) - w_i.$$

Dans l'algorithme de point fixe (2.19), les conditions de stationnarité $\partial L / \partial w_i = 0$ ont été utilisées pour remettre à jour μ_i , tandis que les conditions $\partial L / \partial \mu_i = 0$ ont permis de remettre à jour w_i .

Supposons que le Lagrangien L a un point selle en $(u, w; p, \mu)$ sur $(U^{\text{ad}} \times \mathbb{R}^m) \times (\mathbb{R}^m \times \mathbb{R}^m)$, et plus généralement que les minimisations par rapport à u et w , et les maximisations par rapport

¹⁷opérateur où on ne garde que les blocs indexés ii , $i = 1, \dots, N$,

à p et μ peuvent être accomplies dans un ordre arbitraire (ce qui suppose pratiquement que L est convexe-concave) ; alors, une fois que les sous-problèmes ont résolu $\min_u \max_p L$ à (w^k, μ^k) fixés, fournissant ainsi (u^{k+1}, p^{k+1}) , il reste à résoudre le $\min_w \max_\mu L$, ce qui peut être accompli par une sorte d'algorithme d'Arrow-Hurwicz parallèle, à savoir

$$\begin{aligned} w_i^{k+1} &= w_i^k - \varepsilon^k \left(\frac{\partial L}{\partial w_i}(u^{k+1}, w^k, p^{k+1}, \mu^k) \right)^\top \\ &= w_i^k - \varepsilon^k (p_i^{k+1} - \mu_i^k), \end{aligned} \quad (2.22a)$$

$$\begin{aligned} \mu_i^{k+1} &= \mu_i^k + \rho^k \left(\frac{\partial L}{\partial \mu_i}(u^{k+1}, w^k, p^{k+1}, \mu^k) \right)^\top \\ &= \mu_i^k + \rho^k \left(\sum_{j \neq i} \Theta_{ij}(u_j^{k+1}) - w_i^k \right). \end{aligned} \quad (2.22b)$$

Ces équations sont à comparer aux formules (2.20).

Remarque 2.28. L'équation (2.22b) s'écrit encore

$$\mu_i^{k+1} = \mu_i^k + \rho^k \left(\sum_{j=1}^N \Theta_{ij}(u_j^{k+1}) - \vartheta_i \right),$$

car, d'après (2.15b), $w_i^k = \vartheta_i - \Theta_{ii}(u_i^{k+1})$. Cette équation est la même que l'équation (2.6) de la méthode par les prix.

In [11], une preuve de convergence est donnée essentiellement sous les hypothèses que J est fortement convexe, que Θ est linéaire, et que tous les Θ_{ii} sont surjectives : la convergence est obtenue pour ε^k et ρ^k assez petits (on y reviendra dans la suite du cours) sans que des conditions structurelles comme celles évoquées plus haut pour l'algorithme de point fixe soient requises.

Dans [11], un exemple numérique montre qu'il existe de fait une certaine complémentarité entre les deux types d'algorithmes : l'algorithme de point fixe marche mieux que l'algorithme d'Arrow-Hurwicz pour des couplages faibles entre sous-problèmes (Θ_{ij} , $j \neq i$, est petit comparé à Θ_{ii}), alors que ce dernier devient meilleur pour des interactions fortes (l'autre ne marchant plus éventuellement).

Enfin, dans [11], des versions *séquentielle* et *implicite* de l'algorithme (2.22a)–(2.22b) se sont avérées meilleures que la version parallèle. Une version séquentielle consiste à remettre à jour w d'abord par (2.22a), puis à utiliser la nouvelle valeur w^{k+1} au second membre de (2.22b) à la place de w^k . L'autre version séquentielle procède dans l'ordre inverse. Enfin, la version implicite, la meilleure de toutes d'après les expériences numériques, consiste à mettre μ^{k+1} à la place de μ^k au second membre de (2.22a), et de même w^{k+1} à la place de w^k au second membre de (2.22b), ce qui conduit à un système linéaire implicite en (w^{k+1}, μ^{k+1}) . Sa résolution conduit aux formules explicites suivantes :

$$w_i^{k+1} = w_i^k - \frac{\varepsilon^k}{1 + \varepsilon^k \rho^k} (p_i^{k+1} - \mu_i^k) + \frac{\varepsilon^k \rho^k}{1 + \varepsilon^k \rho^k} \left(\sum_{j \neq i} \Theta_{ij}(u_j^{k+1}) - w_i^k \right), \quad (2.22c)$$

$$\mu_i^{k+1} = \mu_i^k + \frac{\rho^k}{1 + \varepsilon^k \rho^k} \left(\sum_{j \neq i} \Theta_{ij}(u_j^{k+1}) - w_i^k \right) + \frac{\varepsilon^k \rho^k}{1 + \varepsilon^k \rho^k} (p_i^{k+1} - \mu_i^k). \quad (2.22d)$$

2.1.5 Une classification politico-économique des méthodes de décomposition-coordination

Nous nous livrons ici à une interprétation politico-économique des trois méthodes de décomposition-coordination examinées jusqu'à maintenant, en les envisageant sous leur aspect de décentralisation des décisions parmi N acteurs supposés en accord sur (ou contraints à) la recherche d'un optimum collectif.

- On peut interpréter la coordination par les prix comme la situation de l'*économie libérale* basée sur des incitations en termes de prix, sur l'ajustement de ces prix par la loi de l'offre et de la demande, etc.
- À l'opposé (de façon duale), la coordination par les quantités peut être vue comme la situation d'une *économie planifiée* dans laquelle des objectifs en termes de production (ou de consommation) sont assignés par le niveau hiérarchique supérieur.
- Enfin, la coordination par prédiction peut être vue comme un modèle de *système autogestionnaire* dans lequel il n'y a plus de niveau hiérarchique supérieur, et où chaque acteur est responsable d'un secteur particulier d'activité (traduit par l'allocation des contraintes requise par cette approche).

2.2 Deuxième modèle

2.2.1 Introduction du second modèle

On reprend maintenant l'examen des trois méthodes de décomposition-coordination précédentes, mais en les appliquant à un autre type de problème d'optimisation. Ce type de problème est plus proche de la situation où l'on cherche à optimiser un système composé de sous-systèmes interconnectés. L'interaction entre sous-problèmes viendra donc ici de l'existence de variables d'interconnexion plutôt que du partage de ressources communes comme pour le premier modèle. Sur le plan mathématique, il n'y a cependant aucune différence fondamentale entre les deux situations et l'on va s'empresse de transposer le second modèle dans le cadre du premier modèle afin de retrouver les trois méthodes déjà décrites. L'intérêt d'examiner le second modèle est qu'au niveau de l'interprétation et pour certaines considérations pratiques, cet examen révèle des points qui n'ont pu être mis en évidence avec le premier modèle.

Pour motiver le second modèle, considérons le cas simple d'un problème de commande optimale linéaire-quadratique :

$$\dot{x}(t) = Fx(t) + Gu(t) , \quad x(t_0) = \xi , \quad (2.23a)$$

$$\min \int_{t_0}^{t_f} (x^\top(t) Q x(t) + u^\top(t) R u(t)) dt , \quad (2.23b)$$

où $x(t) \in \mathbb{R}^m$ est le vecteur d'état, $u(t) \in \mathbb{R}^n$ est le vecteur de commande, F, G, Q, R sont des matrices de dimensions appropriées (Q et R symétriques). Quant à $\dot{x}(t)$, c'est une notation traditionnelle pour $dx(t)/dt$.

Supposons que x est décomposé en N composantes $x^\top = (x_1^\top, \dots, x_N^\top)$, $x_i \in \mathbb{R}^{m_i}$ et de même $u^\top = (u_1^\top, \dots, u_N^\top)$ avec $u_i \in \mathbb{R}^{n_i}$. Considérant la décomposition par blocs correspondante des matrices, on peut écrire

$$\dot{x}_i = F_{ii}x_i + G_{ii}u_i + v_i , \quad x_i(t_0) = \xi_i , \quad (2.24)$$

avec les variables d'interaction v_i définie par

$$v_i = \sum_{j \neq i} (F_{ij}x_j + G_{ij}u_j) , \quad i = 1, \dots, N . \quad (2.25)$$

Pour conserver l'hypothèse d'un critère (2.23b) additif, on supposera Q et R bloc-diagonales ($Q_{ij} = 0$ et $R_{ij} = 0$ pour $i \neq j$).

À condition initiale ξ_i fixée, l'équation différentielle (2.24) définit une application

$$x_i = P_i(u_i, v_i) , \quad (2.26)$$

Tableau 2.1: Correspondances entre premier et second modèles

	1er modèle	2ème modèle
	u_i	(u_i, v_i)
	J_i	J_i
	ϑ	0
	$\Theta_{ii}(u_i)$	$-v_i$
$i \neq j \rightarrow$	$\Theta_{ij}(u_j)$	$H_{ij}(u_j, v_j)$

où u_i, v_i, x_i sont des fonctions du temps t sur $[t_0, t_f]$. En utilisant (2.26) dans le second membre de (2.25), on peut réécrire (2.25) :

$$v_i = \sum_{j \neq i} H_{ij}(u_j, v_j), \quad i = 1, \dots, N, \quad (2.27a)$$

avec une définition évidente de H_{ij} . De même, en utilisant (2.26) dans (2.23b), cette intégrale peut s'écrire comme une fonctionnelle de (u, v) ayant la forme

$$J(u, v) = \sum_{i=1}^N J_i(u_i, v_i). \quad (2.27b)$$

On se pose donc le problème de la minimisation de (2.27b) sous les contraintes (2.27a) (on supposera que les contraintes sur u sont de la forme $u_i \in U_i^{\text{ad}}$ et qu'il n'y a pas de contrainte d'état). Ceci constitue notre second modèle. Il est facile de voir que celui-ci peut-être identifié au premier modèle (2.1) en établissant les correspondances décrites dans le Tableau 2.1. Avec l'aide de ces correspondances, examinons alors la "traduction" des méthodes de décomposition-coordination sur ce second modèle.

2.2.2 Décomposition par les prix

Le sous-problème i se formule

$$\min_{u_i \in U_i^{\text{ad}}, v_i \in \mathbb{R}^{m_i}} \left(J_i(u_i, v_i) + \sum_{j \neq i} \langle p_j, H_{ji}(u_i, v_i) \rangle - \langle p_i, v_i \rangle \right), \quad (2.28)$$

et l'algorithme de coordination s'écrit

$$p_i^{k+1} = p_i^k + \rho^k \left(\sum_{j \neq i} H_{ij}(u_j^k, v_j^k) - v_i^k \right), \quad (2.29)$$

où (u_i^k, v_i^k) est une solution du problème (2.28) correspondant au système de prix p^k . Les interprétations sont les mêmes qu'au §2.1.2.

Le fait nouveau que révèle l'application de cette méthode par les prix au modèle est le suivant. Considérons à nouveau le problème de commande optimale (2.23) et supposons que la matrice R est définie positive, et même "suffisamment positive" pour que ce problème soit strictement convexe en u (x étant considéré comme fonction de u via (2.23a)) même si Q n'est pas définie positive, voire si elle est non définie. Comme v_i est une fonction affine de x (voir (2.25)), et compte tenu de la définition de $J_i(u_i, v_i)$ dans ce cas, il est clair que J_i risque de ne pas être strictement convexe en v_i , voire d'être non convexe. Dans ces conditions, il se peut qu'il n'y ait pas de solution unique au problème (2.28) ⁽¹⁸⁾ et,

¹⁸voire pas de solution du tout, par exemple si $\inf_{v_i} J_i(u_i, v_i) = -\infty$

corrélativement, que le Lagrangien du problème (2.27) ait un point selle en $(u, v; p)$ non unique en v (ce qui conduit aux difficultés algorithmiques déjà évoquées), voire pas de point selle du tout (ce qui, comme déjà vu, fait échouer la coordination par les prix).

Ainsi, partant d'un problème global bien posé (strictement convexe en u comme on l'a supposé), on aboutit à des sous-problèmes mal conditionnés. Ceci est lié au fait qu'en plus des variables de décision "naturelles" u du problème initial, on a introduit artificiellement d'autres variables de décision v_i au niveau des sous-problèmes. En effet, dans (2.28), la minimisation porte à la fois sur u_i et v_i . Ceci ne se produira pas avec les deux autres méthodes examinées ci-dessous. Ce phénomène souligne un inconvénient de la méthode par les prix¹⁹. De surcroît, l'accroissement du nombre de variables à optimiser au niveau des sous-problèmes peut s'avérer assez pénalisant dans certaines applications sur le plan de la résolution numérique.

2.2.3 Décomposition par les quantités

Grâce aux correspondances du Tableau 2.1, le problème (2.7) se traduit directement ici par

$$\min_{u_i \in U_i^{\text{ad}}, v_i \in \mathbb{R}^{m_i}} J_i(u_i, v_i) \quad (2.30a)$$

sous

$$-v_i = \theta_{ii}, \quad (2.30b)$$

$$H_{ji}(u_i, v_i) = \theta_{ji}, \quad (2.30c)$$

les vecteurs $\theta_i = (\theta_{1i}^\top, \dots, \theta_{Ni}^\top)^\top$ dans \mathbb{R}^m étant ceux qui au §2.1.3 définissent une allocation qui doit vérifier (2.8) (avec ici $\vartheta = 0$). De même, l'algorithme (2.13) se traduit ici par

$$\theta_{ji}^{k+1} = \theta_{ji}^k + \varepsilon^k \left(p_{ji}^k - \frac{1}{N} \sum_{l=1}^N p_{jl}^k \right), \quad i = 1, \dots, N, \quad j = 1, \dots, N, \quad (2.31)$$

où p_{ji}^k est le multiplicateur associé à la contrainte (2.30c) si $j \neq i$, et p_{ii}^k est associé à (2.30b). On note que si J_i et H_{ji} sont différentiables, on a, du fait de la stationnarité en v_i du Lagrangien du problème (2.30),

$$p_{ii}^k = \left(\frac{\partial J_i}{\partial v_i}(u_i^k, v_i^k) \right)^\top + \sum_{j \neq i} \left(\frac{\partial H_{ji}}{\partial v_i}(u_i^k, v_i^k) \right)^\top p_{ji}^k.$$

On note aussi que dans cette méthode, au niveau des sous-problèmes, la minimisation est effectuée par rapport à u_i uniquement car la valeur de v_i est imposée par la coordination (voir (2.30b)). Ceci peut représenter un avantage par rapport à la méthode par les prix. Par contre, et nous retrouvons là un inconvénient déjà discuté au commentaire 2.17-e, les contraintes (2.30c) peuvent être en nombre bien supérieur au nombre de variables de décision contenues dans u_i . Ce sont en effet à la fois les "entrées" v_i mais aussi les "sorties" $H_{ji}(u_i, v_i)$ du sous-système i (à destination des sous-systèmes j) qui sont contraintes. En cela, la méthode par les quantités (ou par allocation) diffèrera de la méthode suivante (par prédiction).

Dans le cas de dimension infinie, des ennuis équivalents peuvent également surgir. Ainsi, en considérant à nouveau le problème de commande optimale (2.23), les contraintes (2.30c) se traduiront par des contraintes égalité sur des expressions linéaires en u_i et x_i , où x_i est la trajectoire solution de (2.24). Il est évident que ceci conduira génériquement à des impossibilités. D'une manière générale, des couplages rigides entre systèmes dynamiques, s'exprimant par des contraintes égalité entre variables d'état sur des intervalles de temps entiers — et non en des instants isolés comme dans la décomposition de trajectoires (voir Figure 1.2), ou par des expressions intégrales de ces variables comme un couplage par des stocks tampons par exemple — ne pourront pas être traités par la méthode par allocation.

¹⁹qui peut être partiellement levé en utilisant un Lagrangien augmenté

2.2.4 Décomposition “par prédiction”

Comme on l’a vu, un trait spécifique de cette méthode est de nécessiter une décomposition de l’espace \mathbb{R}^m des contraintes en N , ou moins de N , composantes. Cependant, dans le cas du second modèle considéré ici, cette décomposition s’impose naturellement et elle est pratiquement définie par les correspondances établies ci-dessus, et en particulier les deux dernières lignes du Tableau 2.1. En accord avec (2.15), le sous-problème i s’écrit

$$\min_{u_i \in U_i^{\text{ad}}, v_i \in \mathbb{R}^{m_i}} \left(J_i(u_i, v_i) + \sum_{j \neq i} \langle \mu_j, H_{ji}(u_i, v_i) \rangle \right) \quad (2.32a)$$

sous

$$-v_i = -w_i. \quad (2.32b)$$

Notons que (2.32b) signifie que, pratiquement, les variables d’interaction v_i sont “prédites” par la coordination (d’où le nom d’*“interaction prediction principle”* donné par Mesarovic et associés [22]). Ceci évite les ennuis potentiels qui naîtraient du traitement de ces variables comme des variables de décisions (comme c’était le cas dans la méthode par les prix). C’est aussi un avantage sur le plan de la complexité des sous-problèmes (moins de variables à optimiser). Par ailleurs, contrairement à ce qui se passe dans la méthode par les quantités, ici seules les “entrées” v_i sont imposées et non les “sorties” (voir (2.30c)).

La méthode par prédiction présente donc certains avantages par rapport aux deux autres méthodes dans le cadre de ce second modèle, d’autant plus que les prix marginaux nécessaires à la coordination sont eux-mêmes évalués très simplement ; si la contrainte (2.32b) est dualisée sous la forme

$$-v_i + w_i = 0,$$

on a en effet

$$p_i^{k+1} = \left(\frac{\partial J_i}{\partial v_i}(u_i^{k+1}, w_i^k) \right)^\top + \sum_{j \neq i} \left(\frac{\partial H_{ji}}{\partial v_i}(u_i^{k+1}, w_i^k) \right)^\top \mu_j^k, \quad (2.33)$$

ce qui signifie que p_i est la dérivée de la fonction coût du sous-problème i par rapport à v_i . Cette interprétation est utile lorsque le sous-problème est résolu par une méthode ne faisant pas directement appel à la dualité et aux multiplicateurs²⁰.

Les algorithmes présentés au §2.1.4 s’adaptent ici sans difficulté. Pour l’algorithme de point fixe avec relaxation (2.20), l’équation (2.20a) devient

$$w_i^{k+1} = w_i^k + \varepsilon^k \left(\sum_{j \neq i} H_{ij}(u_j^{k+1}, w_j^k) - w_i^k \right), \quad (2.34)$$

tandis que (2.20b) reste formellement identique (mais p_i^{k+1} est donné explicitement par (2.33)).

Pour l’algorithme d’Arrow-Hurwicz parallèle (2.22), l’équation (2.22a) reste encore formellement la même, et (2.22b) devient

$$\mu_i^{k+1} = \mu_i^k + \rho^k \left(\sum_{j \neq i} H_{ij}(u_j^{k+1}, w_j^k) - w_i^k \right),$$

à comparer avec l’équation (2.34).

²⁰par exemple, la programmation dynamique lorsqu’il s’agit d’un problème de commande optimale

2.3 Annexe : À propos des séries divergentes

Pour minimiser $J(u) = |u|$, $u \in \mathbb{R}$, on utilise l'algorithme de sous-gradient

$$u^{k+1} = u^k - \varepsilon^k r^k, \quad r^k \in \partial J(u^k), \quad \varepsilon^k > 0.$$

Supposons par exemple que $u^0 > 0$. Tant que $u^k > 0$, $r^k = 1$ et

$$u^k = u^0 - \sum_{\ell=0}^{k-1} \varepsilon^\ell.$$

Si la série des ε^k était de somme finie S , et si $u^0 > S$, la valeur $u^\# = 0$ ne pourrait jamais être atteinte. Ceci justifie la condition de *série divergente*

$$\sum_{k \in \mathbb{N}} \varepsilon^k = +\infty \quad (2.35a)$$

qui assure que l'on aura “assez d'énergie” pour passer par l'optimum quel que soit le point de départ.

Une fois que l'on dépasse $u^\#$, r^k change de signe en gardant la valeur absolue 1. L'algorithme tend à *osciller* et ces oscillations ne s'arrêteraient jamais si on n'imposait pas en plus la condition

$$\lim_{k \rightarrow \infty} \varepsilon^k = 0.$$

Dans la pratique, pour la démonstration de convergence, on impose la condition plus forte

$$\sum_{k \in \mathbb{N}} (\varepsilon^k)^2 < +\infty. \quad (2.35b)$$

2.4 Annexe : Exercices sur la dualité

Exercice 2.29. Cet exercice est un prétexte pour réexaminer certaines interprétations géométriques de la dualité en optimisation. On considère un problème d'optimisation général

$$\min_{u \in U^{\text{ad}} \subset \mathcal{U}} J(u) \text{ sous } \Theta(u) = 0 \in \mathcal{C}, \quad (\text{P}_0)$$

et le problème “perturbé”

$$\min_{u \in U^{\text{ad}} \subset \mathcal{U}} J(u) \text{ sous } \Theta(u) = \vartheta \in \mathcal{C}. \quad (\text{P}_\vartheta)$$

On désigne par $\varphi(\vartheta)$ le coût optimal de ce problème perturbé comme fonction de ϑ , second membre des contraintes (voir (2.10)).

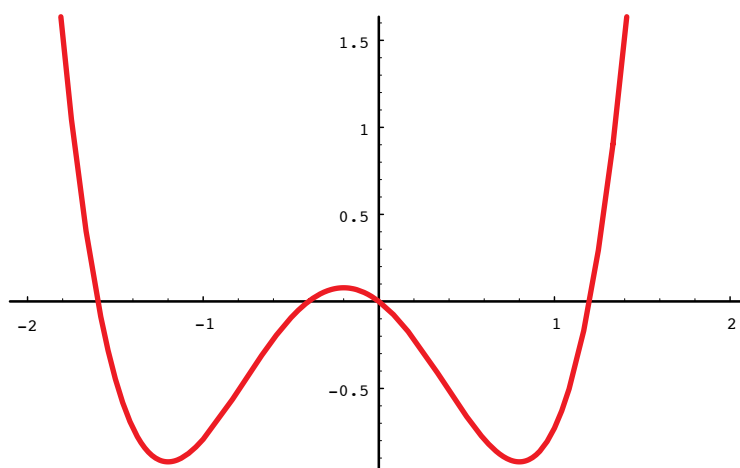
1. Dans l'espace $\mathcal{C} \times \mathbb{R}$, montrer que le “nuage de points” de coordonnées $(\Theta(u), J(u))$ obtenu lorsque u parcourt U^{ad} est bordé inférieurement par le graphe de φ .
2. Un hyperplan de “coefficient directeur” $-p$ (avec $p \in \mathcal{C}^*$, espace dual de \mathcal{C}) et d’“ordonnée à l'origine” a dans l'espace $\mathcal{C} \times \mathbb{R}$ a pour équation

$$y = -\langle p, x \rangle + a.$$

À p fixé, montrer que l'hyperplan “le plus bas” qui rencontre encore le nuage de points précédent (hyperplan d'appui de φ) a pour ordonnée à l'origine $\psi(p)$ où

$$\psi(p) = \min_{u \in U^{\text{ad}}} L(u, p)$$

et L est le Lagrangien du problème (P_0) .

Figure 2.3: Graphe de J (Exemple 2.5)

3. L'enveloppe convexe $\bar{\varphi}$ de φ est par définition l'enveloppe supérieure de ses hyperplans d'appui. Quelle est la valeur de $\bar{\varphi}(0)$?
4. Montrer que φ coïncide avec $\bar{\varphi}$ en 0 si et seulement si L a un point selle sur $U^{\text{ad}} \times \mathbb{C}^*$. Plus généralement, évaluer le “saut de dualité”

$$(\text{écart entre } \min_{u \in U^{\text{ad}}} \max_{p \in \mathbb{C}^*} L(u, p) \text{ et } \max_{p \in \mathbb{C}^*} \min_{u \in U^{\text{ad}}} L(u, p))$$

en termes de φ et $\bar{\varphi}$.

5. On a vu au Lemme 2.2 que s'il existe $p^\sharp \in \mathbb{C}^*$ et $u^\sharp \in \widehat{U}(p^\sharp)$ tel que $\Theta(u^\sharp) = 0$, alors L a un point selle. Alors, p^\sharp réalise le maximum de ψ . En déduire que si L a un saut de dualité, alors ψ est nécessairement non différentiable à son maximum. Quel est son “sous-différentiel” en ce point ? Visualiser les points extrémaux de $\partial\psi(p^\sharp)$ sur un dessin dans le “plan” $\mathbb{C} \times \mathbb{R}$ (on supposera bien sûr que $\mathbb{C} = \mathbb{R}$ pour pouvoir faire le dessin).
6. **Application à l'Exemple 2.5.** Reprendre l'Exemple 2.5 et montrer que l'expression de φ coïncide avec celle de J dont le graphe est représenté sur la Figure 2.3. À partir de cette figure, montrer que $\psi(\cdot)$ est maximale en 0 et tracer l'allure générale du graphe de ψ .

Exercice 2.30. On considère le problème (très simple !), avec $u \in \mathbb{R}$,

$$\min_{u \in [-1, 1]} -u \text{ sous } u = 0.$$

1. Quelles sont les solutions primale u^\sharp et duale p^\sharp (lorsque la contrainte $u = 0$ *uniquement* est dualisée).
2. Ecrire le Lagrangien $L(u, p)$ de ce problème et vérifier les inégalités du point selle.
3. Calculer $\widehat{U}(p)$ pour toutes les valeurs de p .
4. Évaluer la fonction $\psi(p) = \min_{u \in [-1, 1]} L(u, p)$ et dessiner son graphe.
5. Vérifier qu'elle est bien maximale en p^\sharp . Quelle est sa dérivée (ou bien son “sur-différentiel”) en ce point ?

6. Si on arrive à construire, par un algorithme itératif approprié, une suite $\{p^k\}$ qui converge vers p^\sharp , que va-t-il se passer pour les suites $\{u^k\}$ où $u^k \in \widehat{U}(p^k)$? Qu'en déduire quant à la convergence d'un algorithme d'Uzawa ?
7. Reprendre les questions 2 à 6 avec le Lagrangien *augmenté* à la place du Lagrangien ordinaire.

Les deux exercices suivants prolongent l'Exercice 2.29 et portent plus précisément sur l'étude des fonctions φ introduites en (2.10), ou l'équivalent lorsque l'on considère le problème (2.1). On va d'abord observer ce qui se passe dans le cas non convexe puis se concentrer sur le cas convexe pour lequel la fonction φ peut être non différentiable.

Exercice 2.31.

1. On reprend l'exemple considéré à la question 6 de l'Exercice 2.29, c'est-à-dire le problème d'optimisation défini à l'Exemple 2.5. Montrer que le multiplicateur optimal p^\sharp impliqué dans les conditions d'optimalité du premier ordre de ce problème
 - est bien, au signe près, la dérivée de la fonction φ relative à ce problème au point $\vartheta = 0$ (c'est l'*interprétation marginaliste des multiplicateurs*),
 - mais qu'il ne correspond pas à la valeur pour laquelle la fonction ψ (définie par (2.4)) est maximale (cette valeur a été trouvée à la question 6 de l'Exercice 2.29).
2. On se place maintenant dans le cas convexe et on suppose que le Lagrangien du problème d'optimisation admet un point selle. Montrer que l'argument p^\sharp du point selle constitue, au signe près, un sous-gradient de la fonction φ relative à ce problème d'optimisation.

Exercice 2.32. On considère le programme linéaire suivant :

$$\min_{u_1, u_2 \in \mathbb{R}} (2 - (u_1 + u_2)) \quad (2.36a)$$

$$\text{sous } 0 \leq u_1 \leq 1, \quad 0 \leq u_2 \leq 1, \quad (2.36b)$$

$$\text{et } -\frac{1}{2}u_1 + u_2 = \vartheta. \quad (2.36c)$$

On peut considérer que les contraintes (2.36b) définissent les ensembles admissibles $U_i^{\text{ad}}, i = 1, 2$ (contraintes implicites).

1. Étudier d'abord ce problème global et sa solution en fonction des valeurs du paramètre ϑ dans la plage de valeurs où le problème admet une solution. Tracer le graphe de la fonction φ correspondante et montrer qu'elle est non différentiable en $\vartheta = 1/2$ (qui se trouve être l'endroit où son minimum est atteint). Donnez toutes les valeurs possibles du multiplicateur p^\sharp de la contrainte couplante (2.36c) (pour la valeur $\vartheta = 1/2$) et vérifiez que ces valeurs correspondent, au signe près, au sous-différentiel de φ en ce point. Trouver une explication à la non unicité de ces valeurs optimales p^\sharp (c'est-à-dire à la non différentiabilité de φ au point $1/2$).
2. On considère maintenant la décomposition en sous-problèmes correspondant à la méthode par les quantités (voir (2.7)) :

$$\min_{0 \leq u_1 \leq 1} (1 - u_1) \quad \min_{0 \leq u_2 \leq 1} (1 - u_2) \quad (2.37a)$$

$$\text{sous } -\frac{1}{2}u_1 = \theta_1, \quad \text{sous } u_2 = \theta_2, \quad (2.37b)$$

(dont les solutions primales sont évidentes lorsqu'elles existent), avec

$$\theta_1 + \theta_2 = \vartheta \quad (2.38)$$

en raison de (2.8). Quelles plages de valeurs de θ_1 et θ_2 sont acceptables ? Quelles sont les valeurs optimales de (θ_1, θ_2) pour $\vartheta = 1/2$? Quelles sont toutes les valeurs optimales correspondantes possibles p_1^\sharp et p_2^\sharp pour les multiplicateurs des contraintes (2.37b). Quelle est la difficulté par rapport à la condition exigée dans l'énoncé du Lemme 2.12 ou du Lemme 2.14.

2.5 Corrigé des exercices

2.5.1 Corrigé de l'Exercice 2.3

Soit p^\sharp tel qu'il existe $\widehat{u} \in \widehat{U}(p^\sharp)$ vérifiant (2.1b). Par définition, \widehat{u}_i est solution de (2.2) pour tout i , donc

$$J_i(\widehat{u}_i) + \langle p^\sharp, \Theta_i(\widehat{u}_i) \rangle \leq J_i(u_i) + \langle p^\sharp, \Theta_i(u_i) \rangle, \quad \forall u_i \in U_i^{\text{ad}}.$$

En sommant ces inégalités pour $i = 1, \dots, N$, on obtient une inégalité qui peut s'interpréter comme l'inégalité de droite du point selle pour le Lagrangien (2.3) du problème global (2.1).

Quant à l'inégalité de gauche de ce point selle ($L(\widehat{u}, p) \leq L(\widehat{u}, p^\sharp)$, $\forall p$), elle est trivialement vérifiée du fait que l'on a supposé que (2.1b) est vraie pour \widehat{u} . Tout point selle de L étant une solution primale duale de (2.1), on en déduit que \widehat{u} , obtenue comme concaténation de solutions des problèmes (2.2), est une solution de (2.1) à partir du moment où elle vérifie la contrainte (2.1b).

Réciproquement, si (u^\sharp, p^\sharp) est un point selle de L , alors u^\sharp vérifie évidemment (2.1b). Par ailleurs, en considérant l'inégalité de droite du point selle de L avec $u_j = u_j^\sharp$ pour tout j , sauf pour $j = i$, on obtient que u_i^\sharp est solution de (2.2).

2.5.2 Corrigé de l'Exercice 2.6

Reprenons le corrigé précédent dans le cas de contraintes (2.1b) inégalité. La première partie du raisonnement concernant l'inégalité de droite du point selle reste valide sans changement. L'inégalité de gauche doit être démontrée pour tout $p \in (\mathbb{R}^m)^+$ dans le membre de gauche de l'inégalité, et pour p^\sharp dans le membre de droite. Le fait que $\Theta(\widehat{u}) - \vartheta \leq 0$ ⁽²¹⁾ permet d'affirmer que $\langle p, \Theta(\widehat{u}) - \vartheta \rangle \leq 0$ pour tout $p \in (\mathbb{R}^m)^+$, en particulier pour p^\sharp . Mais pour $p = 0$ (valeur admissible), cette expression vaut 0, donc elle doit aussi valoir 0 pour p^\sharp .

La seconde partie de la condition (2.4-d) est donc bien nécessaire pour montrer que la paire (\widehat{u}, p^\sharp) constitue un point selle du Lagrangien L .

Évidemment, on peut se demander si une concaténation \widehat{u} des solutions des problèmes (2.2) vérifiant la contrainte (2.1b) peut être une solution de (2.1) même sans être un point selle de L . Le lecteur est invité à construire des contre-exemples.

2.5.3 Corrigé de l'Exercice 2.9

Soit $\widehat{u}_i(\theta_i^\sharp)$ une solution quelconque de (2.7) correspondant à une solution θ^\sharp de (2.9). La concaténation $\widehat{u}(\theta^\sharp)$ de ces variables est d'abord une solution *admissible* de (2.1) car elle vérifie la contrainte globale (2.1b) (c'est une conséquence évidente du fait que θ^\sharp vérifie (2.8) et que $\widehat{u}_i(\theta_i^\sharp)$ vérifie (2.7b) pour tout i). Avec toute autre solution admissible u de (2.1), on fabrique une allocation θ admissible, c'est-à-dire vérifiant (2.8), en posant $\theta_i = \Theta_i(u_i)$ pour tout i . Soit $\widehat{u}_i(\theta_i)$ une solution correspondante de (2.7). Par définition, on a

$$\varphi_i(\theta_i) = J_i(\widehat{u}_i(\theta_i)) \leq J_i(u_i).$$

²¹Cette inégalité doit se comprendre pour toute coordonnée dans \mathbb{R}^m .

Mais par ailleurs, puisque θ^\sharp est supposée être une solution de (2.9), on a

$$\sum_{i=1}^N J_i(\widehat{u}_i(\theta_i^\sharp)) = \sum_{i=1}^N \varphi_i(\theta_i^\sharp) \leq \sum_{i=1}^N \varphi_i(\theta_i) .$$

De ces deux inégalités, on déduit que

$$\sum_{i=1}^N J_i(\widehat{u}_i(\theta_i^\sharp)) \leq \sum_{i=1}^N J_i(u_i)$$

pour tout u admissible pour le problème global. On a donc bien montré que $\widehat{u}(\theta^\sharp)$ est solution de (2.1).

Réciproquement, soit u^\sharp une solution de (2.1) et θ^\sharp défini par (2.11). Montrons d'abord que $u_i^\sharp = \widehat{u}_i(\theta^\sharp)$. En effet, dans l'inégalité

$$\sum_{j=1}^N J_j(u_j^\sharp) \leq \sum_{j=1}^N J_j(u_j) \quad (2.39)$$

pour tout u vérifiant (2.1b), on peut choisir $u_j = u_j^\sharp$ pour tout j sauf pour $j = i$ pour lequel on choisit n'importe quel u_i tel que $\Theta_i(u_i) = \theta_i^\sharp$, ce qui reste compatible avec (2.1b).

Ceci montre que le premier membre de (2.39) s'écrit aussi $\sum \varphi_i(\theta_i^\sharp)$. Quant au second membre, on pourra l'écrire $\sum \varphi_i(\theta_i)$, pour toute autre allocation θ , en y plaçant des valeurs $\widehat{u}_i(\theta_i)$ correspondantes. Ceci montre que θ^\sharp est solution de (2.9).

2.5.4 Corrigé de l'Exercice 2.13

En sommant les égalités (2.12) pour $i = 1, \dots, N$, et en tenant compte de (2.8) et du fait que tous les p_i^\sharp sont égaux à une valeur commune p^\sharp , on obtient les conditions d'optimalité du premier ordre pour le problème (2.1).

2.5.5 Corrigé de l'Exercice 2.15

Le principe de la démonstration est identique à celui de la démonstration précédente : on écrit les inégalités du point selle pour chaque sous-problème et on les additionne pour tout i en se servant des hypothèses.

2.5.6 Corrigé de l'Exercice 2.16

Le problème global s'écrit :

$$\min_{x_1, y_1, x_2, y_2} (-x_1^2 - y_1^2 - 4x_1y_1 - x_2^2 - y_2^2 - 4x_2y_2) \quad \text{sous} \quad x_1 + y_1 - x_2 - y_2 = 0 .$$

Quant aux sous-problèmes suggérés dans l'énoncé, ils s'écrivent :

$$\begin{aligned} \textbf{Sous-problème 1 :} \quad & \min_{x_1, y_1} (-x_1^2 - y_1^2 - 4x_1y_1) \\ & \text{sous} \quad x_1 + y_1 = 0 , \\ \textbf{Sous-problème 2 :} \quad & \min_{x_2, y_2} (-x_2^2 - y_2^2 - 4x_2y_2) \\ & \text{sous} \quad -x_2 - y_2 = 0 . \end{aligned}$$

On vérifie que $x_i = y_i = p_i = 0, i = 1, 2$, satisfont aux conditions du premier ordre pour ces sous-problèmes, et que les mêmes valeurs nulles pour les variables primales et l'unique variable duale en font de même pour le problème global.

Examinons maintenant les conditions du second ordre. Pour chaque sous-problème, les contraintes étant linéaires se confondent avec leur plan tangent. Les vecteurs $x_1 = a$ quelconque, et $y_1 = -x_1 = -a$, respectivement $x_2 = b$ quelconque, et $y_2 = -x_2 = -b$, sont les vecteurs génériques satisfaisant la contrainte du sous-problème 1, respectivement 2. Lorsque l'on remplace y_i par $-x_i$ dans l'expression du coût J_i , on obtient l'expression $4x_i^2$, évidemment positive pour tout $x_i \neq 0$. Ceci montre que J_i est défini positif le long de la contrainte.

Quant au problème global, il comporte une contrainte linéaire pour quatre variables : l'ensemble admissible, qui se confond avec la variété tangente, comporte trois degrés de liberté. On peut choisir arbitrairement les valeurs $x_1 = a, y_1 = b, x_2 = c$, mais il faut prendre ensuite $y_2 = a + b - c$. On pourrait substituer ces quatre valeurs dans $J_1 + J_2$ et se convaincre que l'expression obtenue n'est pas une forme quadratique définie positive en (a, b, c) . L'énoncé suggère de prendre le cas $b = c = 0$, d'où $x_1 = y_2 = a, x_2 = y_1 = 0$: pour ces valeurs $J_1 + J_2$ vaut $-2a^2$ qui est évidemment négatif pour $a \neq 0$.

2.5.7 Corrigé de l'Exercice 2.19

Il suffit de paraphraser la démonstration du Lemme 2.8 (voir corrigé de l'exercice 2.9, §2.5.3) en effectuant les adaptations rendues nécessaires par le fait que

- des inégalités prennent la place des égalités dans (2.1b) et (2.7b) ;
- la formule (2.14) remplace la formule (2.11) (on va discuter plus loin l'intérêt de cette modification).

Ainsi, dans la première partie de la démonstration, lorsqu'on introduit un u admissible quelconque pour le problème (2.1), il faut définir les θ_i correspondants en utilisant (2.14) (à partir de ce u) au lieu de (2.11). Dans la démonstration de la réciproque, on prendra $u_j = u_j^\#$ pour tout $j \neq i$ et n'importe quel u_i tel que $\Theta_i(u_i) \leq \theta_i^\#$.

La remarque fondamentale est la suivante, pour que le reste de la démonstration fonctionne : pour tout u vérifiant (2.1b) (en particulier pour $u^\#$), il faut fabriquer des θ_i tel que d'une part (2.8) soit vérifiée, et d'autre part il soit vérifié que $\Theta_i(u_i) \leq \theta_i$ (sachant que ceci n'a lieu a priori que pour la somme en i , comme c'est dit par (2.1b)). La formule (2.14) a pour intérêt :

1. d'ajouter au vecteur $\Theta_i(u_i)$ un vecteur *non négatif* (par l'hypothèse que (2.1b) est vérifiée), ce qui rend l'inégalité $\Theta_i(u_i) \leq \theta_i$ évidemment vraie ;
2. de faire en sorte que les θ_i ainsi définis satisfont automatiquement à (2.8), comme on peut le vérifier facilement à partir de (2.14).

2.5.8 Corrigé de l'Exercice 2.21

Les conditions du premier ordre pour le problème (2.15) s'écrivent

$$J'_i(\hat{u}_i) + \sum_{j \neq i} (\Theta'_{ji}(\hat{u}_i))^\top \mu_j + (\Theta'_{ii}(\hat{u}_i))^\top \hat{p}_i = 0 ,$$

$$\Theta_{ii}(\hat{u}_i) = \vartheta_i - w_i .$$

L'utilisation de l'hypothèse (2.16) permet de retrouver la partie i des conditions du premier ordre pour le problème (2.1). La réciproque est tout aussi immédiate.

2.5.9 Corrigé de l'Exercice 2.23

Suivant les indications données, considérons d'abord l'inégalité de droite pour le point selle du Lagrangien relatif au problème (2.15). On obtient : pour tout $u_i \in U_i^{\text{ad}}$

$$J_i(\widehat{u}_i) + \sum_{j \neq i} \langle \mu_j, \Theta_{ji}(\widehat{u}_i) \rangle + \langle \widehat{p}_i, \Theta_{ii}(\widehat{u}_i) \rangle \leq J_i(u_i) + \sum_{j \neq i} \langle \mu_j, \Theta_{ji}(u_i) \rangle + \langle \widehat{p}_i, \Theta_{ii}(u_i) \rangle .$$

En tenant compte de (2.16b) et en sommant pour tout i , on obtient l'inégalité de droite pour le Lagrangien du problème global (2.1).

Pour l'inégalité de gauche, on a : pour tout p_i (avec $p_i \geq 0$ dans le cas de contraintes inégalité),

$$\langle p_i, \Theta_{ii}(\widehat{u}_i) + w_i - \vartheta_i \rangle \leq \langle \widehat{p}_i, \Theta_{ii}(\widehat{u}_i) + w_i - \vartheta_i \rangle .$$

Là encore, en tenant compte de (2.16a) et en sommant pour tout i , on obtient l'inégalité correspondante relative au problème global.

2.5.10 Corrigé de l'Exercice 2.24

Attachons nous à vérifier les conditions du second ordre pour les deux sous-problèmes. Dans le sous-problème (2.17d)–(2.17e), une variation a de x_1 doit s'accompagner d'une variation $-a$ de y_1 pour rester dans le plan tangent aux contraintes. On obtient, en substituant a à x_1 et $-a$ à y_1 dans le Lagrangien, et en ne s'intéressant qu'aux termes de degré 2 (les autres ne contribuant pas aux dérivées du second ordre), l'expression suivante :

$$-a^2 - a^2 - 4a(-a) = 2a^2$$

qui est évidemment positive. Le même type de calcul appliqué au sous-problème (2.17f)–(2.17g) conduit à considérer des variations $x_2 = b$ et $y_2 = -b/2$, ce qui donne, pour les termes de degré 2 :

$$-b^2 - (-b/2)^2 - 4b(-b/2) = 3b^2/4$$

qui est encore positif.

Pour le problème global (2.17a)–(2.17c), les variations qui restent dans le plan tangent aux contraintes sont de la forme²² $(x_1, y_1, x_2, y_2) = (a, b - a, b, 0)$, ce qui donne, pour les termes de degré 2 :

$$-a^2 - (b - a)^2 - 4a(b - a) - b^2 = 2a^2 - 2b^2 - 2ab .$$

Cette expression prend des valeurs négatives, par exemple si $a = 0$ et $b = 1$, ce qui montre que les conditions du second ordre ne sont pas vérifiées pour le problème (2.17a)–(2.17c) alors qu'elles l'étaient pour les problèmes (2.17d)–(2.17g).

2.5.11 Corrigé de l'Exercice 2.25

Les contraintes (2.18b)–(2.18c) définissent une variété de dimension 2 qui peut être paramétrisée par a et b tels que les points admissibles de \mathbb{R}^4 s'expriment par $(x_1, y_1, x_2, y_2) = (a, b, -a, -b)$. En substituant ces valeurs dans la fonction coût (2.18a), on obtient l'expression :

$$a^2 + b^2 + 3ab + a^2 + b^2 - 3(-a)(-b) = 2(a^2 + b^2) .$$

Ceci montre que la fonction coût est définie positive sur la variété admissible, et donc les conditions suffisantes du second ordre sont vérifiées (dans la dérivée seconde du Lagrangien, seule la fonction coût a une contribution non nulle puisque les contraintes sont linéaires).

La contrainte (2.18e) du premier sous-problème définit des déplacements admissibles $(x_1, y_1) = (a, -a)$ et les termes de degré 2 dans (2.18d) donnent alors $a^2 + (-a)^2 + 3a(-a) = -a^2$, ce qui

²²C'est une paramétrisation possible, à 2 paramètres puisqu'il y a 4 variables et 2 contraintes.

montre que la fonction coût (2.18d) est définie *négative* pour les déplacements admissibles. De même, (2.18g) autorise des déplacements $(x_2, y_2) = (b, b)$, ce qui donne, dans (2.18f), les termes du second ordre $b^2 + b^2 - 3b^2 = -b^2$: donc même constat.

On vient donc de mettre en évidence, cette fois-ci, que c'est le problème global qui est *convexe* sur sa variété admissible alors que les deux sous-problèmes sont *concaves* sur leurs variétés admissibles respectives !

2.5.12 Corrigé de l'Exercice 2.27

On suppose que dans (2.1), $N = 2$ et $m = 1$. On écrira donc l'unique contrainte (2.1b) sous la forme $\Theta(u) = \Theta_1(u_1) + \Theta_2(u_2) = 0$. On ne peut donc affecter cette unique contrainte qu'à l'un des deux sous-problèmes, disons le sous-problème 1. Ce sous-problème n'aura donc pas à subir de modification de sa fonction coût (car, en observant le terme marqué essentiel dans (2.15a), Θ_{ji} pour $j = 2$ et $i = 1$ est identiquement nulle). Le sous-problème 2 devra par contre subir une telle modification de coût, mais n'aura pas, par contre, de contrainte du type (2.15b) à gérer. Les deux sous-problèmes s'écriront alors (comparer avec (2.15)) :

$$\begin{aligned} \text{Sous-problème 1} \quad & \min_{u_1 \in U_1^{\text{ad}}} J_1(u_1) \\ \text{sous} \quad & \Theta_1(u_1) = \vartheta - w ; \end{aligned}$$

$$\text{Sous-problème 2} \quad \min_{u_2 \in U_2^{\text{ad}}} (J_2(u_2) + \mu \Theta_2(u_2)) .$$

2.5.13 Corrigé de l'Exercice 2.29

1. Le bord inférieur des points de la forme $(\Theta(u), J(u))$ lorsque u parcourt U^{ad} est obtenu en fixant d'abord l'abscisse, c'est-à-dire en imposant $\Theta(u) = \vartheta$ pour une valeur donnée de ϑ , puis en cherchant l'ordonnée minimale : donc il faut minimiser $J(u)$ sur U^{ad} et sous la contrainte $\Theta(u) = \vartheta$, ce qui est exactement la définition de $\varphi(\vartheta)$.
2. L'hyperplan $y = -\langle p, x \rangle + a$ rencontre le nuage de points s'il existe $u \in U^{\text{ad}}$ tel que x puisse s'écrire $\Theta(u)$, y puisse s'écrire $J(u)$, et (x, y) vérifie l'équation de l'hyperplan. Alors,

$$a = y + \langle p, x \rangle = J(u) + \langle p, \Theta(u) \rangle = L(u, p) .$$

Ensuite, l'hyperplan le plus bas correspond au a minimal, ce qui revient à minimiser le Lagrangien L sur U^{ad} , et donc le a minimal est exactement égal à $\psi(p)$.

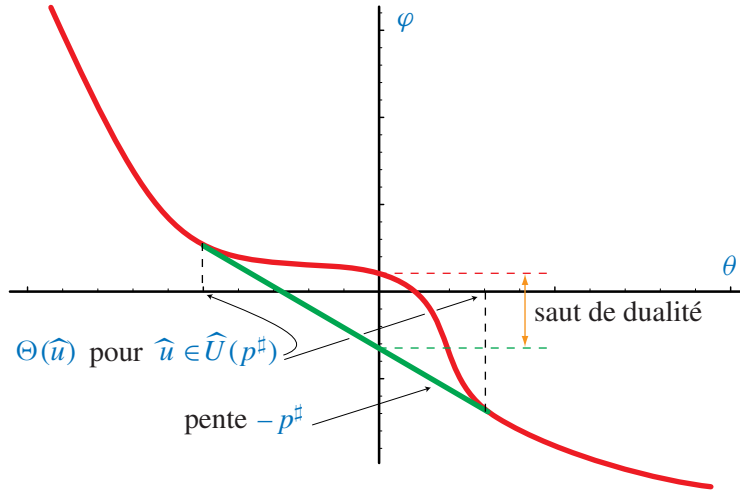
3. En $\vartheta = 0$, les hyperplans passent par leur ordonnée à l'origine a . L'hyperplan d'appui de pente $-p$ passe donc à la hauteur $\psi(p)$, et le plus haut d'entre eux (enveloppe supérieure) passe à $\max_p \psi(p)$. On a donc

$$\bar{\varphi}(0) = \max_{p \in \mathbb{C}^*} \psi(p) .$$

4. On rappelle d'abord que (P_0) est équivalent à $\min_u \max_p L$. En effet, pour calculer cette valeur, on calcule d'abord le max à u fixé, soit

$$\max_{p \in \mathbb{C}^*} L(u, p) = \begin{cases} +\infty & \text{si } \Theta(u) \neq 0 ; \\ J(u) & \text{si } \Theta(u) = 0 . \end{cases}$$

Comme on cherche à minimiser ensuite cette expression en $u \in U^{\text{ad}}$, on commence par éviter la valeur $+\infty$ en respectant la contrainte $\Theta(u) = 0$, et on est ramené enfin à minimiser $J(u)$.

Figure 2.4: Graphe de φ et enveloppe convexe

Ceci étant fait, on obtient donc $\varphi(0) = \min_u \max_p L$. Cette valeur coïncide avec $\bar{\varphi}(0)$, d'après la question précédente, si et seulement si

$$\varphi(0) = \min_{u \in U^{\text{ad}}} \max_{p \in \mathbb{C}^*} L(u, p) = \bar{\varphi}(0) = \max_{p \in \mathbb{C}^*} \psi(p) = \max_{p \in \mathbb{C}^*} \min_{u \in U^{\text{ad}}} L(u, p),$$

ce qui est équivalent à l'existence d'un point selle de L .

D'une manière générale, l'écart entre $\bar{\varphi}(0) = \max_p \min_u L$ et $\varphi(0) = \min_u \max_p L$ (la première valeur étant évidemment inférieure ou égale à la seconde) mesure le "saut de dualité".

- Si L n'a pas de point selle, en considérant $p^\# \in \arg \max_{p \in \mathbb{C}^*} \psi(p)$, pour tout $\hat{u} \in \hat{U}(p^\#)$, nécessairement $\Theta(\hat{u}) \neq 0$ (sinon, L aurait un point selle d'après le Lemme 2.2). Or on sait par ailleurs que $\Theta(\hat{u}) \in \partial \psi(p^\#)$, et comme on a pris un argument $p^\#$ du maximum de ψ , on a aussi nécessairement $0 \in \partial \psi(p^\#)$. On a donc trouvé *deux éléments distincts* dans $\partial \psi(p^\#)$, ce qui prouve que ψ n'est pas différentiable à son maximum.

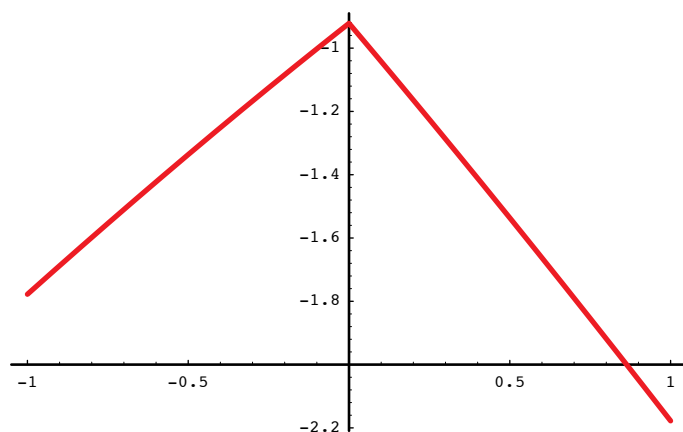
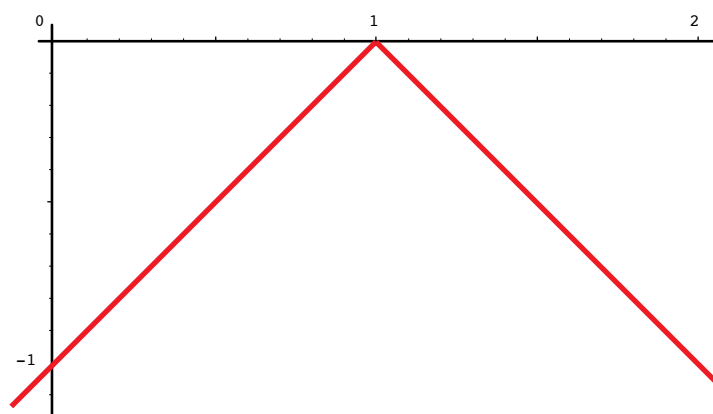
En fait, $\partial \psi(p^\#)$ est obtenu par enveloppe convexe fermée des valeurs $\Theta(\hat{u})$ lorsque \hat{u} parcourt $\hat{U}(p^\#)$. Ces valeurs sont les points où l'hyperplan d'appui du graphe de φ de pente $-p^\#$ touche ce graphe (voir Figure 2.4).

- Dans l'Exemple 2.5, la contrainte étant $u = 0$ (soit $u = \vartheta$ pour le problème perturbé), $\varphi(\vartheta)$ coïncide avec $J(\vartheta)$ dont le graphe est représenté sur la Figure 2.3. On voit que l'enveloppe convexe de cette courbe a une pente horizontale en $\vartheta = 0$ (donc $p^\# = 0$). L'allure de la courbe ψ est représentée sur la Figure 2.5.

2.5.14 Corrigé de l'Exercice 2.30

- Le Lagrangien s'écrit $L(u, p) = -u + pu = (p - 1)u$. La solution primale est évidemment $u^\# = 0$. Le Lagrangien, de dérivée en u égale à $p - 1$, est stationnaire en u au point $u^\#$ (condition de Kuhn et Tucker) si et seulement si $p = p^\# = 1$. L'unique solution primale duale est donc la paire $(u^\#, p^\#) = (0, 1)$.
- Les inégalités du point selle s'écrivent

$$\forall u \in [-1, 1], \forall p \in \mathbb{R}, (p - 1)u^\# \leq (p^\# - 1)u^\# \leq (p^\# - 1)u,$$

Figure 2.5: Allure de ψ pour l'Exemple 2.5Figure 2.6: Graphe de ψ

et sont trivialement vérifiées ici.

3. On a

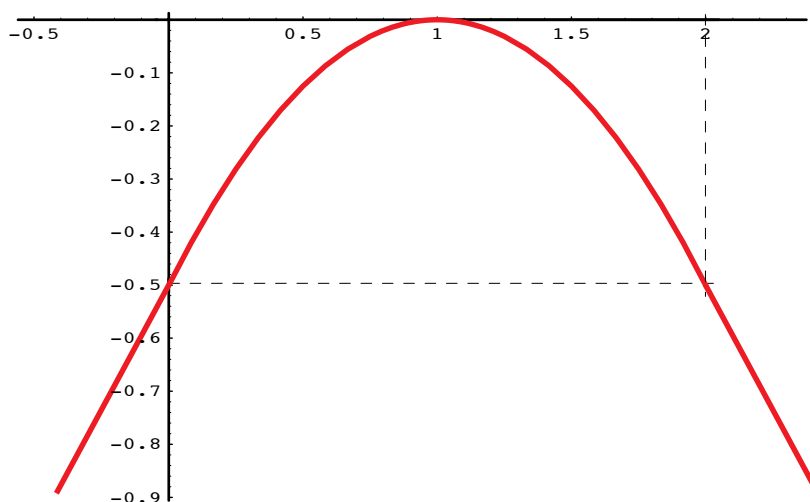
$$\widehat{U}(p) = \begin{cases} 1 & \text{si } p < 1 ; \\ [-1, 1] & \text{si } p = 1 ; \\ -1 & \text{si } p > 1 . \end{cases}$$

4. Compte tenu de la réponse ci-dessus, on a

$$\psi(p) = \begin{cases} p - 1 & \text{si } p \leq 1 ; \\ 1 - p & \text{si } p \geq 1 . \end{cases}$$

Le graphe est représenté sur la Figure 2.6.

5. ψ est bien maximale en $p = 1$ et non différentiable. Son “sous-différentiel” (s’agissant d’une fonction concave) est égal à $[-1, 1]$, c’est-à-dire exactement $\Theta(\widehat{U}(p^\#))$.
6. Si $p^k < 1$, l’algorithme d’Uzawa tendra à faire augmenter p^k mais u^k sera égal à 1 ; inversement, si $p^k > 1$, l’algorithme d’Uzawa tendra à faire diminuer p^k mais u^k sera égal à -1 . On va donc construire des suites $\{u^k\}$ dont les valeurs d’adhérence seront toujours -1 et 1 , alors que la solution optimale est $u^\# = 0$!

Figure 2.7: Graphe de ψ_c

7. Le Lagrangien augmenté est égal à

$$L_c(u, p) = (p - 1)u + cu^2/2$$

pour une constante c positive. On a alors

$$\hat{u}(p) = \begin{cases} 1 & \text{si } p < 1 - c ; \\ (1 - p)/c & \text{si } 1 - c < p < 1 + c ; \\ -1 & \text{si } p > 1 + c . \end{cases}$$

et

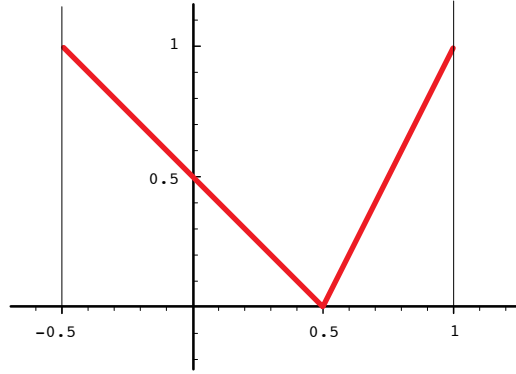
$$\psi_c(p) = \begin{cases} p - 1 + c/2 & \text{si } p \leq 1 - c ; \\ -(p - 1)^2/2c & \text{si } 1 - c \leq p \leq 1 + c ; \\ 1 - p + c/2 & \text{si } p \geq 1 + c . \end{cases}$$

La courbe est représentée sur la Figure 2.7 (pour $c = 1$). Cette fonction est différentiable et son gradient, égal à $\hat{u}(p)$, approche $u^\sharp = 0$ quand p approche $p^\sharp = 1$.

2.5.15 Corrigé de l'Exercice 2.31

- On a vu à la question 6 de l'Exercice 2.29 que la fonction φ relative au problème considéré à l'Exercice 2.5 coïncide avec la fonction J dont le graphe a été représenté sur la Figure 2.3. À l'abscisse $\vartheta = 0$, la pente de cette courbe vaut $-0,768$ qui est bien l'opposé de la valeur du multiplicateur optimal des conditions d'optimalité du premier ordre. Par contre, on a vu aussi dans l'Exercice 2.5 que ψ est maximale en $p = 0$, ce qui n'a rien de commun avec la valeur précédente trouvée pour $\varphi'(0)$. On retiendra donc qu'en l'absence de point selle du Lagrangien, seul le multiplicateur des conditions du premier ordre a une interprétation marginaliste.
- Soit (u^\sharp, p^\sharp) le point selle du Lagrangien du problème (2.1) (avec éventuellement des contraintes inégalité) correspondant à un certain second membre ϑ . On va considérer une autre valeur $\bar{\vartheta}$ de ce second membre, et le point selle $(\bar{u}^\sharp, \bar{p}^\sharp)$ correspondant. L'inégalité de droite du point selle (pour le problème avec ϑ) peut être utilisée pour la valeur \bar{u}^\sharp au second membre, ce qui donne

$$J(u^\sharp) + \langle p^\sharp, \Theta(u^\sharp) - \vartheta \rangle \leq J(\bar{u}^\sharp) + \langle p^\sharp, \Theta(\bar{u}^\sharp) - \vartheta \rangle .$$

Figure 2.8: Graphe de φ pour le problème (2.36)

Le premier membre de cette inégalité est égal à $\varphi(\vartheta)$ (c'est la valeur du premier terme et le second est nul — contrainte égalité, ou “écarts complémentaires” dans le cas inégalité, oblige(nt)). Au second membre, par définition, $J(\bar{u}^\sharp) = \varphi(\bar{\vartheta})$, tandis que $\langle p^\sharp, \Theta(\bar{u}^\sharp) \rangle$ est égal (cas égalité) ou inférieur ou égal (cas inégalité — voyez-vous pourquoi ?) à $\langle p^\sharp, \bar{\vartheta} \rangle$. On obtient donc

$$\varphi(\vartheta) \leq \varphi(\bar{\vartheta}) + \langle p^\sharp, \bar{\vartheta} - \vartheta \rangle ,$$

inégalité qui prouve exactement le résultat recherché.

2.5.16 Corrigé de l'Exercice 2.32

1. Le premier membre de la contrainte (2.36c) est minimal quand u_2 prend sa valeur inférieure 0 (dictée par (2.36b)) et u_1 prend sa valeur supérieure 1. La valeur minimale de ϑ pour laquelle les contraintes (2.36c) restent compatibles avec (2.36b) est donc $-1/2$. Par un raisonnement similaire, on voit que la valeur maximale autorisée pour ϑ est 1. En dehors de l'intervalle $[-1/2, 1]$, il n'y a plus de solution admissible pour le problème (2.36) et φ vaut $+\infty$.

À l'intérieur de cet intervalle, on peut tirer u_1 en fonction de u_2 à l'aide de l'équation (2.36c) ($u_1 = 2u_2 - 2\vartheta$) et le substituer dans le coût (2.36a) qui devient égal à $2 + 2\vartheta - 3u_2$, et les contraintes (2.36b) qui deviennent ensemble :

$$\max(0, \vartheta) \leq u_2 \leq \min(1, \vartheta + 1/2) .$$

Le coût est minimal pour la valeur de u_2 maximale, c'est-à-dire pour $u_2 = \min(1, \vartheta + 1/2)$. Ce coût optimal est par définition égal à $\varphi(\vartheta)$ qui vaut donc :

$$\varphi(\vartheta) = \max(2\vartheta - 1, -\vartheta + 1/2) .$$

Le graphe de cette fonction est représenté sur la Figure 2.8. On observe que cette fonction est minimale, et surtout non différentiable, en $\vartheta = 1/2$ et admet pour sous-différentiel le segment $[-1, 2]$.

Pour $\vartheta = 1/2$, les valeurs optimales de u_1 et u_2 sont, d'après ce qui précède, $u_1 = u_2 = 1$.

Montrons que tout $p^\sharp \in [-2, 1]$ est un multiplicateur optimal de la contrainte (2.36c) lorsque $\vartheta = 1/2$. En effet, en ne dualisant que les contraintes dont on sait qu'elles seront actives à l'optimum (en particulier celles qui définissent la borne supérieure 1 pour u_1 et u_2),²³ le Lagrangien du problème (2.36) s'écrit :

²³De toute façon, les contraintes non actives admettent des multiplicateurs optimaux qui sont nuls

$$L(u, p) = 2 - (u_1 + u_2) + p(-u_1/2 + u_2 - 1/2) + q_1(u_1 - 1) + q_2(u_2 - 1) ,$$

avec $q_1 \geq 0$ et $q_2 \geq 0$. Les conditions de stationnarité de L en u_1 et u_2 donnent :

$$\begin{aligned} \frac{\partial L}{\partial u_1} &= -1 - \frac{p^\sharp}{2} + q_1^\sharp = 0 , \\ \frac{\partial L}{\partial u_2} &= -1 + p^\sharp + q_2^\sharp = 0 , \end{aligned}$$

avec pour seules autres conditions (à partir du moment où $u_1^\sharp = u_2^\sharp = 1$) que $q_1^\sharp \geq 0$ et $q_2^\sharp \geq 0$. On vérifie que ces conditions d'optimalité sont satisfaites pour toute valeur $p^\sharp \in [-2, 1]$, ce qui est conforme avec le fait que $-p^\sharp \in \partial\varphi(1/2)$ (voir point 2 du corrigé de l'Exercice 2.31).

Il est clair que cette non unicité de la valeur optimale des multiplicateurs tient au fait que *trois* contraintes (dans \mathbb{R}^2), parmi toutes les contraintes (2.36b)–(2.36c), se trouvent être actives en même temps (ce qui ne se produit pas pour les autres valeurs autorisées de ϑ — le vérifier), et que, dans ces conditions, les gradients des contraintes actives ne peuvent évidemment pas être linéairement indépendants.

2. La solution primale des sous-problèmes (2.37) est trivialement donnée par $u_1 = -2\theta_1$ et $u_2 = \theta_2$, à condition que $0 \leq -2\theta_1 \leq 1$, c'est-à-dire $-1/2 \leq \theta_1 \leq 0$, et $0 \leq \theta_2 \leq 1$. Mais par ailleurs il faut respecter (2.38), d'où la contrainte supplémentaire sur θ_2 : $-1/2 \leq \vartheta - \theta_2 \leq 0$, donc finalement

$$\max(0, \vartheta) \leq \theta_2 \leq \min(1, \vartheta + 1/2) .$$

Cet intervalle délimite les valeurs de θ_2 pour lesquelles, avec les valeurs correspondantes de θ_1 via (2.38), les sous-problèmes admettront des solutions admissibles.

Pour $\vartheta = 1/2$, la solution optimale du problème (2.36), à savoir $u_1^\sharp = u_2^\sharp = 1$, sera obtenue pour $\theta_1^\sharp = -1/2$ et $\theta_2^\sharp = 1$. La stationnarité des Lagrangiens L_1 et L_2 des deux sous-problèmes, où l'on a comme précédemment dualisé uniquement les contraintes actives à l'optimum, donne :

$$\begin{aligned} \frac{\partial L_1}{\partial u_1} &= -1 - \frac{p_1^\sharp}{2} + q_1^\sharp = 0 , \\ \frac{\partial L}{\partial u_2} &= -1 + p_2^\sharp + q_2^\sharp = 0 , \end{aligned}$$

ce qui donne, avec les conditions de non-négativité des q_i , les valeurs possibles :

$$p_1^\sharp \leq 1 , \quad p_2^\sharp \geq -2 . \quad (2.40)$$

On retrouve, si $p_1^\sharp = p_2^\sharp$ (condition du Lemme 2.12 ou 2.14), la plage commune de valeurs $[-2, 1]$ déjà trouvée à la question précédente, mais ce qui est grave, c'est que toutes les autres valeurs de p_1^\sharp et p_2^\sharp permises par (2.40) peuvent aussi bien être trouvées comme valeurs optimales par les sous-problèmes, ce qui ne permet pas en général de s'apercevoir que la coordination a atteint l'allocation optimale.

Chapitre 3

Le Principe du Problème Auxiliaire

3.1 Introduction

Jusqu'ici dans ce cours, trois méthodes de décomposition-coordination ont été exposées à partir de modèles relativement généraux, mais présentant cependant la particularité d'avoir des fonctions coût et des contraintes additives par rapport à la décomposition en sous-problèmes.

Il a semblé que cette particularité jouait un rôle essentiel dans la démarche de décomposition, et il est vrai qu'elle constitue une forte incitation et un guide pour définir le découpage en sous-problèmes.

Mais par ailleurs, cette caractéristique constitue un frein lorsqu'il s'agit par exemple d'utiliser le Lagrangien augmenté à la place du Lagrangien ordinaire, car ce genre de technique ne respecte pas la structure additive. Il en a été question à plusieurs reprises au Chapitre 2.

Dans ce qui suit, on présente un principe général en optimisation appelé "Principe du Problème Auxiliaire" (PPA). Ce principe a plusieurs objectifs et avantages.

- Il propose un cadre général pour construire des algorithmes itératifs en optimisation. Ceci inclut des algorithmes classiques comme l'algorithme du gradient ou l'algorithme de Newton, ainsi que la plupart des algorithmes de coordination. Dans ce cadre unifié, on réalise donc qu'il n'y a pas de différence essentielle de nature entre tous ces algorithmes.
- Ce cadre permet aussi de se focaliser sur l'étude de convergence d'un *nombre limité* de schémas itératifs de base au travers d'un point de vue général et abstrait. Ceci permet de bien dégager les hypothèses et mécanismes généraux qui assurent la convergence en dehors des particularités de chaque situation spécifique.
- En dégageant les ressorts essentiels, l'application de ces idées au cas des algorithmes de décomposition-coordination permet de s'affranchir des restrictions structurelles (comme l'additivité) évoquées plus haut.

En fait, la théorie de la décomposition-coordination est basée sur deux "piliers" essentiels.

1. Le premier concerne ce que l'on peut appeler la "manipulation des problèmes" à partir de leur forme originale vers d'autres formes équivalentes mais qui se prêtent mieux à l'approche par décomposition. Un exemple de telles manipulations a été donné lors du passage du problème (2.1) à ses formes équivalentes (2.9) ou (2.21). Il s'agit là plus d'"art", de "créativité" ou d'"expérience" que de théorie.
2. L'autre pilier est une bonne compréhension des schémas itératifs, et ceci est le domaine du PPA.

Ce formalisme a été introduit dans [5, 6] dans le contexte de la programmation convexe et différentiable dans des espaces de Hilbert (de dimension éventuellement infinie), puis a été étendu à d'autres contextes : problèmes non différentiables [26, 12, 7], optimisation stochastique [13, 10], inéquations variationnelles [8, 21, 16, 23].

Il n'est bien sûr pas question de donner un exposé complet de tous ces travaux dans le cadre de ce cours. On se contentera d'exposer les idées essentielles du PPA dans le contexte le plus simple : programmation convexe différentiable, avec parfois un aperçu des autres développements.

3.2 Le Principe du Problème Auxiliaire en optimisation sur un ensemble admissible

3.2.1 Cadre général

Dans cette section, on considère un problème général du type suivant :

$$\min_{u \in U^{\text{ad}}} J(u) , \quad (3.1)$$

où U^{ad} est un sous-ensemble convexe fermé de l'espace de Hilbert \mathcal{U} , $J : \mathcal{U} \rightarrow \mathbb{R}$ est convexe, semi-continue inférieurement¹ (s.c.i.) et coercive sur U^{ad} , c'est-à-dire que

$$\lim J(u) = +\infty \text{ quand } \|u\| \rightarrow +\infty \text{ en restant dans } U^{\text{ad}}. \quad (3.2)$$

Dans ces conditions, il existe des solutions u^\sharp au problème (3.1).

Quand J est la somme de deux fonctions convexes s.c.i. \mathcal{J} et \mathfrak{J} et que \mathcal{J} est sous-différentiable, une caractérisation de ces solutions est fournie par l'inéquation variationnelle (avec bien sûr $u^\sharp \in U^{\text{ad}}$)

$$\exists r^\sharp \in \partial \mathcal{J}(u^\sharp) : \forall u \in U^{\text{ad}}, \langle r^\sharp, u - u^\sharp \rangle + \mathfrak{J}(u) - \mathfrak{J}(u^\sharp) \geq 0. \quad (3.3)$$

Cette inéquation se spécialise au cas où \mathcal{J} est Gâteaux-différentiable :

$$\forall u \in U^{\text{ad}}, \langle \mathcal{J}'(u^\sharp), u - u^\sharp \rangle + \mathfrak{J}(u) - \mathfrak{J}(u^\sharp) \geq 0. \quad (3.4)$$

Enfin, si on suppose que $\mathcal{J}(u)$ s'écrit sous la forme $f(\mathcal{G}(u))$ où \mathcal{G} est une fonction du même type que \mathcal{J} et où f est une application convexe *croissante* et différentiable de \mathbb{R} dans \mathbb{R} , on peut donner la forme suivante :

$$\forall u \in U^{\text{ad}}, f'(\mathcal{G}(u^\sharp)) \times (\mathcal{G}(u) - \mathcal{G}(u^\sharp)) + \mathfrak{J}(u) - \mathfrak{J}(u^\sharp) \geq 0. \quad (3.5)$$

Commentaires 3.1. Une fonction $J : \mathcal{U} \rightarrow \mathbb{R}$ est Gâteaux-différentiable en un point u si elle admet des dérivées directionnelles dans toutes les directions d et si ces dérivées dépendent linéairement et continûment de la direction. Une dérivée directionnelle en u dans la direction d est la limite, lorsqu'elle existe, du ratio $(J(u + \varepsilon d) - J(u))/\varepsilon$ lorsque $\varepsilon \rightarrow 0_+$. C'est un scalaire noté $DJ(u; d)$. La dérivée de Gâteaux, lorsqu'elle existe, est donc l'élément de \mathcal{U}^* , espace dual de \mathcal{U} , noté $J'(u)$, tel que $DJ(u; d) = \langle J'(u), d \rangle$. Lorsqu'on dira qu'une fonction est différentiable dans ce cours, on sous-entendra toujours que c'est au sens de Gâteaux. Une notion de différentiabilité plus forte (*Fréchet*) existe mais n'est pas la plus utile en optimisation.

Lorsqu'une fonction J convexe est seulement sous-différentiable, la dérivée directionnelle est donnée par la formule

$$DJ(u; d) = \max_{r \in \partial J(u)} \langle r, d \rangle.$$

¹Une fonction $J : \mathcal{U} \rightarrow \mathbb{R}$ est semi-continue inférieurement si $\liminf_{u \rightarrow u_0} J(u) \geq J(u_0)$. Autre caractérisation : l'épigraphe de J est fermé dans $\mathcal{U} \times \mathbb{R}$.

On parlera de “gradient” (d’une fonction différentiable J) noté $\nabla J(u)$, pour l’élément du primal mis en correspondance avec $J'(u)$ par le théorème de représentation de Riesz. Dans un Hilbert \mathcal{U} , ce théorème met en correspondance tout élément F de \mathcal{U}^* , c’est-à-dire les formes linéaires continues sur \mathcal{U} , avec un élément f de \mathcal{U} de telle sorte que $F(u) = \langle f, u \rangle$ pour tout u . En général, nous ne distinguerons pas un Hilbert de son dual — sauf dans les cas où cette distinction devient très importante — et nous écrirons par exemple l’algorithme du gradient indifféremment

$$u^{k+1} = u^k - \varepsilon \nabla J(u^k) \text{ ou } u^k - \varepsilon J'(u^k) . \quad (3.6)$$

Lorsqu’il sera question de décomposition, on supposera toujours que \mathcal{U} et U^{ad} sont décomposés comme suit :

$$\mathcal{U} = \mathcal{U}_1 \times \cdots \times \mathcal{U}_N ; \quad U^{\text{ad}} = U_1^{\text{ad}} \times \cdots \times U_N^{\text{ad}} ; \quad U_i^{\text{ad}} \subset \mathcal{U}_i , \quad i = 1, \dots, N, \quad (3.7)$$

où \mathcal{U}_i est encore un espace de Hilbert et U_i^{ad} est un sous-ensemble convexe fermé dans \mathcal{U}_i . C’est-à-dire que les contraintes implicites $u \in U^{\text{ad}}$ sont une collection de N contraintes indépendantes (on dira souvent “locales” à chaque sous-problème). Le couplage par les contraintes sera discuté ultérieurement.

De plus, la motivation pour considérer J comme une somme $\mathcal{J} + \mathfrak{J}$ est la suivante :

- \mathfrak{J} sera supposée additive par rapport à la décomposition (3.7) ;
- par conséquent, seule \mathcal{J} sera la source d’un couplage entre sous-problèmes.

3.2.2 Idée et algorithme de base

On considère maintenant une *fonction auxiliaire* K de la même nature que J , mais supposée de plus différentiable. On considère de plus un nombre positif ε , un élément v quelconque dans \mathcal{U} , un sous-gradient $r \in \partial \mathcal{J}(v)$ et enfin la fonction

$$G_{v,r,K,\varepsilon}(u) = K(u) + \langle \varepsilon r - K'(v), u \rangle + \varepsilon \mathfrak{J}(u) .$$

Lemme 3.2. *Soit \hat{u} une solution de*

$$\min_{u \in U^{\text{ad}}} G_{v,r,K,\varepsilon}(u) . \quad (3.8a)$$

Si $\hat{u} = v$, alors v est une solution de (3.1).

Démonstration. Il suffit d’écrire l’inéquation variationnelle du type (3.3) pour le problème (3.8a) et d’utiliser l’hypothèse, ce qui conduit à la même inéquation variationnelle pour le problème (3.1). \square

Ce lemme suggère un algorithme de point fixe dans lequel on peut se permettre a priori de changer la fonction auxiliaire K et le scalaire ε à chaque itération : on considère donc des suites $\{K^k, \varepsilon^k\}_{k \in \mathbb{N}}$.

Algorithme 3.3.

1. Commencer avec $v = u^0$; poser $k = 0$.
2. A l’itération $k + 1$, résoudre

$$\min_{u \in U^{\text{ad}}} G_{u^k, r^k, K^k, \varepsilon^k}(u) \quad (r^k \in \partial \mathcal{J}(u^k)) ; \quad (3.8b)$$

$$\text{soit } (K^k(u) + \langle \varepsilon^k r^k - (K^k)'(u^k), u \rangle + \varepsilon^k \mathfrak{J}(u)) ; \quad (3.8c)$$

soit u^{k+1} une solution.

3. Si $\|u^{k+1} - u^k\|$ est “suffisamment petite”, stop ; sinon, retourner en 2 avec k incrémenté de 1.

La convergence d’un tel algorithme sera étudiée plus loin dans le cas où \mathcal{J} est différentiable. Dans [26, 12, 7], le cas sous-différentiable est étudié, mais on montre que dans ce cas certaines conditions sur la “vitesse de variation” des fonctions auxiliaires K^k doit être respectée, condition qui n’apparaît pas dans le cas différentiable (pour lequel $r^k = \mathcal{J}'(u^k)$).

3.2.3 Utilisations

Avant d’aborder la question de la convergence de l’Algorithme 3.3, examinons les utilisations que l’on peut en faire.

3.2.3.1 Algorithme “prox”

Dans les vingt dernières années, de nombreux développements ont été consacrés, dans la littérature sur la programmation convexe, à une technique connue sous le nom d’*algorithme proximal* (ou “prox” pour les intimes). Cette technique peut être plongée dans la classe générale de l’Algorithme 3.3 en supposant que \mathcal{J} est absent (c’est-à-dire que J est réduit à \mathfrak{J}), et en choisissant $K = (1/2) \|\cdot\|^2$. À l’itération $k + 1$, le problème (3.8c) devient

$$\min_{u \in U^{\text{ad}}} \left(J(u) + \frac{1}{2\varepsilon^k} \|u - u^k\|^2 \right). \quad (3.9)$$

Notre but n’est pas ici de discuter de cet algorithme en particulier, mais juste de mentionner son lien avec le PPA.

3.2.3.2 Algorithme du gradient et du gradient projeté

On garde ici le même choix de la fonction auxiliaire K que ci-dessus, mais on suppose maintenant que J se réduit à \mathcal{J} . De plus, on va supposer dans un premier temps que U^{ad} est égal à tout l’espace \mathcal{U} et que J est différentiable. Alors, (3.8c) peut être résolu explicitement, ce qui conduit à la formule (3.6) de l’algorithme du gradient.

Supposons maintenant que U^{ad} est un sous-ensemble convexe fermé de \mathcal{U} . Le problème (3.8c) conduit alors à la formule de l’algorithme du gradient projeté

$$u^{k+1} = \text{proj}_{U^{\text{ad}}} (u^k - \varepsilon^k \nabla J(u^k)). \quad (3.10)$$

Exercice 3.4. Démontrer cette formule (démontrer que u^{k+1} est bien solution de (3.8c)).

Si $J = \mathcal{J}$ est seulement sous-différentiable, on retrouve un algorithme de sous-gradient (éventuellement projeté) avec des “pas” ε^k suivant la règle (2.35).

Finalement, si J est vue comme la somme de deux fonctions non triviales \mathcal{J} et \mathfrak{J} , toujours avec le même choix de K , on obtient un algorithme qui peut être considéré comme combinant l’idée du gradient ou du sous-gradient (éventuellement projeté) avec celle de la technique “prox”.

3.2.3.3 Algorithme de Newton

On suppose à nouveau que $U^{\text{ad}} = \mathcal{U}$, que J se réduit à \mathcal{J} , et que cette dernière fonction est deux fois continûment différentiable avec une dérivée seconde $J''(u)$ fortement monotone (disons, pour l’instant, définie positive en dimension finie). À l’itération $k + 1$ de l’Algorithme 3.3, on choisit

$$K^k(u) = \frac{1}{2} \langle u, J''(u^k) \cdot u \rangle, \quad (3.11)$$

et $\varepsilon^k = 1$. Alors, (3.8c) peut à nouveau être résolu explicitement par la formule

$$u^{k+1} = u^k - (J''(u^k))^{-1} \cdot \nabla J(u^k) ,$$

qui n'est rien d'autre que l'algorithme de Newton (ou Newton-Raphson).

Si U^{ad} est différent de \mathcal{U} , on ne peut plus exprimer explicitement la solution de (3.8c) (il faut faire appel à une caractérisation par une inéquation variationnelle de type (3.4)), mais, avec le même choix de K^k et de ε^k , l'algorithme ainsi obtenu peut être considéré comme la généralisation de l'algorithme de Newton en présence de contraintes.

Cette utilisation de l'Algorithme 3.3 montre l'intérêt de permettre que la fonction auxiliaire K dépende de l'indice k d'itération. De plus, la souplesse d'utilisation de cet algorithme général est démontrée par la remarque suivante.

Remarque 3.5. Supposons que $J''(u)$ n'est pas partout *fortement* monotone,² et donc non inversible ; alors, le choix (3.11) peut être amélioré en ajoutant à cette fonction auxiliaire un terme fortement convexe, par exemple $(\gamma/2) \|u\|^2$ avec γ positif.

Cette technique pour améliorer le conditionnement du problème auxiliaire (3.8c) est aujourd'hui assez bien connue mais elle se présente comme une possibilité naturelle dans le cadre du PPA qui, comme on le verra lors de l'étude de convergence, réclame une fonction auxiliaire fortement convexe.

3.2.3.4 Décomposition

Nous en venons maintenant à notre but principal, celui de la décomposition. Comme déjà dit plus haut, dans ce contexte, on suppose que \mathcal{U} et U^{ad} sont décomposés comme en (3.7) et que \mathfrak{J} a une forme additive vis-à-vis de cette décomposition, c'est-à-dire qu'elle s'écrit $\mathfrak{J}(u) = \sum_{i=1}^N \mathfrak{J}_i(u_i)$. Par conséquent, le seul couplage entre sous-problèmes provient éventuellement de la fonction \mathcal{J} .

On observe alors que, dans (3.8c), cette fonction \mathcal{J} apparaît seulement par l'entremise d'un terme "linéarisé" en u , et qui est donc additif. *Il suffit donc de choisir la fonction auxiliaire K^k additive pour obtenir un problème auxiliaire (3.8c) qui se décompose en N sous-problèmes indépendants.*

Remarquons que le choix $K(u) = \|u\|^2/2$, conduisant à l'algorithme du gradient ou du sous-gradient projeté, est déjà un choix de fonction auxiliaire additive, ce qui nous autorise à considérer les algorithmes de gradient comme des algorithmes de décomposition. Cependant, la tâche de résolution des sous-problèmes ainsi obtenus est plutôt simpliste (du moins, si la projection sur les ensembles admissibles U_i^{ad} est facile à calculer) puisqu'elle consiste à appliquer la formule (3.10). En contrepartie, on peut s'attendre à ce que le nombre d'itérations de coordination soit important (ce qui arrive en effet avec la méthode du gradient pour des problèmes mal conditionnés).

Le choix (3.11) conduit à un problème auxiliaire comparativement moins trivial à résoudre, et le nombre d'itérations de coordination doit en conséquence s'en trouver allégé. Cependant, (3.11) n'est pas une fonction additive en général : pour obtenir la décomposition, on doit par exemple conserver seulement la partie bloc-diagonale du Hessien $J''(u^k)$ (correspondant à la décomposition (3.7) des espaces) ; on obtient ainsi ce que l'on peut qualifier d'"algorithme de Newton décomposé" (plus connu sous le nom générique de "quasi-Newton").

D'une manière générale, il est intuitivement souhaitable de conserver autant que faire se peut du problème original dans le problème auxiliaire, avec la contrainte d'obtenir un problème auxiliaire décomposable. Quand \mathcal{J} est différentiable, un procédé systématique consiste à définir

$$K_i^k : u_i \mapsto \mathcal{J}(u_1^k, \dots, u_{i-1}^k, u_i, u_i^k, \dots, u_N^k) , \quad (3.12)$$

sous réserve que cette fonction auxiliaire soit fortement convexe (sinon, on est conduit à lui rajouter un terme du type $(\gamma/2) \|u_i\|^2$). On verra plus loin d'autres conséquences de cette technique.

²Voir Annexe, §3.4.

3.2.4 Étude de la convergence

3.2.4.1 Schéma général

L'étude de la convergence des algorithmes décrits dans ce cours passe toujours par le schéma suivant :

1. Trouver une “fonction de Lyapounov” (notons là ℓ), c'est-à-dire une fonction des variables sur lesquelles itère l'algorithme (notons les globalement x), telle que ℓ est bornée inférieurement (en général par 0) et telle que $\ell(x) \rightarrow \infty$ quand $x \rightarrow \infty$.
2. Étudier la variation de cette fonction sur une itération de l'algorithme et montrer que cette variation

$$\Delta_k^{k+1} = \ell(x^{k+1}) - \ell(x^k)$$

est négative c'est-à-dire que la fonction de Lyapounov décroît.

3. On en déduit d'abord que la suite $\{x^k\}$ est bornée (sinon $\ell(x^k)$ tendrait vers l'infini, ce qui est contradictoire avec la décroissance de ces valeurs avec k).
4. Comme elle est bornée inférieurement et décroissante, la suite $\{\ell(x^k)\}$ converge donc (pas nécessairement vers sa borne inférieure) et la différence Δ_k^{k+1} tend vers 0. De ce fait, et si Δ_k^{k+1} peut être considérée comme l'opposée d'une fonction fortement monotone de variables y^k (recoupant totalement ou partiellement les variables x^k), on déduit que les variables y^k tendent vers 0. On essaie alors d'en tirer des conclusions quant à la convergence de l'algorithme.

La possibilité d'étudier la *vitesse de convergence* passe en général par le fait que Δ_k^{k+1} s'exprime directement comme une fonction strictement décroissante de $\ell(x^k)$ (ou $\ell(x^{k+1})$), et que par ailleurs $\ell(x)$ soit encadrée par deux fonctions monotones de $\|x\|$.

3.2.4.2 Convergence de l'Algorithme 3.3

On étudie l'Algorithme 3.3 dans le cas où \mathcal{J} est différentiable. On dira ultérieurement quelques mots du cas non différentiable.

On dispose en fait de deux démonstrations qui diffèrent par le choix de la fonction de Lyapounov. Celle donnée dans [6] utilise la fonction coût elle-même, c'est-à-dire $\mathcal{J} + \mathfrak{J}$, comme fonction de Lyapounov. Elle a l'intérêt de montrer que l'Algorithme 3.3 fait décroître cette fonction coût de façon monotone sous les hypothèses adéquates. Par ailleurs, elle est aussi facile à utiliser avec des fonctions auxiliaires K^k et des “pas” ε^k variables avec k qu'avec des fonctions K et des “pas” ε fixes, ce qui n'est pas le cas pour la démonstration ci-après.

Celle que nous allons présenter utilise une autre fonction, qui est en fait le plus souvent utilisée comme — partie, au moins, de la — fonction de Lyapounov de la plupart des autres démonstrations de convergence. Ce peut être une première raison pour la préférer ici. Mais, il y a d'autres raisons de la privilégier, malgré ses limitations évoquées ci-dessus.

1. Elle est adaptable sans aucune difficulté au cas d'opérateurs non symétriques, c'est-à-dire lorsqu'on cherche à résoudre une inéquation variationnelle du type (3.75) avec une adaptation triviale de l'Algorithme 3.3 au cas où l'opérateur Ψ n'est pas la dérivée d'une fonction (voir [8] pour l'adaptation et [21] pour l'étude de convergence dans le bon cadre d'hypothèses).
2. Elle se prête à l'étude de la vitesse de convergence (sous les hypothèses adéquates) selon la technique utilisée dans [23].

En fait, cette technique de démonstration est devenue possible après la prise de conscience de l'importance de la *propriété de Dunn*³ pour la convergence des algorithmes (voir [21]) d'une part, du fait que l'hypothèse de Lipschitz et l'hypothèse de Dunn sont équivalentes pour les opérateurs symétriques d'autre part (voir Lemme 3.44 en fin de chapitre).

On va énoncer le théorème de convergence “maximal”, c'est-à-dire celui qui résulte de la réunion des conclusions que l'on peut tirer des deux démonstrations (celle de [6, Theorem 2.1] et celle qui va suivre).

Théorème 3.6. *Les hypothèses sont les suivantes :*

- (H1) \mathfrak{J} est une fonction convexe s.c.i. ;
- (H2) \mathcal{J} est une fonction convexe s.c.i. différentiable, de gradient Lipschitzien avec constante A ;
- (H3) le problème (3.1), avec $J = \mathcal{J} + \mathfrak{J}$, admet une solution u^\sharp ;⁴
- (H4) les fonctions K^k sont convexes s.c.i. différentiables, de dérivées uniformément Lipschitziennes et uniformément fortement monotones, c'est-à-dire que si b^k (respectivement, B^k) désigne la constante de forte monotonie (respectivement, de Lipschitz) de $(K^k)'$, alors

$$\exists b, B > 0 : \forall k \in \mathbb{N}, \quad b^k \geq b, \quad B^k \leq B .$$

Alors, les conclusions sont les suivantes :

- (C1) le problème auxiliaire (3.8c) admet une solution unique u^{k+1} ;
- (H5) si

$$\alpha \leq \varepsilon^k \leq \frac{2b^k}{A + \beta} \text{ pour } \alpha, \beta > 0 , \quad (3.13)$$
- (C2) alors $J(u^k)$ décroît strictement (sauf si l'algorithme a convergé en un nombre fini d'itérations) ;
- (C3) $\mathcal{J}'(u^k)$ converge fortement vers la valeur $\mathcal{J}'(u^\sharp)$ (unique même si u^\sharp ne l'est pas — voir Lemme 3.47) ;
- (C4) la suite $\{u^k\}$ est bornée, donc faiblement compacte, et tout point d'accumulation dans la topologie faible est une solution de (3.1) ;
- (H6) si \mathcal{J}' est fortement monotone de constante a sur un borné contenant la suite $\{u^k\}$,
- (C5) la suite $\{u^k\}$ converge fortement vers l'unique solution u^\sharp ;
- (C6) de plus, on a la “majoration a posteriori de l'erreur”

$$\|u^{k+1} - u^\sharp\| \leq \frac{B^k/\varepsilon^k + A}{a} \|u^{k+1} - u^k\| . \quad (3.14)$$

³Voir Annexe, §3.6.

⁴Dans [6], on supposait que J était coercive sur U^{ad} (voir (3.2)), cette hypothèse étant requise en particulier parce que J servait de fonction de Lyapounov dans la démonstration. Ici cette raison n'est plus à invoquer, mais la coercivité, avec les hypothèses précédentes, assure l'existence d'une solution à (3.1).

Démonstration. Comme déjà dit, on se limite ici au cas où les fonctions auxiliaires K^k sont toutes identiques (on notera donc K cette fonction auxiliaire, et b (respectivement, B), la constante de forte monotonie (respectivement, de Lipschitz) de K' . On se limitera alors également à des ε^k tous égaux à un ε fixe. Ces restrictions sont levées dans la démonstration donnée dans [6]. Dans le contexte actuel, l'hypothèse (3.13) prend la forme simplifiée

$$0 < \varepsilon < \frac{2b}{A} . \quad (3.15)$$

(C1) La fonction coût du problème auxiliaire est la somme d'une fonction K s.c.i. et fortement convexe et de fonctions s.c.i. convexes : elle est donc s.c.i. et fortement convexe et satisfait en particulier (3.68), ce qui implique qu'elle est coercive. Le problème (3.8c) admet donc une solution, évidemment unique, caractérisée par l'inéquation variationnelle : $u^{k+1} \in U^{\text{ad}}$,

$$\forall u \in U^{\text{ad}}, \quad \langle K'(u^{k+1}) - K'(u^k) + \varepsilon \mathcal{J}'(u^k), u - u^{k+1} \rangle + \varepsilon (\mathfrak{J}(u) - \mathfrak{J}(u^{k+1})) \geq 0 . \quad (3.16)$$

On se livre maintenant à quelques calculs. On additionne les inéquations variationnelles (3.16) et (3.4) (multipliée par ε) avec respectivement $u = u^\sharp$ et $u = u^{k+1}$, ce qui donne

$$\langle K'(u^{k+1}) - K'(u^k), u^\sharp - u^{k+1} \rangle + \varepsilon \langle \mathcal{J}'(u^k) - \mathcal{J}'(u^\sharp), u^\sharp - u^{k+1} \rangle \geq 0 . \quad (3.17)$$

Choix de la fonction de Lyapounov. On pose

$$\phi^k = \ell(u^k) = K(u^\sharp) - K(u^k) - \langle K'(u^k), u^\sharp - u^k \rangle , \quad (3.18)$$

où u^\sharp est une solution quelconque de (3.1). Comme K est fortement convexe de constante b , d'après (3.68), on a

$$\phi^k \geq \frac{b}{2} \|u^k - u^\sharp\|^2 , \quad (3.19)$$

ce qui montre que ℓ remplit bien les critères d'une fonction de Lyapounov (voir point 1 du §3.2.4.1).

D'une part,

$$\begin{aligned} \langle K'(u^{k+1}) - K'(u^k), u^\sharp - u^{k+1} \rangle &= \langle K'(u^{k+1}), u^\sharp - u^{k+1} \rangle \\ &\quad - \langle K'(u^k), (u^\sharp - u^k) + (u^k - u^{k+1}) \rangle . \end{aligned}$$

Comme K est fortement convexe de constante b , d'après (3.68)

$$-\langle K'(u^k), u^k - u^{k+1} \rangle \leq K(u^{k+1}) - K(u^k) - \frac{b}{2} \|u^{k+1} - u^k\|^2 .$$

Alors

$$\langle K'(u^{k+1}) - K'(u^k), u^\sharp - u^{k+1} \rangle \leq \phi^k - \phi^{k+1} - \frac{b}{2} \|u^{k+1} - u^k\|^2 . \quad (3.20)$$

D'autre part,

$$\begin{aligned} \varepsilon \langle \mathcal{J}'(u^k) - \mathcal{J}'(u^\sharp), u^\sharp - u^{k+1} \rangle &= \varepsilon \langle \mathcal{J}'(u^k) - \mathcal{J}'(u^\sharp), (u^\sharp - u^k) + (u^k - u^{k+1}) \rangle \\ &\leq -\varepsilon \langle \mathcal{J}'(u^k) - \mathcal{J}'(u^\sharp), u^k - u^\sharp \rangle \\ &\quad + \varepsilon \|\mathcal{J}'(u^k) - \mathcal{J}'(u^\sharp)\| \|u^{k+1} - u^k\| \end{aligned}$$

par utilisation de l'inégalité de Schwarz,

$$\begin{aligned} &\leq -\varepsilon \langle \mathcal{J}'(u^k) - \mathcal{J}'(u^\sharp), u^k - u^\sharp \rangle \\ &\quad + \frac{\varepsilon}{2\delta} \|\mathcal{J}'(u^k) - \mathcal{J}'(u^\sharp)\|^2 + \frac{\varepsilon\delta}{2} \|u^{k+1} - u^k\|^2 \end{aligned}$$

pour tout $\delta > 0$,

$$\leq \varepsilon \left(\frac{A}{2\delta} - 1 \right) \langle \mathcal{J}'(u^k) - \mathcal{J}'(u^\sharp), u^k - u^\sharp \rangle + \frac{\varepsilon\delta}{2} \|u^{k+1} - u^k\|^2$$

par utilisation de (3.74) pour $\Psi = \mathcal{J}'$.

En additionnant membre à membre les majorations obtenues et en tenant compte de (3.17), il vient

$$\phi^{k+1} \leq \phi^k + \frac{\varepsilon\delta - b}{2} \|u^{k+1} - u^k\|^2 + \varepsilon \left(\frac{A}{2\delta} - 1 \right) \langle \mathcal{J}'(u^k) - \mathcal{J}'(u^\sharp), u^k - u^\sharp \rangle \quad (3.21)$$

$$\leq \phi^k + \frac{\varepsilon\delta - b}{2} \|u^{k+1} - u^k\|^2 + \varepsilon \left(\frac{1}{2\delta} - \frac{1}{A} \right) \|\mathcal{J}'(u^k) - \mathcal{J}'(u^\sharp)\|^2 \quad (3.22)$$

en utilisant à nouveau (3.74) à condition que $\delta > A/2$.

- (C3) Si de plus $\delta < b/\varepsilon$ — notons que cette hypothèse et la précédente sont bien compatibles avec (3.15) —, l'inégalité (3.22) montre que la suite $\{\phi^k\}$ est décroissante. On est donc dans le cadre de l'application du schéma général du §3.2.4.1 (points 2 à 4). En particulier, $\phi^{k+1} - \phi^k$ tend vers zéro, et d'après (3.22), ceci montre que

$$\|u^{k+1} - u^k\| \rightarrow 0 ; \quad \|\mathcal{J}'(u^k) - \mathcal{J}'(u^\sharp)\| \rightarrow 0 . \quad (3.23)$$

Ceci démontre la conclusion (C3).⁵

- (C4) Par le point 3 du schéma général (§3.2.4.1), on conclut que la suite $\{u^k\}$ est bornée, donc faiblement compacte. Par conséquent, on peut en extraire une sous-suite faiblement convergente. Soit \bar{u} un point d'accumulation (dans la topologie faible) de la suite $\{u^k\}$. D'après l'inéquation variationnelle (3.16) et le fait que K' et \mathcal{J}' sont Lipschitziennes, pour tout $u \in U^{\text{ad}}$,

$$\langle \mathcal{J}'(u^{k+1}), u - u^{k+1} \rangle + \mathfrak{J}(u) - \mathfrak{J}(u^{k+1}) \quad (3.24)$$

$$\begin{aligned} &\geq \left\langle \frac{1}{\varepsilon} (K'(u^k) - K'(u^{k+1})) + \mathcal{J}'(u^{k+1}) - \mathcal{J}'(u^k), u - u^{k+1} \right\rangle \\ &\geq - \left(\frac{B}{\varepsilon} + A \right) \|u^{k+1} - u^k\| \|u - u^{k+1}\| . \end{aligned} \quad (3.25)$$

Par passage à la limite pour une sous-suite $\{u^{k_i}\}$ telle que $u^{k_i} \rightharpoonup \bar{u}$, du fait que \mathfrak{J} est s.c.i., y compris dans la topologie faible (car elle est convexe),⁶ avec l'aide de (3.23) et le fait que u^{k_i} est borné, on obtient

$$\langle \mathcal{J}'(u^\sharp), u - \bar{u} \rangle + \mathfrak{J}(u) - \mathfrak{J}(\bar{u}) \geq 0 , \quad \forall u \in U^{\text{ad}} .$$

On pourra conclure que \bar{u} est une solution en montrant que $\mathcal{J}'(u^\sharp) = \mathcal{J}'(\bar{u})$, ce que nous faisons maintenant.

⁵L'affirmation (C2) résulte de la démonstration de [6].

⁶donc $-\mathfrak{J}(\bar{u}) \geq -\liminf_{k_i \rightarrow \infty} \mathfrak{J}(u^{k_i}) \geq -\limsup_{k_i \rightarrow \infty} \mathfrak{J}(u^{k_i}) = \liminf_{k_i \rightarrow \infty} -\mathfrak{J}(u^{k_i})$

Le fait que \mathcal{J}' est Lipschitzienne entraîne (voir (3.74) en annexe) que

$$\langle \mathcal{J}'(u^k), u^k - \bar{u} \rangle - \langle \mathcal{J}'(\bar{u}), u^k - \bar{u} \rangle \geq \frac{1}{A} \|\mathcal{J}'(u^k) - \mathcal{J}'(\bar{u})\|^2.$$

Par passage à la limite sur la sous-suite $\{u^{k_i}\}$, on vérifie que les deux termes au premier membre de cette inégalité tendent vers zéro. Comme au second membre $\mathcal{J}'(u^k)$ tend vers $\mathcal{J}'(u^\sharp)$ fortement, l'inégalité limite montre que $\mathcal{J}'(u^\sharp) = \mathcal{J}'(\bar{u})$, ce qui termine la démonstration du point (C4).

(C5) La convergence forte résultera du point (C6) ci-après et de (3.23).

(C6) On repart de l'inégalité (3.17) que l'on réécrit

$$\begin{aligned} & \langle K'(u^{k+1}) - K'(u^k), u^\sharp - u^{k+1} \rangle \\ & + \varepsilon \langle \mathcal{J}'(u^k) - \mathcal{J}'(u^{k+1}), u^\sharp - u^{k+1} \rangle \geq \varepsilon \langle \mathcal{J}'(u^\sharp) - \mathcal{J}'(u^{k+1}), u^\sharp - u^{k+1} \rangle. \end{aligned}$$

L'hypothèse de forte monotonie de \mathcal{J}' pour le second membre, et les inégalités de Schwarz et Lipschitz au premier membre conduisent alors à (3.14) (avec ici ε et B indépendants de k) après simplification par $\|u^{k+1} - u^\sharp\|$. \square

Commentaires 3.7.

- (a) L'inégalité (3.14) s'appelle une "majoration a posteriori de l'erreur" parce qu'elle permet, moyennant toutefois une estimation des diverses constantes présentes au second membre, d'estimer à combien on se trouve de l'optimum u^\sharp à partir d'une grandeur "mesurable en ligne", c'est-à-dire au moment du déroulement de l'algorithme, à savoir la distance $\|u^{k+1} - u^k\|$.
- (b) Pour finir la démonstration de (C4), on aurait pu, à partir de (3.25), utiliser une inégalité de la convexité de \mathcal{J} au premier membre, puis le caractère s.c.i. de $\mathcal{J} + \mathfrak{J}$. Mais sous la forme de la démonstration adoptée, on peut vérifier que *celle-ci reste valable sans changement* lorsque \mathcal{J}' est remplacé par un opérateur Ψ monotone *non symétrique* ayant la propriété de Dunn (3.74), *et pas seulement la propriété de Lipschitz*. L'algorithme et sa démonstration restent donc valables pour l'inéquation variationnelle (3.75) dans ce cas. Cette démonstration a été donnée précisément dans ce contexte dans [21].
- (c) L'algorithme du problème auxiliaire appliqué à l'opérateur linéaire non symétrique de l'Exemple 3.46 (voir Annexe, §3.6) avec $K(\cdot) = \|\cdot\|^2/2$ donne

$$u_1^{k+1} = u_1^k - \varepsilon u_2^k; \quad u_2^{k+1} = u_2^k + \varepsilon u_1^k,$$

de sorte que

$$\|u_1^{k+1}\|^2 + \|u_2^{k+1}\|^2 = (1 + \varepsilon^2) (\|u_1^k\|^2 + \|u_2^k\|^2),$$

ce qui signifie que la norme de u^k ne fait que croître, et donc l'algorithme ne converge pas. Cet exemple montre donc que le caractère Lipschitzien d'un opérateur non symétrique est insuffisant pour garantir la convergence et que c'est donc bien la propriété de Dunn qui est décisive dans le cas non symétrique.

- (d) Dans le cas d'un opérateur non symétrique Ψ Lipschitzien de constante A et fortement monotone de constante a , le Lemme 3.44 (Annexe, §3.6) indique que Ψ vérifie aussi l'hypothèse de Dunn avec la constante A^2/a . L'algorithme converge donc si ε reste inférieur à la valeur $2ab/A^2$, résultat déjà obtenu dans [8].

- (e) L'hypothèse de Lipschitz sur \mathcal{J}' est *globale*, ce qui implique en particulier que \mathcal{J} ne peut avoir une croissance *plus forte* que quadratique à l'infini (voir (3.71) en Annexe, §3.5 ; considération qui n'a évidemment d'intérêt que si U^{ad} n'est pas borné). Comme on montre ensuite que la suite $\{u^k\}$ engendrée par l'algorithme est bornée, on peut se demander si une hypothèse de Lipschitz sur tout borné (on dira que \mathcal{J}' est "localement" Lipschitzienne) ne serait pas suffisante. Malheureusement, il semble difficile d'aboutir à la conclusion d'une suite bornée (qui résulte elle-même du fait que la fonction de Lyapounov décroît) sans faire *préalablement* l'hypothèse de Lipschitz *globale*.

3.2.4.3 Un cas de convergence de toute la suite

Grâce à la nouvelle démonstration de convergence ci-dessus, il devient possible d'utiliser un résultat de [7, Lemme 2 de l'Annexe] pour montrer que toute la suite $\{u^k\}$ a le même point d'accumulation (dans la topologie faible), même en l'absence de forte monotonie de \mathcal{J}' : on renonce donc ici à l'hypothèse (H6)), mais on fait la nouvelle hypothèse :

(H6') K' est continue de \mathcal{U} muni de la topologie faible dans \mathcal{U}^* muni de la topologie faible.

Cette hypothèse est satisfaite en particulier si K' est affine (i.e. si K est quadratique).

Lemme 3.8. *Sous l'hypothèse supplémentaire (H6') (à la place de (H6)), la suite $\{u^k\}$ engendrée par l'Algorithme 3.3 a un seul point d'accumulation (dans la topologie faible).*

Démonstration. Supposons qu'il y ait deux points d'accumulation $u_1^\#$ et $u_2^\#$ (qui sont nécessairement tous les deux des solutions de (3.1)). On sait d'après la démonstration précédente que la suite $\{\phi^k\}$ définie par (3.18) converge vers une limite λ . En fait, ϕ^k peut être définie à partir de $u_1^\#$ comme à partir de $u_2^\#$. Nous noterons ϕ_1^k et ϕ_2^k ces deux définitions possibles, et λ_1 , respectivement λ_2 , les limites correspondantes. Considérons une sous-suite $\{u^{k_i}\}$ convergeant faiblement vers $u_1^\#$; on a

$$\begin{aligned} \phi_2^{k_i} &= K(u_2^\#) - K(u^{k_i}) - \left\langle K'(u^{k_i}), u_2^\# - u^{k_i} \right\rangle \\ &= K(u_2^\#) - K(u_1^\#) + \underbrace{K(u_1^\#) - K(u^{k_i}) - \left\langle K'(u^{k_i}), u_1^\# - u^{k_i} \right\rangle}_{\phi_1^{k_i}} - \left\langle K'(u^{k_i}), u_2^\# - u_1^\# \right\rangle. \end{aligned}$$

En passant à la limite, compte tenu des hypothèses faites, et notamment de l'hypothèse de continuité de K' , on obtient

$$\begin{aligned} \lambda_2 &= K(u_2^\#) - K(u_1^\#) + \lambda_1 - \left\langle K'(u_1^\#), u_2^\# - u_1^\# \right\rangle \\ &\geq \lambda_1 + \frac{b}{2} \|u_2^\# - u_1^\#\|^2. \end{aligned}$$

En intervertissant le rôle des indices 1 et 2, on obtient de façon similaire l'inégalité

$$\lambda_1 \geq \lambda_2 + \frac{b}{2} \|u_2^\# - u_1^\#\|^2.$$

Par addition de ces deux dernières inégalités, on prouve que $u_1^\# = u_2^\#$. □

3.2.4.4 Vitesse de convergence dans le cas \mathcal{J} fortement convexe

On définit d'abord la notion de convergence linéaire en moyenne.

Définition 3.9. La suite $\{x^k\}$ converge en moyenne linéairement vers x^\sharp avec le taux $\kappa \in [0, 1[$ si

$$\forall k \in \mathbb{N} : \limsup_{n \rightarrow +\infty} \sqrt[n]{\frac{\|x^{k+n} - x^\sharp\|}{\|x^k - x^\sharp\|}} \leq \kappa .$$

Le lemme suivant est tiré de [23].

Lemme 3.10. Soit $\ell : \mathcal{X} \rightarrow \mathbb{R}$ et x^\sharp un point tel que

$$\exists m > 0, \exists M > 0 : \forall x \in \mathcal{X}, \quad \frac{m}{2} \|x - x^\sharp\|^2 \leq \ell(x) - \ell(x^\sharp) \leq \frac{M}{2} \|x - x^\sharp\|^2 .$$

Soit $\{x^k\}$ une suite telle que

$$\exists \eta \in]0, 1[: \forall k, \quad \ell(x^{k+1}) - \ell(x^\sharp) \leq \eta (\ell(x^k) - \ell(x^\sharp)) .$$

Alors, cette suite $\{x^k\}$ converge en moyenne linéairement vers x^\sharp avec le taux $\sqrt{\eta}$.

Démonstration. On a

$$\frac{m}{2} \|x^{k+n} - x^\sharp\|^2 \leq \ell(x^{k+n}) - \ell(x^\sharp) \leq \eta^n (\ell(x^k) - \ell(x^\sharp)) \leq \eta^n \frac{M}{2} \|x^k - x^\sharp\|^2 ,$$

d'où

$$\forall n, \quad \|x^{k+n} - x^\sharp\| \leq \sqrt{\eta^n \frac{M}{m}} \|x^k - x^\sharp\| .$$

Alors, pour tout $\beta > \eta$, il existe $l \in \mathbb{N}$ tel que, pour tout $n > l$, $(\beta/\eta)^n \geq M/m$, d'où

$$\|x^{k+n} - x^\sharp\| \leq \sqrt{\beta^n} \|x^k - x^\sharp\| ,$$

et donc

$$\limsup_{n \rightarrow +\infty} \sqrt[n]{\frac{\|x^{k+n} - x^\sharp\|}{\|x^k - x^\sharp\|}} \leq \sqrt{\beta} .$$

Ceci étant vrai pour tout $\beta > \eta$ est également vrai pour $\beta = \eta$. □

Théorème 3.11. Lorsque \mathcal{J}' est fortement monotone de constante a , l'Algorithme 3.3 converge linéairement en moyenne avec le taux $\sqrt{1 - (2\varepsilon a/B)(1 - \varepsilon A/2b)}$, le taux optimal estimé (obtenu pour $\varepsilon = b/A$) étant $\sqrt{1 - ab/AB}$.

Démonstration. On va appliquer le Lemme 3.10 à la situation du Théorème 3.6 en utilisant la fonction $\ell(u)$ définie par (3.18). On a posé $\phi^k = \ell(u^k)$ et on observe que $\ell(u^\sharp) = 0$ et que

$$\frac{b}{2} \|u - u^\sharp\|^2 \leq \ell(u) - \ell(u^\sharp) \leq \frac{B}{2} \|u - u^\sharp\|^2 .$$

On repart alors de l'inégalité (3.21) en choisissant $\delta = b/\varepsilon$, ce qui donne

$$\begin{aligned}\phi^{k+1} &\leq \phi^k + \varepsilon \left(\frac{\varepsilon A}{2b} - 1 \right) \langle \mathcal{J}'(u^k) - \mathcal{J}'(u^\sharp), u^k - u^\sharp \rangle \\ &\leq \phi^k + \varepsilon a \left(\frac{\varepsilon A}{2b} - 1 \right) \|u^k - u^\sharp\|^2 \\ &\leq \phi^k + \frac{2\varepsilon a}{B} \left(\frac{\varepsilon A}{2b} - 1 \right) \phi^k \\ &= \left(1 - \frac{2\varepsilon a}{B} \left(1 - \frac{\varepsilon A}{2b} \right) \right) \phi^k.\end{aligned}$$

À partir de là, il suffit d'une part d'appliquer le lemme précédent, d'autre part de montrer que le meilleur taux est atteint pour $\varepsilon = b/A$, c'est-à-dire au milieu de la plage de valeurs autorisées pour ε . \square

3.2.4.5 Quelques mots du cas sous-différentiable

Si \mathcal{J} n'est pas différentiable mais seulement sous-différentiable, cela signifie que la constante de Lipschitz A est essentiellement infinie, ce qui fait que la borne $2b/A$ sur ε est nulle : ceci "explique", ou indique, que la suite $\{\varepsilon^k\}$ doit tendre vers 0, et en fait elle devra satisfaire les conditions (2.35).

De plus, si U^{ad} n'est pas borné, une hypothèse sur le taux de croissance de la norme des sous-gradients de \mathcal{J} à l'infini est requise. On notera en effet que dans le cas d'un gradient Lipschitzien, la croissance de la norme du gradient à l'infini n'est pas plus rapide que linéaire. Dans le cas sous-différentiable, l'hypothèse requise s'écrit

$$\exists c_1, c_2 : \forall u \in U^{\text{ad}}, \forall r \in \partial \mathcal{J}(u), \|r\| \leq c_1 \|u\| + c_2.$$

Une hypothèse de même nature est aussi requise pour \mathfrak{J} (mais elle s'exprime directement sur \mathfrak{J} sans référence à $\partial \mathfrak{J}$ — voir [12] pour les détails).

De telles hypothèses peuvent en fait être évitées par une mise en œuvre différente (et plus coûteuse) de l'algorithme : en supposant que $\mathfrak{J} \equiv 0$ pour simplifier, on remplace ε^k par $\varepsilon^k = \rho^k / \|r^k\|$ dans (3.8c), où $\{\rho^k\}$ est à nouveau une suite satisfaisant (2.35). D'autres mises en œuvre sont également possibles (voir [12]).

Sous les hypothèses décrites sommairement ci-dessus, on obtient essentiellement les mêmes résultats de convergence que dans le cas différentiable, sauf la décroissance monotone de la fonction coût (d'ailleurs, cette fonction de Lyapounov ne peut plus servir de base à la démonstration), et sauf bien sûr la majoration a posteriori de l'erreur (3.14).

3.2.5 Variantes

Dans le cas différentiable, il y a plusieurs autres façons de mettre en œuvre l'algorithme de base et d'assurer les conditions de convergence. Nous en décrivons quelques unes ci-après.

3.2.5.1 Augmentation de la fonction auxiliaire

On a déjà évoqué aux §3.2.3.3 et 3.2.3.4 la possibilité de rajouter à une fonction auxiliaire K constituant un premier choix une autre fonction du type $(\gamma/2) \|\cdot\|^2$ pour $\gamma > 0$. On a vu (cf. (3.9)) que cela rajoute le terme "proximal" $(\gamma/2) \|u - u^k\|^2$ au problème auxiliaire.

À la lumière des conditions de convergence, l'addition de ce terme, qui fait passer la constante b de forte monotonie de K' à la valeur $b + \gamma$, peut faire que, pour γ assez grand, l'intervalle admissible $]0, 2(b + \gamma)/A[$ des valeurs de ε contienne la valeur 1. Autrement dit, l'addition au problème auxiliaire

d'un terme "prox" suffisamment "fort" permet de se passer de ε pour assurer la convergence (c'est-à-dire de prendre $\varepsilon = 1$).

Il est un cas en particulier où le fait de prendre $\varepsilon = 1$ est particulièrement intéressant. C'est le cas où l'on fait un choix "canonique" de la fonction auxiliaire en relation avec \mathcal{J} comme décrit par (3.12). Avec ce choix, et avec $\varepsilon = 1$, les "corrections linéaires" faisant intervenir les dérivées de \mathcal{J} disparaissent et le sous-problème i s'écrit

$$\min_{u \in U^{\text{ad}}} \sum_{i=1}^N \left(\frac{\gamma}{2} \|u_i - u_i^k\|^2 + \mathcal{J}(u_1^k, \dots, u_{i-1}^k, u_i, u_i^k, \dots, u_N^k) + \mathfrak{J}_i(u_i) \right), \quad (3.26)$$

ce que l'on pourrait appeler "algorithme proximal avec décomposition" (à comparer à l'algorithme proximal "pur" décrit au §3.2.3.1 — voir (3.9)).

Exercice 3.12. Avec le choix (3.12) pour la fonction auxiliaire K_i^k et $\varepsilon^k = 1$, montrer que l'on retrouve bien (3.26) à partir de (3.8c).

3.2.5.2 Relaxation

Une autre technique pour assurer la convergence de l'algorithme du problème auxiliaire tout en gardant $\varepsilon = 1$ consiste à recourir à la "sous-" ou "sur-relaxation". On désigne maintenant par $u^{k+1/2}$ la solution du problème auxiliaire (3.8c) avec $\varepsilon^k = 1$, puis on pose

$$u^{k+1} = (1 - \rho^k)u^k + \rho^k u^{k+1/2}, \quad (3.27)$$

qui servira à redéfinir (3.8c) à l'itération suivante. On parle de "sous-relaxation" si $\rho^k < 1$ et de "sur-relaxation" si $\rho^k > 1$.

Il s'avère (voir [6]) que les conditions sur ρ^k sous lesquelles ce nouvel algorithme converge sont exactement les mêmes que celles énoncées au Théorème 3.6 et portant sur ε^k . La seule petite différence est que, si \mathfrak{J} est non différentiable et si U^{ad} n'est pas égal à \mathcal{U} , alors ρ^k doit être tel que u^{k+1} , calculé par (3.27), reste dans U^{ad} (ce qui est de toute façon assuré si $\rho^k \leq 1$).⁷

3.2.5.3 Version séquentielle

Lorsque nous avons considéré l'utilisation de l'Algorithme 3.3 dans l'optique de la décomposition (ce qui passe par le choix de fonctions auxiliaires K^k additives), les sous-problèmes obtenus pouvaient être résolus en *parallèle*. Cette caractéristique est souhaitable si l'on dispose d'un ordinateur parallèle permettant de résoudre les sous-problèmes simultanément. Dans le cas contraire, on préférera une version *séquentielle* de l'Algorithme 3.3.

À l'itération de coordination $k + 1$, les sous-problèmes sont résolus l'un après l'autre, disons, en commençant par le sous-problème 1 et en terminant par le sous-problème N ; une fois les sous-problèmes 1 à $j-1$ résolus, u_i^{k+1} est disponible pour $i = 1, \dots, j-1$; ces solutions sont immédiatement utilisées pour définir le sous-problème j qui s'écrit

$$\min_{u_j \in U_j^{\text{ad}}} \left(K_j^k(u_j) + \left\langle \varepsilon_j^k \frac{\partial \mathcal{J}}{\partial u_j}(u_1^{k+1}, \dots, u_{j-1}^{k+1}, u_j^k, \dots, u_N^k) - (K_j^k)'(u_j^k), u_j \right\rangle + \varepsilon_j^k \mathfrak{J}_j(u_j) \right).$$

Notons qu'un ε_j^k différent est utilisé dans chaque sous-problème. La convergence est démontrée dans [6] sous les conditions $\varepsilon_j^k \in [\alpha_j, \beta_j] \subset]0, 2b_j/A_j]$, où b_j est la constante de forte convexité (uniforme en k) de K_j^k et, de même, A_j est la constante de Lipschitz uniforme de la fonction $u_j \mapsto \partial \mathcal{J}(u)/\partial u_j$.

⁷Cette restriction peut se comprendre en pensant au cas où \mathfrak{J} est justement la fonction indicatrice d'un certain U^{ad} (qui vaut 0 si $u \in U^{\text{ad}}$ et $+\infty$ sinon).

À l'évidence, les constantes b et A considérées dans la version parallèle sont respectivement inférieure et supérieure (ou égales) à $\min_{j=1,\dots,N} b_j$ et $\max_{j=1,\dots,N} A_j$. Ceci signifie que la borne supérieure sur ε^k de la version parallèle est *moins bonne* que les bornes supérieures individuelles des ε_j^k de la version séquentielle. Cette observation illustre indirectement le fait (vérifié en pratique) que la version séquentielle est plus rapide, en nombre d'itérations, que la version parallèle. Par conséquent, la version séquentielle sera préférée si l'on dispose d'un ordinateur mono-processeur. Cependant, la version parallèle peut être plus rapide, en temps de calcul, à partir du moment où l'on dispose d'un ordinateur présentant des possibilités substantielles de calcul parallèle (de type MIMD, comme indiqué au §1.2.2).

3.2.5.4 Linéarisation partielle

En définitive, la décomposition provient du fait que la partie non additive de la fonction coût, à savoir \mathcal{J} , est "linéarisée" dans la formulation du problème auxiliaire. Supposons maintenant que \mathcal{J} est de la forme

$$\mathcal{J}(u) = f\left(\sum_{i=1}^N J_i(u_i)\right),$$

où $J_i : \mathcal{U}_i \rightarrow \mathbb{R}$ est convexe pour tout i , et $f : \mathbb{R} \rightarrow \mathbb{R}$ est différentiable (ou sous-différentiable).⁸

Alors, en suivant le principe de "garder autant que faire se peut du problème original dans le problème auxiliaire", au lieu de (3.8c), on considère le problème auxiliaire suivant

$$\min_{u \in U^{\text{ad}}} \left(K^k(u) - \langle (K^k)'(u^k), u \rangle + \varepsilon^k f' \left(\sum_{i=1}^N J_i(u_i^k) \right) \times \sum_{i=1}^N J_i(u_i) + \varepsilon^k \mathcal{J}(u) \right). \quad (3.28)$$

Cette écriture s'inspire de la façon (3.5) d'écrire les conditions d'optimalité et d'une généralisation correspondante immédiate du Lemme 3.2.

3.2.6 Une application

3.2.6.1 Exposé du problème

On considère un problème du type (3.1), mais avec deux types de variables : d'une part u , elle-même décomposée en variables u_i , $i = 1, \dots, N$, chaque u_i devant appartenir à un sous-ensemble admissible U_i^{ad} , d'autre part v dans un espace \mathcal{V} . De plus, $J(u, v)$ prend la forme suivante :

$$J(u, v) = g(J_1(u_1, v), \dots, J_N(u_N, v)), \quad (3.29)$$

où $g : \mathbb{R}^N \rightarrow \mathbb{R}$ est une fonction croissante de chacun de ses arguments⁹ et chaque J_i est une fonction de $\mathcal{U}_i \times \mathcal{V}$ dans \mathbb{R} .

3.2.6.2 Approche "ad hoc"

Du fait que g est croissante par rapport à chacun de ses arguments, il est clair qu'à v fixé, la minimisation de J par rapport à $u \in U^{\text{ad}}$ se ramène à la résolution des sous-problèmes indépendants, pour $i = 1, \dots, N$,

$$\min_{u_i \in U_i^{\text{ad}}} J_i(u_i, v). \quad (3.30)$$

⁸ f doit aussi être convexe et non décroissante pour que \mathcal{J} soit convexe.

⁹ Ce peut être la somme, le produit si les J_i sont à valeurs positives, ou n'importe quelle combinaison à base de somme, produit, logarithme, exponentielle, etc.

La coordination reviendrait alors à minimiser en v la fonction

$$\varphi : v \mapsto \min_{u \in U^{\text{ad}}} J(u, v) .$$

Supposant g différentiable, ainsi que chaque J_i par rapport à v , et supposant aussi, pour rester simple, que la solution de (3.30) est unique pour tout i (on la notera $\widehat{u}_i(v)$), on sait que

$$\frac{d\varphi}{dv}(v) = \sum_{i=1}^N \frac{\partial g}{\partial x_i} \left(J_1(\widehat{u}_1(v), v), \dots, J_N(\widehat{u}_N(v), v) \right) \frac{\partial J_i}{\partial v}(\widehat{u}_i(v), v) . \quad (3.31)$$

L'algorithme de décomposition-coordination obtenu se décline finalement ainsi : à l'itération k ,

1. résoudre (3.30) pour $v = v^k$; soit u_i^{k+1} la solution pour $i = 1, \dots, N$;
2. remettre à jour v^k par la formule du gradient

$$v^{k+1} = v^k - \rho^k \varphi'(v^k) , \quad (3.32)$$

où la dérivée est évaluée avec (3.31) et $v = v^k$, $\widehat{u}_i(v^k) = u_i^{k+1}$.

3.2.6.3 Approche par le PPA

Nous allons maintenant reprendre cet exemple en le plongeant dans le formalisme général du PPA.

Première étape : choix d'une fonction auxiliaire. Comme il se doit, on choisit une fonction auxiliaire additive par rapport aux u_i ; quant à sa dépendance en v , on la choisira de telle sorte qu'un algorithme de gradient en v émerge de l'application du formalisme général. On pose donc

$$K^k(u, v) = \sum_{i=1}^N K_i^k(u_i) + \frac{1}{2} \|v\|^2 . \quad (3.33)$$

On peut de plus utiliser la technique (3.12) pour définir K_i^k ce qui donne

$$K_i^k(u_i) = g(J_1(u_1^k, v^k), \dots, J_i(u_i, v^k), \dots, J_N(u_N^k, v^k)) . \quad (3.34)$$

Dans le cas où g est la somme (respectivement, le produit) des J_i , $K_i^k(u_i)$ coïncide, à une constante additive (respectivement, multiplicative) près, avec $J_i(u_i, v^k)$.

À cette fonction $K_i^k(\cdot)$, il peut être nécessaire, comme on l'a vu, de rajouter une fonction du type $\gamma \|\cdot\|^2 / 2$.

Deuxième étape : algorithme séquentiel. On utilise maintenant la version *séquentielle* de l'algorithme du PPA, telle que décrite au §3.2.5.3. La notion de séquentialité porte sur le déroulement des minimisations en u d'une part, v d'autre part, et non pas sur les minimisations par rapport à chaque u_i qui restent parallèles.

On a vu au §3.2.5.3 que, dans la version séquentielle, on pouvait utiliser des “pas” ε^k *différents* pour chaque minimisation en séquence. On a vu aussi (voir §3.2.5.1) que dans le cas du choix (3.12), le fait de prendre $\varepsilon^k = 1$ était intéressant, quitte à rajouter un terme $\gamma \|\cdot\|^2 / 2$, avec γ assez grand pour assurer la convergence. On prendra donc le ε^k relatif à u égal à 1, et on désignera par ρ^k celui relatif à v .

Après tous ces choix, l'algorithme obtenu se décline ainsi : à l'itération k ,

1. résoudre, pour $i = 1, \dots, N$

$$\min_{u_i \in U_i^{\text{ad}}} \left(K_i^k(u_i) + \frac{\gamma}{2} \|u_i - u_i^k\|^2 \right), \quad (3.35)$$

où K_i^k est donnée par (3.34) ; soit u_i^{k+1} la solution.

2. remettre à jour v^k par la formule (3.32).

Exercice 3.13. Vérifier qu'en effet, l'application rigoureuse de l'algorithme du PPA conduit bien à ce sous-problème en v .

Commentaires 3.14.

- (a) On obtient donc finalement un algorithme très proche de l'algorithme obtenu par l'approche "ad hoc" qui s'appuyait sur la structure particulière du problème initial. Cependant, l'approche du PPA ne nécessite pas d'introduire la fonction φ et de connaître la règle de calcul de sa dérivée : la remise à jour de v , quoiqu'identique dans les deux approches, découle d'un calcul systématique dans le cas du PPA.
- (b) La remise à jour de v , qui était vue comme une étape de coordination dans l'approche "ad hoc", est plutôt vue ensuite comme l'une des deux étapes d'une version séquentielle de l'algorithme du PPA. Ceci montre que la notion de "coordination" n'est en fait qu'une interprétation plaquée a posteriori sur une approche algorithmique systématique.
- (c) On a vu qu'il est possible de mêler au sein d'un même algorithme des phases parallèles et des phases séquentielles.
- (d) L'application du PPA permet de recourir à des théorèmes généraux de convergence. En particulier, l'introduction d'un terme "prox" dans les sous-problèmes en u_i , qui aurait pu être considérée comme un "truc" supplémentaire dans l'approche "ad hoc", se trouve éventuellement justifiée, et "calibrée" (en ce qui concerne la valeur à donner au coefficient γ) par application des théorèmes de convergence.
- (e) Enfin, "last but not least", si la structure particulière (3.29) que nous avons supposée pour la fonction coût J n'est plus supposée, il est facile de constater que la démarche de *l'algorithme du PPA s'applique sans aucun changement*, sauf bien sûr que les calculs doivent être repris systématiquement pour s'adapter à la nouvelle situation.

Évidemment, pour envisager une décomposition de u en sous-vecteurs u_i alors que v n'est pas décomposé, il faut avoir une motivation pratique, mais elle pourrait venir d'une forme (3.29) de J seulement *approximativement* vérifiée. Le PPA "fait le nécessaire" même dans les cas "non séparables".

3.3 Le Principe du Problème Auxiliaire en optimisation sous contraintes explicites

Jusqu'à maintenant, la partie \mathcal{J} de la fonction coût était, dans le cadre de la décomposition, la seule cause de couplage entre sous-problèmes potentiels. Si les contraintes elles-mêmes sont une cause de couplage, alors, pour construire des algorithmes de décomposition-coordination, il devient nécessaire

- d'une part, de formuler ces contraintes *explicitement* comme des contraintes égalité ou inégalité ;

- d'autre part, de faire appel à la théorie de la dualité afin de transformer, quand c'est possible, les problèmes de minimisation en problèmes de point selle du Lagrangien (ou du Lagrangien augmenté).

Il nous faudra dès lors obtenir des extensions du PPA aux problèmes de point selle.

On considère donc ici des sous-problèmes de la forme suivante :

$$\min_{u \in U^{\text{ad}}} J(u) \quad (3.36a)$$

$$\text{sous } \Theta(u) \in -C. \quad (3.36b)$$

Les hypothèses en vigueur dans toute la fin de ce chapitre sont celles décrites en Annexe (§3.7) pour pouvoir garantir l'existence d'un point selle du Lagrangien.

3.3.1 Le PPA et les problèmes de point selle

Soit $L : \mathcal{U} \times \mathcal{P} \rightarrow \mathbb{R}$ une fonction convexe-concave sur des espaces de Hilbert \mathcal{U} et \mathcal{P} ; on suppose l'existence d'un point selle sur des sous-ensembles admissibles convexes fermés U^{ad} et P^{ad} . Comme précédemment avec la fonction J , on suppose que L est la somme de deux fonctions \mathcal{L} et \mathcal{L} de la même nature que L , \mathcal{L} étant différentiable ou sous-différentiable, tandis que \mathcal{L} sera supposée additive lorsque nous parlerons de décomposition.

Dans la même optique que celle du Lemme 3.2, soit $\Lambda : \mathcal{U} \times \mathcal{P} \rightarrow \mathbb{R}$ une fonction auxiliaire fortement convexe-fortement concave et différentiable ; soit $(v, q) \in \mathcal{U} \times \mathcal{P}$, $r \in \partial_u \mathcal{L}(v, q)$ (sous-différentiel de la fonction convexe $u \mapsto \mathcal{L}(u, q)$ en v), $s \in \partial_p \mathcal{L}(v, q)$ ("sous-différentiel" de la fonction concave $p \mapsto \mathcal{L}(v, p)$ en q), et finalement, un nombre positif ε . Avec ces éléments, cherchons un point selle de

$$\Lambda(u, p) + \langle \varepsilon r - \Lambda'_u(v, q), u \rangle + \langle \varepsilon s - \Lambda'_p(v, q), p \rangle + \varepsilon \mathcal{L}(u, p) \quad (3.37a)$$

sur $U^{\text{ad}} \times P^{\text{ad}}$, ce qui revient aussi à résoudre, pour (\hat{u}, \hat{p}) dans $U^{\text{ad}} \times P^{\text{ad}}$, la paire d'inéquations variationnelles

$$\forall u \in U^{\text{ad}}, \quad \langle \Lambda'_u(\hat{u}, \hat{p}) + \varepsilon r - \Lambda'_u(v, q), u - \hat{u} \rangle + \varepsilon (\mathcal{L}(u, \hat{p}) - \mathcal{L}(\hat{u}, \hat{p})) \geq 0, \quad (3.37b)$$

$$\forall p \in P^{\text{ad}}, \quad \langle \Lambda'_p(\hat{u}, \hat{p}) + \varepsilon s - \Lambda'_p(v, q), p - \hat{p} \rangle + \varepsilon (\mathcal{L}(\hat{u}, p) - \mathcal{L}(\hat{u}, \hat{p})) \leq 0. \quad (3.37c)$$

Il est facile de voir, comme au Lemme 3.2, que si (v, q) est une solution de ce système, c'est aussi un point selle de L sur les mêmes sous-ensembles admissibles.

Des algorithmes analogues à l'Algorithme 3.3 peuvent facilement être imaginés. Mais nous n'allons pas poursuivre le long de cette ligne générale. Nous allons plutôt revenir au problème de trouver le point selle d'un Lagrangien (3.81) associé à un problème de minimisation sous contraintes comme (3.76) et nous allons essayer de retrouver des schémas de décomposition-coordination rencontrés au Chapitre 2.

3.3.2 Le PPA et la décomposition par les prix

3.3.2.1 Obtention de l'algorithme général

Notre propos est ici de retrouver un schéma proche de celui du §2.1.2. On fait les observations suivantes :

1. dans la décomposition par les prix, seul l'espace des variables primales est décomposé en sous-espaces ;
2. le problème du point selle est remplacé, lors du déroulement de l'algorithme, par une alternance de remises à jour des variables primales et des "prix" : ces deux types de variables sont donc traités dans un schéma *séquentiel* ;

3. la remise à jour des prix se fait par un algorithme de type “gradient”.

Le dernier point suggère un choix particulier de la fonction auxiliaire Λ quant à sa dépendance en p (voir §3.2.3.2). Le second point indique que le problème auxiliaire ne sera pas un problème de point selle mais tantôt une minimisation et tantôt une maximisation. Quant au premier point, il nous conduit à ne nous intéresser qu’à la structure par rapport à u de L du point de vue de la décomposition.

Ces observations étant faites, supposons que J dans (3.36a) soit égal à $\mathcal{J} + \mathfrak{J}$, et de même, que Θ dans (3.36b) soit égal à $\Theta + \mathfrak{T}$. Lorsque nous parlerons de décomposition, comme d’habitude, on supposera que \mathfrak{J} et \mathfrak{T} sont des structures additives par rapport à une décomposition de u comme dans (3.7), et il en sera donc de même de $\mathcal{L}(u, p) = \mathfrak{J}(u) + \langle p, \mathfrak{T}(u) \rangle$. Évidemment, $\mathcal{L}(u, p) = \mathcal{J}(u) + \langle p, \Theta(u) \rangle$.

Alors, dans (3.37), choisissons

$$\Lambda(u, p) = K(u) - \frac{1}{2\alpha} \|p\|^2, \quad (3.38)$$

où K est fortement convexe et différentiable et α est un nombre positif (le second terme ci-dessus est alors fortement concave en p). Selon le schéma où u et p sont traités séquentiellement, on aboutit à l’algorithme suivant.

Algorithme 3.15. 1. Commencer avec u^0 et p^0 ; poser $k = 0$.

2. À l’itération $k + 1$, résoudre

$$\min_{u \in U^{\text{ad}}} \left(K(u) + \langle \varepsilon \mathcal{J}'(u^k) - K'(u^k), u \rangle + \varepsilon \mathfrak{J}(u) + \varepsilon \langle p^k, \Theta'(u^k) \cdot u + \mathfrak{T}(u) \rangle \right); \quad (3.39a)$$

soit u^{k+1} une solution.

3. Calculer

$$p^{k+1} = \text{proj}_{C^*} (p^k + \rho(\Theta + \mathfrak{T})(u^{k+1})), \quad (3.39b)$$

où ρ est égal à $\alpha\varepsilon$.

4. Si $\|u^{k+1} - u^k\| + \|p^{k+1} - p^k\|$ est “suffisamment petite”, stop ; sinon, retourner en 2 avec k incrémenté de 1.

Exercice 3.16. Retrouver (3.39) en partant de (3.37) et des choix effectués (fonction auxiliaire (3.38), schéma séquentiel).

3.3.2.2 Utilisations

Considérons la répartition de L en $\mathcal{L} + \mathfrak{L}$ suivante :

$$\mathcal{L} = \mathcal{J}(u) ; \quad \mathfrak{L} = \mathfrak{J}(u) + \langle p, \Theta(u) \rangle ,$$

autrement dit, Θ est considérée comme identiquement nulle et Θ se résume à \mathfrak{T} . Dans (3.39a), on choisit K égal à $\varepsilon \mathcal{J}$;¹⁰ alors l’Algorithme 3.15 se réduit à l’algorithme d’Uzawa qui consiste à minimiser le Lagrangien en u à p fixé, puis à remettre à jour p par la formule (3.39b). Comme on l’a vu au §2.1.2, cette approche conduit à une décomposition “par les prix” si J et Θ sont additives par rapport à la décomposition (3.7), c’est-à-dire si \mathcal{J} (la partie non additive de J) est aussi identiquement nulle.

Si \mathcal{J} n’est pas fortement convexe (et en particulier s’il est nul !), il est nécessaire d’“augmenter” le choix $K = \varepsilon \mathcal{J}$ avec par exemple un terme $\|\cdot\|^2/2$, conduisant au terme “prox” $\|u - u^k\|^2/2$ supplémentaire dans l’algorithme d’Uzawa (terme qui préserve la décomposition).

¹⁰sous réserve des conditions de convergence discutées ultérieurement, en particulier la forte convexité de K

À l'opposé, si \mathfrak{J} et \mathfrak{T} sont identiquement nuls (donc L se réduit à \mathcal{L}) et si K est choisie comme la fonction $\|\cdot\|^2/2$, on obtient l'algorithme d'*Arrow-Hurwicz* qui revient à une alternance de pas de gradient projeté dans l'espace primal et dans l'espace dual selon le schéma suivant :

$$u^{k+1} = \text{proj}_{U^{\text{ad}}} (u^k - \varepsilon \nabla_u L(u^k, p^k)) , \quad (3.40a)$$

$$p^{k+1} = \text{proj}_{C^*} (p^k + \rho \nabla_p L(u^{k+1}, p^k)) . \quad (3.40b)$$

Comme l'algorithme du gradient (voir §3.2.3.2), cet algorithme peut être considéré comme un algorithme très simple de décomposition. Mais d'autres algorithmes de décomposition peuvent être obtenus avec d'autres choix de fonctions auxiliaires K additives (en supposant toujours que \mathfrak{J} et \mathfrak{T} sont additives avec (3.7)). En effet, sous les hypothèses structurelles en vigueur, (3.39a) se décompose en N sous-problèmes indépendants.

On vient donc de généraliser la méthode de décomposition par les prix au cas où la fonction coût et les contraintes, par la présence des termes \mathcal{J} et Θ , ne sont pas additives, à condition que ces termes soient *différentiables* (on dira plus loin quelques mots du cas sous-différentiable). En jouant sur la décomposition de J et Θ en $\mathcal{J} + \mathfrak{J}$ et $\Theta + \mathfrak{T}$, respectivement, et en jouant sur le choix de K en relation avec \mathcal{J} (et éventuellement $\langle p^k, \Theta(\cdot) \rangle$), c'est en fait tout un continuum d'algorithmes que l'on peut ainsi obtenir.

Remarque 3.17. On aurait pu, comme à l'Algorithme 3.3, envisager de prendre des fonctions auxiliaires K dépendantes de l'indice d'itération k . Cependant, on ne dispose pas de démonstration de convergence pour ce cas (ce qui n'empêche pas d'en utiliser dans la pratique, sous réserve de respecter les conditions de convergence portant sur K de façon uniforme en k).

3.3.2.3 Convergence

L'Exercice 2.30 en fin de Chapitre 2 a montré que l'algorithme d'Uzawa appliqué au Lagrangien d'un problème d'optimisation dont la fonction coût J est simplement convexe (même avec une dérivée J' Lipschitzienne) ne permet pas d'obtenir la convergence des variables primales (en plus de problèmes de non différentiabilité de la fonction duale ψ définie par (2.4)). On va donc ici faire une hypothèse de *forte* convexité pour cette fonction coût J .

On montre d'abord que cette hypothèse — plus une hypothèse de Lipschitz sur la fonction contrainte Θ — assure non seulement la différentiabilité de ψ (ce qui serait déjà assuré par la *stricte* convexité de J), mais aussi le fait que ψ' est Lipschitzienne.

Lemme 3.18. *Si J est fortement convexe de constante a et si Θ est Lipschitzienne de constante τ , alors ψ , définie par (2.4), est différentiable et de dérivée Lipschitzienne de constante τ^2/a .*

Démonstration. On observe d'abord que pour une fonction f fortement convexe de constante a , et si u^\sharp désigne son unique argument du minimum (sur U^{ad}), pour $u \in U^{\text{ad}}$ et $\alpha \in [0, 1]$, on a

$$\begin{aligned} f(u^\sharp) &\leq f(\alpha u^\sharp + (1 - \alpha)u) \\ &\leq \alpha f(u^\sharp) + (1 - \alpha)f(u) - \alpha(1 - \alpha)\frac{a}{2} \|u - u^\sharp\|^2 . \end{aligned}$$

En faisant passer $\alpha f(u^\sharp)$ à l'extrême gauche, en simplifiant par le facteur $(1 - \alpha)$ qui apparaît en facteur des deux côtés extrêmes, puis en posant $\alpha = 1$, on obtient

$$f(u) - f(u^\sharp) \geq \frac{a}{2} \|u - u^\sharp\|^2 . \quad (3.41)$$

Considérons maintenant deux valeurs (admissibles) p et q des variables duales et les $\arg \min u$ et v correspondants de L (voir définition (2.4) de ψ). Si J est fortement convexe, c'est aussi le cas de $L(\cdot, p)$

pour tout p admissible car, comme on l'a vu, $\langle p, \Theta(\cdot) \rangle$ est convexe si $p \in C^*$ et Θ est C -convexe. Alors, par application de (3.41),

$$L(v, p) - L(u, p) \geq \frac{a}{2} \|v - u\|^2.$$

En sommant cette inégalité avec l'inégalité analogue où le rôle de (u, p) est échangé avec celui de (v, q) , on obtient

$$\langle p - q, \Theta(v) - \Theta(u) \rangle \geq a \|v - u\|^2.$$

D'une part, l'hypothèse de Lipschitz sur Θ permet d'obtenir

$$\|v - u\|^2 \geq \frac{1}{\tau^2} \|\Theta(v) - \Theta(u)\|^2,$$

d'autre part, on note que $\Theta(u) = \psi'(p)$ et $\Theta(v) = \psi'(q)$. L'inégalité ci-dessus devient alors l'expression de la propriété de Dunn pour ψ avec la constante τ^2/a (en tenant compte du fait que ψ est une fonction concave). Ceci achève la démonstration. \square

Remarque 3.19. En reprenant cette démonstration avec l'hypothèse

$$J(\alpha u + (1 - \alpha)v) \leq \alpha J(u) + (1 - \alpha)J(v) - \alpha(1 - \alpha) \frac{1}{2d} \|\Theta(u) - \Theta(v)\|^2,$$

ce qui est une sorte de “forte convexité de J par rapport à Θ ”, on aboutit à une autre démonstration du résultat ci-dessus avec la constante de Dunn de ψ évaluée à d . Les hypothèses faites dans le lemme impliquent la propriété ci-dessus avec la constante d évaluée à τ^2/a . La réciproque étant fausse a priori, la propriété ci-dessus semble plus faible que les hypothèses du lemme. Le même genre de remarque s'appliquerait à tout ce qui va suivre.

La propriété de Lipschitz de ψ' avec la constante τ^2/a montre que l'algorithme d'Uzawa, qui n'est rien d'autre qu'un algorithme de gradient projeté pour maximiser ψ sur C^* , converge si son pas ρ est inférieur à $2a/\tau^2$: ceci est une simple application du Théorème 3.6 (voir plus précisément (3.15)) en observant que l'algorithme du gradient correspond au choix $K(\cdot) = (1/2) \|\cdot\|^2$, fonction pour laquelle $b = 1$. C'est une borne similaire que l'on va retrouver pour ρ dans le théorème de convergence relatif à l'Algorithme 3.15.

Théorème 3.20. *Les hypothèses sont les suivantes :*

- (H1) \mathfrak{J} est une fonction convexe s.c.i. ;
- (H2) \mathcal{J} est une fonction convexe s.c.i. différentiable, de gradient Lipschitzien avec constante A ;
- (H3) la fonction $J = \mathcal{J} + \mathfrak{J}$ est fortement convexe de constante a ;
- (H4) la fonction Θ est C -convexe et différentiable ; elle est telle que la fonction convexe $u \mapsto \langle p, \Theta(u) \rangle$, est de dérivée Lipschitzienne de constante G uniformément en p ;¹¹
- (H5) la fonction \mathfrak{T} est C -convexe ;
- (H6) la fonction $\Theta = \Theta + \mathfrak{T}$ est Lipschitzienne de constante τ ;
- (H7) le Lagrangien du problème (3.36) admet un point selle (u^\sharp, p^\sharp) sur $U^{\text{ad}} \times C^*$;¹²

¹¹On reviendra plus loin — voir Commentaires 3.23-c — sur cette hypothèse très forte.

¹²Une hypothèse de coercivité de J sur U^{ad} — voir (3.2) — et une hypothèse de qualification des contraintes — voir (3.83) — seraient requises en plus des hypothèses ci-dessus pour garantir cette existence. Évidemment, sous l'hypothèse de forte convexité retenue ici, u^\sharp est unique (mais pas nécessairement p^\sharp).

(H8) la fonction K est convexe s.c.i. différentiable, de dérivée Lipschitzienne de constante B et fortement monotone de constante b .

Alors, les conclusions sont les suivantes :

(C1) le problème auxiliaire (3.39a) admet une solution unique u^{k+1} ;

(H9) si

$$0 < \varepsilon \leq \frac{b}{A+G} \text{ et } 0 < \rho < \frac{a}{\tau^2} , \quad (3.42)$$

(C2) alors u^k converge (en norme) vers u^\sharp ;

(C3) la suite $\{p^k\}$ est bornée et tout point d'accumulation de cette suite, dans la topologie faible, constitue avec u^\sharp un point selle de L .

Démonstration.

(C1) On ne revient pas sur l'existence et l'unicité de u^{k+1} comme solution de (3.39a), ce point étant justifié par des arguments similaires à ceux de la démonstration du Théorème 3.6.

(C2) On fait le choix de la fonction de Lyapounov suivante :

$$\ell(u, p) = \underbrace{K(u^\sharp) - K(u) - \langle K'(u), u^\sharp - u \rangle}_{\phi_1(u)} + \underbrace{\frac{\varepsilon}{2\rho} \|p - p^\sharp\|^2}_{\phi_2(p)} , \quad (3.43)$$

où (u^\sharp, p^\sharp) est un point selle de L . On observe que le terme $\phi_1(u)$ est celui déjà utilisé dans (3.18) et que $\phi_2(p)$ est un terme analogue dérivé de la fonction auxiliaire (3.38). On a

$$\ell(u, p) \geq (b/2) \|u - u^\sharp\|^2 + (\varepsilon/2\rho) \|p - p^\sharp\|^2 ,$$

ce qui qualifie ce choix comme fonction de Lyapounov.

L'inéquation variationnelle associée au problème (3.39a) s'écrit

$$\begin{aligned} \forall u \in U^{\text{ad}}, \quad & \langle K'(u^{k+1}) - K'(u^k) + \varepsilon \mathcal{J}'(u^k), u - u^{k+1} \rangle + \varepsilon (\mathfrak{J}(u) - \mathfrak{J}(u^{k+1})) \\ & + \varepsilon \langle p^k, \Theta'(u^k).(u - u^{k+1}) + \mathfrak{T}(u) - \mathfrak{T}(u^{k+1}) \rangle \geq 0 . \end{aligned} \quad (3.44)$$

On l'utilise avec $u = u^\sharp$ et on additionne cette inéquation (divisée par ε) à celle caractérisant u^\sharp (renforcée par le caractère fortement convexe de $L(\cdot, p)$), où l'on pose $u = u^{k+1}$, à savoir,

$$\begin{aligned} & (\mathcal{J} + \mathfrak{J})(u^{k+1}) - (\mathcal{J} + \mathfrak{J})(u^\sharp) \\ & + \langle p^\sharp, (\Theta + \mathfrak{T})(u^{k+1}) - (\Theta + \mathfrak{T})(u^\sharp) \rangle \geq \frac{a}{2} \|u^{k+1} - u^\sharp\|^2 , \end{aligned} \quad (3.45)$$

ce qui donne

$$\begin{aligned} & \frac{1}{\varepsilon} \langle K'(u^{k+1}) - K'(u^k), u^\sharp - u^{k+1} \rangle \\ & + \langle \mathcal{J}'(u^k), u^\sharp - u^{k+1} \rangle + (\mathcal{J}(u^{k+1}) - \mathcal{J}(u^\sharp)) + \langle p^k, \Theta'(u^k).(u^\sharp - u^{k+1}) + \Theta(u^{k+1}) - \Theta(u^\sharp) \rangle \\ & + \langle p^k - p^\sharp, \Theta(u^\sharp) - \Theta(u^{k+1}) \rangle \geq \frac{a}{2} \|u^{k+1} - u^\sharp\|^2 . \end{aligned} \quad (3.46)$$

Par ailleurs, l'inégalité de gauche du point selle de L est équivalente à l'équation de point fixe suivante, valable pour tout ρ positif :

$$p^\sharp = \text{proj}_{C^*} (p^\sharp + \rho(\Theta + \mathfrak{T})(u^\sharp)) . \quad (3.47)$$

Exercice 3.21. Démontrer l'équivalence de (3.47) avec l'inégalité de gauche du point selle.

De cette équation et de (3.39b), on déduit, parce que la projection est un opérateur non expansif,

$$\|p^{k+1} - p^\sharp\| \leq \|p^k - p^\sharp + \rho((\Theta + \mathfrak{T})(u^{k+1}) - (\Theta + \mathfrak{T})(u^\sharp))\|.$$

Exercice 3.22. Montrer que la projection proj_C sur un sous-ensemble convexe fermé C est un opérateur non expansif, c'est-à-dire que

$$\|\text{proj}_C x - \text{proj}_C y\| \leq \|x - y\|. \quad (3.48)$$

Cette inégalité est élevée au carré et divisée par 2ρ , ce qui donne

$$\begin{aligned} \frac{1}{2\rho} \|p^{k+1} - p^\sharp\|^2 &\leq \frac{1}{2\rho} \|p^k - p^\sharp\|^2 + \langle p^k - p^\sharp, (\Theta + \mathfrak{T})(u^{k+1}) - (\Theta + \mathfrak{T})(u^\sharp) \rangle \\ &\quad + \frac{\rho}{2} \|(\Theta + \mathfrak{T})(u^{k+1}) - (\Theta + \mathfrak{T})(u^\sharp)\|^2, \end{aligned}$$

puis elle est ajoutée à (3.46). On obtient finalement, après remultiplication par ε ,

$$\begin{aligned} &\underbrace{\langle K'(u^{k+1}) - K'(u^k), u^\sharp - u^{k+1} \rangle}_{(a)} \\ &\quad + \underbrace{\varepsilon \langle \mathcal{J}'(u^k), u^\sharp - u^{k+1} \rangle + \varepsilon (\mathcal{J}(u^{k+1}) - \mathcal{J}(u^\sharp))}_{(b)} \\ &\quad + \underbrace{\varepsilon \langle p^k, \Theta'(u^k) \cdot (u^\sharp - u^{k+1}) + \Theta(u^{k+1}) - \Theta(u^\sharp) \rangle}_{(c)} \\ &\quad + \underbrace{\frac{\varepsilon}{2\rho} (\|p^k - p^\sharp\|^2 - \|p^{k+1} - p^\sharp\|^2)}_{(d)} + \underbrace{\frac{\varepsilon\rho}{2} \|(\Theta + \mathfrak{T})(u^{k+1}) - (\Theta + \mathfrak{T})(u^\sharp)\|^2}_{(e)} \\ &\quad - \underbrace{\varepsilon \frac{a}{2} \|u^{k+1} - u^\sharp\|^2}_{(f)} \geq 0. \quad (3.49) \end{aligned}$$

Le terme (a) est traité comme dans la preuve du Théorème 3.6 (voir (3.20)), ce qui permet de le comparer à la variation de ϕ_1 . Le terme (d) s'exprime directement comme la variation de ϕ_2 . Le terme (e) est majoré par $(\varepsilon\rho\tau^2/2) \|u^{k+1} - u^\sharp\|^2$ grâce la propriété de Lipschitz de Θ . Examinons le traitement du terme (b). Le traitement du terme (c) pourra suivre la même procédure où l'on remplace $\mathcal{J}(\cdot)$ par $\langle p^k, \Theta(\cdot) \rangle$. On a

$$\begin{aligned} \langle \mathcal{J}'(u^k), u^\sharp - u^{k+1} \rangle &= \langle \mathcal{J}'(u^k), u^\sharp - u^k \rangle + \langle \mathcal{J}'(u^k), u^k - u^{k+1} \rangle \\ &\leq \mathcal{J}(u^\sharp) - \mathcal{J}(u^k) - \frac{1}{2A} \|\mathcal{J}'(u^k) - \mathcal{J}'(u^\sharp)\|^2 \\ &\quad + \mathcal{J}(u^k) - \mathcal{J}(u^{k+1}) + \frac{A}{2} \|u^{k+1} - u^k\|^2 \end{aligned} \quad (3.50)$$

par utilisation de (3.72) et (3.71), respectivement. On en déduit que le terme (b) est majoré par

$$(\varepsilon/2) \left(A \|u^{k+1} - u^k\|^2 - (1/A) \|\mathcal{J}'(u^k) - \mathcal{J}'(u^\sharp)\|^2 \right).$$

On peut faire des calculs similaires s'appliquant à $\langle p^k, \Theta(\cdot) \rangle$ au lieu de $\mathcal{J}(\cdot)$, la constante G remplaçant donc A . Le rassemblement de toutes ces majorations dans (3.49) aboutit, en posant

$$\Delta_k^{k+1} = \phi_1(u^{k+1}) + \phi_2(p^{k+1}) - \phi_1(u^k) - \phi_2(p^k),$$

à

$$\Delta_k^{k+1} \leq \frac{\varepsilon}{2}(\rho\tau^2 - a) \|u^{k+1} - u^\sharp\|^2 + \frac{1}{2}(\varepsilon(A + G) - b) \|u^{k+1} - u^k\|^2 ,$$

où on a omis, au second membre, certains termes négatifs qui ne seront pas utiles.

En tenant compte des conditions (3.42), on voit que le second membre est strictement négatif (sauf si l'algorithme a convergé en un nombre fini d'itérations). On est donc dans le cadre du schéma général (§3.2.4.1) et l'on peut affirmer que la fonction de Lyapounov décroît, la suite $\{(u^k, p^k)\}$ est bornée, et la différence de deux termes successifs tend vers 0, ce qui montre la convergence forte de u^k vers u^\sharp .

- (C3) La suite $\{p^k\}$ étant bornée admet des points d'accumulation dans la topologie faible. Soit \bar{p} un tel point. Montrons que c'est, associé à u^\sharp , un point selle de L . Comme C^* est convexe fermé, donc aussi faiblement fermé, \bar{p} appartient à C^* .

On considère maintenant une sous-suite $\{p^{k_i}\}$ convergeant faiblement vers \bar{p} . Sans entrer dans les détails, compte tenu de la convergence forte de u^k vers u^\sharp , des hypothèses de Lipschitz sur les diverses dérivées impliquées, et du caractère convexe s.c.i., donc faiblement s.c.i., des diverses fonctions coûts apparaissant dans (3.44), il doit être clair qu'en passant à la limite dans cette inéquation, on obtient l'inégalité de droite du point selle pour le couple (u^\sharp, \bar{p}) .

Pour prouver l'autre inégalité du point selle, on observe d'abord que la suite $\{\ell(u^k, p^k) = \phi_1(u^k) + \phi_2(p^k)\}$ a une limite, et comme $\phi_1(u^k)$ (voir (3.43)) tend vers 0 en raison de la convergence forte de u^k vers u^\sharp , la suite $\{\phi_2(p^k) = (\varepsilon/2\rho) \|p^k - p^\sharp\|^2\}$ a donc aussi une limite.

En second lieu, puisque nécessairement $\Theta(u^\sharp) \in -C$ et que $\bar{p} \in C^*$, alors $\langle \bar{p}, \Theta(u^\sharp) \rangle \leq 0$. On va montrer l'inégalité opposée, ce qui montrera que $\langle \bar{p}, \Theta(u^\sharp) \rangle = 0$. Pour cela, on observe que (3.39b) est équivalent à l'inéquation variationnelle : $p^{k+1} \in C^*$ et, pour tout $p \in C^*$,

$$\langle p - p^{k+1}, p^k + \rho\Theta(u^{k+1}) - p^{k+1} \rangle \leq 0 .$$

On pose $p = p^\sharp$ et l'on obtient

$$\begin{aligned} \rho \langle p^{k+1} - p^\sharp, \Theta(u^{k+1}) \rangle &\geq \langle p^{k+1} - p^\sharp, (p^{k+1} - p^\sharp) - (p^k - p^\sharp) \rangle \\ &\geq \frac{1}{2} (\|p^{k+1} - p^\sharp\|^2 - \|p^k - p^\sharp\|^2) . \end{aligned}$$

En passant à la limite pour la sous-suite telle que p^{k_i+1} tend faiblement vers \bar{p} , on observe que le second membre tend vers 0 (car $\|p^k - p^\sharp\|^2$ a une limite), tandis que le premier membre tend vers $\langle \bar{p}, \Theta(u^\sharp) \rangle$. On en déduit l'inégalité recherchée.

Pour terminer, on observe que le fait que $\langle \bar{p}, \Theta(u^\sharp) \rangle = 0$ est équivalent à l'inégalité de gauche du point selle puisque, pour tout $p \in C^*$, $\langle p, \Theta(u^\sharp) \rangle \leq 0 = \langle \bar{p}, \Theta(u^\sharp) \rangle$. \square

Commentaires 3.23.

- (a) Dans [12], un théorème et une démonstration quelque peu différents sont donnés. Les hypothèses diffèrent en ce sens que la forte convexité est supposée pour \mathcal{J} et non pas, comme ici, pour $\mathcal{J} + \mathfrak{J}$. Les conclusions sont les mêmes, mais avec la borne $2a/\tau^2$ pour ρ au lieu de celle donnée dans (3.42). La démonstration diffère d'une part par le choix de la fonction de Lyapounov¹³, d'autre part par la façon de mener les calculs : au lieu de (3.45), on utilise plutôt, comme dans la démonstration du Théorème 3.6, l'inéquation variationnelle faisant intervenir $\mathcal{J}'(u^\sharp)$.

¹³qui comporte, par rapport à (3.43), un terme supplémentaire $-\varepsilon(a/2) \|u - u^\sharp\|^2$; on vérifiera que ce terme négatif conserve malgré tout les propriétés requises à la nouvelle fonction de Lyapounov ;

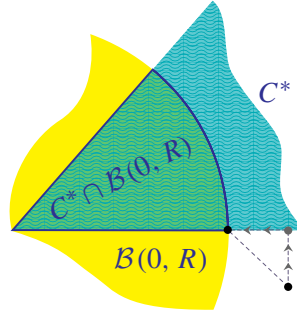


Figure 3.1: Projection sur l'intersection d'un cône et d'une boule

- (b) Inversement, on pourrait reprendre la démonstration du Théorème 3.6 dans le cas où on avait supposé que \mathcal{J} est fortement convexe, mais en faisant maintenant l'hypothèse que c'est $\mathcal{J} + \mathfrak{J}$ qui est fortement convexe : il faudrait alors conduire les calculs (pour les termes qui restent pertinents dans le cadre du Théorème 3.6) d'une façon analogue à celle utilisée dans la démonstration ci-dessus et cela conduirait à la borne b/A sur ε (au lieu de $2b/A$ — voir (3.15)).
- (c) Revenons maintenant sur l'hypothèse (H4) du Théorème 3.20 (qui fait l'objet de la note 11 de bas de page), hypothèse quelque peu irréaliste. En effet, il faut s'attendre à ce que la fonction $u \mapsto \langle p, \Theta(u) \rangle$ ait une dérivée dont la constante de Lipschitz est plutôt proportionnelle à $\|p\|$ (c'est-à-dire que G devient $g \|p\|$ pour $g > 0$). Dans la démonstration, à la place de G , on verrait alors apparaître $g \|p^k\|$, et ceci ne permettrait pas de conclure à la décroissance de la fonction de Lyapounov sans prouver au préalable que p^k reste borné, ce qui s'avère impossible a priori.

On a donc recours au procédé suivant pour contourner la difficulté. On modifie l'Algorithme 3.15, et plus particulièrement la formule (3.39b), en remplaçant la projection sur C^* par la projection sur $C_R^* = C^* \cap \mathcal{B}(0, R)$ où $\mathcal{B}(0, R)$ est la boule fermée de centre 0 et de rayon R dans \mathcal{C}^* , et où R doit être choisi assez grand pour que cette boule contienne au moins un p^\sharp . On observe que

1. cette condition nécessite de connaître une estimation de la norme d'un multiplicateur p^\sharp optimal, mais compte tenu de l'interprétation économique des multiplicateurs, cette hypothèse n'est pas irréaliste dans la pratique ;
2. si la condition ci-dessus est vérifiée, alors le Lagrangien du problème (3.36) admet encore un point selle sur $U^{\text{ad}} \times C_R^*$ et la formule (3.47) est aussi vraie avec la projection sur C_R^* ; les calculs de la démonstration peuvent alors être adaptés sans difficulté ;
3. la projection sur l'intersection du cône C^* et de la boule $\mathcal{B}(0, R)$ n'est pas beaucoup plus difficile à calculer que la projection sur C^* : on commence par faire le calcul de cette projection sur C^* et si le résultat est de norme supérieure à R , on ramène cette norme à R par homothétie (voir Figure 3.1) ;
4. avec ce nouvel algorithme, on a automatiquement que $\|p^k\| \leq R$.

Dans ces conditions, on peut énoncer les conditions de convergence (3.42) avec G remplacé par gR .

On observe qu'un certain compromis est à trouver : si on veut augmenter les chances qu'il existe un p^\sharp dans $\mathcal{B}(0, R)$, on a tendance à prendre R grand, mais alors la borne sur ε diminue.

3.3.3 Le PPA et la décomposition par prédiction

Le but est maintenant de retrouver la méthode de décomposition par prédiction présentée au §2.1.4. En fait, nous avons vu deux algorithmes distincts : l'algorithme de point fixe (2.19) et l'algorithme d'Arrow-Hurwicz (2.22), ainsi que diverses variantes. Nous allons montrer comment ces algorithmes peuvent être retrouvés, et étendus, en utilisant le PPA.

3.3.3.1 Obtention de l'algorithme de point fixe

À chaque itération de la méthode par prédiction, c'est un problème d'optimisation sous contrainte qui est résolu, autant dire, dans notre contexte, un problème de point selle d'un Lagrangien. Cette observation à deux conséquences :

1. d'une part, contrairement à la démarche adoptée précédemment, les variables primales et duales vont être traitées *en parallèle*, et non plus *séquentiellement*, afin d'obtenir un problème auxiliaire sous forme de problème de point selle ;
2. d'autre part, il faut faire en sorte que ce point selle puisse se réinterpréter comme un problème d'optimisation sous contrainte, ce qui nécessite que la fonction dont on cherche le point selle ait une forme de Lagrangien.

Afin de répondre à cette seconde préoccupation, une fonction auxiliaire ressemblant à un Lagrangien doit être utilisée. Au lieu de (3.38), on choisit maintenant la fonction auxiliaire

$$\Lambda(u, p) = K(u) + \langle p, \Omega(u) \rangle \quad (3.51)$$

où

- $K : \mathcal{U} \rightarrow \mathbb{R}$ est encore une fois une fonction de la même nature que J ;
- $\Omega : \mathcal{U} \rightarrow \mathcal{C}$ est une fonction de la même nature que Θ .

Une fois le choix de cette fonction auxiliaire effectué, on revient à l'expression (3.37a) dont on cherche le point selle, ce qui revient à résoudre le couple d'inéquations variationnelles (3.37b)–(3.37c), où, bien entendu, \mathcal{L} et \mathcal{L} sont les parties du Lagrangien du problème (3.36) relatives à (\mathcal{J}, Θ) et $(\mathfrak{J}, \mathfrak{T})$, respectivement. On se place aussi dans le cas où \mathcal{J} et Θ sont différentiables. À l'itération $k + 1$, on remplace, dans (3.37), v par u^k , q par p^k , r par $\mathcal{L}'_u(u^k, p^k)$ et s par $\mathcal{L}'_p(u^k, p^k) = \Theta(u^k)$. Avec la réinterprétation du problème auxiliaire, on obtient finalement l'algorithme suivant.

Algorithme 3.24. 1. Commencer avec u^0 et p^0 ; poser $k = 0$.

2. A l'itération $k + 1$, résoudre

$$\min_{u \in U^{\text{ad}}} \left(K(u) + \varepsilon \mathfrak{J}(u) + \langle \varepsilon \mathcal{J}'(u^k) - K'(u^k), u \rangle + \langle p^k, (\varepsilon \Theta'(u^k) - \Omega'(u^k)) \cdot u \rangle \right) \quad (3.52a)$$

sous

$$\Omega(u) + \varepsilon \mathfrak{T}(u) \in -C + \Omega(u^k) - \varepsilon \Theta(u^k) . \quad (3.52b)$$

soit u^{k+1} une solution et p^{k+1} un multiplicateur optimal associé à la contrainte (3.52b).

3. Si $\|u^{k+1} - u^k\| + \|p^{k+1} - p^k\|$ est “suffisamment petite”, stop ; sinon, retourner en 2 avec k incrémenté de 1.

Exercice 3.25. En suivant la démarche évoquée ci-dessus, retrouver (3.52).

3.3.3.2 Utilisations

Utilisation en décomposition. Comme toujours, le problème auxiliaire est décomposable si on choisit une fonction auxiliaire additive par rapport à une décomposition des espaces donnée, et si on a “linéarisé” les fonctions du problème original qui ne sont pas elles-mêmes additives (celles qui sont réputées additives sont en lettres gothiques). La différence fondamentale entre la situation traitée ici et celle de la décomposition par les prix traitée précédemment est que, pour cette dernière, on ne s’intéressait qu’à une décomposition des variables primales selon (3.7). Ici, on s’intéresse non seulement à une décomposition des variables primales *mais aussi* à une décomposition des variables duales, c’est-à-dire des contraintes explicites (3.36b), ou bien encore, de l’espace \mathcal{C} des contraintes. Mais comme pour l’ensemble admissible U^{ad} qu’il fallait supposer lui-même décomposé selon une décomposition compatible avec celle de \mathcal{U} (voir (3.7)), il faut ici supposer que le cône positif C , ou de façon équivalente le cône dual C^* , est décomposable selon une décomposition compatible : on suppose donc que

$$C^* = \mathcal{C}_1^* \times \cdots \times \mathcal{C}_N^* ; \quad C^* = C_1^* \times \cdots \times C_N^* , \quad (3.53)$$

où C_i^* est un cône convexe fermé dans \mathcal{C}_i^* , pour $i = 1, \dots, N$. Évidemment, cette condition est trivialement vérifiée dans le cas des contraintes égalité ($C = \{0\}$ ou $C^* = \mathcal{C}^*$).

Remarque 3.26. Il est important de remarquer que la façon de numérotter les variables primales de $i = 1$ à N est a priori indépendante de celle qui permet de numérotter les variables duales.¹⁴ En associant un indice i de u_i avec un indice i de p_i , on réalise en fait implicitement l’“allocation des contraintes aux sous-problèmes” discutée au §2.1.4.

Remarque 3.27. Pour le choix habituel du cône positif dans \mathbb{R}^m , à savoir $(\mathbb{R}^+)^m$, la décomposition de C^* en $C_1^* \times \cdots \times C_N^*$ est acquise.

Une fois posées les décompositions (3.7) et (3.53), la partie \mathcal{L} du Lagrangien $L(u, p)$ qui est additive est celle qui peut s’écrire $\sum_i \mathcal{L}_i(u_i, p_i)$: elle sera composée

- de la partie \mathfrak{J} de la fonction coût qui est elle-même additive par rapport à (3.7) ;
- et de la partie \mathfrak{T} de la fonction “contrainte” qui est *bloc-diagonale* par rapport à (3.7) et (3.53), et pas seulement additive comme dans la décomposition par les prix.

Il reste alors à choisir la fonction auxiliaire (3.51) également additive, ce qui implique que K soit additive et que Ω soit bloc-diagonale.

Alors, le problème auxiliaire (3.52) se sépare en N sous-problèmes indépendants. On notera que par rapport à la situation considérée au Chapitre 2, on a une fois de plus obtenu cette décomposition en présence d’un couplage entre sous-problèmes à la fois par la partie \mathcal{J} de la fonction coût et par la partie non découplée des contraintes, à savoir la partie Θ .

Remarque 3.28. Comme déjà observé au §2.1.4, il n’y a pas de difficulté à faire cohabiter des sous-problèmes en (u_i, p_i) avec des sous-problèmes en u_j seulement, c’est-à-dire à considérer une décomposition de \mathcal{C}^* en moins de N sous-espaces.

Autres utilisations. En dehors de l’utilisation en décomposition, il existe d’autres utilisations possibles de l’algorithme 3.24, dont certaines généralisent celles qui ont été décrites au §3.2.3. Nous nous contenterons ici de mentionner ce qui peut être considéré comme l’extension de la méthode de Newton évoquée au §3.2.3.3, extension au cas avec contraintes explicites. En plus du choix (3.11) de la fonction

¹⁴En réalité, comme il a été dit au commentaire 2.26-g du Chapitre 2, il n’est pas nécessaire qu’il y ait autant d’éléments dans la décomposition duale que dans la décomposition primale.

auxiliaire K^k (qui constitue, avec les termes linéaires correctifs apportés par le PPA, une approximation quadratique de la fonction coût originale), on cherche à approximer linéairement les contraintes par le choix suivant de la fonction auxiliaire Ω^k (en supposant à nouveau que Θ se réduit à Θ dans le problème original) :

$$\Omega^k(u) = \Theta'(u^k) \cdot u . \quad (3.54)$$

On obtient donc un problème auxiliaire à coût quadratique et contraintes (explicites) linéaires qui, en l'absence de contraintes implicites liées à U^{ad} , peut être résolu analytiquement. Cette famille d'algorithmes est aujourd'hui désignée par le sigle SQP (Sequential Quadratic Programming) dans la littérature. Le même type d'approche a été appliquée depuis assez longtemps à la résolution des problèmes de commande optimale (sous le nom de “quasi-linéarisation”) en prenant appui sur le fait que les problèmes de commande optimale linéaires-quadratiques obtenus à chaque itération ont une solution analytique (via la fameuse équation de Riccati).

3.3.3.3 Convergence

On ne dispose d'une étude de convergence de l'Algorithme 3.24 que dans la situation assez restrictive suivante (voir [6]).

- Les fonctions coûts \mathcal{J} et K sont quadratiques tandis que \mathfrak{J} est identiquement nulle.¹⁵ On posera donc, avec deux opérateurs linéaires auto-adjoints et monotones \mathbf{K} et \mathbf{J} de \mathcal{U} dans \mathcal{U} (ou \mathcal{U}^*) et un point \mathbf{j} dans \mathcal{U} (ou \mathcal{U}^*),

$$K(u) = \frac{1}{2} \langle u, \mathbf{K}u \rangle , \quad J(u) = \mathcal{J}(u) = \frac{1}{2} \langle u, \mathbf{J}u \rangle + \langle \mathbf{j}, u \rangle .$$

- Les contraintes (3.52b) sont des contraintes égalité, c'est-à-dire que $C = \{0\}$.
- Les fonctions Ω et Θ sont affines tandis que \mathfrak{T} est identiquement nulle. On posera donc, avec deux opérateurs linéaires \mathbf{O} et \mathbf{T} de \mathcal{U} dans \mathcal{C} et un point \mathbf{t} dans \mathcal{C} , ainsi qu'un scalaire $\gamma > 0$ (dont on verra l'utilité plus loin),

$$\Omega(u) = \gamma \mathbf{O}u , \quad \Theta(u) = \Theta(u) = \mathbf{T}u + \mathbf{t} .$$

- Le sous-ensemble U^{ad} est égal à tout l'espace \mathcal{U} .

Sous ces hypothèses, les conditions d'optimalité du problème (3.52) s'écrivent :

$$\begin{aligned} \mathbf{K}(u^{k+1} - u^k) + \varepsilon(\mathbf{J}u^k + \mathbf{j}) + \gamma \mathbf{O}^*(p^{k+1} - p^k) + \varepsilon \mathbf{T}^* p^k &= 0 , \\ \gamma \mathbf{O}(u^{k+1} - u^k) + \varepsilon(\mathbf{T}u^k + \mathbf{t}) &= 0 , \end{aligned}$$

où, par exemple, \mathbf{T}^* désigne l'opérateur adjoint de \mathbf{T} . Par ailleurs, les conditions d'optimalité caractérisant une solution (u^\sharp, p^\sharp) du problème (3.36) s'écrivent :

$$(\mathbf{J}u^\sharp + \mathbf{j}) + \mathbf{T}^* p^\sharp = 0 , \quad (3.55a)$$

$$\mathbf{T}u^\sharp + \mathbf{t} = 0 . \quad (3.55b)$$

La combinaison de ces deux systèmes d'équations donne :

$$\mathbf{K}(u^{k+1} - u^k) + \varepsilon(\mathbf{J}(u^k - u^\sharp) + \mathbf{j}) + \gamma \mathbf{O}^*(p^{k+1} - p^k) + \varepsilon \mathbf{T}^*(p^k - p^\sharp) = 0 , \quad (3.56a)$$

$$\gamma \mathbf{O}(u^{k+1} - u^k) + \varepsilon \mathbf{T}(u^k - u^\sharp) = 0 . \quad (3.56b)$$

¹⁵Cette dernière restriction, comme celle sur \mathfrak{T} , pourraient probablement être levées assez facilement mais nous conservons ici la version la plus simple.

Théorème 3.29. *Sous les hypothèses suivantes :*

(H1) \mathbf{K} et \mathbf{J} sont fortement monotones et Lipschitziens ;

(H2) \mathbf{O} et \mathbf{T} sont surjectifs ;

alors

(C1) les solutions (u^{k+1}, p^{k+1}) de (3.56) et (u^\sharp, p^\sharp) de (3.55) sont uniques.

Si, de plus

(H3) l'opérateur (symétrique)

$$\mathbf{TJ}^{-1}\mathbf{O}^* + \mathbf{OJ}^{-1}\mathbf{T}^* \quad (3.57)$$

est fortement monotone ;

alors,

(C2) ε peut être choisi assez petit et γ assez grand pour que

$$2\mathbf{K} - \varepsilon\mathbf{J} \quad (3.58a)$$

et

$$\gamma(\mathbf{TJ}^{-1}\mathbf{O}^* + \mathbf{OJ}^{-1}\mathbf{T}^*) - \mathbf{TJ}^{-1}(\mathbf{K} + \varepsilon\mathbf{J}/2)\mathbf{J}^{-1}\mathbf{T}^* \quad (3.58b)$$

soient fortement monotones, et, dans ces conditions,

(C3) l'itération (3.56) produit une suite $\{(u^k, p^k)\}$ qui converge fortement vers (u^\sharp, p^\sharp) .

La démonstration de ce théorème peut être trouvée dans [6]. Elle est basée, comme les démonstrations précédentes, sur l'étude d'une fonction de Lyapounov, en l'occurrence

$$\ell(u, p) = \frac{1}{2} \langle u - u^\sharp, \mathbf{J}(u - u^\sharp) \rangle + \langle p - p^\sharp, \mathbf{T}(u - u^\sharp) \rangle + \langle p - p^\sharp, \mathbf{TJ}^{-1}\mathbf{T}^*(p - p^\sharp) \rangle. \quad (3.59)$$

Les deux premiers termes ci-dessus sont construits à partir du Lagrangien du problème (3.36) (on a ramené à 0 sa valeur et celles de ses dérivées en (u^\sharp, p^\sharp)) ; le dernier terme est ajouté pour rendre ℓ fortement convexe en (u, p) .

Exercice 3.30. Montrer que ℓ est fortement convexe en (u, p) . Quel est le rapport du dernier terme avec $\psi(p)$ définie par (2.4) ?

Commentaires 3.31.

- (a) L'hypothèse (H3) est une hypothèse géométrique *préalable* sans laquelle on ne peut garantir que l'algorithme converge pour un bon choix de ε (et de γ). Cette situation est nouvelle par rapport à celle des deux précédents algorithmes étudiés dans ce chapitre. Pour ceux-ci, des hypothèses classiques de forte monotonie et de Lipschitz étaient suffisantes pour permettre de fixer une plage de valeurs admissibles pour les pas de l'algorithme.

On doit attribuer la nécessité de cette condition géométrique préalable au fait que la fonction auxiliaire (3.51) qui a été choisie ici n'est pas fortement concave en p comme l'était par exemple (3.38).

On rappelle que le choix (3.51) a été dicté par la nécessité de retrouver un problème auxiliaire de point selle qui se réinterprète comme un problème de minimisation sous contrainte.

- (b) Sous l'hypothèse (H1), il doit être clair que l'opérateur (3.58a) peut effectivement être rendu fortement monotone par un choix approprié de ε .
- (c) Sous l'hypothèse (H3), il doit être également clair qu'on peut assurer la forte monotonie de l'opérateur (3.58b) pour γ assez grand.

Exercice 3.32. Montrer que le choix dicté par le commentaire 3.31-b ci-dessus correspond exactement à (3.15).

3.3.3.4 Discussion de la condition géométrique

Les commentaires ci-dessus montre que le point crucial est la vérification de la condition géométrique (H3).

On observe d'abord que cette condition est vérifiée si on peut choisir \mathbf{O} égal à \mathbf{T} (la monotonie de (3.57) est alors évidente, et la *forte* monotonie résulte de l'hypothèse de surjectivité de ces opérateurs). Mais la nécessité de décomposer le problème auxiliaire impose de choisir \mathbf{O} bloc-diagonal comme discuté au §3.3.3.2. Par exemple, au §2.1.4, \mathbf{O} était proportionnel à la partie bloc-diagonale de \mathbf{T} . Ainsi si \mathbf{T} est bloc-diagonal, ou très proche de cette structure parce que les termes “hors-diagonaux” sont très petits devant les blocs de la diagonale, on voit que l'hypothèse (H3) est vérifiée sans difficulté. C'est la situation de *faible couplage* entre sous-problèmes par les contraintes.

Cependant, cette situation ne semble pas être la seule qui soit favorable au respect de l'hypothèse (H3). Voici un autre exemple. Supposons que $\mathbf{T} = \mathbf{O} + \alpha \mathbf{O} \mathbf{S} \mathbf{J}$, où α est un nombre positif, $\mathbf{S} : \mathcal{U} \rightarrow \mathcal{U}$ est un opérateur “antisymétrique”, en ce sens que $\mathbf{S}^* = -\mathbf{S}$, alors que \mathbf{O} est supposé bien sûr bloc-diagonal. C'est donc le terme $\alpha \mathbf{O} \mathbf{S} \mathbf{J}$ qui est la cause du couplage entre sous-problèmes. Le calcul de l'expression (3.57) donne dans ce cas

$$(\mathbf{O} + \alpha \mathbf{O} \mathbf{S} \mathbf{J}) \mathbf{J}^{-1} \mathbf{O}^* + \mathbf{O} \mathbf{J}^{-1} (\mathbf{O} + \alpha \mathbf{O} \mathbf{S} \mathbf{J})^* = 2\mathbf{O} \mathbf{J}^{-1} \mathbf{O}^*$$

en prenant en compte l'antisymétrie de \mathbf{S} . On voit donc à nouveau que l'hypothèse (H3) est vérifiée. Ceci étant vrai pour toute valeur positive de α , on constate que le couplage par la contrainte lié au terme $\alpha \mathbf{O} \mathbf{S} \mathbf{J}$ peut devenir arbitrairement grand sans remettre en cause cette conclusion.

D'une manière générale, la façon de distribuer les contraintes aux sous-problèmes discutée au §2.1.4 est en relation directe avec la géométrie de \mathbf{O} et donc avec la satisfaction de la contrainte (H3).

Pour pouvoir énoncer le Théorème 3.29, on a supposé que $U^{\text{ad}} = \mathcal{U}$, sachant que U^{ad} a toujours servi à exprimer des contraintes “locales” (ou décomposées). En présence de contraintes locales, à condition qu'il s'agisse de contraintes égalité (seul cas couvert par le théorème), il faut traiter ces contraintes comme les autres pour pouvoir se replacer dans le cadre du théorème, c'est-à-dire qu'il faut les incorporer à la définition des opérateurs \mathbf{T} et \mathbf{O} . L'exemple qui suit montre que la présence de telles contraintes locales peut contribuer à faire échouer l'hypothèse (H3).

Exemple 3.33. On considère le problème suivant

$$\begin{aligned} & \min \frac{1}{2} (x_1^2 + x_2^2 + x_3^2) \\ & \text{sous} \\ & x_1 - x_2 + 2x_3 = 0, \\ & x_1 + 3x_2 + x_3 = 0, \end{aligned}$$

avec la décomposition $u_1 = (x_1, x_2)$ et $u_2 = x_3$. On alloue la première contrainte au premier sous-problème et la seconde au second. Ceci conduit à définir

$$\mathbf{T} = \begin{pmatrix} 1 & -1 & 2 \\ 1 & 3 & 1 \end{pmatrix}, \quad \mathbf{O} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

alors que \mathbf{J} est l'identité. L'expression (3.57) est égale à

$$\begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix},$$

matrice évidemment définie positive.

Ajoutons maintenant au problème ci-dessus la contrainte $x_2 = 0$, contrainte évidemment allouée au premier sous-problème. Alors,

$$\mathbf{T} = \begin{pmatrix} 1 & -1 & 2 \\ 0 & 1 & 0 \\ 1 & 3 & 1 \end{pmatrix}, \quad \mathbf{O} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

ce qui donne maintenant pour (3.57) :

$$\begin{pmatrix} 4 & -2 & 0 \\ -2 & 2 & 3 \\ 0 & 3 & 2 \end{pmatrix}.$$

Cette matrice n'est pas définie positive, ne serait-ce que parce que son déterminant, égal à -28 , est négatif.

3.3.3.5 Obtention de l'algorithme d'Arrow-Hurwicz

Comme nous l'avons vu au §2.1.4, l'algorithme de coordination “à la Arrow-Hurwicz” (2.22) repose initialement sur une manipulation du problème (2.1) qui remplace ce problème par le problème (2.21) équivalent. Nous considérons maintenant une manipulation du même genre à partir du problème (3.36), en nous limitant cependant au cas de contraintes égalité ($C = \{0\}$) afin de simplifier le propos. On introduit une nouvelle application $\Xi : \mathcal{U} \rightarrow \mathcal{C}$ et une nouvelle variable primale $w \in \mathcal{C}$: il est clair que (3.36) est équivalent à

$$\min_{u \in U^{\text{ad}}, w \in \mathcal{C}} J(u) \quad (3.60a)$$

sous

$$\Xi(u) + w = 0, \quad (3.60b)$$

$$\Theta(u) - \Xi(u) - w = 0. \quad (3.60c)$$

Le Lagrangien correspondant s'écrit

$$L(u, w; p, \mu) = J(u) + \langle p, \Xi(u) + w \rangle + \langle \mu, \Theta(u) - \Xi(u) - w \rangle. \quad (3.61)$$

Dans ce Lagrangien, il convient, comme d'habitude, de distinguer la partie \mathcal{L} qui sera “linéarisée” dans le problème auxiliaire et la partie \mathcal{L} qui sera conservée telle quelle. Pour faire ce choix, il convient de ne pas perdre de vue que les variables (u, p) seront traitées simultanément à un niveau de l'algorithme (ou, si l'on préfère, à une étape dans un schéma séquentiel) alors que (w, μ) seront traitées à un autre niveau (ou à une autre étape du schéma séquentiel). De ces considérations, on déduit entre autres les choix suivants.

- Sachant qu'on choisira bien sûr Ξ bloc-diagonal par rapport à une décomposition (3.7)–(3.53), le terme $\langle p, \Xi(u) \rangle$ est additif en (u_i, p_i) et a donc sa place dans \mathcal{L} .

En fait, on choisira

$$\Xi = \mathfrak{T} + (1/\varepsilon)\Omega, \quad (3.62)$$

où \mathfrak{T} est comme d'habitude la partie bloc-diagonale de Θ , $\Omega : \mathcal{U} \rightarrow \mathcal{C}$ est une nouvelle application auxiliaire, et l'introduction du nombre positif ε à ce stade est une astuce non essentielle, mais destinée simplement à faciliter la comparaison avec l'algorithme par prédiction précédemment rencontré.

- Le terme $\langle p, w \rangle$, bien qu'additif en (p_i, w_i) , doit être rangé dans \mathcal{L} parce que les variables p et w ne sont pas traitées au même stade de l'algorithme.

- On peut en dire de même des termes “croisés” en (u, μ) .
- Plus intéressant est le cas du terme $\langle \mu, w \rangle$: puisque ce terme est additif en (w_i, μ_i) et que les deux variables seront traitées au même stade de l’algorithme, on devrait inclure ce terme dans la partie \mathcal{L} . Si on le fait, on obtiendra cependant un problème “couplé” en (w, μ) : ceci conduit directement à la version que nous avons qualifiée d’“implicite” au Chapitre 2 (voir (2.22c)–(2.22d) à la fin du §2.1.4).

Si on conserve $\langle \mu, w \rangle$ dans la partie linéarisée \mathcal{L} , on obtiendra deux sous-problèmes indépendants en w et en μ : ceci conduit directement à la version parallèle (2.22). C’est cette version que nous allons d’abord étudier.

Ces considérations étant faites, on est donc amené à décomposer L dans (3.61) de la façon suivante :

$$L = \mathcal{L} + \mathcal{L}, \quad (3.63a)$$

avec

$$\mathcal{L} = \mathcal{J}(u) + \langle p, w \rangle + \langle \mu, \Theta(u) - (1/\varepsilon)\Omega(u) - w \rangle, \quad (3.63b)$$

$$\mathcal{L} = \mathfrak{J}(u) + \langle p, \mathfrak{T}(u) + (1/\varepsilon)\Omega(u) \rangle. \quad (3.63c)$$

On considère alors la fonction auxiliaire :

$$\Lambda(u, w; p, \mu) = K(u) + \frac{1}{2\alpha} \|w\|^2 - \frac{1}{2\beta} \|\mu\|^2, \quad (3.64)$$

où α et β sont des nombres positifs (comparer cette fonction auxiliaire à (3.38) et (3.51)). Observons que si K est fortement convexe, cette fonction est également fortement convexe en (u, w) et fortement concave en μ , mais elle ne l’est toujours pas en p (dont elle ne dépend pas en fait). Ce choix est encore dicté par la préoccupation de pouvoir réinterpréter le problème auxiliaire en (u, p) comme un problème de minimisation sous contrainte.

La différence notable avec la situation de l’algorithme précédent est qu’ici, \mathcal{L} (la partie linéarisée de L) ne dépend de p que par le terme $\langle p, w \rangle$ dont la géométrie est “plus simple” que celle du terme $\langle p, \Theta(u) \rangle$ qu’il fallait précédemment ranger dans \mathcal{L} . Ceci peut expliquer, dans le théorème de convergence qui va suivre, la disparition d’une hypothèse géométrique du type (H3) du Théorème 3.29.

Il faut maintenant mener des calculs analogues à ceux qui sont effectués dans (3.37) en ne perdant pas de vue le fait que les variables sont maintenant (u, p, w, μ) et que le calcul du point selle de (3.37a) est traité séquentiellement : d’abord on recherche un point selle en (u, p) à (w^k, μ^k) fixés, ce qui donne une solution (u^{k+1}, p^{k+1}) , puis on cherche un point selle en (w, μ) avec ces valeurs fixées de (u, p) . On aboutit finalement à l’algorithme suivant (Arrow-Hurwicz parallèle).

Algorithme 3.34. 1. Commencer avec u^0, p^0, w^0, μ^0 ; poser $k = 0$.

2. A l’itération $k + 1$, résoudre d’abord

$$\min_{u \in U^{\text{ad}}} \left(K(u) + \varepsilon \mathfrak{J}(u) + \langle \varepsilon \mathcal{J}'(u^k) - K'(u^k), u \rangle + \langle \mu^k, (\varepsilon \Theta'(u^k) - \Omega'(u^k)) \cdot u \rangle \right) \quad (3.65a)$$

sous

$$\Omega(u) + \varepsilon \mathfrak{T}(u) + \varepsilon w^k = 0. \quad (3.65b)$$

soit u^{k+1} une solution et p^{k+1} un multiplicateur optimal associé à la contrainte (3.65b).

3. Remettre ensuite à jour (w, μ) par

$$w^{k+1} = w^k - \rho_1(p^{k+1} - \mu^k), \quad (3.65c)$$

$$\mu^{k+1} = \mu^k + \rho_2(\Theta(u^{k+1}) - (1/\varepsilon)\Omega(u^{k+1}) - w^k), \quad (3.65d)$$

où $\rho_1 = \alpha\varepsilon$ et $\rho_2 = \beta\varepsilon$.

4. Si $\|u^{k+1} - u^k\| + \|p^{k+1} - p^k\| + \|w^{k+1} - w^k\| + \|\mu^{k+1} - \mu^k\|$ est “suffisamment petite”, stop ; sinon, retourner en 2 avec k incrémenté de 1.

Exercice 3.35. En suivant la démarche évoquée ci-dessus, retrouver (3.65).

Remarque 3.36. Comme déjà remarqué au Chapitre 2, en tenant compte de (3.65b), on peut réécrire (3.65d) de la façon suivante :

$$\mu^{k+1} = \mu^k + \rho_2((\Theta + \mathfrak{T})(u^{k+1})),$$

ce qui rappelle la remise à jour dans la méthode par les prix.

Exercice 3.37. En faisant “basculer” le terme $-\langle \mu, w \rangle$ de \mathcal{L} à \mathfrak{L} (voir (3.63)), reconsidérer les calculs ayant conduit à l’Algorithme 3.34 et retrouver la version implicite (2.22c)–(2.22d) décrite au Chapitre 2.

L’étude de convergence de l’Algorithme 3.34 a été menée dans [11] dans le cas de contraintes égalité, et donc pour des opérateurs Θ et Ω affines, et en l’absence des termes \mathfrak{J} et \mathfrak{T} . Sans énoncer un théorème précis (voir la référence [11]), on peut résumer les hypothèses et conclusions de la façon suivante.

- La fonction coût $J = \mathcal{J}$ est supposée *fortement convexe*, s.c.i. et différentiable de dérivée Lipschitzienne de constante A .
- La fonction $\Theta = \Theta$ est affine et continue.
- La fonction coût auxiliaire K satisfait les hypothèses habituelles (en particulier, elle est fortement convexe de constante b).
- La fonction auxiliaire Ω est affine, continue et *surjective*.
- Dans ces conditions, il existe une solution *unique* (u^{k+1}, p^{k+1}) au problème (3.65a)–(3.65b).
- Pour assurer la convergence, les conditions sur les pas ε , ρ_1 , ρ_2 sont assez compliquées. On commence par imposer à ε la condition $\varepsilon \in]0, b/A[$. Puis il faut choisir ρ_1 dans une plage de valeurs entre 0 et une borne supérieure dépendant de ε . Enfin, il faut imposer à ρ_2 d’appartenir à une plage de valeurs entre 0 et une borne supérieure dépendant de ε et ρ_1 . L’expression explicite de ces bornes supérieures n’est pas facile à donner et résulte de conditions rencontrées dans la démonstration. Ce que l’on peut dire est que la borne supérieure sur ρ_1 s’approche de 0 si ε est choisi très près de sa borne inférieure ou supérieure, et il en va de même pour ρ_2 si ε ou ρ_1 s’approche de ses bornes. On a donc intérêt à chaque fois à viser le milieu des intervalles.
- Moyennant les conditions ci-dessus, la suite $\{u^k\}$ tend fortement vers la solution $u^\#$ unique du problème (3.36) et par conséquent w^k tend vers $-(1/\varepsilon)\Omega(u^\#)$ (d’après (3.65b)).¹⁶
- Les suites $\{p^k\}$ et $\{\mu^k\}$ restent bornées, ont les mêmes points d’accumulation dans la topologie faible, et ces points d’accumulation sont des multiplicateurs optimaux des contraintes du problème (3.36).
- Enfin, si $\Theta = \Theta$ est *surjective*, on peut garantir l’unicité du multiplicateur optimal dans (3.36) et la convergence de p^k et μ^k vers ce multiplicateur a lieu dans la topologie *forte*.

¹⁶On rappelle que le cas considéré ici est celui où \mathfrak{T} est identiquement nulle.

3.4 Annexe : Monotonie, forte monotonie et forte convexité

Définition 3.38. Un opérateur Ψ de l'espace de Hilbert \mathcal{U} dans son dual, éventuellement multivoque, est *fortement monotone* de constante a si

$$\exists a > 0 : \forall u, v, \forall p \in \Psi(u), q \in \Psi(v), \langle p - q, u - v \rangle \geq a \|u - v\|^2. \quad (3.66)$$

Il est *simplement monotone* si l'inégalité est vérifiée au moins avec $a = 0$. Il est *strictement monotone* si l'inégalité de la monotonie simple est une inégalité stricte dès que $u \neq v$.

Le lien avec la convexité est bien connu : une fonction de \mathcal{U} dans \mathbb{R} , différentiable, est (strictement) convexe si et seulement si sa dérivée est un opérateur (strictement) monotone. De même, le sous-différentiel d'une fonction (strictement) convexe est un opérateur multivoque (strictement) monotone.

Le cas d'une dérivée ou d'un sous-différentiel *fortement* monotone correspond à celui d'une fonction *fortement* convexe.

Définition 3.39. On dit qu'une fonction $J : \mathcal{U} \rightarrow \mathbb{R}$ est *fortement convexe* si il existe $a > 0$ tel que pour tout $\alpha \in [0, 1]$, tout u et tout v ,

$$J(\alpha u + (1 - \alpha)v) \leq \alpha J(u) + (1 - \alpha)J(v) - \alpha(1 - \alpha)\frac{a}{2} \|v - u\|^2. \quad (3.67)$$

Lemme 3.40. La propriété (3.66), écrite pour $\Psi = \partial J$, et la propriété (3.67) sont toutes deux équivalentes à l'inégalité suivante : pour tout u , tout v , et tout $r \in \partial J(u)$,

$$J(v) \geq J(u) + \langle r, v - u \rangle + \frac{a}{2} \|v - u\|^2, \quad (3.68)$$

avec évidemment la même valeur de a dans les trois cas.

Démonstration. Dans cette démonstration, (3.68) joue un rôle central et il est plus facile de démontrer que les deux autres propriétés lui sont équivalentes que de chercher à faire une démonstration “circulaire”.

(3.68) \Leftrightarrow (3.67) L'implication de gauche à droite est obtenue en utilisant une première fois (3.68) avec $u \leftarrow \alpha u + (1 - \alpha)v$, $r \in \partial J(\alpha u + (1 - \alpha)v)$ et $v \leftarrow v$, puis une seconde fois avec la même substitution pour u et r mais avec la substitution $v \leftarrow u$. La première inégalité ainsi obtenue étant multipliée par $1 - \alpha$ et la seconde par α , on les additionne et on trouve (3.67).

Réciproquement, partant de (3.67) que l'on peut réécrire

$$(1 - \alpha)(J(v) - J(u)) \geq J(\alpha u + (1 - \alpha)v) - J(u) + \alpha(1 - \alpha)\frac{a}{2} \|v - u\|^2,$$

il vient, avec $r \in \partial J(u)$,

$$(1 - \alpha)(J(v) - J(u)) \geq (1 - \alpha) \langle r, v - u \rangle + \alpha(1 - \alpha)\frac{a}{2} \|v - u\|^2,$$

que l'on divise membre à membre par $1 - \alpha$, puis l'on prend $\alpha = 1$ pour obtenir (3.68).

(3.68) \Leftrightarrow (3.66) L'implication de gauche à droite est obtenue en écrivant (3.68) une seconde fois après avoir interverti le rôle de u et v ($s \in \partial J(v)$ prenant donc la place de r dans cette deuxième version), puis en additionnant les deux versions, ce qui donne (3.66) pour $\Psi = \partial J$.

Réciproquement, on considère la fonction

$$f(\beta) = J((1 - \beta)u + \beta v) . \quad (3.69)$$

Lorsque β varie de 0 à 1, $w(\beta) = (1 - \beta)u + \beta v$ varie de u à v sur la droite joignant ces deux points. La fonction f n'est pas différentiable si J ne l'est pas, mais elle est convexe et admet des dérivées directionnelles de part et d'autre de chaque point. Au point β dans la direction $+1$, on a

$$\begin{aligned} Df(\beta; 1) &= \lim_{\varepsilon \rightarrow 0_+} \left(J(w(\beta + \varepsilon)) - J(w(\beta)) \right) / \varepsilon \\ &\geq \langle r_\beta, v - u \rangle \quad \left(r_\beta \in \partial J(w(\beta)) \right) \\ &\geq \langle r, v - u \rangle + a\beta \|v - u\|^2 \quad \left(r \in \partial J(u) \right), \end{aligned}$$

en utilisant (3.66). D'après cette inégalité, il est clair que

$$\begin{aligned} f(1) - f(0) &\geq \int_0^1 (\langle r, v - u \rangle + a\beta \|v - u\|^2) d\beta \\ &= \langle r, v - u \rangle + a \|v - u\|^2 / 2, \end{aligned}$$

ce qui établit (3.68). □

3.5 Annexe : Propriété de Lipschitz

Définition 3.41. Un opérateur Ψ de l'espace de Hilbert \mathcal{U} dans son dual vérifie la *propriété de Lipschitz*, ou est *Lipschitzien*, si

$$\exists A > 0 : \forall u, v \in \mathcal{U}, \quad \|\Psi(u) - \Psi(v)\| \leq A \|u - v\|. \quad (3.70)$$

Pour un opérateur linéaire, A est un majorant de la norme d'opérateur définie par :

$$\|\Psi\| = \sup_{\|u\| \leq 1} \|\Psi(u)\|.$$

Évidemment, la définition (3.70) n'a de sens que pour un opérateur univoque. On l'applique maintenant à la dérivée d'une fonction.

Lemme 3.42.

1. Si J est différentiable et de dérivée Lipschitzienne avec la constante A , alors, pour tous u, v ,

$$J(v) \leq J(u) + \langle J'(u), v - u \rangle + \frac{A}{2} \|v - u\|^2. \quad (3.71)$$

2. Le fait que J' soit Lipschitzienne avec la constante A est équivalente au fait que J et J' vérifient l'inégalité

$$J(v) \geq J(u) + \langle J'(u), v - u \rangle + \frac{1}{2A} \|J'(v) - J'(u)\|^2. \quad (3.72)$$

Démonstration.

1. On utilise à nouveau la fonction $f(\cdot)$ définie par (3.69). Cette fois-ci, cette fonction est différentiable et l'on a

$$\begin{aligned} f'(\beta) &= \langle J'(w(\beta)), v - u \rangle \\ &= \langle J'(u), v - u \rangle + \langle J'(w(\beta)) - J'(u), v - u \rangle \\ &\leq \langle J'(u), v - u \rangle + \beta A \|v - u\|^2 \end{aligned}$$

en utilisant successivement l'inégalité de Schwarz et l'hypothèse. Ensuite, par intégration de $\beta = 0$ à $\beta = 1$, on obtient le résultat désiré.

2. La propriété de Lipschitz de J' entraînant (3.71), on utilise cette inégalité avec $v - (J'(v) - J'(u))/A$ à la place de v et v à la place de u , ce qui donne

$$J\left(v - \frac{1}{A}(J'(v) - J'(u))\right) \leq J(v) + \frac{1}{A}\left(-\langle J'(v), J'(v) - J'(u) \rangle + \frac{1}{2}\|J'(v) - J'(u)\|^2\right).$$

Par ailleurs, du fait de la convexité de J ,

$$J\left(v - \frac{1}{A}(J'(v) - J'(u))\right) \geq J(u) + \left\langle J'(u), v - u - \frac{1}{A}(J'(v) - J'(u)) \right\rangle.$$

En combinant ces deux inégalités, on obtient (3.72).

Réciproquement, supposant (3.72), on peut récrire cette inégalité en intervertissant le rôle de u et v , puis sommer les deux versions pour obtenir

$$\langle J'(v) - J'(u), v - u \rangle \geq \frac{1}{A} \|J'(v) - J'(u)\|^2. \quad (3.73)$$

En utilisant ensuite l'inégalité de Schwarz pour majorer le premier membre de (3.73), on obtient la propriété de Lipschitz après simplification. \square

On ne peut pas vraiment énoncer une réciproque de la première partie de ce lemme. On peut seulement montrer que (3.71) entraîne la propriété que, pour tous u, v ,

$$\langle J'(u) - J'(v), u - v \rangle \leq A \|u - v\|^2,$$

propriété a priori plus faible que la propriété de Lipschitz.

3.6 Annexe : Propriété de Dunn

Définition 3.43. Un opérateur Ψ de l'espace de Hilbert \mathcal{U} dans son dual vérifie la *propriété de Dunn* avec la constante A si

$$\exists A > 0 : \forall u, v \in \mathcal{U}, \quad \langle \Psi(u) - \Psi(v), u - v \rangle \geq \frac{1}{A} \|\Psi(u) - \Psi(v)\|^2. \quad (3.74)$$

Cette propriété est en fait tout simplement la propriété de forte monotonie avec la constante $1/A$ pour l'opérateur (éventuellement multivoque) Ψ^{-1} (il suffit, pour le voir, de poser $p = \Psi(u)$ et $q = \Psi(v)$, donc $u \in \Psi^{-1}(p)$ et $v \in \Psi^{-1}(q)$). Elle a été nommée ainsi parce que J.C. Dunn [15] a montré que si Ψ est la dérivée J' d'une fonction convexe *deux fois* différentiable, et si J' est Lipschitzienne de constante A , alors Ψ vérifie (3.74).

En fait, comme on va le voir dans la suite, l'hypothèse de régularité C^2 de J est inutile. De plus, dans le cas d'un opérateur "non symétrique", c'est-à-dire non égal à la dérivée d'une fonction J (ou "ne dérivant pas d'un potentiel"), la propriété de Dunn est plus forte, mais aussi plus intéressante du point de vue de la convergence des algorithmes, que la propriété de Lipschitz qu'elle entraîne (en particulier, cela montre que Ψ est nécessairement univoque).

Il est cependant possible de montrer une forme d'équivalence entre les deux propriétés (mais avec des constantes différentes) pour des opérateurs fortement monotones.

Lemme 3.44.

1. Si Ψ vérifie la propriété de Dunn avec la constante A , alors Ψ est monotone et Lipschitzienne de constante A .
2. Si Ψ est fortement monotone de constante a et Lipschitzienne de constante A , alors Ψ vérifie la propriété de Dunn avec la constante A^2/a .
3. Si Ψ est la dérivée J' d'une fonction J , et si J' est Lipschitzienne de constante A , J' vérifie aussi la propriété de Dunn avec la même constante.
4. Les propriétés de Dunn et de Lipschitz de J' et l'inégalité (3.72) sont trois énoncés équivalents.

Démonstration.

1. La propriété (3.74) entraîne évidemment la monotonie de Ψ . De plus, par l'inégalité de Schwarz,

$$\langle \Psi(u) - \Psi(v), u - v \rangle \leq \|\Psi(u) - \Psi(v)\| \|u - v\| ,$$

et donc

$$\frac{1}{A} \|\Psi(u) - \Psi(v)\|^2 \leq \langle \Psi(u) - \Psi(v), u - v \rangle .$$

En simplifiant par $\|\Psi(u) - \Psi(v)\|$, on obtient l'inégalité de Lipschitz pour Ψ avec la constante A .

2. Supposons maintenant que Ψ est fortement monotone et Lipschitzienne de constantes respectives a et A :

$$\begin{aligned} \langle \Psi(u) - \Psi(v), u - v \rangle &\geq a \|u - v\|^2 , \\ \|\Psi(u) - \Psi(v)\|^2 &\leq A^2 \|u - v\|^2 . \end{aligned}$$

En combinant ces deux inégalités,

$$\langle \Psi(u) - \Psi(v), u - v \rangle \geq \frac{a}{A^2} \|\Psi(u) - \Psi(v)\|^2 ,$$

ce qui montre que Ψ a la propriété (3.74) avec la constante A^2/a .

3. On a déjà vu dans la preuve du Lemme 3.42 que la propriété de Lipschitz entraîne (3.72) qui entraîne (3.73) qui n'est autre que la propriété de Dunn de J' .
4. Cette équivalence des trois propositions est un simple corollaire de tout ce qui précède. □

Commentaires 3.45. Ce lemme affirme donc l'équivalence des propriétés de Dunn et Lipschitz dans le cas symétrique. Dans le cas non symétrique, la propriété de Dunn est plus forte que (puisqu'elle implique) la propriété de Lipschitz. Une réciproque est possible seulement dans le cas fortement monotone, mais dans ce cas la constante de Lipschitz A est en général inférieure à la constante de Dunn (estimée) A^2/a (car bien sûr $A/a \geq 1$). L'exemple suivant montre que pour un opérateur non symétrique, la propriété de Lipschitz n'implique pas la propriété de Dunn en général.

Exemple 3.46. La matrice $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ qui représente la rotation de $\pi/2$ dans le plan est évidemment un opérateur linéaire Lipschitzien, mais cet opérateur ne vérifie pas la propriété (3.74).

Soit \mathfrak{J} une fonction à valeurs réelles et U^{ad} un sous-ensemble convexe fermé de \mathcal{U} . On considère l'inéquation variationnelle : trouver $u^\sharp \in U^{\text{ad}}$ tel que

$$\forall u \in U^{\text{ad}}, \quad \langle \Psi(u^\sharp), u - u^\sharp \rangle + \mathfrak{J}(u) - \mathfrak{J}(u^\sharp) \geq 0. \quad (3.75)$$

On sait que si Ψ est fortement monotone, u^\sharp est unique. De façon analogue, si Ψ vérifie la propriété de Dunn, c'est $\Psi(u^\sharp)$ qui est unique même si u^\sharp ne l'est pas.

Lemme 3.47. Soit u^\sharp et u^* deux solutions de (3.75). Alors $\Psi(u^\sharp) = \Psi(u^*)$.

Démonstration. On a simultanément

$$\langle \Psi(u^\sharp), u^* - u^\sharp \rangle + \mathfrak{J}(u^*) - \mathfrak{J}(u^\sharp) \geq 0$$

et

$$\langle \Psi(u^*), u^\sharp - u^* \rangle + \mathfrak{J}(u^\sharp) - \mathfrak{J}(u^*) \geq 0,$$

d'où, par addition,

$$\langle \Psi(u^\sharp) - \Psi(u^*), u^* - u^\sharp \rangle \geq 0,$$

mais avec (3.74),

$$-\frac{1}{A} \|\Psi(u^\sharp) - \Psi(u^*)\|^2 \geq 0,$$

ce qui établit le résultat. □

3.7 Annexe : Problèmes sous contraintes et dualité

On se reportera à la référence [9] pour un traitement plus exhaustif.

3.7.1 Formulation et relation d'ordre dans un espace de Hilbert

On considère des problèmes de la forme suivante :

$$\min_{u \in U^{\text{ad}}} J(u) \quad (3.76a)$$

$$\text{sous } \Theta(u) \in -C, \quad (3.76b)$$

où

- J est une fonction s.c.i. d'un espace de Hilbert \mathcal{U} dans \mathbb{R} ;
- U^{ad} est un sous-ensemble convexe fermé de \mathcal{U} ;
- Θ est une application de \mathcal{U} dans un autre espace de Hilbert \mathcal{C} ;
- C est un cône convexe fermé de \mathcal{C} tel que

$$C \cap (-C) = \{0\} . \quad (3.77)$$

En introduisant C , le but est de munir l'espace \mathcal{C} d'une *relation d'ordre* notée \succeq ainsi définie :

$$\theta_1 \succeq \theta_2 \Leftrightarrow \theta_1 - \theta_2 \in C .$$

Cette relation est bien

- réflexive : $\theta \succeq \theta$ car $0 \in C$;
- transitive : si $\theta_1 - \theta_2 \in C$ et $\theta_2 - \theta_3 \in C$, alors $\theta_1 - \theta_3 \in C$ (en effet, dans un cône convexe, si x_1 et x_2 sont dans C , $y_1 = (1/\alpha)x_1$ et $y_2 = (1/(1-\alpha))x_2$ le sont aussi, pour $\alpha \in]0, 1[$, et donc $\alpha y_1 + (1-\alpha)y_2 = x_1 + x_2$ l'est aussi) ;
- antisymétrique : si $\theta_1 - \theta_2 \in C$ et $\theta_2 - \theta_1 \in C$, alors $\theta_1 - \theta_2 = 0$, ceci grâce à (3.77).

Un exemple classique est le cas où $\mathcal{C} = \mathbb{R}^m$ et l'on prend

$$C = (\mathbb{R}^+)^m = \{x \in \mathbb{R}^m \mid x_i \geq 0, i = 1, \dots, m\} .$$

Cette relation d'ordre est compatible avec la structure d'espace vectoriel, c'est-à-dire, si $\theta_1 \succeq 0$ et $\theta_2 \succeq 0$, alors $\theta_1 + \theta_2 \succeq 0$ (comme déjà vu ci-dessus pour un cône convexe) et $\alpha\theta \succeq 0$ pour $\alpha \geq 0$ (c'est la propriété même de cône). De plus, la limite d'une suite d'éléments positifs est un élément positif (car C est fermé).

Le cône C est appelé "*cône positif primal*". La contrainte (3.76b) doit être comprise comme un ensemble de contraintes inégalité, éventuellement en dimension infinie. En fait, le cas des contraintes égalité est aussi couvert par cette écriture : il suffit de considérer $C = \{0\}$ qui est bien encore un cône convexe. Plus généralement, le mélange de contraintes inégalité et égalité se traite en considérant un cône du type $C = C_i \times C_e$ avec $C_e = \{0\}$: C_e correspond au sous-espace dans lequel évolue le second membre des contraintes égalité.

Remarque 3.48. Notons qu'en plus des contraintes (3.76b), il existe toujours des contraintes implicites du type $u \in U^{\text{ad}}$, mais seules les premières vont être dualisées.

3.7.2 Convexité et croissance en relation avec un cône positif

À partir de la relation d'ordre sur \mathcal{C} définie par le cône C , une notion de convexité pour des fonctions à valeurs dans \mathcal{C} peut être définie. On dit que $\Theta : \mathcal{U} \rightarrow \mathcal{C}$ est *C-convexe* si

$$\forall u, v \in \mathcal{U}, \quad \forall \alpha \in [0, 1], \quad \alpha\Theta(u) + (1-\alpha)\Theta(v) - \Theta(\alpha u + (1-\alpha)v) \in C . \quad (3.78)$$

Notons que si $C = \{0\}$, ceci signifie simplement que Θ est affine.

De même, pour une fonction ϕ de \mathcal{C} dans \mathbb{R} , on dit que ϕ est *C-non décroissante* si

$$\theta_1 - \theta_2 \in C \Rightarrow \phi(\theta_1) \geq \phi(\theta_2) . \quad (3.79)$$

Exercice 3.49. Montrer que la composition $\phi \circ \Theta$ d'une fonction *C-convexe* Θ avec une fonction *C-non décroissante* et convexe ϕ est convexe.

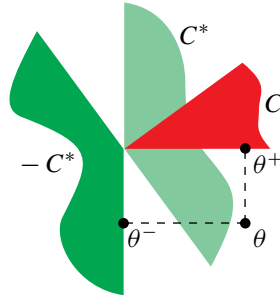


Figure 3.2: Cône, cône adjoint et décomposition orthogonale

3.7.3 Cône positif dual

De même qu'une relation d'ordre a été introduite dans l'espace \mathcal{C} par l'intermédiaire d'un cône convexe fermé C , une relation d'ordre peut être introduite dans l'espace dual \mathcal{C}^* des multiplicateurs associés aux contraintes (3.76b). Bien qu'un espace de Hilbert soit identifiable à son dual, il n'est pas approprié d'utiliser nécessairement *le même* cône positif dans \mathcal{C} et \mathcal{C}^* . Une fois introduit C , cône positif primal, on introduit le “cône positif dual”, ou “cône adjoint”, ou encore “cône conjugué”, noté C^* et ainsi défini :

$$C^* = \{ p \in \mathcal{C}^* \mid \langle p, \theta \rangle \geq 0, \forall \theta \in C \} . \quad (3.80)$$

Cette définition est illustrée par la Figure 3.2. On appelle parfois $-C^*$ le “cône polaire” de C . Dans le cas classique où $C = (\mathbb{R}^+)^m$, C^* coïncide avec C (si \mathbb{R}^n est identifié à son dual). Mais ce n'est pas le cas en général, et c'est pourquoi nous préférons les notations $\theta \in C$ ou $\theta \in C^*$, plutôt que $\theta \geq 0$ qui serait ambiguë lorsque l'espace \mathcal{C} est identifié à son dual.

La Figure 3.2 illustre également la décomposition orthogonale et unique de tout élément $\theta \in \mathcal{C}$ en deux éléments :¹⁷ $\theta^+ = \text{proj}_C \theta$ et $\theta^- = \text{proj}_{-C^*} \theta$ tels que $\langle \theta^+, \theta^- \rangle = 0$.

Exercice 3.50. Montrer que si $\Theta : \mathcal{U} \rightarrow \mathcal{C}$ est C -convexe et si $p \in C^*$, alors la fonction à valeurs réelles $u \mapsto \langle p, \Theta(u) \rangle$ est convexe (au sens ordinaire du terme).

3.7.4 Lagrangien et dualité

Un outil de base de la dualité est le *Lagrangien* associé au problème (3.76), à savoir

$$L(u, p) = J(u) + \langle p, \Theta(u) \rangle . \quad (3.81)$$

On l'appelle aussi “Lagrangien ordinaire” par opposition au Lagrangien augmenté dont il est aussi question dans ce cours.

Un *point selle* de L sur $U^{\text{ad}} \times C^*$ est un point (u^\sharp, p^\sharp) dans ce sous-ensemble et tel que

$$\forall (u, p) \in U^{\text{ad}} \times C^*, \quad L(u^\sharp, p) \leq L(u^\sharp, p^\sharp) \leq L(u, p^\sharp) . \quad (3.82)$$

Exercice 3.51. Démontrer que l'ensemble des points selle de L est un “rectangle” en ce sens que si (u_1^\sharp, p_1^\sharp) et (u_2^\sharp, p_2^\sharp) sont des points selle, alors (u_1^\sharp, p_2^\sharp) et (u_2^\sharp, p_1^\sharp) le sont aussi.

Exercice 3.52. Démontrer que si (u^\sharp, p^\sharp) est un point selle de L sur $U^{\text{ad}} \times C^*$, alors u^\sharp est une solution de (3.76).

¹⁷Bien sûr, on doit ici encore identifier \mathcal{C}^* à \mathcal{C} par le théorème de représentation de Riesz pour que la somme des deux éléments ait un sens.

Remarque 3.53. Le fait que p est restreint à C^* est l'équivalent en dimension infinie du fait que les multiplicateurs sont non négatifs pour des contraintes inégalité. On doit en effet noter que les contraintes égalité correspondent au cas où $C = \{0\}$, et que dans ce cas, en appliquant la définition (3.80), on a $C^* = \mathbb{C}^*$, ce qui signifie qu'il n'y a pas de contrainte de signe sur les multiplicateurs.

La réciproque de l'énoncé de l'Exercice 3.52 repose sur les conditions *suffisantes* sous lesquelles une fonction admet un point selle. Essentiellement, il y a trois types d'hypothèses :

- des hypothèses de convexité-concavité ;
- des hypothèses de semi-continuité inférieure-supérieure destinées à garantir l'atteignabilité des minima-maxima ;
- des hypothèses de coercivité destinées à garantir qu'ils sont atteints à distance finie.

Dans le cas particulier du Lagrangien du problème d'optimisation (3.76), on est donc conduit à supposer que

- U^{ad} est convexe fermé ainsi que le cône C ;
- J est convexe, s.c.i. et coercive (voir (3.2)) ;
- Θ est C -convexe¹⁸ (et il faut faire une hypothèse de continuité pour garantir que $u \mapsto \langle p, \Theta(u) \rangle$ est s.c.i pour tout $p \in C^*$ — dans la pratique, pour la convergence des algorithmes, on supposera même que Θ est Lipschitzienne) ;
- une certaine *condition de qualification des contraintes* est satisfaite (voir ci-dessous).

Le rôle de cette qualification des contraintes est d'assurer la coercivité de L dans l'espace dual \mathbb{C}^* , c'est-à-dire par rapport à p . Elle s'énonce

$$0 \in \text{int}(\Theta(U^{\text{ad}}) + C) . \quad (3.83)$$

Moyennant les hypothèses ci-dessus, on peut affirmer l'existence d'un point selle du Lagrangien du problème (3.76). L'Exercice 2.29 en fin de Chapitre 2 a cependant montré que ces hypothèses ne sont pas *nécessaires* pour qu'un point selle existe.

3.8 Corrigé des exercices

3.8.1 Corrigé de l'Exercice 3.4

Avec $K(\cdot) = 1/2 \|\cdot\|^2$ et \mathfrak{J} identiquement nul, le problème (3.8c) peut être écrit

$$\min_{u \in U^{\text{ad}}} \left(\frac{1}{2} \|u - u^k + \varepsilon^k r^k\|^2 \right) - \frac{1}{2} \|u^k - \varepsilon^k r^k\|^2 .$$

Par définition de la projection sur un convexe (point à distance minimale d'un point donné et appartenant au convexe), on voit que la solution de ce problème est donnée par la formule (3.10) avec éventuellement r^k remplacé par $\nabla J(u^k)$.

¹⁸donc, comme on l'a vu, affine lorsque $C = \{0\}$ — cas des contraintes égalité

3.8.2 Corrigé de l'Exercice 3.12

Il suffit de noter qu'en raison de (3.12),

$$(K_i^k)'(u_i) = \mathcal{J}'_{u_i}(u_1^k, \dots, u_{i-1}^k, u_i, u_{i+1}^k, \dots, u_N^k),$$

et que cette expression peut s'écrire $\mathcal{J}'_{u_i}(u^k)$ pour $u_i = u_i^k$. Or ceci est la valeur qu'il faut donner à la coordonnée r_i^k de r^k dans (3.8c). Les corrections linéaires apparaissant dans ce problème auxiliaire s'annulent donc pour le choix $\varepsilon^k = 1$.

3.8.3 Corrigé de l'Exercice 3.13

En raison du choix (3.34) pour la dépendance en u_i de la fonction auxiliaire (augmenté de $(\gamma/2) \|u_i\|^2$) et en prenant $\varepsilon^k = 1$ pour calculer les corrections linéaires en u , on a à nouveau une simplification de la correction linéaire en u_i , sauf pour le terme $-\gamma \langle u_i^k, u_i \rangle$, qui, combiné avec $(\gamma/2) \|u_i\|^2$, donne le dernier terme de (3.35) à une constante près, à savoir $(\gamma/2) \|u_i^k\|^2$; ce problème (3.35) constitue donc bien le sous-problème auxiliaire en u_i . Pour ce qui est du sous-problème auxiliaire en v (résolu une fois que u^{k+1} a été calculé, selon le schéma séquentiel), la dépendance de la fonction auxiliaire (3.33) par rapport à v et la correction linéaire en v (avec le "pas" ρ^k) conduit à

$$\min_v \left(\frac{1}{2} \|v\|^2 + \langle \rho^k J'_v(u^{k+1}, v^k) - v^k, v \rangle \right).$$

La résolution explicite de ce problème conduit bien à la formule (3.32) si l'on note que $\varphi'(v^k)$ est bien évalué à $J'_v(u^{k+1}, v^k)$ d'après (3.31).

3.8.4 Corrigé de l'Exercice 3.16

Compte tenu des valeurs prises ici par \mathcal{L} , \mathcal{L} et Λ (voir (3.38)), l'inéquation variationnelle (3.37b) donne ici, avec $\hat{u} = u^{k+1}$, $v = u^k$, $q = p^k$, exactement (3.44), ce qui correspond au problème (3.39a). On doit ensuite considérer l'inéquation variationnelle (3.37c) avec $\hat{u} = v = u^{k+1}$ en raison du schéma séquentiel adopté, et $\hat{p} = p^{k+1}$, $q = p^k$; ceci conduit à

$$\forall p \in C^*, \left\langle -\frac{1}{\alpha} p^{k+1} + \varepsilon(\Theta(u^{k+1}) + \mathfrak{T}(u^{k+1})) + \frac{1}{\alpha} p^k, p - p^{k+1} \right\rangle \leq 0,$$

ce qui équivaut à (3.39b), compte tenu du fait que $\rho = \alpha \varepsilon$.

3.8.5 Corrigé de l'Exercice 3.21

La projection p^\sharp calculée par (3.47) est caractérisée par l'inéquation variationnelle

$$\forall p \in C^*, \langle p^\sharp + \rho(\Theta + \mathfrak{T})(u^\sharp) - p^\sharp, p - p^\sharp \rangle \leq 0,$$

ce qui est exactement l'inégalité de gauche du point selle.

3.8.6 Corrigé de l'Exercice 3.22

On a, pour tout $z \in C$,

$$\begin{aligned} \langle x - \text{proj}_C x, z - \text{proj}_C x \rangle &\leq 0, \\ \langle y - \text{proj}_C y, z - \text{proj}_C y \rangle &\leq 0. \end{aligned}$$

On pose $z = \text{proj}_C y$ dans la première inéquation et $z = \text{proj}_C x$ dans la seconde et on additionne les inéquations ainsi obtenues, ce qui donne

$$\|\text{proj}_C y - \text{proj}_C x\|^2 \leq \langle y - x, \text{proj}_C y - \text{proj}_C x \rangle ,$$

d'où l'on déduit facilement (3.48).

3.8.7 Corrigé de l'Exercice 3.25

L'application de (3.37) avec le choix (3.51) de la fonction auxiliaire Λ et avec $v = u^k, q = p^k, \hat{u} = u^{k+1}, \hat{p} = p^{k+1}$, avec

$$r = \mathcal{L}'_u(u^k, p^k) = \mathcal{J}'(u^k) + (\Theta'(u^k))^* \cdot p^k ,$$

où $*$ désigne l'opérateur adjoint, et avec

$$s^k = \mathcal{L}'_p(u^k, p^k) = \Theta(u^k) ,$$

donne lieu aux inéquations variationnelles suivantes

$$\begin{aligned} \forall u \in U^{\text{ad}}, \quad & \langle K'(u^{k+1}) + \varepsilon \mathcal{J}'(u^k) - K'(u^k), u - u^{k+1} \rangle \\ & + \langle p^{k+1}, \Omega'(u^{k+1}) \cdot (u - u^{k+1}) \rangle + \langle p^k, (\varepsilon \Theta'(u^k) - \Omega'(u^k)) \cdot (u - u^{k+1}) \rangle \\ & + \varepsilon (\mathfrak{J}(u) - \mathfrak{J}(u^{k+1}) + \langle p^{k+1}, \mathfrak{T}(u) - \mathfrak{T}(u^{k+1}) \rangle) \geq 0 , \end{aligned} \quad (3.84a)$$

$$\forall p \in C^*, \quad \langle \Omega(u^{k+1}) - \Omega(u^k) + \varepsilon (\Theta(u^k) + \mathfrak{T}(u^{k+1})) , p - p^{k+1} \rangle \leq 0 . \quad (3.84b)$$

Ces inéquations variationnelles sont exactement les conditions d'optimalité du problème (3.52).

3.8.8 Corrigé de l'Exercice 3.30

La fonction de Lyapounov (3.59) est une forme quadratique construite sur la matrice d'opérateurs

$$\frac{1}{2} \begin{pmatrix} \mathbf{J} & \mathbf{T}^* \\ \mathbf{T} & 2\mathbf{TJ}^{-1}\mathbf{T}^* \end{pmatrix} .$$

Une condition nécessaire et suffisante pour que cette forme quadratique soit fortement convexe est que les “mineurs principaux”, c'est-à-dire \mathbf{J} et $2\mathbf{TJ}^{-1}\mathbf{T}^* - \mathbf{TJ}^{-1}\mathbf{T}^* = \mathbf{TJ}^{-1}\mathbf{T}^*$, soient fortement convexes, ce qui est garanti par l'hypothèse (H1) sur \mathbf{J} et l'hypothèse de surjectivité (H2) sur \mathbf{T} . Les deux premiers termes de (3.59) correspondent au Lagrangien du problème original. Quant au dernier terme, destiné à “convexifier” le Lagrangien, il peut s'interpréter de la façon suivante. Pour le calcul de $\psi(p)$ (voir (2.4)), il faut minimiser le Lagrangien en u à p donné : on trouve ici

$$\hat{u}(p) = -\mathbf{J}^{-1}(\mathbf{T}^* p + \mathbf{j}) \quad \text{et} \quad \psi(p) = -\frac{1}{2} \langle p, \mathbf{TJ}^{-1}\mathbf{T}^* p \rangle + \text{termes linéaires et constants}.$$

À ces termes près (qui ne sont pas essentiels dans la définition d'une fonction de Lyapounov), on peut considérer que (3.59) est égale à $L(u, p) - 2\psi(p)$. Cette expression est étudiée dans un cadre plus général (non nécessairement quadratique) dans [7, Conclusion] où l'on montre qu'elle a en effet un certain nombre de propriétés intéressantes qui la qualifient comme une “bonne” fonction de Lyapounov.

3.8.9 Corrigé de l'Exercice 3.32

Si b est la constante de forte monotonie de \mathbf{K} , alors $\langle u, \mathbf{K}u \rangle \geq b \|u\|^2$, et si A est la constante de Lipschitz de \mathbf{J} , alors $\langle u, \mathbf{J}u \rangle \leq A \|u\|^2$, d'où il doit être clair que l'on peut garantir que l'opérateur (3.58a) est fortement monotone si (3.15) est vérifiée.

3.8.10 Corrigé de l'Exercice 3.35

On repart à nouveau de (3.37) avec L donné par (3.61) et décomposé selon (3.63), Λ donné par (3.64), et en se plaçant au pas k de l'algorithme. On ne doit pas perdre de vue que les inéquations variationnelles (3.37b)–(3.37c) doivent maintenant être écrites deux fois : une fois pour le couple (u, p) , les variables (w, μ) étant alors fixées à leurs valeurs courantes à l'indice k , et une fois pour le couple (w, μ) , les variables (u, p) étant alors fixées à leurs valeurs courantes à l'indice $k + 1$ (en raison du schéma séquentiel adopté). Ceci donne successivement :

$$\forall u \in U^{\text{ad}}, \left\langle K'(u^{k+1}) - K'(u^k) + \varepsilon \mathcal{J}'(u^k), u - u^{k+1} \right\rangle + \varepsilon \left\langle \mu^k, (\Theta'(u^k) - (1/\varepsilon)\Omega'(u^k)) \cdot (u - u^{k+1}) \right\rangle \\ + \varepsilon \left(\mathfrak{J}(u) - \mathfrak{J}(u^{k+1}) + \left\langle p^{k+1}, \mathfrak{T}(u) - \mathfrak{T}(u^{k+1}) + (1/\varepsilon)(\Omega(u) - \Omega(u^{k+1})) \right\rangle \right) \geq 0,$$

$$\varepsilon(w^k + \mathfrak{T}(u^{k+1}) + (1/\varepsilon)\Omega(u^{k+1})) = 0,$$

puis

$$\frac{w^{k+1} - w^k}{\alpha} + \varepsilon(p^{k+1} - \mu^k) = 0,$$

$$-\frac{\mu^{k+1} - \mu^k}{\beta} + \varepsilon(\Theta(u^{k+1}) - (1/\varepsilon)\Omega(u^{k+1}) - w^k) = 0.$$

Ces conditions sont directement interprétables en termes de (3.65) moyennant les définitions de ρ_1 et ρ_2 .

3.8.11 Corrigé de l'Exercice 3.37

Le déplacement du terme $-\langle \mu, w \rangle$ de \mathcal{L} vers \mathcal{L} n'affecte que les deux dernières équations ci-dessus où il faut remplacer le dernier terme w^k , respectivement μ^k , par w^{k+1} , respectivement μ^{k+1} . Ceci conduit à des formules implicites dont la résolution mène aux formules (2.22c)–(2.22d) à condition d'adapter les notations (remplacer (ε^k, ρ^k) par (ρ_1, ρ_2) et $\sum_{j \neq i} \Theta_{ij}(u_j^{k+1})$ par $\Theta(u^{k+1}) - (1/\varepsilon)\Omega(u^{k+1})$).

3.8.12 Corrigé de l'Exercice 3.49

En appliquant ϕ à $\theta_1 = \alpha\Theta(u) + (1 - \alpha)\Theta(v)$ et à $\theta_2 = \Theta(\alpha u + (1 - \alpha)v)$, on a, du fait de (3.78) et (3.79) :

$$\phi \circ \Theta(\alpha u + (1 - \alpha)v) \leq \phi(\alpha\Theta(u) + (1 - \alpha)\Theta(v)) \\ \leq \alpha\phi \circ \Theta(u) + (1 - \alpha)\phi \circ \Theta(v),$$

cette dernière inégalité du fait de la convexité de ϕ . Ceci établit le résultat souhaité.

3.8.13 Corrigé de l'Exercice 3.50

On fait le produit de dualité du premier membre de (3.78) avec p . Du fait de (3.80), ce produit est non négatif, ce qui établit le résultat désiré.

3.8.14 Corrigé de l'Exercice 3.51

On a les deux inégalités : $\forall u \in U^{\text{ad}}, \forall p \in C^*$,

$$L(u_1, p) \underbrace{\leq}_a L(u_1, p_1) \underbrace{\leq}_b L(u, p_1) , \quad (3.85)$$

$$L(u_2, p) \underbrace{\leq}_a L(u_2, p_2) \underbrace{\leq}_b L(u, p_2) . \quad (3.86)$$

Alors,

$$\begin{aligned} L(u_2, p_1) &\leq L(u_2, p_2) && \text{d'après (3.86a),} \\ &\leq L(u_1, p_2) && \text{d'après (3.86b),} \\ &\leq L(u_1, p_1) && \text{d'après (3.85a),} \\ &\leq L(u_2, p_1) && \text{d'après (3.85b),} \end{aligned}$$

d'où l'on déduit l'égalité des quatre valeurs $L(u_i, p_j)$, $i = 1, 2$, $j = 1, 2$. Alors, $\forall u \in U^{\text{ad}}, \forall p \in C^*$,

$$\begin{aligned} L(u_2, p) &\leq L(u_2, p_1) && \text{d'après (3.86a) et } L(u_2, p_1) = L(u_2, p_2) , \\ &\leq L(u, p_1) && \text{d'après (3.85b) et } L(u_2, p_1) = L(u_1, p_1) , \end{aligned}$$

ce qui montre que (u_2, p_1) est un point selle.

3.8.15 Corrigé de l'Exercice 3.52

Avec l'inégalité de gauche de (3.82) où l'on choisit $p = 0$ dans le membre gauche, on déduit que

$$\langle p^\sharp, \Theta(u^\sharp) \rangle \geq 0 .$$

Mais on va montrer que $\Theta(u^\sharp) \in -C$, et comme $p^\sharp \in C$ par hypothèse, l'inégalité inverse est aussi vraie, et donc

$$\langle p^\sharp, \Theta(u^\sharp) \rangle = 0 . \quad (3.87)$$

Le fait que $\Theta(u^\sharp) \in -C$ se démontre de la façon suivante : si ce n'est pas vrai, alors en vertu de la décomposition orthogonale illustrée par la Figure 3.2, $\text{proj}_C \Theta(u^\sharp) \neq 0$. On peut alors choisir $p = \alpha \text{proj}_C \Theta(u^\sharp)$, où α est un scalaire positif, dans le membre le plus à gauche de (3.82), et, en laissant tendre α vers $+\infty$, on obtient une contradiction. Donc $\text{proj}_C \Theta(u^\sharp) = 0$, ce qui équivaut à $\Theta(u^\sharp) \in -C$ (et donc u^\sharp est admissible pour le problème (3.76)).

On va maintenant exploiter l'inégalité de droite de (3.82). À cause de (3.87), cette inégalité s'écrit

$$\forall u \in U^{\text{ad}}, \quad J(u^\sharp) \leq J(u) + \langle p^\sharp, \Theta(u) \rangle .$$

En particulier, puisque $p^\sharp \in C$, si $\Theta(u) \in -C$, le dernier terme est non positif, et on a donc montré que $J(u^\sharp) \leq J(u)$ pour tout u satisfaisant les contraintes du problème (3.76), et donc u^\sharp , qui est admissible, est en fait optimal.

Chapitre 4

Régularisation et Lagrangien augmenté

4.1 Lien entre Lagrangien augmenté et régularisation

4.1.1 Non convexité et convexité plus ou moins forte

Comme on l’a vu aux Chapitres 2 et 3, l’utilisation de la décomposition par les prix repose sur l’existence d’un point selle du Lagrangien du problème d’optimisation sous contrainte (3.36). Un saut de dualité rend donc l’utilisation de cette méthode impossible. Comme on l’a vu à l’Exercice 2.29, le saut de dualité s’accompagne, de plus, de la non différentiabilité de la fonction duale ψ définie par (2.4) au point où elle atteint son maximum. Or l’algorithme d’Uzawa s’interprète comme un algorithme de gradient ou de sous-gradient (projeté) pour maximiser cette fonction ψ .

La différentiabilité de la fonction duale ψ est garantie lorsque la fonction coût primaire J est *strictement* convexe. Mais le Théorème 3.6 a montré qu’un algorithme de gradient par exemple, pour être garanti de converger avec des pas fixes, ou en tout cas ne tendant pas vers 0, nécessite que le gradient soit Lipschitzien. Le Lemme 3.18 a montré que la fonction ψ a une dérivée Lipschitzienne dans le cas où la fonction coût primaire J est *fortement* convexe. C’est cette hypothèse qui a été requise au Théorème 3.20 (voir (H3)) pour montrer la convergence de l’Algorithme 3.15 dont Uzawa est un cas particulier.

Ainsi, bien qu’on puisse garantir l’existence d’un point selle du Lagrangien dans le cas d’un problème de minimisation sous contrainte *simplement convexe* (moyennant quelques autres hypothèses techniques — voir Annexe, §3.7), les algorithmes nécessitent, quant à eux, une hypothèse de *forte convexité*. L’Exercice 2.30 a montré que l’absence de forte convexité ne permet pas d’espérer la convergence d’algorithmes du type Uzawa en ce qui concerne plus particulièrement les variables primales, même si on parvenait à faire converger les variables duales vers leur valeur optimale par une stratégie adaptée concernant le choix des pas de l’algorithme d’Uzawa (pas “en série divergente” — voir Annexe, §2.3).

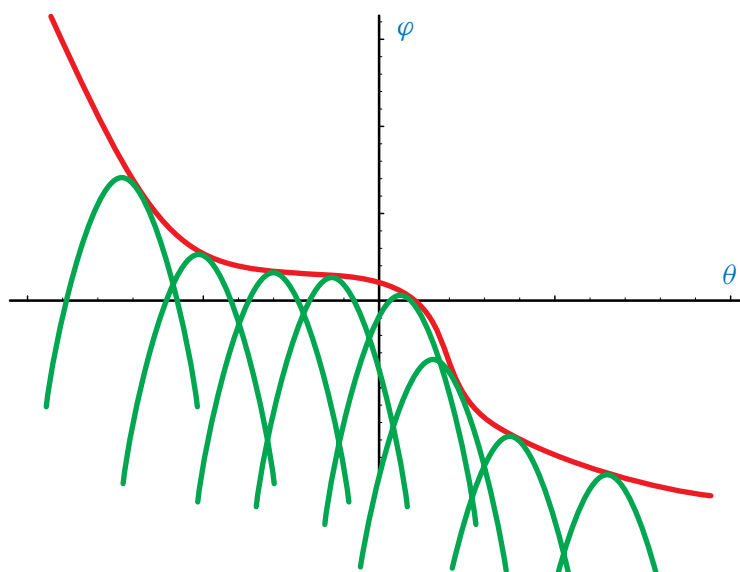
4.1.2 Lagrangien augmenté

La technique du Lagrangien augmenté (aussi appelée “multiplier method” en Anglais) a été introduite pour surmonter les sauts de dualité survenant dans le cas non convexe (voir [3]) : une interprétation géométrique va nous permettre de comprendre comment. Nous allons pour cela reprendre l’Exercice 2.29.

On a vu qu’un saut de dualité survient lorsque la fonction “perturbation” (cas de contraintes égalité)

$$\varphi(\theta) = \left\{ \min J(u) \mid u \in U^{\text{ad}} \text{ et } \Theta(u) = \theta \right\} \quad (4.1)$$

ne coïncide pas en $\theta = 0$ avec son enveloppe convexe. Comme l’a montré la démarche de l’exercice, le fait que l’enveloppe convexe de φ joue un rôle provient du fait que le Lagrangien $J(u) + \langle p, \Theta(u) \rangle$ est interprété comme l’ordonnée à l’origine d’un hyperplan de coefficient directeur $-p$ dans l’espace $\mathcal{C} \times \mathbb{R}$

Figure 4.1: Graphe de φ et enveloppe de paraboles concaves

passant par le point $(\Theta(u), J(u))$. Minimiser le Lagrangien, c'est-à-dire calculer la fonction duale $\psi(p)$, c'est rechercher l'ordonnée à l'origine de l'hyperplan de "pente" $-p$ s'appuyant sur l'épigraphe de φ . Toutes ces interprétations géométriques sont illustrées par la Figure 2.4.

Dans cette optique, l'idée du Lagrangien augmenté consiste, pour "ausculter" l'épigraphe de φ , à remplacer les "objets plats" que sont les hyperplans par des objets concaves, et suffisamment "pointus" pour venir se loger dans les "trous" éventuels de l'épigraphe. Ainsi, si on remplace le Lagrangien ordinaire (3.81) par le Lagrangien augmenté

$$L_c(u, p) = J(u) + \langle p, \Theta(u) \rangle + \frac{c}{2} \|\Theta(u)\|^2, \quad (4.2)$$

où c est un nombre positif, cela revient, selon le point de vue de la question 2 de l'exercice, à remplacer l'hyperplan d'équation $y = -\langle p, x \rangle + a$ par le paraboloïde concave

$$y = -\langle p, x \rangle - \frac{c}{2} \|x\|^2 + a,$$

qui est d'autant plus "pointu" que c est grand.

Il est facile de constater que toutes les questions de l'Exercice 2.29 peuvent être reprises avec ce nouvel élément d'appui à la place de l'hyperplan précédent.

La Figure 4.1 illustre ce qui va se passer dans cette nouvelle situation : si c est assez grand, la famille de paraboles d'appui indexée par leur "pente à l'origine" $-p$ va parvenir à "suivre" l'épigraphe de φ , y compris dans sa partie concave. Ainsi, si l'enveloppe de cette famille de paraboles d'appui coïncide en 0 avec $\varphi(0)$, on aura surmonté le saut de dualité et on aura réalisé la condition de point selle

$$\max_p \min_u L_c(u, p) = \min_u \max_p L_c(u, p).$$

Par contre, si c n'est pas assez grand pour les concavités de l'épigraphe de φ , la Figure 4.2 montre que le saut de dualité subsiste (mais est atténué par rapport à ce qu'il serait avec le Lagrangien ordinaire).

En fait, pour certains problèmes, il sera impossible de réduire le saut de dualité à 0 avec cette technique même pour c arbitrairement grand.

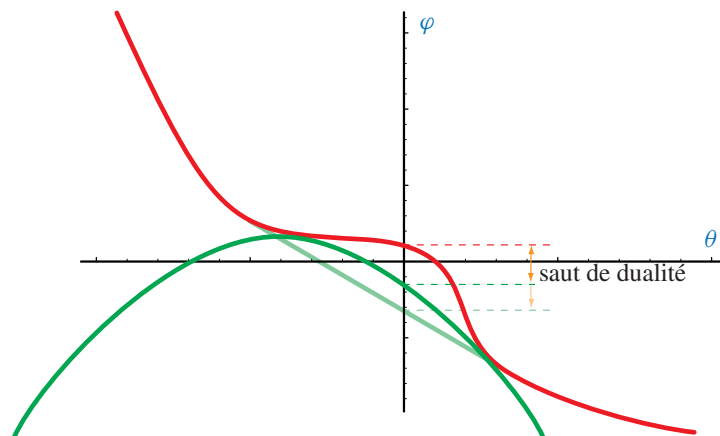


Figure 4.2: Le saut de dualité subsiste si c n'est pas assez grand (mais il est plus petit que pour $c = 0$)

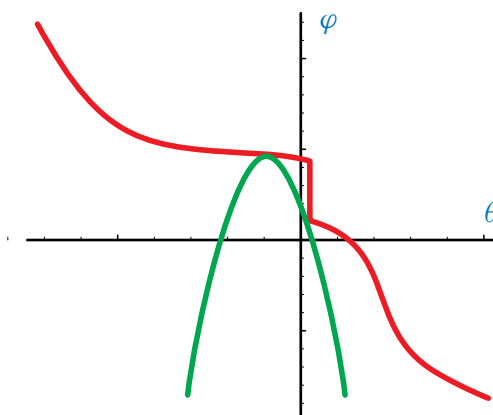


Figure 4.3: Cas de problèmes en variables entières ou mixtes

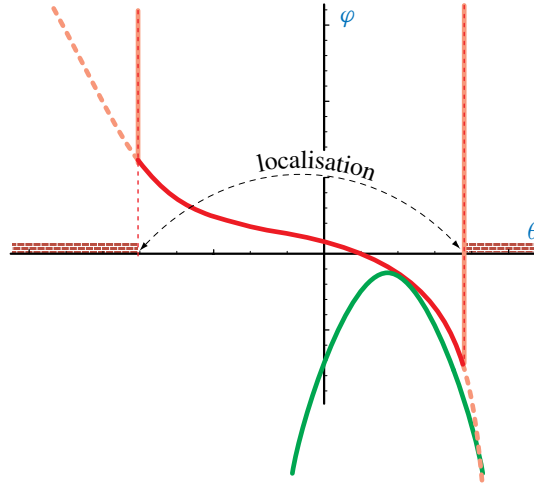


Figure 4.4: Décroissance plus rapide que quadratique et effet de la “localisation”

- Il y a d’abord les problèmes en variables entières ou mixtes pour lesquels des discontinuités de φ ne sont pas exclues (changement brutal de la solution entière pour une faible variation de la perturbation θ), ce qui peut conduire à la situation illustrée par la Figure 4.3 : aucune “parabole” ne peut “combler le trou” causé par une discontinuité verticale au voisinage du point $\theta = 0$.

Ceci dit, on observera que même dans ces cas, le saut de dualité sera moindre avec le Lagrangien augmenté qu’avec le Lagrangien ordinaire.¹

- Il y a ensuite le cas des problèmes où φ peut tendre vers $-\infty$ plus vite qu’une fonction quadratique de la perturbation θ . Alors, aucune “parabole” ne peut rester entièrement sous le graphe de φ comme l’illustre la Figure 4.4.

Mais cette figure illustre aussi le fait que si on limite la plage de variation des variables primales (en remplaçant U^{ad} par $U^{\text{ad}} \cap \mathcal{V}(u^0)$, où $\mathcal{V}(u^0)$ désigne un voisinage convexe fermé d’une estimation u^0 de la solution $u^\#$), ce qui aura pour effet de limiter aussi la plage atteignable pour θ , remplaçant $\varphi(\theta)$ par la valeur $+\infty$ en dehors de cette plage atteignable,² alors les “branches infinies” du graphe de φ sont “redressées” vers le haut, et le problème d’une décroissance plus rapide que quadratique disparaît.

C’est pourquoi, dans la théorie du Lagrangien augmenté pour le cas *non convexe* (voir [3]), on ne prétend retrouver qu’un point selle “local” pour c assez grand : le mot “local” doit être compris au sens ci-dessus, c’est-à-dire celui d’une limitation des variables primales à un voisinage de la solution.

Le fait de retrouver un point selle, même seulement local, est évidemment un progrès décisif pour appliquer une méthode duale à un problème d’optimisation sous contrainte non convexe. Mais, même dans le cas convexe, l’utilisation du Lagrangien augmenté est un très grand pas en avant.

¹La “relaxation lagrangienne” est une technique très répandue pour approcher la solution de problèmes en variables entières ou mixtes, en particulier parce qu’elle conduit souvent à une décomposition de la minimisation primale. Beaucoup d’auteurs dans la littérature, s’ils n’ignorent pas que le Lagrangien augmenté est préférable du point de vue de la réduction du saut de dualité — facilitant d’autant la “fabrication à la main” d’une solution admissible à partir de la solution obtenue par relaxation lagrangienne —, pensent que le recours à cette technique est interdit par la préoccupation de préserver la possibilité de décomposition, parce qu’ils méconnaissent en fait le contenu de ce qui va suivre dans ce chapitre.

² $+\infty$ est la valeur optimale du coût pour un problème de minimisation sous contrainte dont l’ensemble des solutions admissibles est vide.

- On va pouvoir résoudre la difficulté de convergence des variables primales rencontrées à l'Exercice 2.30 lorsque la fonction coût n'est ni fortement, ni même strictement convexe, mais seulement convexe. Ceci tient à la propriété de "stabilité en u " du Lagrangien augmenté, déjà évoquée au Commentaire 2.4-c.
- Par ailleurs, même dans le cas fortement convexe, tous les essais numériques confirment la plus grande rapidité de convergence des algorithmes par dualité utilisant le Lagrangien augmenté plutôt que le Lagrangien ordinaire.

Les deux points ci-dessus ont une explication commune. Elle a été donnée par Rockafellar dans le cas convexe : on montre que la fonction duale

$$\psi_c(p) = \min_{u \in U^{\text{ad}}} L_c(u, p), \quad (4.3)$$

analogue de la fonction ψ introduite dans (2.4), mais ici à partir du Lagrangien augmenté, est précisément reliée à ψ par une *régularisation de Yosida-Moreau*. On va voir que cette régularisation transforme une fonction concave ψ non nécessairement différentiable en une autre fonction concave ψ_c , différentiable, de gradient Lipschitzien, et admettant le même maximum (du point de vue de la valeur de la fonction ψ et du point de vue des arguments $p^\#$ réalisant le maximum).

Remarque 4.1. On a parfois tendance à voir le Lagrangien augmenté comme le mélange d'une technique de dualité (Lagrangien ordinaire) et de pénalisation, le terme supplémentaire $(c/2) \|\Theta(u)\|^2$ jouant ce rôle. Si cette interprétation était la bonne, l'équivalent du Lagrangien augmenté pour le cas de contraintes inégalité devrait amener à remplacer, dans le cas d'une contrainte *scalaire* $\Theta(u) \leq 0$, le terme quadratique par

$$\frac{c}{2} \left(\max(0, \Theta(u)) \right)^2,$$

et dans le cas général $\Theta(u) \in -C$, par

$$\frac{c}{2} \|\text{proj}_{C^*} \Theta(u)\|^2,$$

car d'après la décomposition orthogonale illustrée par la Figure 3.2 (et plutôt avec les cônes opposés), ce terme constitue bien une pénalisation de la contrainte.

Or, les formules ci-dessus ne permettent pas de retrouver un Lagrangien augmenté stable en u pour le cas de contraintes inégalité, et en fait ne permettent pas de préserver l'interprétation de la régularisation de ψ en ψ_c . L'exercice ci-dessous indique une façon de retrouver les bonnes formules qui seront explicitées plus loin.

Exercice 4.2. Dans le cas de contraintes inégalité pour le problème (3.36) (c'est-à-dire avec $C \neq \{0\}$), on peut se ramener au cas égalité en introduisant une "variable d'écart" $\xi \in -C$ supplémentaire (par rapport à laquelle il faudra aussi minimiser) et en transformant la contrainte en égalité : $\Theta(u) - \xi = 0$.

Former le Lagrangien augmenté correspondant à ce nouveau problème avec contrainte égalité. Montrer que la minimisation en $\xi \in -C$ de cette expression peut être effectuée par des formules analytiques qui fournissent alors l'expression, ne dépendant plus que de u , qui est la "bonne" expression du Lagrangien augmenté dans le cas inégalité. On pourra, si on le souhaite, commencer par le cas d'une contrainte scalaire.

4.1.3 Régularisation de Yosida-Moreau

Bien que nous nous en servions plutôt pour des fonctions concaves, on développe ici le cas convexe.

Définition 4.3. La *régularisée de Yosida-Moreau* d'une fonction convexe $J : \mathcal{U} \rightarrow \mathbb{R}$ est la fonction

$$J_c(u) = \inf_{v \in \mathcal{U}} \left(\frac{1}{2c} \|u - v\|^2 + J(v) \right), \quad (4.4)$$

où c est un scalaire positif.

Le lemme suivant est facile à démontrer.

Lemme 4.4. La fonction J_c est convexe et différentiable. De plus, $J_c(u) \leq J(u)$ pour tout u dans \mathcal{U} , mais les deux fonctions coïncident sur

$$\arg \min_{u \in \mathcal{U}} J(u) = \arg \min_{u \in \mathcal{U}} J_c(u). \quad (4.5)$$

Exercice 4.5. Donner une démonstration de ce lemme.

Exercice 4.6. Calculer explicitement la transformée de Yosida-Moreau de $J : \mathbb{R} \rightarrow \mathbb{R}$ où $J(u) = |u|$. Tracer le graphe.

Remarque 4.7. Le fait peut être le plus surprenant est que J_c , bien que calculée comme l'enveloppe inférieure d'une famille de fonctions convexes, est malgré tout convexe. Ceci peut être déduit d'un résultat plus général qui affirme que si une fonction $f(x, y)$ est *conjointement* convexe en (x, y) , alors la fonction "marginale" $g(x) = \inf_y f(x, y)$ est convexe en x .

Exercice 4.8. Donner une démonstration de l'affirmation ci-dessus.

Remarque 4.9. Il faut insister sur le fait que la formule (4.4) et le Lemme 4.4 sont énoncés pour des minimisations dans *tout* l'espace \mathcal{U} . Si on vise plutôt à résoudre un problème de minimisation de J sur un sous-ensemble convexe fermé U^{ad} , alors il faut remplacer J par $J + \mathcal{I}_{U^{\text{ad}}}$ où $\mathcal{I}_{U^{\text{ad}}}$ est la fonction indicatrice de U^{ad} valant 0 si $u \in U^{\text{ad}}$ et $+\infty$ sinon. La définition (4.4) de la régularisée devient

$$J_c(u) = \inf_{v \in U^{\text{ad}}} \left(\frac{1}{2c} \|u - v\|^2 + J(v) \right), \quad (4.6)$$

alors que (4.5) doit être remplacée par

$$\arg \min_{u \in U^{\text{ad}}} J(u) = \arg \min_{u \in U^{\text{ad}}} J_c(u). \quad (4.7)$$

Autrement dit, le calcul de la régularisation s'opère sur U^{ad} mais la minimisation de la fonction régularisée peut avoir lieu sur tout l'espace (alors que *le schéma inverse ne marche pas !*).

On va montrer que la régularisée est non seulement différentiable, mais de dérivée Lipschitzienne. On a donc la possibilité de remplacer la minimisation d'une fonction convexe éventuellement non différentiable par celle d'une fonction de gradient Lipschitzien, ce qui est évidemment plus favorable sur le plan algorithmique. Malheureusement, en général, le calcul (4.4) ou (4.6) de la régularisée est pratiquement aussi compliqué que la résolution du problème de minimisation original. On va cependant voir que l'idée est exploitable dans le cas de la fonction concave ψ définie par (2.4), non pas que cette fonction et sa régularisée soient faciles à calculer directement, mais cela reviendra à remplacer le Lagrangien ordinaire dans (2.4) par le Lagrangien augmenté.

4.2 Un point de vue sur la régularisation

4.2.1 Bref rappel sur la transformée de Fenchel(-Legendre) et sur l'inf-convolution

Sur les sujets de ce paragraphe, on pourra consulter [24] pour plus de détails.

Définition 4.10. Pour une fonction $J : \mathcal{U} \rightarrow \mathbb{R}$, la *transformée de Fenchel*, notée $J^* : \mathcal{U}^* \rightarrow \mathbb{R}$, est définie par

$$J^*(v) = \sup_{u \in \mathcal{U}} (\langle v, u \rangle - J(u)) . \quad (4.8)$$

Comme enveloppe supérieure d'une famille de fonctions affines en v , la transformée de Fenchel est convexe et s.c.i. De plus, si J est convexe s.c.i., alors $(J^*)^* = J$.

Il y a un autre point de vue sur la transformée de Fenchel si on s'intéresse aux dérivées, ou plus généralement au sous-différentiel. Supposant J convexe et sous-différentiable, si le sup est atteint en un point $\hat{u}(v)$ dans (4.8), alors

$$v \in \partial J(\hat{u}(v)) ,$$

(condition de stationnarité du problème), et donc $\hat{u}(v) \in (\partial J)^{-1}(v)$ (inverse ensembliste). Par ailleurs, la dérivée partielle de l'expression à maximiser dans (4.8) par rapport à v , évaluée en $\hat{u}(v)$, constitue un sous-gradient de J^* . Ce calcul redonne précisément $\hat{u}(v)$, et donc $\hat{u}(v) \in \partial J^*(v)$. On peut montrer avec rigueur que

$$\partial J^* = (\partial J)^{-1} . \quad (4.9)$$

La transformée de Fenchel, au niveau des sous-différentiels, n'est donc rien d'autre que l'inversion.

La régularisée (4.4) a été en fait définie à partir d'une opération de portée plus générale introduite ci-après.

Définition 4.11. L'*inf-convolution* de deux fonctions J et K de \mathcal{U} dans \mathbb{R} est la fonction

$$(J \square K)(u) = \inf_v (J(u - v) + K(v)) . \quad (4.10)$$

Pour la même raison que celle invoquée à la Remarque 4.7, cette fonction est convexe lorsque J et K le sont.

Exercice 4.12. Démontrer que la transformée de Fenchel d'une inf-convolution est la somme des transformées de Fenchel (ou si l'on préfère, la transformée de Fenchel convertit les inf-convolutions en sommes³).

Lemme 4.13. Si J et K sont deux fonctions convexes s.c.i. et sous-différentiables, et si⁴

$$\text{dom } J^* \cap \text{int dom } K^* \neq \emptyset , \quad (4.11)$$

alors

$$\partial(J \square K) = ((\partial J)^{-1} + (\partial K)^{-1})^{-1} ; \quad (4.12)$$

autrement dit, l'inf-convolution réalise, au niveau des sous-différentiels, la moyenne harmonique.

³Ce résultat est à rapprocher du fait analogue que la transformée de Fourier transforme les convolutions en produits.

⁴D'une manière générale, $\text{dom } J$ désigne le domaine de J , c'est-à-dire l'ensemble des u tels que $\text{dom } J(u) < +\infty$.

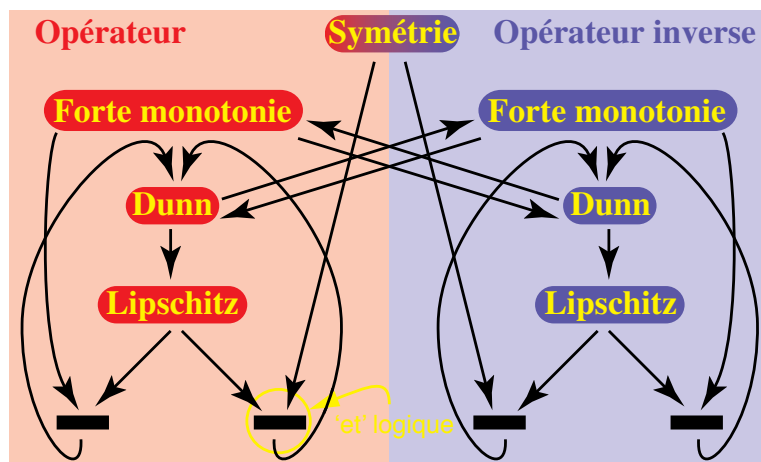


Figure 4.5: Graphe d'implications des propriétés d'un opérateur et son inverse

Démonstration. La démonstration procède d'un "aller-retour" avec la transformée de Fenchel. En effet,

$$(\partial(J \square K))^{-1} = \partial((J \square K)^*)$$

d'après (4.9),

$$= \partial(J^* + K^*)$$

d'après l'Exercice 4.12,

$$= \partial(J^*) + \partial(K^*)$$

grâce à l'hypothèse technique (4.11) qui assure que le sous-différentiel de la somme est égal à la somme des sous-différentiels,

$$= (\partial J)^{-1} + (\partial K)^{-1}$$

encore d'après (4.9). Il suffit de reprendre les inverses (ou les transformées de Fenchel en ce qui concerne les fonctions elles-mêmes) pour trouver (4.12). \square

Remarque 4.14. Sachant que $\text{dom } J^* \supset \text{im } \partial J$, on peut remplacer l'hypothèse (4.11) par l'hypothèse (a priori plus forte)

$$\text{im } \partial J \cap \text{int im } \partial K \neq \emptyset. \quad (4.13)$$

4.2.2 Transmission de propriétés par l'inf-convolution

Le Lemme 4.13 a des conséquences très intéressantes en ce qui concerne les diverses propriétés passées en revue dans les annexes du Chapitre 3 (Lipschitz, Dunn, forte monotonie). La Figure 4.5 résume les implications entre propriétés des opérateurs et de leurs inverses. Pour alléger la figure, on n'a pas fait mention des diverses constantes, mais on rappelle que

- si a et D sont les constantes de forte monotonie et de Dunn d'un opérateur, D^{-1} et a^{-1} sont les constantes correspondantes pour l'opérateur inverse ;

- si D est la constante de Dunn, c'est aussi la constante de Lipschitz ;
- si a est la constante de forte monotonie et si A est la constante de Lipschitz, A^2/a est la constante de Dunn (estimée) ;
- dans le cas symétrique, la constante de Lipschitz est aussi la constante de Dunn.

À ces considérations, il faut ajouter les remarques évidentes que

- si un opérateur Φ est fortement monotone de constante a et si Ψ est simplement monotone, la somme $\Phi + \Psi$ est fortement monotone de constante a ; évidemment, si Ψ est aussi fortement monotone de constante b , alors la somme est fortement monotone de constante $a + b$;
- pour garantir que $\Phi + \Psi$ est Lipschitzien, il faut que Φ et Ψ soient *tous les deux* Lipschitziens ; une estimation de la constante de Lipschitz de la somme est la somme des constantes de Lipschitz.

Lemme 4.15. *On considère deux fonctions convexes s.c.i. sous-différentiables J et K de \mathcal{U} dans \mathbb{R} et on suppose l'une des hypothèses (4.11) ou (4.13) satisfaite.*

1. *Si K est différentiable et de dérivée Lipschitzienne de constante B , alors $J \square K$ est aussi différentiable et de dérivée Lipschitzienne avec la même constante.*
2. *Si de plus J est différentiable de dérivée Lipschitzienne de constante A , alors la constante de Lipschitz de $(J \square K)'$ est estimée à $(A^{-1} + B^{-1})^{-1}$.*
3. *Si J et K sont toutes les deux fortement convexes de constantes respectives a et b , alors $J \square K$ est fortement convexe de constante $(a^{-1} + b^{-1})^{-1}$.*

Démonstration.

1. D'après l'hypothèse et les considérations précédentes, $(K')^{-1}$ est fortement monotone de constante B^{-1} , et donc également $(\partial J)^{-1} + (K')^{-1}$, ce qui implique que $((\partial J)^{-1} + (K')^{-1})^{-1}$ vérifie la propriété de Dunn avec la constante B , mais comme l'opérateur est symétrique, il est Lipschitzien avec la même constante. Enfin, le Lemme 4.13 permet de conclure.
2. On reprend la démonstration ci-dessus avec l'hypothèse supplémentaire sur J' .
3. D'après l'hypothèse, ∂J et ∂K sont fortement monotones de constantes respectives a et b , donc $(\partial J)^{-1}$ et $(\partial K)^{-1}$ vérifient la propriété de Dunn avec les constantes respectives a^{-1} et b^{-1} , et elles sont Lipschitziennes avec les mêmes constantes. On poursuit comme ci-dessus en faisant la somme et en reprenant l'inverse. \square

4.2.3 Zéro d'un opérateur et inf-convolution

Définition 4.16. On appelle *zéro* d'un opérateur Ψ (de \mathcal{U} dans \mathcal{U}^*) tout point u^\sharp tel que $0 \in \Psi(u^\sharp)$.

En fait, ce qu'on appelle donc un zéro c'est plutôt tout élément de l'image de zéro par Ψ^{-1} .

Exercice 4.17. Supposons que Ψ est de la forme $\Phi + \partial \mathcal{I}_{U^{\text{ad}}}$ où Φ est monotone et U^{ad} est un sous-ensemble convexe fermé de \mathcal{U} . Montrer que les zéros de Ψ sont les solutions de l'inéquation variationnelle

$$u^\sharp \in U^{\text{ad}}, \quad \exists r \in \Phi(u^\sharp) : \quad \forall u \in U^{\text{ad}}, \quad \langle r, u - u^\sharp \rangle \geq 0.$$

Lemme 4.18. *On considère deux fonctions convexes s.c.i. sous-différentiables J et K de \mathcal{U} dans \mathbb{R} et on suppose l'une des hypothèses (4.11) ou (4.13) satisfaite. Les zéros de ∂J sont les mêmes que les zéros de $\partial(J \square K)$ si et seulement si $\partial K(0) = \{0\}$.*

Démonstration. Un zéro u^\sharp de $\partial(J \square K)$ est tel que

$$u^\sharp \in (\partial(J \square K))^{-1}(0) = (\partial J)^{-1}(0) + (\partial K)^{-1}(0) .$$

Pour que le dernier terme ne provoque aucune translation ni aucune dilatation de l'autre terme auquel il est ajouté, il faut et il suffit qu'il se réduise à $\{0\}$, ce qui est équivalent à la condition énoncée. \square

4.2.4 Apport de la régularisation de Yosida-Moreau

Comme on l'a vu, la régularisation (4.4) est un cas particulier d'inf-convolution entre une fonction J et la fonction $(1/2c) \|\cdot\|^2$. Cette dernière fonction a toutes les propriétés requises (différentiabilité et dérivée Lipschitzienne, forte convexité, dérivée nulle en 0, $\text{dom } K' = \mathcal{U}$). De ces observations, et de tout ce qui précède, on tire le théorème suivant.

Théorème 4.19.

1. *La transformée de Yosida-Moreau J_c d'une fonction J convexe, s.c.i. et sous-différentiable est différentiable et de dérivée Lipschitzienne de constante au moins égale à $1/c$ (et à $A/(1 + cA)$ si J a une dérivée Lipschitzienne de constante A).*
2. *Si J est fortement convexe de constante a , alors J_c est aussi fortement convexe de constante $a/(1 + ca)$.*
3. *Le minimum de J_c est atteint aux mêmes points que le minimum de J sur \mathcal{U} .*

Ce théorème, qui renforce le Lemme 4.4, est un simple corollaire de tout ce qui précède dans cette section.

Remarque 4.20. La première partie de l'énoncé du point 1 du Théorème 4.19 (relatif à la constante $1/c$ de J'_c) est valable même si la régularisation a lieu sur U^{ad} (ce qui revient, comme on l'a déjà vu, à remplacer J par $J + \mathcal{I}_{U^{\text{ad}}}$). Mais la deuxième partie de l'énoncé 1 (avec la constante $A/(1 + cA)$) n'est valide que si la régularisation a lieu sur \mathcal{U} car évidemment $J + \mathcal{I}_{U^{\text{ad}}}$ n'a pas une dérivée Lipschitzienne.

Exercice 4.21. Redémontrer ce théorème en se servant seulement de la définition de la régularisée et des inégalités définissant les propriétés invoquées dans le théorème.

Remarque 4.22. Insistons sur le fait que si le problème est de minimiser J sur un sous-ensemble convexe fermé U^{ad} et non pas sur tout \mathcal{U} , alors il faut appliquer la théorie à $J + \mathcal{I}_{U^{\text{ad}}}$. Il en résulte que la régularisée se calcule sur U^{ad} comme indiqué à la Remarque 4.9 (voir (4.6)). Par contre, comme indiqué par (4.7), on peut alors chercher le minimum de J_c sur tout \mathcal{U} , et non pas seulement sur U^{ad} (même si le théorème implique que ce minimum sera finalement atteint sur U^{ad}).

Le mot de “régularisation” est déjà justifié par le fait que J_c est différentiable et de dérivée Lipschitzienne même si J n'est que sous-différentiable. Mais on va voir que l'amélioration concerne aussi ce qu'on peut appeler le “conditionnement” de J dans le cas où celui-ci peut être défini.

Définition 4.23. On appelle *conditionnement* d'une fonction J , fortement convexe de constante a , et différentiable de dérivée Lipschitzienne de constante A , le ratio A/a .

Cette définition se justifie par les résultats sur la vitesse de convergence d'algorithmes appartenant à la famille de l'Algorithme 3.3 (dont l'algorithme du gradient), résultats énoncés au Théorème 3.11. On a vu que le taux optimal de la convergence linéaire en moyenne était d'autant meilleur que le conditionnement (toujours supérieur ou égal à 1) était proche de 1.

Géométriquement, les “ensembles de niveau” de la fonction J (ensembles du type $\{u \mid J(u) \leq \alpha\}$) ressemblent plus ou moins à des ellipsoïdes, et ces ellipsoïdes ont une excentricité plus ou moins forte en fonction du conditionnement, le cas idéal étant constitué par les fonctions du type $\gamma \|\cdot\|^2$ dont les ensembles de niveau sont “sphériques” et dont le conditionnement vaut 1.

Lemme 4.24. *Lorsque le conditionnement de J est défini, le conditionnement de la régularisée J_c est meilleur que celui de J . Ce conditionnement s'améliore de façon monotone lorsque c varie de 0 à $+\infty$: il est inchangé lorsque c tend vers 0 et il tend vers le conditionnement idéal 1 lorsque c tend vers $+\infty$. Pour deux fonctions J ayant la même constante de forte convexité, l'amélioration du conditionnement pour une valeur fixée de c est, en valeur relative, meilleure pour la fonction la moins bien conditionnée. Ou bien, à conditionnement égal, elle est meilleure pour la fonction avec la constante de forte convexité la meilleure (c'est-à-dire la plus grande).*

Démonstration. Notons $\omega = A/a$ le conditionnement de J et ω_c celui de J_c . Des résultats du Théorème 4.19, on obtient

$$\omega_c = \frac{A/(1+cA)}{a/(1+ca)} = \omega \frac{1+ca}{1+cA} = \omega \frac{1+ca}{1+c\omega a},$$

d'où l'on déduit que ω_c/ω est inférieur ou égal à 1 et est une fonction décroissante de c , de a et de ω avec les limites annoncées pour $c \rightarrow 0$ et $c \rightarrow +\infty$. \square

4.3 Retour sur les Lagrangiens augmentés

4.3.1 Une nouvelle définition du Lagrangien augmenté

Le §4.1.2 a donné une justification géométrique de la formule (4.2) au moins dans le cas des contraintes égalité. Cette justification géométrique est illustrée par la Figure 4.1. L'Exercice 4.2 a suggéré une façon de trouver la bonne formule pour le cas des contraintes inégalité.

Nous allons aborder ici le problème sous un angle totalement différent, approche évoquée après la formule (4.3). Évidemment, cette nouvelle approche ne fonctionne que dans le cas convexe, mais on a déjà dit que même dans ce cas un problème de “stabilité en u ” du Lagrangien ordinaire peut se poser, problème automatiquement réglé par le Lagrangien augmenté.

Essayons donc de calculer la régularisée de Yosida-Moreau de la fonction duale ψ définie par (2.4). Il faut bien sûr dans ce cas adapter la définition de la régularisée au cas concave.

Avant d'entreprendre ce calcul, on doit souligner que ψ définie par (2.4) doit être maximisée sur C^* dans le cas de contraintes inégalité : il faut donc considérer, selon les Remarques 4.9 et 4.22 (voir en particulier (4.6)) la régularisée de la fonction

$$\bar{\psi}(p) = \psi(p) - \mathcal{I}_{C^*}(p) = \begin{cases} \min_{u \in U^{\text{ad}}} L(u, p) & \text{si } p \in C^*, \\ -\infty & \text{sinon.} \end{cases} \quad (4.14)$$

Alors

$$\begin{aligned} \psi_c(p) &= \max_{q \in C^*} \left(\bar{\psi}(q) - \frac{1}{2c} \|p - q\|^2 \right) \\ &= \max_{q \in C^*} \left(\min_{u \in U^{\text{ad}}} L(u, q) - \frac{1}{2c} \|p - q\|^2 \right) \end{aligned}$$

d'après (4.14)

$$= \min_{u \in U^{\text{ad}}} \left(J(u) + \max_{q \in C^*} \left(\langle q, \Theta(u) \rangle - \frac{1}{2c} \|p - q\|^2 \right) \right)$$

en utilisant l'expression de L et la possibilité de faire commuter les min et max dans le cas convexe-concave.

La formule (4.3) se trouve donc justifiée si on adopte la définition suivante du Lagrangien augmenté :

$$L_c(u, p) = J(u) + \zeta_c(p, \Theta(u)) , \quad (4.15a)$$

avec

$$\zeta_c(p, \theta) = \max_{q \in C^*} \left(\langle q, \theta \rangle - \frac{1}{2c} \|p - q\|^2 \right) . \quad (4.15b)$$

On peut facilement, au moins dans le cas des contraintes égalité où $C^* = \mathcal{C}^*$, vérifier que l'on retrouve bien l'expression (4.2). Le lemme ci-dessous fournit d'autres expressions de la fonction ζ_c introduite par (4.15b), de ses dérivées, et énonce certaines de ses propriétés.

Lemme 4.25.

1. La fonction ζ_c introduite par (4.15b) peut aussi être définie par

$$\zeta_c(p, \theta) = \min_{\xi \in -C} \left(\langle p, \theta - \xi \rangle + \frac{c}{2} \|\theta - \xi\|^2 \right) \quad (4.15c)$$

$$= \frac{1}{2c} \left(\|\text{proj}_{C^*}(p + c\theta)\|^2 - \|p\|^2 \right) . \quad (4.15d)$$

2. Elle est différentiable et

$$(\zeta_c)'_p(p, \theta) = \frac{1}{c} (\text{proj}_{C^*}(p + c\theta) - p) , \quad (4.15e)$$

$$(\zeta_c)'_{\theta}(p, \theta) = \text{proj}_{C^*}(p + c\theta) . \quad (4.15f)$$

3. Elle est concave en p et convexe en θ .

4. Elle est C -non décroissante en θ .

Démonstration. On pourra aussi consulter le corrigé de l'Exercice 4.2 qui donne d'autres indications sur la démonstration ci-dessous.

1. Le problème (4.15b) est un problème de maximisation sous la contrainte $q \in C^*$ que l'on dualise avec le multiplicateur $\xi \in -C$. Plus précisément, on se convainc que (4.15b) est équivalent à

$$\zeta_c(p, \theta) = \max_{q \in C^*} \inf_{\xi \in -C} \left(\langle q, \theta \rangle - \frac{1}{2c} \|p - q\|^2 - \langle q, \xi \rangle \right)$$

par le fait que l' \inf_{ξ} vaut $-\infty$ si $q \notin C^*$. Par dualité (le problème de maximisation étant concave sous contrainte convexe et "qualifiée"), on peut intervertir le max et l'inf dans l'expression ci-dessus qui deviennent respectivement un $\sup_{q \in C^*}$ (et même un $\max_{q \in C^*}$) et un $\min_{\xi \in -C}$. Enfin, le max étant celui d'une fonction quadratique sans contrainte se résoud explicitement, ce qui conduit finalement à l'expression alternative (4.15c) de ζ_c .

Par ailleurs, en partant par exemple de l'expression (4.15b), on réécrit celle-ci sous la forme

$$\max_{q \in C^*} \frac{1}{2c} (\|p + c\theta\|^2 - \|p + c\theta - q\|^2 - \|p\|^2), \quad (4.16)$$

d'où il résulte immédiatement que le max est atteint pour $q^\sharp = \text{proj}_{C^*}(p + c\theta)$ et que la valeur optimale est égale à (4.15d) en utilisant la décomposition orthogonale de $p + c\theta$ sur les cônes C^* et $-C$ illustrée par la Figure 3.2 (qui montre plutôt la décomposition analogue sur C et $-C^*$).

2. La dérivée partielle en p de l'expression sous le max dans (4.16) est égale à $(q - p)/c$ et cette quantité, évaluée à l'unique $\arg \max q^\sharp$ qui a été calculé ci-dessus (l'unicité montrant par ailleurs la différentiabilité de ζ_c en p), conduit à (4.15e) (selon un résultat souvent utilisé dans ce cours). On montre de la même façon la formule (4.15f).
3. La fonction ζ_c est concave en p comme enveloppe inférieure de fonctions affines en p d'après (4.15c). Elle est convexe en θ comme enveloppe supérieure de fonctions affines en θ d'après (4.15b).
4. La fonction est C -non décroissante en θ car sa dérivée en θ appartient à C^* d'après (4.15f). \square

Le point 4 du lemme permet de montrer que L_c est convexe-concave en (u, p) (en utilisant notamment le résultat de l'Exercice 3.49).

Remarque 4.26. Dans le cas $\mathcal{C} = \mathbb{R}$ (donc $C^* = \mathbb{R}^+$), on obtient l'expression suivante de ζ_c :

$$\zeta_c(p, \theta) = \frac{1}{2c} \left((\max(0, p + c\theta))^2 - p^2 \right).$$

Se reporter à ce sujet à la Remarque 4.1.

4.3.2 Propriétés du Lagrangien augmenté

Lemme 4.27.

1. Le Lagrangien augmenté L_c admet sur $U^{\text{ad}} \times \mathcal{C}^*$ les mêmes points selle que L sur $U^{\text{ad}} \times C^*$.
2. Le Lagrangien augmenté L_c est “stable en u ”.

Démonstration.

1. Si on désigne par $(U^\sharp \times P^\sharp)$ l'ensemble des points selle de L et par $(U_c^\sharp \times P_c^\sharp)$ le même ensemble relatif à L_c (on a vu que ces ensembles ont bien une forme “produit” — voir Exercice 3.51), alors

$$\begin{aligned} P^\sharp &= \arg \max_{p \in C^*} \psi(p) \\ &= \arg \max_{p \in \mathcal{C}^*} \psi_c(p) \end{aligned}$$

d'après le point 3 du Théorème 4.19 appliqué à ψ et ψ_c ,

$$= P_c^\sharp.$$

Par ailleurs,

$$\begin{aligned} U_c^\sharp &= \arg \min_{u \in U^{\text{ad}}} \left(\max_{p \in \mathcal{C}^*} L_c(u, p) \right) \\ &= \arg \min_{u \in U^{\text{ad}}} \left(\max_{p \in \mathcal{C}^*} \max_{q \in C^*} \left(L(u, q) - \frac{1}{2c} \|p - q\|^2 \right) \right) \\ &= \arg \min_{u \in U^{\text{ad}}} \left(\max_{q \in C^*} L(u, q) \right) \end{aligned}$$

en prenant d'abord le max en p qui est atteint pour $p = q$,

$$= U^\sharp .$$

2. Soit $\widehat{u} \in \arg \min_{u \in U^{\text{ad}}} L_c(u, p^\sharp)$ avec $p^\sharp \in P_c^\sharp$. Ceci est équivalent à l'inégalité de droite du point selle de L_c pour le couple (\widehat{u}, p^\sharp) . Il s'agit de montrer l'autre inégalité du point selle, ce qui prouvera enfin que $\widehat{u} \in U^\sharp$.

Puisque $p^\sharp \in P_c^\sharp = \arg \max_{p \in \mathcal{C}^*} \psi_c(p)$ et *puisque ψ_c est différentiable*, on a évidemment $\psi'_c(p^\sharp) = 0$. On sait que du fait que $\widehat{u} \in \arg \min_{u \in U^{\text{ad}}} L_c(u, p^\sharp)$, alors

$$\psi'_c(p^\sharp) = \frac{\partial L_c}{\partial p}(\widehat{u}, p^\sharp) = 0 . \quad (4.17)$$

La dernière inégalité prouve que la fonction $p \mapsto L_c(\widehat{u}, p)$ est stationnaire, donc maximale car concave, ce qui prouve l'autre inégalité du point selle. \square

Commentaires 4.28.

1. On voit, dans la dernière partie de la démonstration, le rôle fondamental joué par la différentiabilité de ψ_c . En effet, si ψ_c était seulement sous-différentiable, on pourrait seulement dire que

$$0 \in \partial \psi_c(p^\sharp) \quad \text{et} \quad \frac{\partial L_c}{\partial p}(\widehat{u}, p^\sharp) \in \partial \psi_c(p^\sharp) ,$$

mais cela ne permettrait pas de conclure que

$$0 = \frac{\partial L_c}{\partial p}(\widehat{u}, p^\sharp) .$$

Or la différentiabilité de ψ_c provient de l'interprétation en terme de régularisée de ψ .

2. On ne peut pas dire que le Lagrangien augmenté réalise une “forte convexification” du Lagrangien ordinaire. En effet, si c'était le cas, l'arg min en u de $L_c(\cdot, p^\sharp)$, noté $\widehat{u}(p^\sharp)$, devrait être unique. Or on vient de voir que si le problème original n'a pas de solution unique, mais bien un ensemble U^\sharp optimal, alors L_c “ne perd” aucune de ces solutions. Le “miracle” de la différentiabilité de ψ_c ne vient pas de l'unicité de $\widehat{u}(p^\sharp)$ mais de celle de l'expression

$$\frac{\partial L_c}{\partial p}(\widehat{u}(p^\sharp), p^\sharp) = \frac{1}{c} \left(\text{proj}_{C^*} \left(p^\sharp + c \Theta(\widehat{u}(p^\sharp)) \right) - p^\sharp \right)$$

(d'après (4.15e)), dont on sait maintenant qu'elle est nulle pour tout $\widehat{u}(p^\sharp)$ d'après (4.17).

4.4 Algorithme

4.4.1 Obtention de l'algorithme

La démarche d'obtention d'un algorithme du problème auxiliaire suit le schéma habituel :

1. répartition de l'expression du Lagrangien augmenté L_c entre une partie “additive en u ” (en supposant la décomposition (3.7)), notée \mathcal{L}_c , et une partie “non additive” mais différentiable, notée \mathcal{L}_c ;
2. choix d'une fonction auxiliaire et application des calculs du type (3.37), en n'oubliant pas, ici encore, le traitement séquentiel des variables primales et duales.

Du point de vue de ce qui va naturellement dans \mathcal{L}_c , si $J = \mathcal{J} + \mathfrak{J}$, alors \mathfrak{J} en fait bien sûr partie. Par contre, si $\Theta = \Theta + \mathfrak{T}$, on ne profite pas immédiatement de la partie \mathfrak{T} du fait qu'elle est composée avec la fonction ζ_c (voir (4.15)) qui est non linéaire en θ et qui ne préserve donc pas la structure additive. On n'a donc pas un intérêt immédiat à distinguer les parties Θ et \mathfrak{T} .

Cependant, si on applique une technique de “linéarisation partielle” à la fonction $\zeta_c(\cdot, p)$ comme dans (3.28), on pourra tout de même profiter de la présence éventuelle d'un terme \mathfrak{T} additif.

Du point de vue du choix de la fonction auxiliaire, il n'y a pas de raison a priori de faire un choix différent du choix (3.38). Les calculs (3.37), avec les adaptations décrites ci-dessus, conduisent finalement à l'algorithme suivant.

Algorithme 4.29. 1. Commencer avec u^0 et p^0 ; poser $k = 0$.

2. À l'itération $k + 1$, résoudre

$$\min_{u \in U^{\text{ad}}} \left(K(u) + \langle \varepsilon \mathcal{J}'(u^k) - K'(u^k), u \rangle + \varepsilon \mathfrak{J}(u) + \varepsilon \langle (\zeta_c)'_{\theta}(p^k, \Theta(u^k)), \Theta'(u^k) \cdot u + \mathfrak{T}(u) \rangle \right) ; \quad (4.18a)$$

soit u^{k+1} une solution.

3. Calculer

$$p^{k+1} = p^k + \rho (\zeta_c)'_p(p^k, \Theta(u^{k+1})), \quad (4.18b)$$

où ρ est égal à $\alpha \varepsilon$.

4. Si $\|u^{k+1} - u^k\| + \|p^{k+1} - p^k\|$ est “suffisamment petite”, stop ; sinon, retourner en 2 avec k incrémenté de 1.

Remarque 4.30. 1. On note, dans (4.18b), l'absence d'une projection sur C^* du fait que la maximisation est opérée sur tout \mathcal{C}^* . Mais, avec (4.15e), cette formule se développe en

$$p^{k+1} = \left(1 - \frac{\rho}{c}\right) p^k + \frac{\rho}{c} \text{proj}_{C^*}(p^k + c \Theta(u^{k+1})). \quad (4.18c)$$

Avec $\rho = c$, on retrouve exactement (3.39b).

2. Par ailleurs, avec la formule (4.15f), le terme $(\zeta_c)'_{\theta}(p^k, \Theta(u^k))$ intervenant dans (4.18a) est égal à

$$p^{k+1/2} = \text{proj}_{C^*}(p^k + c \Theta(u^k)).$$

Ceci correspond à une sorte de pas de l'algorithme de remise à jour de p qui est effectué à partir de p^k et de u^k (c'est-à-dire avant que u^{k+1} ne soit calculé). En quelque sorte, l'utilisation d'un Lagrangien augmenté conduit à une “double remise à jour” des prix, une fois avec la formule ci-dessus qui utilise u^k , une autre fois avec la formule (4.18c) qui utilise u^{k+1} (mais les deux fois à partir de p^k).

4.4.2 Convergence

Terminons en donnant rapidement un aperçu des conditions de convergence énoncées dans [12]. Dans cette référence, la présence d'un terme Θ n'a pas été étudiée explicitement : on le supposera donc nul ici. Pour le reste, les hypothèses et notations sont pratiquement les mêmes qu'au Théorème 3.20 à l'importante différence près que *l'hypothèse (H3) sur la forte convexité de J n'est plus indispensable*.

Les conditions trouvées sur les pas ε et ρ sont

$$0 < \varepsilon < \frac{b}{A + c\tau^2} ; \quad (4.19a)$$

$$0 < \rho < 2c \text{ (cas } C = \{0\}); \quad 0 < \rho \leq c \text{ (cas } C \neq \{0\}). \quad (4.19b)$$

Sous ces conditions, la suite $\{(u^k, p^k)\}$ engendrée par l'algorithme est bornée et tout point d'accumulation de cette suite (dans la topologie faible) est un point selle.

Commentaires 4.31.

1. On peut s'attendre à une convergence forte de $\{u^k\}$ sous l'hypothèse supplémentaire de forte convexité de J .
2. Pour l'algorithme d'Uzawa pur appliqué à L_c (c'est bien sûr un cas particulier de l'Algorithme 4.29), la borne sur ρ (borne stricte) est $2c$ même pour le cas inégalité ($C \neq 0$).
3. En fait, on peut montrer que la plage de valeurs acceptables pour le couple (ε, ρ) est plus large que celle définie par (4.19a)–(4.19b). On a

$$(\varepsilon, \rho) \in \bigcup_{\beta \in [0,1]} \left] 0, \frac{b}{A + (1 + \beta)c\tau^2} \right[\times \left] 0, \left(1 + \frac{\beta}{2}\right)c \right]. \quad (4.19c)$$

4. Les conditions de convergence révèlent un compromis dans le choix de c . On sait, d'après le Lemme 4.24, que plus c est grand, plus la régularisée ψ_c sera bien conditionnée, ce qui est favorable au choix de pas ρ grands. C'est bien ce que confirme (4.19b). Pour l'algorithme d'Uzawa pur (qui ne conduit malheureusement pas à la décomposition, même dans le cas de J et Θ additives), on aura donc a priori intérêt à choisir c arbitrairement grand.

Cependant, cette attitude ignore la difficulté plus ou moins grande de minimiser L_c en u à p fixé, c'est-à-dire le conditionnement de ce problème. Il est à craindre que ce conditionnement primal se dégrade lorsque c augmente exagérément au profit d'un meilleur conditionnement du problème dual. Ce compromis est révélé par la considération simultanée de (4.19a) et (4.19b) (ou celle de (4.19c)) dès que la minimisation en u fait l'objet, comme la maximisation en p , d'itérations explicites dans le cadre de l'Algorithme 4.29 (dont l'algorithme d'Arrow-Hurwicz appliqué à L_c est un autre cas "extrême").

Ce compromis entre conditionnement primal et conditionnement dual dans le choix de c a été confirmé par toutes les expériences numériques (nombreuses !) conduites à ce jour.

4.5 Corrigé des exercices

4.5.1 Corrigé de l'Exercice 4.2

Le Lagrangien augmenté pour un problème avec contrainte égalité de la forme $\Theta(u) - \xi = 0$ s'écrit

$$L_c(u, \xi, p) = J(u) + \langle p, \Theta(u) - \xi \rangle + \frac{c}{2} \|\Theta(u) - \xi\|^2.$$

Cette forme quadratique en ξ doit être minimisée par rapport à cette variable sous la contrainte $\xi \in -C$, ce qui donne lieu à l'inéquation variationnelle :

$$\forall \xi \in -C, \langle c(\widehat{\xi} - \Theta(u)) - p, \xi - \widehat{\xi} \rangle \geq 0 \quad \text{soit} \quad \langle \Theta(u) + p/c - \widehat{\xi}, \xi - \widehat{\xi} \rangle \leq 0.$$

La solution explicite s'écrit

$$\widehat{\xi} = \text{proj}_{-C} (\Theta(u) + p/c).$$

Selon la décomposition orthogonale illustrée par la Figure 3.2, ou plutôt la décomposition sur les cônes $-C$ et C^* , on a

$$\Theta(u) + p/c = \widehat{\xi} + \text{proj}_{C^*} (\Theta(u) + p/c) \Rightarrow \Theta(u) - \widehat{\xi} = \text{proj}_{C^*} (\Theta(u) + p/c) - p/c.$$

En reportant cette expression dans le Lagrangien L_c , on obtient

$$\begin{aligned} L_c(u, \widehat{\xi}(u, p), p) &= J(u) + \left\langle p, \text{proj}_{C^*} (\Theta(u) + \frac{p}{c}) - \frac{p}{c} \right\rangle \\ &\quad + \frac{c}{2} \left\| \text{proj}_{C^*} (\Theta(u) + \frac{p}{c}) - \frac{p}{c} \right\|^2 \\ &= J(u) + \frac{c}{2} \left\| \text{proj}_{C^*} (\Theta(u) + \frac{p}{c}) \right\|^2 - \frac{1}{2c} \|p\|^2. \end{aligned}$$

C'est l'expression du Lagrangien augmenté pour le cas des contraintes inégalité, qui se spécialise au cas égalité quand $C = \{0\}$, donc $C^* = C^*$, et proj_{C^*} devient l'identité.

4.5.2 Corrigé de l'Exercice 4.5

La fonction que l'on doit minimiser au second membre de (4.4) — notons la $f(u, v)$ — est une fonction différentiable et évidemment fortement convexe en v . Elle atteint donc son minimum en v en un point unique (notons $\widehat{v}(u)$ cet arg min). Par un théorème classique, la dérivée du résultat, c'est-à-dire de $J_c(u)$, est la dérivée partielle de $f(u, v)$ en u , soit $(u - v)/c$, évaluée au point $\widehat{v}(u)$.

Par ailleurs,

- d'une part, $J_c(u) \leq J(u)$ pour tout u puisque $J(u) = f(u, u)$ qui est forcément supérieur ou égal à $\inf_v f(u, v)$; en particulier, $\inf_u J_c(u) \leq \inf_u J(u)$;
- d'autre part, $f(u, v) \geq J(v)$ puisque le terme $\|u - v\|^2 / 2c$ est non négatif, donc $\inf_{u,v} f(u, v) = \inf_u J_c(u) \geq \inf_v J(v)$.

De ces considérations, on déduit que $\inf_u J_c(u) = \inf_v J(v)$. Ensuite, il doit être clair que si l' \inf_v est atteint pour J en un certain point, il est atteint pour le même v et pour $u = v$ en ce qui concerne f . La réciproque est tout aussi évidente.

4.5.3 Corrigé de l'Exercice 4.6

On doit minimiser la fonction : $v \mapsto |v| + \|v - u\|^2 / 2c$.

- Si $\widehat{v}(u) > 0$, alors en ce point la fonction $v \mapsto |v|$ a pour dérivée 1 et donc $1 + (\widehat{v}(u) - u)/c = 0$, d'où $\widehat{v}(u) = u - c$, mais la condition $\widehat{v}(u) > 0$ impose que $u > c$; alors $J_c(u) = |u - c| + c/2 = u - c/2$.
- Si $\widehat{v}(u) < 0$, alors en ce point la fonction $v \mapsto |v|$ a pour dérivée -1 et donc $-1 + (\widehat{v}(u) - u)/c = 0$, d'où $\widehat{v}(u) = u + c$, mais la condition $\widehat{v}(u) < 0$ impose que $u < -c$; alors $J_c(u) = |u + c| + c/2 = -u - c/2$.

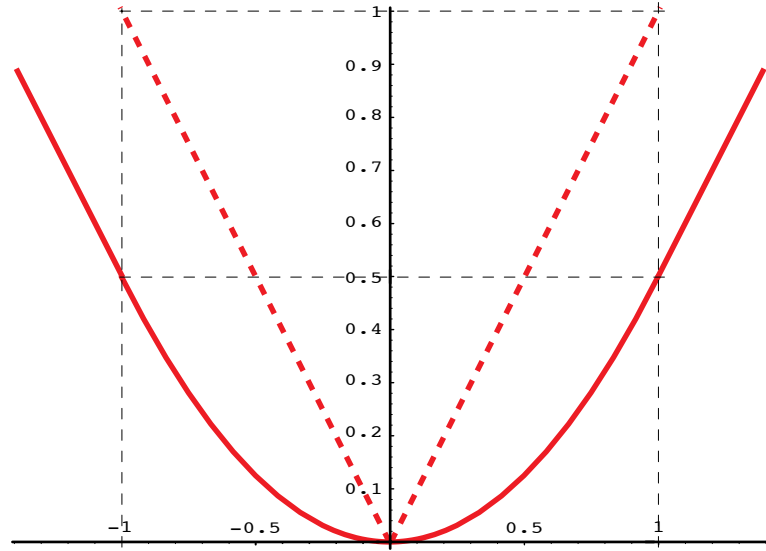


Figure 4.6: Graphe de la régularisée de Yosida de la fonction “valeur absolue”

- Enfin, si $\widehat{v}(u) = 0$, le sous-différentiel de la fonction $v \mapsto |v|$ en ce point est le segment $[-1, 1]$ et il faut que $(\widehat{v}(u) - u)/c \in [-1, 1]$, c’est-à-dire que $u \in [-c, c]$; alors, $J_c(u) = \|u\|^2 / 2c$.

En résumé,

$$J_c(u) = \begin{cases} -u - c/2 & \text{si } u < -c ; \\ \|u\|^2 / 2c & \text{si } -c \leq u \leq c ; \\ u - c/2 & \text{si } u > c . \end{cases}$$

Le graphe de cette fonction est représenté sur la Figure 4.6. On notera la similitude avec la Figure 2.7.

4.5.4 Corrigé de l’Exercice 4.8

Pour deux valeurs données x_1 et x_2 et un nombre $\alpha \in [0, 1]$, on a

$$g(\alpha x_1 + (1 - \alpha)x_2) = \inf_y f(\alpha x_1 + (1 - \alpha)x_2, y) = \inf_{y_1, y_2} f(\alpha x_1 + (1 - \alpha)x_2, \alpha y_1 + (1 - \alpha)y_2) ,$$

car on ne perd ni ne gagne rien à exprimer y comme $\alpha y_1 + (1 - \alpha)y_2$. Mais, grâce à la convexité conjointe de f en (x, y) , on a

$$\begin{aligned} g(\alpha x_1 + (1 - \alpha)x_2) &\leq \inf_{y_1, y_2} (\alpha f(x_1, y_1) + (1 - \alpha)f(x_2, y_2)) \\ &= \alpha \inf_{y_1} f(x_1, y_1) + (1 - \alpha) \inf_{y_2} f(x_2, y_2) \\ &= \alpha g(x_1) + (1 - \alpha)g(x_2) . \end{aligned}$$

4.5.5 Corrigé de l’Exercice 4.12

On a

$$(J \square K)^*(p) = \sup_u \left(\langle p, u \rangle - \inf_v (J(u - v) + K(v)) \right)$$

d'après (4.8) et (4.10),

$$= \sup_u \sup_v (\langle p, u \rangle - J(u - v) - K(v))$$

du fait que $-\inf f = \sup(-f)$,

$$= \sup_w \sup_v (\langle p, w + v \rangle - J(w) - K(v))$$

en changeant u en $v + w$,

$$\begin{aligned} &= \sup_w (\langle p, w \rangle - J(w)) + \sup_v (\langle p, v \rangle - K(v)) \\ &= (J^* + K^*)(p) . \end{aligned}$$

4.5.6 Corrigé de l'Exercice 4.17

Si $0 \in (\Phi + \partial \mathcal{I}_{U^{\text{ad}}})(u^\sharp)$, il existe $r \in \Phi(u^\sharp) \cap (-\partial \mathcal{I}_{U^{\text{ad}}})(u^\sharp)$. Comme le domaine de définition de $\mathcal{I}_{U^{\text{ad}}}$ est U^{ad} , il faut d'abord que $u^\sharp \in U^{\text{ad}}$. Ensuite, par définition du sous-différentiel d'une fonction convexe,

$$\forall u, \quad \mathcal{I}_{U^{\text{ad}}}(u) - \mathcal{I}_{U^{\text{ad}}}(u^\sharp) \geq \langle -r, u - u^\sharp \rangle .$$

Comme $\mathcal{I}_{U^{\text{ad}}}(u) = +\infty$ si $u \notin U^{\text{ad}}$, cette inégalité est non triviale seulement si $u \in U^{\text{ad}}$, et elle devient

$$\forall u \in U^{\text{ad}}, \quad \langle r, u - u^\sharp \rangle \geq 0 .$$

Cette inégalité est celle recherchée en se souvenant que $r \in \Phi(u^\sharp)$.

4.5.7 Corrigé de l'Exercice 4.21

Considérons deux valeurs u_1 et u_2 et notons v_1 et v_2 les valeurs correspondantes (uniques) où le minimum est atteint dans (4.4). Pour un certain $r_1 \in \partial J(v_1)$, respectivement $r_2 \in \partial J(v_2)$, on a les inéquations

$$\begin{aligned} \langle r_1 + (v_1 - u_1)/c, v_2 - v_1 \rangle &\geq 0 , \\ \langle r_2 + (v_2 - u_2)/c, v_1 - v_2 \rangle &\geq 0 , \end{aligned}$$

d'où, par addition,

$$\langle r_1 - r_2 + ((v_1 - u_1) - (v_2 - u_2))/c, v_2 - v_1 \rangle \geq 0 , \quad (4.20)$$

soit encore

$$\langle ((u_1 - v_1) - (u_2 - v_2))/c, u_1 - u_2 \rangle \geq \langle r_1 - r_2, v_1 - v_2 \rangle + c \|((u_1 - v_1) - (u_2 - v_2))/c\|^2 . \quad (4.21)$$

On sait que $J'_c(u) = (u - \widehat{v}(u))/c$. Par ailleurs, le premier terme au second membre de l'inégalité (4.21) est non négatif du fait de la monotonie de ∂J . Par conséquent, (4.21) s'interprète comme l'inégalité de Dunn avec la constante $1/c$ pour J'_c . En particulier, J'_c est Lipschitzien de constante $1/c$.

Si de plus J' existe et est Lipschitzien de constante A , et si $U^{\text{ad}} = \mathcal{U}$, alors les inéquations ci-dessus sont des équations, et en particulier $(u_i - v_i)/c = r_i = J'(v_i)$ pour $i = 1, 2$. Le premier terme au second membre de (4.21) est minoré par $(1/A) \|J'(v_1) - J'(v_2)\|^2$ (propriété de Dunn de J') qui est égal à $(1/A) \|((u_1 - v_1) - (u_2 - v_2))/c\|^2$. Alors, (4.21) se réécrit

$$\langle J'_c(u_1) - J'_c(u_2), u_1 - u_2 \rangle \geq ((1/A) + c) \|J'_c(u_1) - J'_c(u_2)\|^2 ,$$

d'où la constante de Dunn et de Lipschitz $A/(1 + cA)$.

Supposons maintenant ∂J fortement monotone de constante a (et U^{ad} à nouveau un convexe fermé quelconque). Revenons à (4.20) qui donne

$$\langle u_2 - u_1, v_2 - v_1 \rangle \geq (1 + ca) \|v_2 - v_1\|^2 ,$$

d'où l'on déduit que

$$\|v_1 - v_2\| \leq (1/(1 + ca)) \|u_1 - u_2\| \quad (4.22)$$

(propriété de Lipschitz de $u \mapsto \widehat{v}(u)$). Alors,

$$\begin{aligned} \langle J'_c(u_1) - J'_c(u_2), u_1 - u_2 \rangle &= (1/c) \langle (u_1 - v_1) - (u_2 - v_2), u_1 - u_2 \rangle \\ &= (1/c) (\|u_1 - u_2\|^2 - \langle v_1 - v_2, u_1 - u_2 \rangle) \\ &\geq (1/c) (\|u_1 - u_2\|^2 - \|v_1 - v_2\| \|u_1 - u_2\|) \\ &\geq (a/(1 + ca)) \|u_1 - u_2\|^2 , \end{aligned}$$

en utilisant (4.22).

Bibliographie

- [1] K.J. Arrow, L. Hurwicz, “Decentralization and Computation in Resource Allocation”, In : “*Essays in Economics and Econometrics*”, R.W. Pfouts, Ed., University of North Carolina Press, 34–104, 1960.
- [2] A. Bensoussan, J.-L. Lions, R. Temam, “Sur les méthodes de décomposition, de décentralisation et de coordination et applications”, In: “*Sur les Méthodes Numériques en Sciences Physiques et Économiques*”, J.-L. Lions, G.I. Marchouk, Eds., Dunod, Paris, 1974.
- [3] D.P. Bertsekas, “*Constrained optimization and Lagrange multiplier methods*”, Academic Press, 1982.
- [4] C.B. Brosilow, L.S. Lasdon, J.D. Pearson, “Feasible Optimization Methods for Interconnected Systems”, *Proceedings Joint Automatic Control Conference*, Troy, New-York, 1965.
- [5] G. Cohen, “Optimization by decomposition and coordination: a unified approach”, *IEEE Transactions on Automatic Control*, **AC-23**, 222–232, 1978.
- [6] G. Cohen, “Auxiliary Problem Principle and Decomposition of Optimization Problems”, *Journal of optimization Theory and Applications*, **32**, 277–305, 1980.
- [7] G. Cohen, “*Décomposition et coordination en optimisation déterministe différentiable et non différentiable*”, Thèse de Doctorat ès-Sciences Mathématiques, Université de Paris IX-Dauphine, 1984.
- [8] G. Cohen, “Auxiliary Problem Principle extended to variational inequalities”, *Journal of Optimization Theory and Applications*, **59**, 325–333, 1988.
- [9] G. Cohen, “*Convexité et Optimisation*”, cours ENPC, accessible par
<http://www-rocq.inria.fr/metalau/cohen/documents/Ponts-cours-A4-NB.pdf>
ou
<http://cermics.enpc.fr/~cohen-g/documents/Ponts-cours-A4-NB.pdf>, 2000.
- [10] G. Cohen, J.-C. Culioli, “Decomposition-coordination algorithms in stochastic optimization”, *SIAM Journal of Control and Optimization*, **28**, 1372–1403, 1990.
- [11] G. Cohen, B. Miara, “Optimization with an auxiliary constraint and decomposition”, *SIAM Journal of Control and Optimization*, **28**, 137–157, 1990.
- [12] G. Cohen, D.L. Zhu, “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*”, **I**, 203–266, JAI Press Inc., Greenwich, Connecticut, 1984.

- [13] J.-C. Culioli, “*Algorithmes de décomposition-coordination en optimisation stochastique*”, Thèse de Mathématiques et Automatique, Centre Automatique et Systèmes, École des Mines de Paris, 1987.
- [14] G.B. Dantzig, P. Wolfe, “The Decomposition Algorithm for Linear Program”, *Econometrica*, **29**, 767–778, 1961.
- [15] J.C. Dunn, “Convexity, monotonicity, and gradient process in Hilbert space”, *Journal of Mathematical Analysis and Applications*, **53**, 145–158, 1976.
- [16] N. El Farouq, “*Algorithmes de résolution d’inéquations variationnelles*”, Thèse de Mathématiques et Automatique, Centre Automatique et Systèmes, École des Mines de Paris, 1993.
- [17] W. Findeisen, F.N. Bailey, N. Brdys, K. Malinowski, P. Tatjewski, A. Wozniak, “*Control and Coordination in Hierarchical Systems*”, Wiley, New-York, 1980.
- [18] L.S. Lasdon, “*Optimization Theory for Large Systems*”, Mac Millan, Toronto, 1970.
- [19] L.S. Lasdon, J.D. Schoeffler, “A Multilevel Technique for Optimization”, *Proceedings Joint Automatic Control Conference*, Troy, New-York, 1965.
- [20] J. Marschak, R. Radner, “*The Economic Theory of Teams*”, Yale University Press, New Haven, Connecticut, 1971.
- [21] M.A. Mataoui, “*Contributions à la décomposition et à l’agrégation des problèmes variationnels*”, Thèse de Mathématiques et Automatique, Centre Automatique et Systèmes, École des Mines de Paris, 1990.
- [22] M.D. Mesarovic, D. Macko, Y. Takahara, “*Theory of Hierarchical Multilevel Systems*”, Academic Press, New-York, 1970.
- [23] A. Renaud, “*Algorithmes de régularisation et de décomposition pour les problèmes variationnels monotones*”, Thèse de Mathématiques et Automatique, Centre Automatique et Systèmes, École des Mines de Paris, 1993.
- [24] R.T. Rockafellar, “*Convex Analysis*”, Princeton University Press, Princeton, 1970.
- [25] D.A. Wismer (Ed.), “*Optimization Methods for Large Scale Systems with Applications*”, Mac Graw-Hill, 1971.
- [26] D.L. Zhu, “*Optimisation sous-différentiable et méthodes de décomposition*”, Thèse de Mathématiques et Automatique, Centre Automatique et Systèmes, École des Mines de Paris, 1982.